JOURNAL OF MATHEMATICAL PHYSICS VOLUME 11, NUMBER 12 DECEMBER 1970

Singularity Structure of Causal Distributions and Restricted Equal-Time Limits*

P. J. OTTERSON
Department of Physics, New York University, New York, New York
AND
J. SUCHER
Center for Theoretical Physics, Department of Physics and Astronomy
University of Maryland, College Park, Maryland

(Received 31 March 1970)

We study aspects of the space–time singularity structure of several classes of causal distributions, including the cases usually encountered in perturbation theory. Various definitions of restricted equal-time limit are considered which allow for the presence of highly singular Schwinger terms. It is shown that, with one definition, every causal distribution has a restricted equal-time limit. A form of sum rule valid even in the presence of singular Schwinger terms is given.

1. INTRODUCTION

The e.t. (equal-time) limits of matrix elements of current commutators continue to be of interest from the viewpoint of current algebra. In the simplest case of a vacuum-to-vacuum matrix element

$$C(x) = \langle 0|\,[j_A(x/2), j_B(-x/2)]\,|0\rangle,$$

one has available the Källén–Lehmann representation

$$C(x) = \int da\rho(a)\Delta(x; a). \tag{1.1}$$

A discussion of the e.t. limit of (1.1) (with Δ replaced by $\partial_0\Delta$) has been given by Brandt.[1] In the present paper we introduce a technique which enables us to exhibit explicitly the space–time singularity structure of $C(x)$ for a large class of spectral functions; the question of e.t. limits can then be discussed more directly. We also consider various definitions of such limits and the derivation of sum rules when Schwinger terms are present, i.e., when the e.t. limit exists only in a restricted sense, made precise below.

In Sec. IIA we show how *any* $C(x)$ of the form (1.1), with $\rho(a)$ a tempered distribution, may be written as a multiple derivative of an ordinary function of x defined by an integral having the form of a Fourier–Bessel transform. These integrals are evaluated for two classes of spectral functions $\rho(a)$: The first class shows how nonintegrable $\rho(a)$ can lead to $C(x)$ which vanish in an arbitrarily large neighborhood of the $t = 0$ hyperplane; this example is used later to cast light on the difference between various definitions of e.t. limit.[1–4] The second class incorporates all the cases encountered in perturbation theoretic calculations of spectral functions. In Sec. IIB, a generalization to the case of a Jost–Lehmann–Dyson representation is given.

In Sec. III we develop techniques and formulas for carrying out the needed distribution theoretic differentiations; the results are summarized in Tables I and II. The use of these tables is illustrated by examples from quantum electrodynamics.

In Sec. IV various definitions of e.t. limit are considered; it is emphasized that a sharp definition of Schwinger terms may be given in terms of the concept of restricted e.t. limit, in which the limit is defined only on a subspace of $\mathcal{S}(R_3)$. It is shown that every causal distribution has at least a restricted e.t. limit, in the sense of one of the definitions.

Concluding remarks concerned with the derivation of sum rules in the presence of Schwinger terms are contained in Sec. V.

A word about notation: If D is a distribution, $D(x)$ the associated generalized function, and $u(x)$ a test function, we shall often indicate the action of D on u by placing $D(x)$ in boldface square brackets and $u(x)$ in boldface round brackets to the right, thus:

$$[D](u) = [D(x)](u(x)) = \int dx\, D(x)u(x).$$

2. DISTRIBUTIONS DEFINED BY SPECTRAL REPRESENTATIONS

A. Källén–Lehmann Representation

The mathematical basis for this section is provided by the following relation:

Identity: Let $\Delta(x; a)$ be the usual causal solution of the homogeneous, mass-$a^{\frac{1}{2}}$ Klein–Gordon equation. Let d_x denote a differentiation operator or 1, and let $\rho(a)$ be a tempered distribution. Then, when b is not in the support of $\rho(a)$, we have

$$\int da\rho(a)d_x\Delta(x; a) = d_x(\Box + b)^n \int da\rho(a)(b - a)^{-n}\Delta(x; a). \tag{2.1}$$

Proof: Let $u(x)$ be a test function. Using $(\Box + a) \times \Delta(x; a) = 0$, we have[5]

$$\begin{aligned}
&\left[\int da\rho(a)d_x\Delta(x;a)\right](u)\\
&\quad=\int da\rho(a)\{[d_x\Delta(x;a)](u)\}\\
&\quad=\int da\rho(a)(b-a)^{-n}[\Delta(x;a)](\pm d_x(\Box+b)^n(u))\\
&\quad=\left[\int da\rho(a)(b-a)^{-n}\Delta(x;a)\right](\pm d_x(\Box+b)^n(u))\\
&\quad=\left[d_x(\Box+b)^n\int da\rho(a)(b-a)^{-n}\Delta(x;a)\right](u).
\end{aligned}$$

Equation (2.1) may be used to obtain explicit expressions for distributions specified by Källén–Lehmann spectral functions. Suppose, for simplicity, that $\rho(a)$ vanishes for $a \leq 0$ and that $d_x = 1$. Then (2.1) (for $n \geq 1$) becomes[6]

$$\begin{aligned}
&\int da\rho(a)\Delta(x;a)\\
&=(-\Box)^n\int daa^{-n}\rho(a)\left[-\left(\frac{a}{x^2}\right)^{\frac{1}{2}}\frac{\epsilon(x_0)\theta(x^2)}{4\pi}J_1[a(x^2)^{\frac{1}{2}}]\right].
\end{aligned}\tag{2.2}$$

Although the integrals (2.2) are to be interpreted in the sense of distribution theory, it can be shown that if $\rho(a)$ is a function, the integral on the right-hand side of (2.2) may be interpreted, for each x, as an ordinary Lebesgue integral, provided that it converges for large a. Since $\rho(a)$ is tempered, one can always choose n sufficiently large to ensure convergence and perform the differentiation after the integration.

This procedure may be used to convert *any* spectral integral to a classical integral, since any tempered distribution $\rho(a)$ may be expressed in the form

$$\rho(a) = \left(-\frac{d}{da}\right)^n\tilde{\rho}(a),$$

where $\tilde{\rho}(a)$ is a continuous function. Then

$$\begin{aligned}
&\int da\rho(a)\Delta(x;a)\\
&\quad=(\Box+b)^m\int da\tilde{\rho}(a)(b-a)^{-m}\left[\left(\frac{d}{da}\right)^n\Delta(x;a)\right].
\end{aligned}\tag{2.1'}$$

The use of $\tilde{\rho}$ permits integration over singularities in $\rho(a)$ occurring at *finite* values of a; the factor $(b-a)^{-m}$ forces convergence of the integral at infinity.

As an illustration of the use of (2.1), consider the class of spectral functions $\rho(a) = a^\kappa \sin(a^{\frac{1}{2}}\beta)\theta(a)$, with κ a nonnegative integer. Let

$$E(x;\kappa,\beta) \equiv \int_0^\infty daa^\kappa \sin(\beta\sqrt{a})\Delta(x;a). \tag{2.3}$$

Then, using (2.2), we have

$$\begin{aligned}
&E(x;\kappa,\beta)\\
&=(-\Box)^{\kappa+1}\int_0^\infty da\left(-a^{-\frac{1}{2}}\frac{\epsilon(x_0)\theta(x^2)}{4\pi(x^2)^{\frac{1}{2}}}J_1(ax^2)^{\frac{1}{2}}\sin\beta\sqrt{a}\right)\\
&=(-\Box)^{\kappa+1}\left[\frac{\epsilon(x_0)\theta(x^2)\theta(x^2-\beta^2)}{2\pi(x^2-\beta^2)^{\frac{1}{2}}}\left(\frac{\beta}{x^2}\right)\right].
\end{aligned}\tag{2.3'}$$

A remarkable feature of this example is that the support of the distribution $E(x;\kappa,\beta)$ excludes the hyperplane $t = 0$. Thus, there exist distributions, all derivatives of which have well-defined e.t. limits (in a sense to be defined in Sec. 3), for which the corresponding spectral functions are nonintegrable ordinary functions.

We turn now to another class of spectral functions, which covers most of the examples known from perturbation theory. Consider $\rho(a) = \theta(a-b)a^\alpha \times (\log b^{-1}a)^\kappa$, with $b > 0$ and κ a nonnegative integer. Again using (2.1), one finds

$$\begin{aligned}
E(x;\alpha,b,\kappa) &= \int_b^\infty daa^\alpha[\log(ab^{-1})]^\kappa\Delta(x;a)\\
&=(-\Box)^n\left(\frac{\epsilon(x_0)\theta(x^2)}{4\pi}I(x;n,\alpha,\kappa,b)\right),
\end{aligned}\tag{2.4}$$

where, with $\eta = a/b$,

$$\begin{aligned}
&I(x;n,\alpha,\kappa,b)\\
&\quad=-\int_1^\infty d\eta\frac{b^{\alpha-n+2}\eta^{\alpha-n+\frac{1}{2}}}{(x^2b)^{\frac{1}{2}}}(\log\eta)^\kappa J_1(\eta x^2b)^{\frac{1}{2}}\\
&\quad=-\int_1^\infty d\eta\frac{b^{\alpha-n+2}\eta^{\alpha-n+\frac{1}{2}}}{(x^2b)^{\frac{1}{2}}}(\log\eta)^\kappa\\
&\qquad\times\int_{-i\infty}^{i\infty}ds\frac{\Gamma(-s)[\frac{1}{2}(\eta x^2b)^{\frac{1}{2}}]^{1+2s}}{2\pi i\Gamma(s+2)}.
\end{aligned}$$

(The last equality is the Mellin–Barnes identity for the Bessel function J_1. The path of integration in s passes just to the left of the origin.)

Now pick $n > \alpha + 2$, so that the last integral exists over the real (is, η) plane and also over $d\eta$ for each fixed s. Using the Fubini theorem to invert the order

of integration, we obtain

$$I(x; n, \alpha, \kappa, b) = -\int_{-i\infty}^{i\infty} \frac{ds\Gamma(-s)(\frac{1}{2}bx^2)^{s\alpha-n+2}}{4i\pi\Gamma(s+2)} I(s, n, \alpha, \kappa, b), \quad (2.5)$$

where

$$I(s, n, \alpha, \kappa) = \int_1^\infty d\eta\eta^{\alpha-n+1+s}(\log\eta)^\kappa = \int_0^\infty dy e^{(\alpha-n+s+2)y} y^\kappa = \frac{\kappa!}{(n-\alpha-s-2)^{\kappa+1}}.$$

$I(s, n, \alpha, \kappa)$ is evaluated for s purely imaginary; the resulting function is then analytically continued into the right half s plane. [The poles of $I(s, n, \alpha, \kappa)$ are ultimately responsible for Schwinger terms, as we shall see.]

Finally, the integral over s may be performed by closing the contour at infinity in the right half s plane and evaluating the residues. (The new arc contributes nothing to the integral, for the magnitude of the new integrand is strictly less than that occurring in the original Mellin–Barnes identity.) Summarizing the results of these integrations, we have

$$E(x; \alpha, \kappa, b) = -(-\Box)^n\left(\frac{\epsilon(x_0)\theta(x^2)}{4\pi} I(x; \alpha, \kappa, b, n)\right), \quad (2.6)$$

where, in the case of no overlap between the poles of $\Gamma(-s)$ and $(n-\alpha-s-2)^{-\kappa-1}$,

$$I(x; \alpha, \kappa, b, n) = \frac{k!}{2} \sum_{\substack{m=0,1\ldots \\ m\neq n-\alpha-2}}^{\infty} \frac{(-\frac{1}{2}bx^2)^{m+b\alpha-n+2}}{m!\,(m+1)!\,(n-\alpha-m-2)^{\kappa+1}} + h$$

with

$$h = \frac{(-1)^\kappa}{2} b^{\alpha-n+2} \frac{1}{\kappa}\left[\left(\frac{\partial}{\partial s}\right)^\kappa \frac{\Gamma(-s)(\frac{1}{2}bx^2)^s}{\Gamma(s+2)}\right]_{s=n-\alpha-2}.$$

If α is an integer, the poles overlap and

$$h = \frac{-1}{(\kappa+1)}\left[\left(\frac{\partial}{\partial s}\right)^{\kappa+1} \times \left(\frac{(s-l)\Gamma(s-l)(\frac{1}{2}bx^2)^s}{\Gamma(s+2)}\right)\right]_{s=l=n-\alpha-2}.$$

These formulas are summarized in Table I.

B. JLD Representation

To analyze the general matrix element of a current commutator, it is convenient to use the Jost–Lehmann–Dyson (JLD) representation. The techniques of the preceding subsections may be easily generalized to cover a large class of JLD spectral functions.

We note that if one takes the Fourier transform of the usual momentum space JLD representation, one obtains, through formal convolution (see, however, Sec. 5 for critical remarks on this procedure) a representation of the form

$$D(x) \equiv \langle a|\,[j_A(\tfrac{1}{2}x), j_B(-\tfrac{1}{2}x)]\,|b\rangle = \int dM^2\rho(x; M^2)\Delta(x; M).$$

TABLE I. Spectral functions and corresponding distributions as derivatives of ordinary functions.

Spectral function $\rho(a)$	Distribution $E(x) = \int da\rho(a)\Delta(x; a)$
$\theta(a)a^k \sin(\beta\sqrt{a})$	$E(x; \kappa, \beta) = -(-\Box)^{\kappa+1}\left(\frac{\epsilon(x_0)\theta(x^2-\beta^2)}{2\pi(x^2-\beta^2)^{\frac{1}{2}}}\frac{\beta}{x^2}\right), \quad \kappa \geq 0$
$\theta(a-b)a^\alpha\left(\log\frac{a}{b}\right)^k$	$E(x; \alpha, \kappa, b) = -(-\Box)^n\left(\frac{\epsilon(x_0)\theta(x^2)}{4\pi}\right) I(x; \alpha, \kappa, b, n), \quad n \geq 1, n > \alpha+2$
	$I(x; \alpha, \kappa, b, n) = h(x; \alpha, \kappa, b, n) + \frac{1}{2}\Gamma(k) \sum_{\substack{m=0,1,\ldots \\ m\neq n-\alpha-2}}^{\infty} \frac{(-\frac{1}{2}bx^2)^m b^{\alpha-n+2}}{m!\,(m+1)!\,(n-\alpha-m-2)^{\kappa+1}}$
	$h(x; \alpha, \kappa, b, n) = \frac{(-)^\kappa b^{\alpha-n+2}}{2}$ $\frac{1}{(\kappa)}\left[\left(\frac{\partial}{\partial s}\right)^\kappa\left(\frac{\Gamma(-s)(\frac{1}{2}bx^2)^s}{\Gamma(s+2)}\right)\right]_{s=n-\alpha-2\neq \text{integer}}$ or $\frac{-1}{(\kappa+1)}\left[\left(\frac{\partial}{\partial s}\right)^{\kappa+1}\left(\frac{(s-l)\Gamma(s-l)(\frac{1}{2}bx^2)^s}{\Gamma(s+2)}\right)\right]_{s=l}$ $l = n-\alpha-2 = \text{integer}$

Assuming for simplicity that supp ρ excludes the point $M^2 = 0$ and applying the identity (2.1) with $d_x = 1$, we obtain the relation

$$D(x) = \int dM^2 \rho(x; M^2)(-M^2)^{-n} \Box^n \Delta(x; M^2)$$

or, on integration by parts,

$$D(x) = \sum_{k=0}^{2} \sum_{\mu,\nu} g_{\mu\nu} \binom{2}{k} (\partial_{x_\mu})^k \times \int dM^2 (-M^2)^{-n} [\partial_{x_\nu}^{2-k} \rho(x; M^2)] \Box^{n-1} \Delta(x; M^2).$$

This can be repeated until all the powers of $\Box$ are removed. Now note that if $\rho(x; M^2)$ is smooth in M^2 and, together with its derivatives with respect to x, is polynomially bounded in M^2, one may evaluate the integrals classically, as in 2A, by taking n large enough, and the differentiations may be performed afterwards. We now turn to the problems encountered in calculating such derivatives.

3. DIFFERENTIATION OF CAUSAL LORENTZ INVARIANT DISTRIBUTIONS

A. Computations

In the preceding section, we expressed several distributions as derivatives of functions. The present section is devoted to explicitly calculating relevant (distribution-theoretic) derivatives. We begin with two simple examples.

Example 1:

$$\Box\, [\epsilon(x_0)\theta(x^2)] = 4\epsilon(x_0)\delta(x^2). \tag{3.1}$$

Proof: Let $u = u(t, r, \Omega)$ be a test function. By definition,

$$[\Box\epsilon(x_0)\theta(x^2)](u) = \int_{-\infty}^{\infty} dt \epsilon(t) \int_0^{|t|} r^2\, dr \int d\Omega (\Box u)(t, r, \Omega).$$

Now we take u to be a product test function: $u(t, r, \Omega) = v(t)w(r, \Omega)$. (This involves no loss of generality—to show that two distributions agree, it is sufficient to show that they agree on a dense subspace of test functions, such as the linear span of analytic product functions.) Let $w(r, \Omega)$ be expanded in spherical harmonics. Only the spherical symmetric part remains after the integration over angles; let this part be denoted $w(r)$. We have, on integration by parts,

$$[\Box\epsilon(x_0)\theta(x^2)](v(t)w(r)) = 4\pi \int dt v(t)[d_t^2 \epsilon(t) I(t; w) - \epsilon(t) I(t; \nabla^2 w)],$$

where

$$I(t; w) = \int_0^{|t|} r^2 w(r)\, dr$$

and

$$I(t; \nabla^2 w) = \int_0^{|t|} r^2 (r^{-1} d_r^2 r w)\, dr.$$

Now

$$d_t^2[\epsilon(t) I(t; w)] = 2tw(|t|) + t^2 w'(|t|)\epsilon(t)$$

and (integrating twice by parts)

$$\epsilon(t) I(t, \nabla^2 w) = t^2 w'(|t|)\epsilon(t).$$

Thus

$$\begin{aligned}[\Box\epsilon(x_0)\theta(x^2)](v(t)w(r)) &= 8\pi \int_{-\infty}^{\infty} dt [tv(t)w(|t|)] \\ &= 4 \int_{-\infty}^{\infty} dt v(t)\epsilon(t) \int_0^{\infty} r^2\, dr \int d\Omega \left(\frac{\delta(|t| - r)}{2|t|} \right) w(r) \\ &= 4[\epsilon(x_0)\delta(x^2)](v(t)w(r)),\end{aligned}$$

and so

$$\Box[\epsilon(x_0)\theta(x^2)] = 4\epsilon(x_0)\delta(x^2).$$

Example 2:

$$\Box[\epsilon(x_0)\delta(x^2)] = 0. \tag{3.2}$$

Proof: With conventions as above,

$$\begin{aligned}[\Box\epsilon(x_0)\delta(x^2)](v(t)w(r)) &= 8\pi \int dt v(t)(d_t^2 t w(|t|) - t(r^{-1} d_r^2 r w)_{r=|t|}) \\ &= 0.\end{aligned}$$

This second identity may be obtained formally by applying the operator $\Box = (d_t^2 - r^{-1} d_r^2 r)$ to the object

$$\delta(t^2 - r^2) = \tfrac{1}{2}\epsilon(t) r^{-1} \delta(|t| - r).$$

Such formal manipulations are suspect; a formal calculation of $(d_t^2 - r^{-1} d_r^2 r)[\frac{1}{2} t^{-1} \delta(|t| - r)]$ yields the nonsensical answer $t^{-3}\delta(t - r)$.

The relation

$$\Box[\epsilon(x_0)\theta(x^2)(x^2)^n] = 4n(n + 2)\epsilon(x_0)\theta(x^2)(x^2)^{n-1}, \quad n \geq 1, \tag{3.3}$$

may be verified by techniques similar to those above.

The following trick is useful for differentiating more complicated distributions.

Trick: Consider the generalized function $I(t)$ defined, for $t > 0$, by

$$I(t) = \int_0^t f(t, r)\, dr = \lim_{\epsilon \to 0_+} \int_0^{t-\epsilon} f(t, r)\, dr,$$

where $f(t, r)$ is smooth and finite except possibly at $r = t$. [As a function of the variables $r \pm t$, f may have an integrable singularity in the variable $r - t$, as does the function $f(r, t) = \log(t - r)$.] Then, assuming for simplicity that $f(t, 0) = 0$ and using a comma to indicate differentiation, we have

$$d_t I(t) = \int_0^t (f_{,t} + f_{,r})\, dr, \tag{3.4}$$

although $f_{,t}$ and $f_{,r}$ may be separately nonintegrable near $r = t$.

Proof: Recall that the derivative of a limit of distributions is the limit of differentiated distributions. Then

$$\begin{aligned} d_t I(t) &= \lim_{\epsilon \to 0_+} d_t \int_0^{t-\epsilon} f(t, r)\, dr \\ &= \lim_{\epsilon \to 0_+} \left[\left(\int_0^{t-\epsilon} f_{,t}(t, r)\, dr \right) + f(t, t - \epsilon) \right] \\ &= \lim_{\epsilon \to 0_+} \left(\int_0^{t-\epsilon} (f_{,t} + f_{,r})\, dr \right), \end{aligned}$$

which proves (3.4).

This trick may be used to obtain a fairly explicit general expression for distributions of the form $\Box[\epsilon(x_0)\theta(x^2)f(x^2)]$:

$$\begin{aligned} &[\Box\epsilon(x_0)\theta(x^2)f(x^2)](u) \\ &\quad = \int_{-\infty}^{\infty} dt\epsilon(t) \int_0^{|t|} dr \int d\Omega [4r^2 u(|t| - r)^2 f'' \\ &\qquad + 4(r^2 u)_{,r}(|t| - r)f' + 2(ru)_{,r} f], \end{aligned} \tag{3.5}$$

where primes denote differentiation of a function with respect to its argument.

Proof: Let $u(x) = v(t)w(r, \Omega)$. Then

$$\begin{aligned} &[\Box\epsilon(x_0)\theta(x^2)f(x^2)](v(t)w(r, \Omega)) \\ &\quad = \int_{-\infty}^{\infty} dt\epsilon(t)\{v_{,tt}(t)I(t; rw) - v(t)I[t; (rw)_{,rr}]\}, \end{aligned}$$

where

$$I(t; rw) = \int d\Omega \int_0^{|t|} (rf)(rw)\, dr.$$

Noting that $I(0; rw) = I_{,t}(0; rw) = 0$, we get, on integration by parts,

$$\begin{aligned} &[\Box\epsilon(x_0)\theta(x^2)f(x^2)](v(t)w(r)) \\ &\quad = \int dt\epsilon(t)v(t)\{I_{,tt}(t; rw) - I[t; (rw)_{,rr}]\}. \end{aligned}$$

Now let

$$g(t, r) = \int d\Omega f(t^2 - r^2)(r^2 w)$$

so that

$$I(t; rw) = \int_0^{|t|} g(t, r)\, dr,$$

and apply the trick twice to calculate $I_{,tt}$:

$$\begin{aligned} I_{,tt} &= \int_0^{|t|} (g_{,rr} + 2g_{,r,t} + g_{,tt})\, dr \\ &= \int_0^{|t|} dr \int d\Omega [4r^2 w(|t| - r)^2 f'' + 4(|t| - r)(r^2 w)_{,r} f' \\ &\quad + 2(rw)_{,r} f + (rw)_{,rr}(rf)]. \end{aligned}$$

The integral of the last term in the integrand is just $I[t; (rw)_{,rr}]$, and thus

$$\begin{aligned} &[\epsilon(x_0)\theta(x^2)f(x^2)](v(t)w(r, \Omega)) \\ &\quad = \int_{-\infty}^{\infty} dt v(t)\epsilon(t) \int d\Omega \int_0^{|t|} dr [4(r^2 w)(|t| - r)^2 f'' \\ &\qquad + 4(r^2 w)_{,r}(|t| - r)f' + 2(rw)_{,r} f], \end{aligned}$$

which is equivalent to (3.5) for $u = v(t)w(r, \Omega)$.

We conclude this section by calculating

$$\Box^n[\epsilon(x_0)\theta(x^2)\log x^2].$$

We begin by substituting $f(x^2) = \log x^2$ into the general expression above and obtain, for $n = 1$,

$$[\Box\epsilon(x_0)\theta(x^2)\log x^2](u) = \int_{-\infty}^{\infty} dt\epsilon(t)I(t; u), \tag{3.6}$$

where

$$\begin{aligned} I(t; u) = \int d\Omega \int_0^{|t|} dr \Big(&- \frac{4r^2 u}{(|t| + r)^2} + \frac{4(r^2 u)_{,r}}{(|t| + r)} \\ &+ 2(ru)_{,r} \log(t^2 - r^2) \Big). \end{aligned} \tag{3.6'}$$

A fairly tedious calculation which consists in applying the trick several times to the above expression yields[7]

$$\begin{aligned} &[\Box^2\epsilon(x_0)\theta(x^2)\log x^2](u) \\ &\quad = 4\int_{-\infty}^{\infty} dt[u(t, |t|)P(t^{-1}) + \epsilon(t)u_{,r}(t, |t|)], \end{aligned} \tag{3.7}$$

where

$$u(t, r) = \int d\Omega u(t, r, \Omega).$$

This expression, together with Table I, will be used to derive the Schwinger terms in second-order quantum electrodynamics.

Equation (3.6) may be rewritten in manifestly covariant form. In order to motivate the definitions to come, recall the formulas for 1-dimensional distributions

$$\delta[f(\xi)] = [f'(\xi)]^{-1}\delta(\xi - \xi_r), \tag{3.8a}$$

where ξ_r is a zero of $f(\xi)$, and

$$\delta^{(n)}[f(\xi)] = [f'(\xi)]^{-1}\left(-\frac{d}{d\xi}[f'(\xi)]^{-1}\right)^n \delta(\xi - \xi_r), \tag{3.8b}$$

valid when $f(\xi)$ is monotonic increasing and appropriately differentiable. In particular, taking $f = x^2 = t^2 - r^2$ and viewing $\delta'(t^2 - r^2)$ as a t-parametrized family of distributions acting on functions of r, we find

$$[\delta'(t^2 - r^2)] = \tfrac{1}{4}r^{-3}\delta(|t| - r) - \tfrac{1}{4}r^{-2}d_r\delta(|t| - r).$$

For fixed $t \neq 0$, acting on a test function $w(r)$ (remembering the r^2 factor of integration), we see that this gives

$$[\delta'(t^2 - r^2)](w(r))$$
$$= \int_0^\infty dr r^2\left[\frac{1}{4r^3}\delta(|t| - r) - \frac{1}{4r^2}d_r\delta(|t| - r)\right](w(r))$$
$$= \frac{1}{4}\frac{w(|t|)}{|t|} + w_{,r}(|t|) = \frac{d}{dr^2}\frac{rw(r)}{2}\bigg|_{r=|t|}.$$

Thus, we *define* the 4-dimensional distribution $\epsilon(x_0)\delta'(x^2)$ by its action on a test function $u(t, r, \Omega)$ via

$$[\epsilon(x_0)\delta'(x^2)](u)$$
$$\equiv \tfrac{1}{4}\int_{-\infty}^{\infty} dt[u(t, |t|)P(t^{-1}) + \epsilon(t)u_{,r}(t, |t|)], \tag{3.9}$$

where

$$u(t, r) = \int d\Omega u(t, r, \Omega).$$

[Acting on product test functions $v(t)w(r)$ with $v(0) = 0$, this agrees with the t-parametrized interpretation above—thus the definition is reasonable. It is the handling of the singularity at $t = 0$ which is a matter of definition.] On comparing (3.7) and (3.9), we arrive at the manifestly covariant expression

$$\Box^2[\epsilon(x_0)\theta(x^2)\log x^2] = 16\epsilon(x_0)\delta'(x^2). \tag{3.10}$$

Similarly, viewing $\delta^{(n)}(t^2 - r^2)$ as a t-parametrized family of distributions, we find, acting on a test function $w(r)$,

$$[\delta^{(n)}(t^2 - r^2)](w(r)) = \left(\frac{d}{dr^2}\right)^n\left(\frac{rw}{2}\right)\bigg|_{r=|t|}.$$

Thus

$$d_t^2[[\delta^{(n)}(t^2 - r^2)](w(r))]$$
$$= (d_r)^2(d_{r^2})^n(\tfrac{1}{2}rw)\big|_{r=t}$$
$$= [2d_{r^2} + 4r^2(d_{r^2})^2](d_{r^2})^n(\tfrac{1}{2}rw)\big|_{r=|t|},$$

and

$$[\nabla^2[\delta^{(n)}(t^2 - r^2)]](w)$$
$$= \left(\frac{d}{dr^2}\right)^n d_r^2(\tfrac{1}{2}rw)\big|_{r=|t|}$$
$$= (d_{r^2})^n(2d_{r^2} + 4r^2d_{r^2}^2)(\tfrac{1}{2}rw)\big|_{r=t},$$

which implies

$$[\Box\delta^{(n)}(t^2 - r^2)](w(r))$$
$$\equiv d_t^2[\delta^{(n)}(t^2 - r^2)](w) - [\nabla^2\delta^{(n)}(t^2 - r^2)](w)$$
$$= 4(r^2d_{r^2}^{n+2} - d_{r^2}^n r^2 d_{r^2}^2)(\tfrac{1}{2}rw)\big|_{r=t}$$
$$= -4nd_{r^2}^{n+1}(\tfrac{1}{2}rw)\big|_{r=t}$$
$$= -4n[\delta^{(n+1)}(t^2 - r^2)](w(r)).$$

This makes reasonable the inductive *definition* (for 4-dimensional distributions now)

$$\epsilon(x_0)\delta^{(n+1)}(x^2) \equiv -(4n)^{-1}\Box\epsilon(x_0)\delta^{(n)}(x^2), \quad n \leq 1, \tag{3.11}$$

with which we have

$$\Box^n\epsilon(x_0)\theta(x^2)\log x^2 = (-4)^n(n-2)!\,\epsilon(x_0)\delta^{(n-1)}(x^2), \quad n \geq 2. \tag{3.12}$$

The formulas derived in this section are summarized in Table II.

B. Use of Tables I and II

By combining the second entry of Table I with the formulas of Table II, one is able to analyze in detail the strongest e.t. singularities of any Källén–Lehmann spectral representation for which the weight function behaves asymptotically like a polynomial in $\log a$ and fractional powers of a. (This covers all known cases in perturbation theory.) The idea is to subtract out the polynomial and use the tables to analyze its contribution. The remainder, corresponding to an integrable spectral function, has the e.t. behavior of the integrand. The work is greatly simplified by the observation that only the lowest fractional or logarithmic power of x^2 is important.

As an explicit example, consider the vacuum expectation values of current commutators in scalar and spinor electrodynamics, with

$$C_{\mu\nu} = \langle 0|\,[j_\mu(\tfrac{1}{2}x), j_\nu(\tfrac{1}{2}x)]\,|0\rangle.$$

TABLE II.

a. Various explicit distribution theoretic derivatives.

(1) $\Box[\epsilon(x_0)\theta(x^2)(x^2)^n] = 4n(n+1)\epsilon(x_0)\theta(x^2)(x^2)^{(n-1)}, \quad n \geq 1$

(2) $\Box[\epsilon(x_0)\theta(x^2)(x^2)^n(\log x^2)^m] = 4\epsilon(x_0)\theta(x^2)(x^2)^{n-1} \times [n(n+1)(\log x^2)^m + (2n+1)m(\log x^2)^{m-1} + m(m-1)(\log x^2)^{m-2}], \quad n \geq 1, m \geq 0$

(3) $\Box[\epsilon(x_0)\theta(x^2)] = 4\epsilon(x_0)\delta(x^2)$

(4) $\Box[\epsilon(x_0)\delta(x^2)] = 0$

(5) $\Box^n[\epsilon(x_0)\theta(x^2)\log x^2] = (-4)^n(n-2)!\,\epsilon(x_0)\delta^{(n-1)}(x^2), \quad n \geq 2$

(6) $\Box[\epsilon(x_0)\delta^{(n)}(x^2)] = -4n\epsilon(x_0)\delta^{(n+1)}(x^2)$

b. Distributions acting on the test function $u = u(x) = u(t, r, \Omega)$.

(1) $[\epsilon(x_0)\theta(x^2)f(x^2)](u) = \int_{-\infty}^{\infty} dt\epsilon(t)\int_0^{|t|} r^2\,dr\int d\Omega u f(x^2)$ (f integrable)

(2) $[\epsilon(x_0)\delta(x^2)](u) = \int_{-\infty}^{\infty} dt\int d\Omega \tfrac{1}{2}tu(t, |t|, \Omega)$

(3) $[\epsilon(x_0)\delta(x^2)](u) = \tfrac{1}{4}\int_{-\infty}^{\infty} dt\int d\Omega[u(t, |t|, \Omega)P(t^{-1}) + \epsilon(t)u_{,r}(t, |t|, \Omega)]$

(4) $[\Box\epsilon(x_0)\theta(x^2)\log x^2](u) = \int_{-\infty}^{\infty} dt\epsilon(t)I(|t|; w)$, where

$$I(|t|; w) = \int d\Omega\int_0^{|t|} dr\left(-\frac{4r^2u}{(|t|+r)^2} + \frac{4(r^2u)_{,r}}{(|t|+r)} + 2(ru)_{,r}\log(t^2 - r^2)\right)$$

In the scalar case one has

$$C_{\mu\nu} = \frac{ie^2}{48\pi^2}\int_{4m^2}^{\infty} da a^{-\frac{1}{2}}(a - 4m^2)^{\frac{3}{2}} \times (g_{\mu\nu} - a^{-1}\partial_{x_\mu}\partial_{x_\nu})\Delta(x; a) + \mathcal{O}(e^4),$$

and in the spinor case[1] one has

$$C_{\mu\nu} = \frac{ie^2}{12\pi^2}\int_{4m^2}^{\infty} da(a^{\frac{1}{2}} + 2m^2a^{-\frac{1}{2}})(a - 4m^2)^{\frac{1}{2}} \times (g_{\mu\nu} - a^{-1}\partial_{x_\mu}\partial_{x_\nu})\Delta(x; a) + \mathcal{O}(e^4).$$

From Table I we find

$$C_{0k}(x) \sim \Box^2\frac{d}{dt}\frac{d}{dx_k}[\epsilon(x_0)\theta(x^2)\log x^2 + \mathcal{O}(x^2)]. \quad (3.13)$$

Using Table II, we note

$$\Box^2[\epsilon(x_0)\theta(x^2)\log x^2] \sim \epsilon(x_0)\delta'(x^2).$$

Let $u = v(t)w(\mathbf{x})$ be a testing function for which $r^{-2}d_kw(\mathbf{x})$ is finite for $r \to 0$. Then

$$[C_{0k}](u) \propto \int dtv(t)h(t; w),$$

where, with $w_{,k} \equiv \int (\partial w/\partial x_k)_{r=|t|}\,d\Omega$,

$$h(t; w) = \frac{w_{,k,r}(|t|)}{|t|} - \frac{w_{,k}(|t|)}{t^2} + w_{,k,r,r}(|t|).$$

After a calculation similar to one done in Ref. 3, one then finds

$$h(0; w) = [-\tfrac{1}{2}\nabla^2 d_k\delta(\mathbf{x})](w). \quad (3.14)$$

Thus the restricted e.t. limit of $C_{0k}(x)$ (see Sec. 4) contains a Schwinger term (found in the spinor case in Ref. 1) proportional to $d_k\nabla^2\delta(\mathbf{x})$. The coefficients are $(ie^2/12\pi^2)$ and $(ie^2/48\pi^2)$ for spinor and scalar electrodynamics, respectively. The result for the scalar case was also derived in Ref. 3, using a different method.

It should be emphasized that, although the distribution $d_k\nabla^2\delta(\mathbf{x})$ can be extended (in a variety of ways) to act on any $w \in \mathcal{S}(R_3)$, the limit of $h(t; w)$ as $t \to 0$ does *not* exist unless $w(\mathbf{x})$ is restricted as above.

4. EQUAL-TIME LIMITS AND SCHWINGER TERMS

In this section we consider various definitions of equal-time limits of 4-dimensional distributions D and use examples from Sec. 1 to illustrate their different features.

(1) If a distribution $D(t)$ in a single variable t is equivalent to a continuous function $f(t)$ in a neighborhood N of the point $t = 0$, then a natural definition of the "equal time-limit" of D is simply the number $f(0)$. A generalization of this idea to a 4-dimensional distribution $D(x) = D(\mathbf{x}; t)$ is readily available if D,

when acting on product testing functions $u(x) = v(t)w(x)$ with supp $v(t) \subset N$, has the form

$$[D](v(t)w(\mathbf{x})) = \int dt v(t) H(w; t), \tag{4.1}$$

with H a continuous function of $t \in N$, for any $w \in \mathcal{S}(R^3)$. Also, if the map $w \to H(w; t)$ is a distribution for each $t \in N$, then we define (D1):

$$\left[\lim_{t\to 0} D\right]_1 (w) = H(w; 0). \tag{4.2}$$

When the conditions on H hold only for $w(\mathbf{x})$ in a proper subspace $\mathcal{S}_n \subset \mathcal{S}(R^3)$, $\mathcal{S}_n$ consisting of testing functions $w(\mathbf{x})$ such that $\lim [r^{-n} w(\mathbf{x})]$ as $r \to 0$ is finite, we shall say that the corresponding map $w(\mathbf{x}) \to H(w; 0)$ is a *restricted* e.t. limit of D and write $(\mathrm{D1})_n$:

$$\left[\lim_{t\to 0}\Big|_n D\right]_1 (w) = H(w, 0), \tag{4.3}$$

the bar and subscript indicating that the limit exists only on the subspace $\mathcal{S}_n$. If n is the smallest positive integer for which H exists, we refer to H as a *Schwinger term.*

The techniques of Sec. 2 were developed with this definition in mind—they enable one to compute the continuous function $H(w, t)$ explicitly. (Cf. Tables I and II and Sec. 3B.)

(2) It is clear that when a distribution D has a restricted e.t. limit in the sense of definition $(\mathrm{D1})_n$, then

$$\left[\lim_{t\to 0}\Big|_n D\right]_1 (w(\mathbf{x})) = \lim_{m\to\infty} D[v_m(t)w(\mathbf{x})] \tag{4.4}$$

for every sequence $v_m(t) \to \delta(t)$ in C'_N. (C_N is the space of continuous function on $\bar{N}$ and prime denotes dual.) This property may be used as a definition: Suppose that a distribution D has the property that there exists a neighborhood N of the origin $t = 0$ such that, for each sequence of testing functions $v_m(t) \to \delta(t)$ with support $v_m(t) \subset N$, $\lim [D(v_m(t)w(\mathbf{x}))]$ as $m \to \infty$ exists and defines a linear functional on $\mathcal{S}_n$, independent of sequence taken. We may then define $(\mathrm{D2})_n$:

$$\left[\lim_{t\to 0}\Big|_n D\right]_2 (w(\mathbf{x})) = \lim_{m\to\infty} D[v_m(t)w(\mathbf{x})]. \tag{4.5}$$

It has been shown[7] that if a distribution has an e.t. limit in the sense of $(\mathrm{D2})_n$, then it may be represented by a continuous function in the sense of $(\mathrm{D1})_n$. Thus definition $(\mathrm{D2})_n$, which might appear to have wider applicability, is in fact equivalent to $(\mathrm{D1})_n$.

A stronger requirement than that occurring in definition (D2), namely that the right-hand side of (4.5) exists and is sequence independent for $N = (-\infty, \infty)$, has been considered in the literature.[1] With such a modified definition (D2′), the existence of an equal-time limit is determined by the (Lebesgue) integrability of the spectral function in the case of a Källén–Lehmann representation. However, as the first example of Sec. 2 shows, there are distributions with nonintegrable spectral weights which vanish in a neighborhood of $t = 0$ and hence have well-defined equal-time limits (zero) in the sense (D2), but not in the sense (D2′). This is because $E(x; \kappa, \beta)$ [Eq. (1.3)] has strong singularities at points $t \neq 0$ and sequences $\{v_m(t)\}$ which tend to $\delta(t)$ in $\mathcal{S}'(R)$ may be contrived so as to pick up contributions from these singularities. Thus the definition (D2′), unlike (D2), does not have a purely local character.

(3) The range of applicability of (D2) may be extended by placing additional restrictions on the class of sequences considered. One might require that the supports shrink to zero, as in the following definition,[2] using a dilatation sequence. Let $v(t)$ be a testing function with the properties support $v \subset (-1, 1)$ and $\int v(t)\, dt = 1$, and let $v_m(t) = mv(mt)$. Then, if a distribution D is such that $\lim [D(v_m(t)w(\mathbf{x}))]$ exists, is independent of the choice of v, and defines a linear functional on $\mathcal{S}_n$, we may define a restricted e.t. limit by $(\mathrm{D3})_n$:

$$\left[\lim_{t\to 0}\Big|_n D\right]_3 (w(\mathbf{x})) = \lim_{m\to\infty} D[mv(mt)w(\mathbf{x})]. \tag{4.6}$$

It is not difficult to find distributions which have e.t. limits in the sense of (D3), but not in the sense of (D1) or (D2). For example, let

$$D(x) = \sum_{n=1}^{\infty} n^{-2}\theta(t - n^{-1})\delta(\mathbf{x} - t).$$

Moreover, under (D3), unlike under (D1) or (D2), the following theorem holds[8].

Theorem: Every causal distribution has a restricted e.t. limit, in the sense of $(\mathrm{D3})_n$.

Proof: The symbolic function $D(x)$ associated with a tempered distribution D may be written in the form $D(x) = \partial^k F(x)$, where $F(x)$ is a continuous function and ∂^k is a product of partial derivatives, i.e., $\partial^k = \prod_{i=0}^{3} (\partial/\partial x^i)^{k_i}$.

Let $h(\mathbf{x})$ be a test function in $\mathcal{S}(3)$, with supp $h \subset \{\mathbf{x} \,|\mathbf{x}| < 2\}$ and $h(x) = 1$ for $|\mathbf{x}| < \frac{3}{2}$. Since D is causal and supp $v \in (-1, 1)$, D gives the *same* result

acting on $[v_m(t)w(\mathbf{x})]h(\mathbf{m})$ as on $v_m(mt)w(\mathbf{x})$. Thus, with $|k| = \sum k_i$,

$$[D](mv(mt)w(\mathbf{x}))$$

$$= (-1)^{|k|}\iint d\mathbf{x}\, dt F(\mathbf{x}, t)\partial^k mv(mt)h(m\mathbf{x})w(\mathbf{x}) \quad (4.7)$$

$$= (-1)^{|k|}m^{k-3}\iint d\mathbf{x}\, dt F\left(\frac{\mathbf{x}}{m}, \frac{t}{m}\right)\partial^k v(t)h(\mathbf{x})w\left(\frac{\mathbf{x}}{m}\right), \quad (4.8)$$

where, to obtain the second line, we have introduced new variables $\mathbf{x}' = m\mathbf{x}$, $t' = mt$ and then dropped the primes. If we choose $w(\mathbf{x}) \in \mathcal{S}_n(3)$, then the integral in (4.8) will be of order m^{-n} for large m, since the integration is over a compact set and F is continuous. Hence, the rhs of (4.7) will tend to zero as $m \to \infty$, for $n \geq k - 2$. It follows that D has at least a restricted e.t. limit (equal to zero) of index $k - 2$.

We now prove a fact which is intuitively obvious—that, if D is a causal distribution (i.e., vanishes outside the light cone) and has a restricted e.t. limit, then the support of the restricted limit must consist of the origin alone. The proof is in terms of definition (D1); analogous proofs hold for the other definitions.

Proof: Suppose that a point $\mathbf{x} = \mathbf{b} \neq 0$ is in the support of $(\lim_{t\to 0}|_n D)$. Let $N_b = \{\mathbf{x} \mid |\mathbf{x} - \mathbf{b}| < \frac{1}{4}|\mathbf{b}|\}$. Let $H(w; t)$ be the continuous function and N the neighborhood of $t = 0$, corresponding to D. Since $\mathbf{b}$ is in the support of the restricted limit of D, there exists a test function $w_b(\mathbf{x})$ with support contained in N_b (thus w_b is in $\mathcal{S}_n$ for each n) such that $H(w_b; 0) \neq 0$. Since $H(w_b; t)$ is a continuous function of t in N, there exists a neighborhood T_b of $t = 0$ such that $H(w_b; t) > 0$ for t in T_b. Let $v_b(t)$ be a positive test function with support contained in $N \cap T_b \cap (-\frac{1}{4}|b|, \frac{1}{4}|b|)$. Then $v_b(t)w_b(\mathbf{x})$ is a test function in $\mathcal{S}(R^4)$ with support outside the light cone, and

$$[D](v_b(t)w_b(\mathbf{x})) = \int dt v_b(t)H(w_b; t) \neq 0.$$

This contradicts the assumption that D is causal.

Since a distribution with support consisting of a single point may be expressed as a finite sum of derivatives of delta functions, the e.t. limit of a causal distribution may be written

$$\left[\lim_{t\to 0} D\right] = \sum_{n=0}^{N} a_n \partial_{\mathbf{x}}^n \delta(\mathbf{x}), \quad (4.9)$$

where the a_n are constants, provided that the *unrestricted* ($n = 0$) e.t. limit exists.

5. CONCLUDING REMARKS

In conclusion we make a number of remarks related to the content of the preceding sections.

A. Failure of Convolution Assumption

In the derivation of sum rules in current algebra, one often uses the procedure of multiplying by $\delta(t)$ a current-commutator matrix element

$$C(x) = \langle\alpha|\,[j_A(\tfrac{1}{2}x), j_B(-\tfrac{1}{2}x)]\,|\beta\rangle$$

followed by Fourier transformation of the product, using "the convolution theorem" to compute the transform.[9]

To see the danger of this procedure, let us apply it to the distribution $(\partial/\partial t)E(x)$, where $E(x)$ is defined by Eq. (2.3), with, say, $k = \beta = 1$, so that

$$E(x) = \int_0^\infty da\sigma(a)\Delta(x; a)$$

with $\sigma(a) = a \sin \sqrt{a}$. Since, according to (2.3′), the support of $E(x)$ and hence of $\partial_t E(x)$ excludes the interval $-1 < t < 1$, we have $\delta(t)\partial_t E(x) \equiv 0$. On the other hand, the Fourier transformation of $\partial_t E(x)$ is $i\epsilon(k_0)k_0\sigma(k_0^2 - k^2)$ so that a formal application of the convolution theorem would give

$$0 = \int dk_0' |k_0'|\, \sigma(k_0'^2 - k^2) = 2\int da\sigma(a).$$

However, $\int da\sigma(a)$ is not well defined because $\sigma(a)$ is not integrable to infinity.

The source of this difficulty is that the usual convolution theorems do not apply to arbitrary pairs of distributions, even when they can be multiplied, as for the case at hand. A generalization of the concept of convolution which covers such cases is possible and will be discussed elsewhere.[10]

B. Sum Rules in the Presence of Schwinger Terms

If the unrestricted e.t. limit of $C(x)$ exists, we may write, using (4.9),

$$C(\mathbf{x}, 0) = \sum a_n \partial_x^n \delta(\mathbf{x}) \quad (5.1)$$

so that the procedure described above yields

$$\int dk_0 \hat{C}(\mathbf{k}, k_0) = \sum a_n(-i\mathbf{k})^n \quad (5.2)$$

with an obvious symbolic notation. If only a restricted e.t. limit exists, then $P(x)C(x)$, where P is a suitable polynomial in the components x^μ of x, will have an

e.t. limit, like the rhs of (5.1), so that multiplication by $\delta(t)$,followed by Fourier transformations, yields

$$\int dk_0 P\left(-i\frac{\partial}{\partial k}\right)\hat{C}(\mathbf{k}, k_0) = -\sum_n a_n(-i\mathbf{k})^n. \quad (5.3)$$

Equation (5.3) constitutes a generalization of (5.2) to the case when "real" Schwinger terms are present.

This is the case, for example, in photo-meson scattering, with a $g\phi^4$ strong interaction, where the relevant e.t.c has a $P(1/t)$ singularity in order g^2e^2.[7] It would be interesting to see whether such generalized sum rules can be useful in the investigation of physical scattering amplitudes.

ACKNOWLEDGMENTS

It is a pleasure to thank Dr. R. Brandt and Dr. C. Orzalesi for many conversations on the topics of this paper. We are also grateful to Dr. C.-H. Woo for a number of helpful comments.

* Supported in part by funds from the National Science Foundation and in part by the U.S. Air Force under grant AFOSR 68-1453A.

[1] R. A. Brandt, Phys. Rev. **166**, 1745 (1968).

[2] B. Schroer and B. Stichel, Commun. Math. Phys. **3**, 258 (1966).

[3] P. Otterson, J. Math. Phys. **10**, 8 (1969).

[4] D. Boulware and R. Jackiw, Phys. Rev. **186**, 1442 (1969) and references quoted there.

[5] That $\Delta(x; a)$ acting on testing functions $u(x)$ gives rise to a testing function in the variable a follows from the work of W. Güttinger and A. Rieckers, Commun. Math. Phys. **7**, 190 (1968). For precise conditions on $\rho(a)$ and for a general discussion of Lorentz invariant distributions, see the review by W. Güttinger in Fortschr. Phys. **14**, 483 (1966), especially Chap. 9; our method for carrying out differentiations of distributions differs from that used in this reference.

[6] We use the familiar decomposition

$$\Delta(x; a) = (2\pi)^{-1}\epsilon(x_0)\delta(x^2) - (4\pi)^{-1}\epsilon(x_0)\theta(x^2)(a/x^2)^{\frac{1}{2}}J_1[a(x^2)^{\frac{1}{2}}]$$

and the relation $\Box\epsilon(x_0)\delta(x^2) = 0$.

[7] P. Otterson, Ph.D. thesis, University of Maryland, 1969 (unpublished).

[8] A similar theorem has recently been given by A. M. Völkel (University of Nijmegen preprint).

[9] See, e.g., C. G. Bollini and J. D. Giambiagi, Nucl. Phys. **87**, 965 (1967); D. Amati, R. Jengo, and E. Remiddi, Nuovo Cimento **51A**, 999 (1967).

[10] P. Otterson (N.Y.U.) Preprint 7/70

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 11, NUMBER 12 DECEMBER 1970

Finding Hidden Symmetry by Numerical Methods*

OLE J. HEILMANN†
Department of Mathematics, Massachusetts Institute of Technology
Cambridge, Massachusetts 02139

(Received 16 June 1970)

Given a Hermitian matrix that depends upon some parameter, one often wants to be able to find all the parameter independent symmetries of the matrix. A method for accomplishing this is given here. The only numerical procedure involved is the diagonalization of Hermitian matrices.

I. INTRODUCTION

Normally, in quantum mechanics, symmetry is considered as something which is easy to find and which, when it has been found, is used to simplify the problem of diagonalizing the Hamiltonian. The method to be suggested here is the opposite: First diagonalize the Hamiltonian, and then use the results of the diagonalization to find the symmetry. One might wonder whether this can be of any use. The answer is that if one can find all the symmetry by just looking at the Hamiltonian, then one, of course, does not need any numerical methods to find it. One might, however, come across the situation where the results of the diagonalization of the Hamiltonian make one think that perhaps one has not found all the symmetry. This is actually the usual reaction if one finds degeneracy or violations of the noncrossing rule. Since the knowledge of the symmetry often will provide physical insight into the problem, and since the symmetry, when it has once been found, might be simple to verify and use in the solution of other problems, one would in such situations like to have a method which is more powerful than the trial and error method for finding symmetry. It was actually the occurrence of such a situation, when diagonalizing the Hubbard Hamiltonian for benzene,[1] that caused the evolution of the method to be presented in this article. It turned out to be a very useful tool in the benzene case and to provide great help in pointing the way to the right conjecture. It was even possible to use the method to prove rigorously that there was not any additional symmetry. Furthermore, the method can be used not only to find rigorous symmetry, but also approximate symmetry, which in turn could be used to find useful approximation methods.

II. THE NECESSARY THEORY OF MATRICES

The discussion will be limited to Hermitian operators on a finite-dimensional space, since some of the theorems to be used are not generally valid for operators in infinite dimensions. In any event, the subsequent algorithmic solution will apply only to finite matrices.

The first step will be to state a set of definitions and theorems on matrices which will be used in the following. The theorems will not be proved, since they are supposed to be either easy to look up in a textbook or trivial to prove. Matrices will in general be $n \times n$. Capital letters will stand for matrices and the corresponding small letter with two indices for the elements of the matrix. I will be used for the identity matrix. The superscript H will be used for Hermitian conjugation.

Theorem 1: Hermitian and unitary matrices can be diagonalized by unitary transformations.

Theorem 2: Two diagonalizable matrices commute iff they can be diagonalized with the same similarity transformation.

Theorem 3: If A is any $n \times m$ matrix, then there exist unitary matrices U (of size $n \times n$) and W (of size $m \times m$) such that

$$UAW^H = D, \tag{1}$$

where D is a $n \times m$ matrix of the form

$$D = \begin{Vmatrix} D_1 & 0 \\ 0 & 0 \end{Vmatrix}. \tag{2}$$

D_1 is a diagonal matrix whose diagonal elements are the nonzero singular values of A (i.e., the square root of the nonzero eigenvalues of AA^H or A^HA). U can be found as the unitary transformation which diagonalizes AA^H (the nonzero eigenvalues being ordered as in D_1),

$$UAA^HU^H = D', \tag{3}$$

while W is the product of two unitary matrices, V' and V,

$$W = V'V, \tag{4}$$

where V is the unitary transformation which diagonalizes A^HA (again with the same ordering of the nonzero eigenvalues as in D_1):

$$VA^HAV^H = D''. \tag{5}$$

The matrix V' is given as follows: Let C be the matrix

$$C = UAV^H = \begin{pmatrix} C_1 & 0 \\ 0 & 0 \end{pmatrix}, \tag{6}$$

where C_1 is a nonsingular matrix of the same dimension as D_1. Then V' is the $m \times m$ matrix

$$V'^H = \begin{pmatrix} C_1^{-1}D_1 & 0 \\ 0 & 0 \end{pmatrix}. \tag{7}$$

The transformations U and W are unique except for the following possibilities:

(1) U can be multiplied by any $n \times n$ unitary transformation of the form

$$\begin{pmatrix} I & 0 \\ 0 & U_2 \end{pmatrix}, \tag{8}$$

where I is of the same dimension as D_1.

(2) W can be multiplied by any $m \times m$ unitary transformation of the form

$$\begin{pmatrix} I & 0 \\ 0 & W_2 \end{pmatrix}, \tag{9}$$

where I is of the same dimension as D_1.

(3) If in D_1 some of the diagonal elements, say the k first, are equal, then U and W can be multiplied simultaneously by matrices of the form

$$\begin{pmatrix} V_k & 0 \\ 0 & I \end{pmatrix}, \tag{10}$$

V_k being any k-dimensional unitary transformation, and the whole matrix being $n \times n$ when multiplied with U and $m \times m$ when multiplied with W.

Definition 1: The graph G_A corresponding to a matrix A is defined as follows: Take n vertices numbered from 1 to n, and connect vertices j and k by a bond, if $|a_{jk}| + |a_{kj}| \neq 0$. Two vertices will be considered connected if there exists a path in the graph which connects them.

Definition 2: If all the vertices in the graph G_A are connected, then A is nonreducible (nonreducible is the same as irreducible for Hermitian matrices).

Theorem 4: If a diagonal matrix commutes with a nonreducible matrix, then the diagonal matrix is a constant times the identity matrix.

Definition 3: The eigenspace of matrix A belonging to an eigenvalue λ is the set of all eigenvectors with the eigenvalue λ.

Theorem 5: If the elements of a Hermitian matrix are holomorphic functions of a real parameter, then the eigenvalues are holomorphic functions of the parameter, and the eigenvalues are either equal for all values of the parameter or are at most equal at a finite number of points in any finite interval of the parameter.

III. DEFINITION OF SYMMETRY

The next step will be to investigate what it means to find all the symmetry of a Hermitian matrix. The following definition seems reasonably in agreement with the normal use of the word when one does not include antiunitary symmetry like time reversal.

Definition 4: Finding all the symmetry of a Hermitian matrix means finding all unitary transformations which leave it invariant.

This is, of course, equivalent to finding all the unitary matrices which commute with the given Hermitian matrix. This, however, has really no independent interest, since one has the following theorem.

Theorem 6: Finding all the symmetry of a Hermitian matrix is equivalent to finding all the eigenspaces of the matrix.

Proof: Consider a representation in which the Hermitian matrix is diagonal. Obviously, any change of sign of any eigenvector and any permutation of eigenvectors belonging to the same eigenspace are unitary transformations which commute with the given matrix. It is also obvious that knowing that these unitary transformations commute with a given matrix is enough to find the eigenspaces. QED

If, however, the Hermitian matrix depends on a parameter, then it becomes interesting to find all the symmetry, if it is defined as follows.

Definition 5: Finding all the symmetry of a matrix which depends on a parameter means finding all parameter-independent unitary transformations which leave the matrix invariant for all values of the parameter.

From Theorem 2 one easily deduces the following.

Theorem 7: Finding all the symmetry of a matrix, which depends on a parameter, is equivalent to finding the subspaces which are invariant subspaces independent of the parameter.

IV. FINDING THE SYMMETRY IN THE CASE OF NO DEGENERATION

The problem in the remaining part of this article is how to find these invariant subspaces for a Hermitian matrix $H(p)$. Such a matrix can be written as

$$H(p) = \sum_{j=1}^{m} f_j(p) H^{(j)}, \tag{11}$$

where the functions $f_1(p)_1 f_2(p), \cdots, f_m(p)$ are real functions, which are linearly independent, and the matrices $H^{(1)}, H^{(2)}, \cdots, H^{(m)}$ are Hermitian matrices. There is, of course, not a unique choice for the expansion; in general, it would be preferable to get m as small as possible, and for the use of numerical methods it will be preferable to use some reasonable choice of scaling. If one wants to find exact symmetry, then trying to get the norms of the matrices $H^{(j)}$ approximately equal would be reasonable, and if one wants to find approximate symmetry, then it is usually desirable to get the norms of the functions $f_j(p)$ to be approximately equal. But for the theory it does not matter which choice one takes.

We will first consider the relatively simple case, in which there exist a value p_0 such that none of the eigenvalues of $H(p)$ are degenerate. According to Theorem 5 it will always be possible to find such a value if none of the eigenvalues are permanently degenerate. Then the representation in which $H(p_0)$ is diagonal is uniquely determined. Changing to this representation, any unitary matrix U, which commutes with $H(p)$, is also diagonal according to Theorem 2. Next, form the matrix A whose elements are the sum of the numerical values of the corresponding elements of the matrices $H^{(1)}, H^{(2)}, \cdots, H^{(m)}$,

$$a_{kl} = \sum_{j=1}^{m} |h_{kl}^{(j)}|. \tag{12}$$

The following theorem solves the problem.

Theorem 8 (Main Theorem): If $H(p)$ is a Hermitian matrix given by (11), where the functions $f_1(p), f_2(p), \cdots, f_m(p)$ are linearly independent, if $H(p)$ is diagonal with nondegenerate eigenvalues for some value of p, and if the matrix A is formed according to (12), then a subspace is an invariant subspace of $H(p)$ independent of p if it is spanned by such a set of basic vectors of the representation that the corresponding vertices in the graph G_A are not connected to any other vertices.

Proof: By performing a permutation transformation, A can always be brought to the form

$$A = \begin{pmatrix} A_{11} & 0 & \cdots & 0 \\ 0 & A_{22} & \cdots & 0 \\ \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot \\ 0 & 0 & \cdots & A_{kk} \end{pmatrix}, \tag{13}$$

where the submatrices $A_{11}, A_{22}, \cdots, A_{kk}$ are nonreducible. The disconnected parts of G_A will then be precisely the parts which correspond to $A_{11}, A_{22}, \cdots, A_{kk}$, respectively. After the same permutation transformation, the form of $H(p)$ will be

$$H(p) = \begin{pmatrix} H_{11}(p) & 0 & \cdots & 0 \\ 0 & H_{22}(p) & \cdots & 0 \\ \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot \\ 0 & 0 & \cdots & H_{kk}(p) \end{pmatrix}, \tag{14}$$

and, since the functions $f_1(p), \cdots, f_m(p)$ were linearly independent, there exists a value of p for which the submatrices $H_{11}(p), H_{22}(p), \cdots, H_{kk}(p)$ are nonreducible. Theorem 8 now follows immediately from Theorems 2 and 4.

V. FINDING THE SYMMETRY IN THE GENERAL CASE

We will now proceed to outline a method which will also work for degenerate eigenvalues. The method introduced for nondegenerate eigenvalues will not work. Since the set of eigenvectors for $H(p_0)$ is not given uniquely, the matrix A is not given uniquely, and the connectedness of the graph G_A will, in general, depend upon the choice of representation. We can therefore not be sure of finding all the invariant subspaces by this method. The proof of Theorem 8, however, allows for the following corollary.

Corollary: An invariant subspace for $H(p)$ given by (11) is invariant independent of p iff it is an invariant subspace for all the matrices $H^{(1)}, H^{(2)}, \cdots, H^{(m)}$.

The vectors which span the invariant subspaces of $H(p)$ can consequently be chosen to be eigenvectors of $H^{(1)}$, and so the first step of the general method will be to diagonalize $H^{(1)}$ and change to a representation in which $H^{(1)}$ is diagonal. By choosing a suitable

expansion in (11) and a suitable initial representation, it would in many cases be possible to start with an $H^{(1)}$ which is already diagonal and thereby greatly reduce the necessary computations. The eigenvectors, which correspond to nondegenerate eigenvalues, are now fixed, and we only have to worry about rotations within the eigenspaces of larger dimensions than one. By a suitable permutation, the eigenvectors belonging to the same eigenspaces can be given consecutive numbers. If a matrix U commutes with $H^{(1)}$, it is necessarily block diagonal in this representation:

$$U = \begin{pmatrix} U_1 & 0 & 0 & \cdots & 0 \\ 0 & U_2 & 0 & \cdots & 0 \\ 0 & 0 & U_3 & \cdots & 0 \\ \cdot & \cdot & \cdot & & \\ \cdot & \cdot & \cdot & & \\ \cdot & \cdot & \cdot & & \\ 0 & 0 & 0 & & U_k \end{pmatrix}. \tag{15}$$

We proceed to $H^{(2)}$, which, with the same division into submatrices, looks like

$$H^{(2)} = \begin{pmatrix} H_{11}^{(2)} & H_{12}^{(2)} & H_{13}^{(2)} & \cdots & H_{1k}^{(2)} \\ H_{21}^{(2)} & H_{22}^{(2)} & H_{23}^{(2)} & \cdots & H_{2k}^{(2)} \\ H_{31}^{(2)} & H_{32}^{(2)} & H_{33}^{(2)} & \cdots & H_{3k}^{(3)} \\ \cdot & \cdot & \cdot & & \\ \cdot & \cdot & \cdot & & \\ \cdot & \cdot & \cdot & & \\ H_{k1}^{(2)} & H_{k2}^{(2)} & H_{k3}^{(2)} & & H_{kk}^{(2)} \end{pmatrix}. \tag{16}$$

Then, if U and $H^{(2)}$ commute, this implies that U_1 and $H_{11}^{(2)}$ commute, U_2 and $H_{22}^{(2)}$ commute, etc. So the next step will be to treat $H_{11}^{(2)}, H_{22}^{(2)}, \cdots, H_{kk}^{(2)}$ as $H^{(1)}$. That means diagonalizing and performing a suitable permutation. We then again have U in the form (15) (except that we may have more and smaller blocks), and the matrices $H_{11}^{(2)}, H_{22}^{(2)}, \cdots, H_{kk}^{(2)}$ are now of the form $\alpha_1 I, \alpha_2 I_2, \cdots, \alpha_k I$. This takes care of the diagonal part of $H^{(2)}$. The off-diagonal parts give rise to the following types of conditions:

$$U_1 H_{12}^{(2)} = H_{12}^{(2)} U_2. \tag{17}$$

Multiplying Eq. (17) by its Hermitian conjugate, one gets

$$U_1 H_{12}^{(2)} H_{12}^{(2)H} U_1^H = H_{12}^{(2)} H_{12}^{(2)H}, \tag{18}$$

$$U_2^H H_{12}^{(2)H} H_{12}^{(2)} U_2 = H_{12}^{(2)H} H_{12}^{(2)}. \tag{19}$$

Using the usual arguments together with Theorem 3 with $H_{12}^{(2)}$ for A, one finds that one should apply precisely that transformation to $H^{(2)}$, which carries $H_{12}^{(2)}$ into the form D of Eq. (2), and that the ambiguity we are left with in the choice of basic vectors is precisely the one stated at the end of Theorem 3. The division of $H^{(2)}$ (and U) into submatrices is then adjusted, and we proceed to the next off-diagonal submatrix of $H^{(2)}$. We will, however, have to keep track of not only the subspaces in which free rotation are still allowed, but also which of the subspaces are coupled together so that they have to be rotated simultaneously. After we have finished with the off-diagonal submatrices of $H^{(2)}$, all of them are either square matrices proportional to the identity matrix or else are zero. The subspaces which are coupled together will, after some permutation, correspond to diagonal blocks of submatrices in $H^{(2)}$ which are of the form

$$\begin{pmatrix} b_{11}I, & b_{12}I, & b_{13}I, & \cdots & b_{1l}I \\ b_{21}I, & b_{22}I, & b_{23}I, & \cdots & b_{2l}I \\ b_{31}I, & b_{32}I, & b_{33}I, & \cdots & b_{3l}I \\ \cdot & \cdot & \cdot & & \cdot \\ \cdot & \cdot & \cdot & & \cdot \\ \cdot & \cdot & \cdot & & \cdot \\ b_{l1}I, & b_{l2}I, & b_{l3}I, & \cdots & b_{ll}I \end{pmatrix} \tag{20}$$

and where the matrix B is irreducible. The unitary transformations, which commute with this matrix must be of the form

$$\begin{pmatrix} U' & 0 & 0 & \cdots & 0 \\ 0 & U' & 0 & \cdots & 0 \\ 0 & 0 & U' & \cdots & 0 \\ \cdot & \cdot & \cdot & & \cdot \\ \cdot & \cdot & \cdot & & \cdot \\ \cdot & \cdot & \cdot & & \cdot \\ 0 & 0 & 0 & & U' \end{pmatrix}. \tag{21}$$

When the procedure used on $H^{(2)}$ is continued to $H^{(3)}, H^{(4)}, \cdots$, one will eventually get to the point where one can stop, either because all the basic vectors are fixed, in which case the method for the nondegenerate case can be used, or because the procedure has been applied to the last of the matrices, $H^{(m)}$. In this case one can also, of course, use the method for the nondegenerate case to find the diagonal matrices which commute with $H(p)$, but besides these matrices there will also be unitary matrices with diagonal blocks of the form (21), corresponding to the rotations which have not been fixed during the procedure. The commutation of these matrices with $H(p)$ will imply permanent degeneracy, which is a consequence of the symmetry of $H(p)$.

This finishes the description of the general method for finding all the symmetry of $H(p)$. We conclude with some remarks about the numerical stability

of the method and how the results should be interpreted. The only numerical procedure involved (aside from matrix multiplication) is diagonalization of Hermitian matrices. This is a very stable numerical procedure, except that in the case of nearly degenerate eigenvalues only the invariant subspace corresponding to these eigenvalues is well determined. This difficulty might be overcome, however, by treating nearly degenerate eigenvalues as if they were degenerate. The only real numerical problem is that elements of A which should be zero will in general not come out of the computations as zeros, but as small numbers, and the level one chooses for the cutoff might be of importance for the degree of symmetry one finds. Since, however, there generally will be many paths in the graph G_A, from one vertex to another, one might expect that with a reasonable value for the cutoff level it would be possible to vary this level up and down within a factor of ten without changing the symmetry one finds. Furthermore, the method being numerical, it cannot give a rigorous proof of the existence of symmetry; it only can tell what one should try to prove. Finally, even the existence of unitary transformations which nearly commute with $H(p)$ might also provide important physical insight and suggest useful approximations.

ACKNOWLEDGMENT

The author is indebted to Professor E. Lieb for suggestions and encouragement.

* Work partially supported by National Science Foundation Grant GP-9414 and by Statens Naturvidenskabelige Forskningsråd Grant 551-208/69.

† On leave from University of Copenhagen. Permanent address: Chemistry Laboratory III, The H. C. Ørsted Institute, Universitetsparken 5, 2100 Copenhagen Ø, Denmark.

[1] O. J. Heilmann and E. H. Lieb, "Violation of the Noncrossing Rule: The Hubbard Hamiltonian for Benzene" (to be published).

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 11, NUMBER 12 DECEMBER 1970

Ground State Functional of the Linearized Gravitational Field*

KAREL KUCHAŘ†
Joseph Henry Laboratories, Princeton University, Princeton, New Jersey 08540
(Received 30 April 1970)

The ground state functional of the linearized Einstein theory of gravitation is given as a functional of the gauge invariant Ricci tensor, and compared with the corresponding electromagnetic expression. The connection of the canonically quantized nonlinear theory of gravitation with the linearized theory is exhibited. Time is treated as a momentum variable rather than as a superspace coordinate, which leads to an "extrinsic time representation" $h^{TT}{}_{ik}, h_i, t = -\frac{1}{2}\Delta^{-1}\pi^T$. The state functional of the linearized theory is shown to be the initial value of the state functional of the canonical theory on a constant extrinsic time hypersurface in the lowest order of a perturbation expansion. By means of the Einstein–Schrödinger equation, this functional can be integrated off this initial hypersurface.

1. INTRODUCTION

One of the elementary questions we can ask about an arbitrary quantized field is how the field quantities fluctuate around their expectation values. Although all expectation values are subject to fluctuations, it is simplest to consider the field in a state for which the expectation values of the field quantities vanish and in which there is therefore no field whatsoever from the classical standpoint. This state corresponds to zero values of total energy and momentum and represents the ground state of the field, the vacuum.

Wheeler has many times discussed the importance of quantum fluctuations of geometry at distances comparable with Planck's length $l_0 = (16\pi\hbar G c^{-3})^{\frac{1}{2}} \approx 10^{-32}$ cm. He arrived at a qualitative picture of the gravitational vacuum as a foamlike structure violently resonating at small distances among various states with widely different geometries and topologies.[1–4] On the basis of dimensional analysis and analogies with electromagnetism, he estimated how the typical components of the metric g, affine connection Γ, and Riemann curvature tensor R fluctuate in regions of linear dimensions l:

$$\Delta g \sim l_0/l, \quad \Delta\Gamma \sim l_0/l^2, \quad \Delta R \sim l_0/l^3. \tag{1}$$

Particle Representation and Field Representation

To characterize the fluctuations in more detail, we would like to know the probability that the field is distributed in space in a definite manner, i.e., that the field quantities are definite functions of the coordinates after the measurement. The "particle representation," in which the field is specified by occupation numbers, i.e., the numbers of field quanta with given momenta and energies, is not suitable for studying fluctuations of field quantities. A natural representation in this case is the "field representation," in which the field quantities themselves, as functions of the space coordinates, serve as canonical coordinates. In this representation, the state of the field is described by a functional of the field quantities, and this functional is directly the probability amplitude of the field quantities. The field representation has an intimate connection with Feynman's method of path integrals, and was studied by a number of authors.[5–7]

When the field equations are invariant with respect to a gauge group acting on the potentials of the theory, as in electrodynamics or in the Yang–Mills theory, the potentials themselves have no direct physical significance, and we would like to express the state functional as a functional of gauge invariant field variables. As a matter of fact, all components of the field tensor cannot simultaneously enter into the state functional, because from the point of view of the canonical formalism they represent conjugate variables. For example, in electrodynamics we can choose the magnetic field strengths $\mathbf{B}(\mathbf{x})$ [subject to the condition $\operatorname{div}\mathbf{B}(\mathbf{x}) = 0$] as canonical coordinates, and electric field strengths $\mathbf{E}(\mathbf{x})$ [subject to the condition $\operatorname{div}\mathbf{E}(\mathbf{x}) = 0$] as canonical momenta. The state functional can then be expressed either as a functional of $\mathbf{B}$ or as a functional of $\mathbf{E}$, but not as a functional of all components of the electromagnetic field tensor F_{ik}.

Canonical Formulation of Einstein's Theory and Linearized Theory

Einstein's theory of gravitation can be cast into canonical form, if we take the components of the metric of a spacelike hypersurface as canonical coordinates and the components of the extrinsic curvature tensor (or rather their combinations) as canonical momenta. The state functional in the metric representation becomes, therefore, a functional of the metric $g_{ik}(\mathbf{x})$. The role of a gauge transformation is played by the transformation of the metric induced by the new choice of coordinates. The state functional must not change under these transformations, and therefore depends only on the intrinsic geometry of the spacelike hypersurface, not on its particular representation by metric components. The set of all 3-space geometries is called superspace. In the metric representation, the state functional satisfies a functional differential equation, which can be called the Einstein–Schrödinger equation, because it is analogous

to the Schrödinger equation in the quantum mechanics of particles. No exact solutions of this equation are known, and, in spite of considerable effort by DeWitt,[8] its interpretation is still not quite clear in all respects.

For weak gravitational fields, the classical Einstein theory of gravitation can be replaced by a linearized theory, which is closely analogous to classical electrodynamics. In the linearized theory of gravitation, small deviations of the metric are studied with respect to the fixed flat space–time background. The system of coordinates is chosen in such a way that it does not differ too much from a Galilean system adapted to the background. Small changes of coordinates, which are permissible, induce gauge transformations of the metric. The components of the linearized Riemann tensor do not change under gauge transformations and therefore represent field variables. The linearized gravitational field can be quantized in the same way as the electromagnetic field, through a decomposition into independent harmonic oscillators. This procedure corresponds to the "particle representation," in which the state of the gravitational field is described by numbers of gravitons of given momentum and energy.

The relation between the quantized linearized theory of gravitation and the canonical formulation of the full nonlinear theory via the Einstein–Schrödinger equation in superspace is not quite straightforward. In the linearized theory, the state functional is defined only on privileged sections through the flat space–time background, namely on a system of parallel hyperplanes. Only one time parameter is necessary to specify the chosen hyperplane within the system. In the canonical formulation of the full nonlinear theory, any spacelike section is permissible, and a time function of the Tomonaga–Schwinger formalism is necessary for its specification. This time function and the conjugate energy density form a pair of canonically conjugate quantities, which are implicitly mixed with the other conjugate pairs of the canonical formalism. Relatively little attention has been paid to the process of "linearization" of the Einstein–Schrödinger equation. We discuss this question in more detail in Sec. 3.

Aims of the Paper

To give a full description of Wheeler's gravitational vacuum, we should find a state functional satisfying the Einstein–Schrödinger equation in the superspace of geometries with different topologies. It will apparently require a long time and many preliminary efforts before such a complete description can be given. However, we can try to make at least some steps forward on the path which leads from a qualitative picture to the final formal result. The electromagnetic ground state functional in the **B** representation is well known.[9] Let us therefore pursue the similarities between electromagnetism and linearized gravitation and find the corresponding state functional of the linearized gravitational field. This task is, of course, much less ambitious than Wheeler's original program. We may assume that this functional gives a fairly good description of the probability of small fluctuations of geometry from the flat background. We may also assume that, for large fluctuations in extended regions of space, the state functional is suppressed to such an extent that the differences between the exact functional and our functional are not very important. However, the linearized theory cannot give correct information about large fluctuations of geometry in regions whose linear dimensions are comparable to Planck's length, and these fluctuations are just the most interesting ones. Moreover, because the fluctuations are studied on the background of *a priori* chosen flat space, we have no chance to speak about fluctuations in topology.

The ground state functional of the linearized gravitational field is given in Sec. 2. To find out what guideline this functional gives to the solution of the corresponding problem in the canonical formulation of the nonlinear theory, we investigate in Sec. 3 the transition from this formulation to the quantum theory of the linearized gravitational field. In Sec. 4, we check that our ground state functional can be obtained in this way as an approximate solution of the Einstein–Schrödinger equation in a suitable representation.

Notation

Let us explain our notation. Greek indices run through the values 0, 1, 2, 3; Latin indices through the values 1, 2, 3. The space–time metric has the signature $-, +, +, +$. Partial differentiation is denoted by a comma; covariant differentiation with respect to the metric of 3-dimensional space by a stroke. The Riemann tensor, the Ricci tensor, and the scalar curvature of 3-space are defined in accordance with conventions proposed by Misner, Thorne, and Wheeler,[10] namely:

$$R^i{}_{klm} = \Gamma^i{}_{km,l} - \Gamma^i{}_{kl,m} + \Gamma^n{}_{km}\Gamma^i{}_{nl} - \Gamma^n{}_{kl}\Gamma^i{}_{nm},$$
$$R_{ik} = R^l{}_{ilk}, \quad R = R^i{}_i,$$

and the same convention is applied to the corresponding space–time tensors. The left superscripts 4 are used to distinguish 4-dimensional quantities from the 3-dimensional ones, e.g., ${}^4g^{ik}$ is the contravariant space–time metric tensor, whereas g^{ik} is the contravariant metric tensor of 3-dimensional space. The

determinant of g_{ik} is denoted by g. Linearized tensors bear the superscript (1), e.g., $R^{(1)}{}_{iklm}$ is the linearized Riemann tensor. The 3-dimensional Fourier transform of a function $f(\mathbf{x})$ is denoted by a bar:

$$\bar{f}(\mathbf{k}) = (2\pi)^{-\frac{3}{2}} \int d^3x\, f(\mathbf{x}) e^{-i\mathbf{k}\cdot\mathbf{x}},$$

$$f(\mathbf{x}) = (2\pi)^{-\frac{3}{2}} \int d^3k\, \bar{f}(\mathbf{k}) e^{i\mathbf{k}\cdot\mathbf{x}}.$$

The Laplace operator in flat space is denoted by Δ, and its inverse by Δ^{-1}. Asterisks are used for complex conjugation. We employ absolute units, in which $\hbar = c = 16\pi G = 1$ (G is Newton's gravitational constant, c is the velocity of light, $2\pi\hbar$ is Planck's constant). Planck's length $l_0 = (16\pi G\hbar c^{-3})^{\frac{1}{2}} \approx 10^{-32}$ cm becomes thereby the unit of length. Rationalized units are used for the electromagnetic field.

2. GROUND STATE FUNCTIONAL IN LINEARIZED THEORY

Although our aim is to get the ground state functional of the linearized gravitational field in the "field representation," it is intuitively helpful to start from its form in the "particle representation" and only afterwards to perform a canonical transformation to the field variables. The similarities between the electromagnetic and linearized gravitational fields are far reaching and are best exhibited in a table (see Table I). In the following, we explain the steps

TABLE I. Comparison of procedures leading to the ground state functionals of the electromagnetic field and the linearized Einstein gravitational field.

Row No.		Electromagnetic field	Gravitational field
1	Lagrangian density	$\mathfrak{L}_E = -\frac{1}{4}(A_{\kappa,\iota} - A_{\iota,\kappa})(A^{\kappa}{}_{,\iota} - A^{\iota}{}_{,\kappa})$	$\mathfrak{L}_G = -\frac{1}{4}h_{\iota\kappa,\lambda}h^{\iota\kappa}{}_{,\lambda} - \frac{1}{2}h^{\iota}_{\iota,\lambda}h^{\lambda\kappa}{}_{,\kappa} + \frac{1}{4}h^{\iota}_{\iota,\lambda}h^{\kappa}_{\kappa,\lambda} + \frac{1}{2}h^{\kappa}{}_{\iota,\kappa}h^{\iota\lambda}{}_{,\lambda}$
2	Field potentials	A_ι	$h_{\iota\kappa} = g_{\iota\kappa} - \eta_{\iota\kappa}$
3	Gauge transformation	$A_\iota \to A_\iota + \xi_{,\iota}$	$h_{\iota\kappa} \to h_{\iota\kappa} + \xi_{\iota,\kappa} + \xi_{\kappa,\iota}$
4	Gauge invariant field intensities	$F_{\iota\kappa} = A_{\kappa,\iota} - A_{\iota,\kappa}$	${}^4R^{(1)}{}_{\iota\kappa\lambda\mu} = \frac{1}{2}(h_{\iota\mu,\kappa\lambda} + h_{\kappa\lambda,\iota\mu} - h_{\iota\lambda,\kappa\mu} - h_{\kappa\mu,\iota\lambda})$
5	Field equations	$F_\iota{}^{\kappa}{}_{,\kappa} = 0$	${}^4R^{(1)}{}_{\iota\kappa} - \frac{1}{2}{}^4R^{(1)}\eta_{\iota\kappa} = 0$
6	Lorentz-invariant gauge conditions	$A^{\iota}{}_{,\iota} = 0$ (Lorentz condition)	$(h^{\kappa}_{\iota} - \frac{1}{2}h^{\lambda}_{\lambda}\delta^{\kappa}_{\iota})_{,\kappa} = 0$ (Einstein–Hilbert condition)
7	Vacuum gauge	$A_{i,i} = 0,\ A_0 = 0$ (Coulomb conditions)	$h_{ik,k} = 0,\quad h_{ii} = 0,\quad h_{0\iota} = 0$
8	Potentials in vacuum gauge	$A^{\mathrm{T}}{}_i$	$h^{\mathrm{TT}}{}_{ik}$
9	Space parts of field tensors as functions of potentials	$\mathbf{B} = \operatorname{curl}\mathbf{A}$	$R^{(1)}{}_{ik} = \frac{1}{2}(h_{il,kl} + h_{kl,il} - h_{ik,ll} - h_{ll,ik})$
10	Integrability conditions	$\operatorname{div}\mathbf{B} = 0$	$(R^{(1)}{}_{ik} - \frac{1}{2}R^{(1)}\delta_{ik})_{,k} = 0$
11	How to construct A^{T} and $h^{\mathrm{TT}}{}_{ik}$ out of gauge-invariant quantities	Choose $\mathbf{B}$ that satisfies (10E); then $\mathbf{A}^{\mathrm{T}} = -\Delta^{-1}\operatorname{curl}\mathbf{B}$	(I) Choose $R^{(1)}{}_{ik}$ that satisfies (10G); then $h^{\mathrm{TT}}{}_{ik} = -2\Delta^{-1}(R^{(1)}{}_{ik} - \frac{1}{4}R^{(1)}\delta_{ik} - \frac{1}{4}\Delta^{-1}R^{(1)}{}_{,ik})$ (II) Choose $R^{(1)}{}_{ik}$ that satisfies (10G) and the equation $R^{(1)} = 0$; then $h^{\mathrm{TT}}{}_{ik} = -2\Delta^{-1}R^{(1)}{}_{ik}$
12	Reduced Lagrangian	$L_E = \frac{1}{2}\int d^3x(A^{\mathrm{T}}{}_{i,0}A^{\mathrm{T}}{}_{i,0} - A^{\mathrm{T}}{}_{i,k}A^{\mathrm{T}}{}_{i,k})$	$L_G = \frac{1}{4}\int d^3x(h^{\mathrm{TT}}{}_{ik,0}h^{\mathrm{TT}}{}_{ik,0} - h^{\mathrm{TT}}{}_{ik,l}h^{\mathrm{TT}}{}_{ik,l})$
13	Field momenta	$\pi^{\mathrm{T}i} = \frac{\delta L_E}{\delta A^{\mathrm{T}}{}_{i,0}} = A^{\mathrm{T}}{}_{i,0}$	$\pi^{\mathrm{TT}ik} = \frac{\delta L_G}{\delta h^{\mathrm{TT}}{}_{ik,0}} = \frac{1}{2}h^{\mathrm{TT}}{}_{ik,0}$

TABLE I (*continued*).

Row No.		Electromagnetic field	Gravitational field
14	Hamiltonian	$\frac{1}{2}\int d^3x(\pi^{\mathrm{T}i}\pi^{\mathrm{T}i} + A^{\mathrm{T}}{}_{i,k}A^{\mathrm{T}}{}_{i,k})$	$\int d^3x(\pi^{\mathrm{TT}ik}\pi^{\mathrm{TT}ik} + \frac{1}{4}h^{\mathrm{TT}}{}_{ik,l}h^{\mathrm{TT}}{}_{ik,l})$
15	Canonical coordinates and momenta of field oscillators in k-representation	$Q_i^{(+)} = (\sqrt{2})^{-1}[\bar{A}^{\mathrm{T}}{}_i(\mathbf{k}) + \bar{A}^{\mathrm{T}*}{}_i(\mathbf{k})]$ $Q_i^{(-)} = (i\sqrt{2})^{-1}[\bar{A}^{\mathrm{T}}{}_i(\mathbf{k}) - \bar{A}^{\mathrm{T}*}{}_i(\mathbf{k})]$ $P_i^{(\pm)} = Q_i^{(\pm)}{}_{,0}$	$Q_{ik}^{(+)} = \frac{1}{2}[\bar{h}^{\mathrm{TT}}{}_{ik}(\mathbf{k}) + \bar{h}^{\mathrm{TT}*}{}_{ik}(\mathbf{k})]$ $Q_{ik}^{(-)} = (2i)^{-1}[\bar{h}^{\mathrm{TT}}{}_{ik}(\mathbf{k}) - \bar{h}^{\mathrm{TT}*}{}_{ik}(\mathbf{k})]$ $P_{ik}^{(\pm)} = Q_{ik}^{(\pm)}{}_{,0}$
16	Hamiltonian as a function of oscillator variables	$\int_{k_3>0} d^3k\, \frac{1}{2}(P_i^{(+)}P_i^{(+)} + P_i^{(-)}P_i^{(-)} + k^2Q_i^{(+)}Q_i^{(+)} + k^2Q_i^{(-)}Q_i^{(-)})$	$\int_{k_3>0} d^3k\, \frac{1}{2}(P_{ik}^{(+)}P_{ik}^{(+)} + P_{ik}^{(-)}P_{ik}^{(-)} + k^2Q_{ik}^{(+)}Q_{ik}^{(+)} + k^2Q_{ik}^{(-)}Q_{ik}^{(-)})$
17	Ground state functional as a direct product of ground state functions of field oscillators	$\Psi = \mathcal{N}\exp\left(-\frac{1}{2}\int_{k_3>0} d^3k \times k(Q_i^{(+)}Q_i^{(+)} + Q_i^{(-)}Q_i^{(-)})\right)$	$\Psi = \mathcal{N}\exp\left(-\frac{1}{2}\int_{k_3>0} d^3k \times k(Q_{ik}^{(+)}Q_{ik}^{(+)} + Q_{ik}^{(-)}Q_{ik}^{(-)})\right)$
18	Ground state functional as a functional of $\bar{A}^{\mathrm{T}}{}_i(\mathbf{k})$ or $\bar{h}^{\mathrm{TT}}{}_{ik}(\mathbf{k})$	$\Psi = \mathcal{N}\exp\left(-\frac{1}{2}\int d^3k\, k\bar{A}^{\mathrm{T}}{}_i(\mathbf{k})\bar{A}^{\mathrm{T}*}{}_i(\mathbf{k})\right)$	$\Psi = \mathcal{N}\exp\left(-\frac{1}{4}\int d^3k\, k\bar{h}^{\mathrm{TT}}{}_{ik}(\mathbf{k})\bar{h}^{\mathrm{TT}*}{}_{ik}(\mathbf{k})\right)$
19	Ground state functional expressed by means of first derivatives of field potentials	$\Psi = \mathcal{N}\exp\left(-\frac{1}{4\pi^2}\int d^3x\int d^3x' \times \frac{\mathbf{B}(\mathbf{x})\cdot\mathbf{B}(\mathbf{x}')}{\lvert\mathbf{x}-\mathbf{x}'\rvert^2}\right)$	$\Psi = \mathcal{N}\exp\left(-\frac{1}{8\pi^2}\int d^3x\int d^3x' \times \frac{h^{\mathrm{TT}}{}_{ik,l}(\mathbf{x})h^{\mathrm{TT}}{}_{ik,l}(\mathbf{x}')}{\lvert\mathbf{x}-\mathbf{x}'\rvert^2}\right)$
20	Ground state functional expressed by means of second derivatives of field potentials	$\Psi = \mathcal{N}\exp\left(\frac{1}{8\pi^2}\int d^3x\int d^3x' \times B_{i,k}(\mathbf{x})B_{i,l}(\mathbf{x}')n_kn_l\right)$ $\mathbf{n} = \frac{\mathbf{x}-\mathbf{x}'}{\lvert\mathbf{x}-\mathbf{x}'\rvert}$	$\Psi = \mathcal{N}\exp\left(-\frac{1}{2\pi^2}\int d^3x\int d^3x' \times R^{(1)}{}_{ik}(\mathbf{x})R^{(1)}{}_{il}(\mathbf{x}')n_kn_l\right)$ $R^{(1)} = 0$

and results given in this table. To refer to the table, we use the row number and the letters E and G for electromagnetic and gravitational columns, respectively.

Ground State Functionals in Particle Representation

The Lagrangian densities of the electromagnetic (1E) and linearized gravitational (1G) fields are expressed as functions of the vector potential A_ι, (2E), or of the symmetric tensor potential $h_{\iota\kappa} = g_{\iota\kappa} - \eta_{\iota\kappa}$, $\eta_{\iota\kappa} = \eta^{\iota\kappa} = \mathrm{diag}\ (-1, 1, 1, 1)$, (2G). Greek indices are lowered and raised by means of $\eta_{\iota\kappa}$ and $\eta^{\iota\kappa}$. The electromagnetic Lagrangian density does not change under the gauge transformation (3E), and the gravitational Lagrangian density changes only by a divergence under the gauge transformation (3G), which is induced by a linearized transformation of coordinates

$$x^\iota \to x'^\iota = x^\iota + \xi^\iota(x^\kappa).$$

The field equations (5E, G) are therefore invariant under these gauge transformations and can be written by means of the gauge invariant field variables (4E, G).

The freedom existing in the choice of the electromagnetic and gravitational potential can be limited by gauge conditions. The well-known Lorentz-invariant gauge conditions (6E, G) separate the components of the potentials in the field equations and reduce these equations to the form of wave equations for the potentials. These equations insure that an additional gauge transformation, generated by functions ξ and ξ^ι which satisfy homogeneous wave equations and therefore do not disturb the gauge conditions (6E, G), simplifies the potentials even further [(7E, G)]. The scalar electromagnetic potential vanishes, as well as the scalar and vector gravitational potentials, so that the electromagnetic potential is fully determined by a space vector A_i and the gravitational potential by a

space tensor h_{ik}. Moreover, the vector A_i is transversal ($A_{i,i} = 0$), and the tensor h_{ik} is transversal ($h_{ik,k} = 0$) and traceless ($h_{ii} = 0$). We accept the notation (8E, G) for potentials in the gauge (7E, G), introduced by Arnowitt, Deser, and Misner.[11,12]

Is it possible to recover the transversal components $A^{\mathrm{T}}(\mathbf{x})$ and $h^{\mathrm{TT}}{}_{ik}(\mathbf{x})$ directly from the gauge invariant spatial field intensities $\mathbf{B}(\mathbf{x})$ and $R^{(1)}{}_{iklm}(\mathbf{x})$. Let us note at first that, in a 3-dimensional space, the Ricci tensor $R_{ik}(\mathbf{x})$ completely determines the Riemann curvature tensor $R_{iklm}(\mathbf{x})$, so that we can use $R_{ik}(\mathbf{x})$ instead of $R_{iklm}(\mathbf{x})$. However, the vector field $\mathbf{B}(\mathbf{x})$ or the tensor field $R_{ik}(\mathbf{x})$ cannot be chosen quite arbitrarily, because they must satisfy the divergence conditions (10E, G). These conditions are just the integrability conditions, ensuring that the equations (9E, G) for the potentials A_i and h_{ik} have a solution. The scalar condition (10E) reduces the three components of $\mathbf{B}$ to what are effectively two independent components, from which two independent components of $\mathbf{A}^{\mathrm{T}}$ can be extracted, as shown in (11E). The vector condition (10G) reduces the six independent components of the symmetric tensor $R^{(1)}{}_{ik}$ to what are effectively three independent components, whereas $h^{\mathrm{TT}}{}_{ik}$ has only two independent components. Two ways are then open. We can either form such a combination of the $R^{(1)}{}_{ik}$ that the remaining independent component automatically drops out of this combination, as in (11G I), or we can restrict R_{ik} further by the additional condition $R^{(1)} = 0$ and construct $h^{\mathrm{TT}}{}_{ik}$ as in (11G II). In fact, the linearized Einstein equations (5G) *in vacuo* imply that the 3-space scalar curvature $R^{(1)}$ vanishes. We can prove (11G I) by substituting into (9G) a decomposition of the potential h_{ik} into transversal traceless, transversal, and longitudinal components [see Eqs. (20)–(26) in the next section]. We can also check directly that the potentials $h^{\mathrm{TT}}{}_{ik}$ in (11G I) satisfy the conditions $h^{\mathrm{TT}}{}_{ii} = h^{\mathrm{TT}}{}_{ik,k} = 0$ by virtue of the Bianchi identities (10G). The additional condition $R^{(1)} = 0$ in an alternative procedure (11G II) simply insures that the transversal part h^{T} of the potential vanishes [see Eq. (24)].

Using the gauge conditions (7E, G) and discarding divergencies in the Lagrangian densities (1E, G), we bring the Lagrangians into reduced forms (12E, G). From these forms, the canonical momenta (13E, G) and the Hamiltonians (14E, G) are deduced by the standard procedure.

Let us now pass to the Fourier transforms of the potentials, and introduce canonical coordinates $Q^{(\pm)}$ and canonical momenta $P^{(\pm)}$ in the $\mathbf{k}$ representation by (15E) and (15G). The Hamiltonians (14E, G) then assume the form of Hamiltonians for a system of independent harmonic oscillators of unit mass and frequencies k [(16E, G)]. Let us note that the canonical coordinates $Q^{(\pm)}(\mathbf{k})$ introduced in (15E, G) are the coefficients of a decomposition of the potentials $A^{\mathrm{T}}{}_i(\mathbf{x})$ or $h^{\mathrm{TT}}{}_{ik}(\mathbf{x})$ into a system of standing waves $Q^{(+)}(\mathbf{k}) \cos \mathbf{k}\cdot\mathbf{x}$, $Q^{(-)}(\mathbf{k}) \sin \mathbf{k}\cdot\mathbf{x}$. This choice of canonical coordinates guarantees that the "magnetic" energy $\frac{1}{2}\int d^3x\, A^{\mathrm{T}}{}_{i,k} A^{\mathrm{T}}{}_{i,k}$ (or $\frac{1}{4}\int d^3x\, h^{\mathrm{TT}}{}_{ik,l} h^{\mathrm{TT}}{}_{ik,l}$) in the Hamiltonians (14E, G) goes over into potential energy of a system of harmonic oscillators (16E, G), and the "electric" energy $\frac{1}{2}\int d^3x\, \pi^{\mathrm{T}i}\pi^{\mathrm{T}i}$ (or $\int d^3x \times \pi^{\mathrm{TT}ik}\pi^{\mathrm{TT}ik}$) goes over into the kinetic energy of a system of harmonic oscillators. On the other hand, a more frequently used decomposition into running waves mixes these energies together. The choice (15E, G) of canonical coordinates is essential if we want to express the state functionals (17E, G) in the Q representation directly by means of the field variables $\mathbf{B}(\mathbf{x})$ or $R^{(1)}{}_{ik}(\mathbf{x})$, respectively. The coordinates $Q^{(\pm)}(\mathbf{k})$ are not all independent, because $Q^{(\pm)}(-\mathbf{k}) = \pm Q^{(\pm)}(\mathbf{k})$. We can take $Q^{(\pm)}(\mathbf{k})$ for $k_3 \geq 0$ as independent coordinates. That is why the integrals in (16E, G) and (17E, G) are taken only over the region $k_3 \geq 0$ of the $\mathbf{k}$ space.

The ground state functional of the electromagnetic or gravitational field (17E, G) is a direct product of the ground state wavefunctions of harmonic oscillators (15E, G). We can write the state functionals also by means of the Fourier transforms $\bar{A}^{\mathrm{T}}{}_i(\mathbf{k})$ or $\bar{h}^{\mathrm{TT}}{}_{ik}(\mathbf{k})$ of the potentials, as in (18E, G). The coefficients $\mathcal{N}$ are normalization constants.

Ground State Functionals in Field Representation

Let us now try to pass to the "field representation" of the state functionals (18E, G). For the electromagnetic field we can easily express the integral in the exponent of (18E) in terms of the gauge invariant magnetic field intensity $\mathbf{B}(\mathbf{x})$. By virtue of the relation

$$\overline{|\mathbf{x}|^{-2}} = \tfrac{1}{2}(2\pi)^{\frac{1}{2}}k^{-1},$$

we get

$$\int d^3x \int d^3x' \frac{B_i(\mathbf{x})B_i(\mathbf{x}')}{|\mathbf{x}-\mathbf{x}'|^2} = (2\pi)^{\frac{3}{2}}\int d^3k\, \tfrac{1}{2}(2\pi)^{\frac{1}{2}}k^{-1}\bar{B}_i(\mathbf{k})\bar{B}^*{}_i(\mathbf{k}) = 2\pi^2\int d^3k\, k\bar{A}^{\mathrm{T}}{}_i(\mathbf{k})\bar{A}^{\mathrm{T}*}{}_i(\mathbf{k}).$$

This yields the well-known gauge invariant form of the electromagnetic ground state functional (19E).[9] In the same way, we can express the gravitational ground state functional by means of a quadratic combination

of first derivatives of the transversal traceless components of the gravitational potential (19G). Expressing the potentials $h^{TT}{}_{ik}$ through $R^{(1)}{}_{ik}$ according to (11G), we get Ψ as a functional of gauge invariant intensities. However, there is a fourfold or a sixfold integral over the $\mathbf{x}$ coordinates in the exponential, depending on whether (11G II) or (11G I) is used.

However, the gravitational ground state functional can be expressed by yet another and simpler way in a form

$$\Psi = \mathcal{N} \exp\left(-\int d^3x \int d^3x' \times K_{iklm}(\mathbf{x}-\mathbf{x}')R^{(1)}{}_{ik}(\mathbf{x})R^{(1)}{}_{lm}(\mathbf{x}')\right). \quad (2)$$

Let us suppose that $R^{(1)}{}_{ik}$ satisfies $R^{(1)} = 0$. Because the linearized Ricci tensor is a combination of the second derivatives of potentials, its Fourier transform depends on the square of the wave vector $\mathbf{k}$:

$$\bar{R}^{(1)}{}_{ik}(\mathbf{k}) = \tfrac{1}{2}k^2 \bar{h}^{TT}{}_{ik}(\mathbf{k}).$$

The Fourier transform of the kernel $K_{iklm}(\mathbf{x}-\mathbf{x}')$ must therefore contain k^{-3} in order that (2) coincide with (18G). This is true for a nonradial tensorial kernel containing a unit vector $\mathbf{n} = (\mathbf{x}-\mathbf{x}')/|\mathbf{x}-\mathbf{x}'|$. We prove in the Appendix that

$$\overline{n_i n_k} = [-(2\pi)^{\frac{3}{2}}\delta(\mathbf{k}) + \tfrac{1}{2}(2\pi)^{\frac{1}{2}}k^{-3}]\delta_{ik} + [4(2\pi)^{\frac{3}{2}}\delta(\mathbf{k}) - \tfrac{3}{2}(2\pi)^{\frac{1}{2}}k^{-3}]\frac{k_i}{k}\frac{k_k}{k}. \quad (3)$$

Investigating the double integral

$$\int d^3x \int d^3x'\, R^{(1)}{}_{ik}(\mathbf{x})R^{(1)}{}_{il}(\mathbf{x}')n_k n_l = (2\pi)^{\frac{3}{2}}\int d^3k\, \bar{R}^{(1)}{}_{ik}(\mathbf{k})\bar{R}^{(1)*}{}_{il}(\mathbf{k})\overline{n_k n_l} = \tfrac{1}{4}(2\pi)^{\frac{3}{2}}\int d^3k\, k^4 \bar{h}^{TT}{}_{ik}(\mathbf{k})\bar{h}^{TT*}{}_{il}(\mathbf{k})\overline{n_k n_l} = \tfrac{1}{2}\pi^2\int d^3k\, k\bar{h}^{TT}{}_{ik}(\mathbf{k})\bar{h}^{TT*}{}_{ik}(\mathbf{k}),$$

we learn that the only contribution from $\overline{n_k n_l}$ comes from the term $\frac{1}{2}(2\pi)^{\frac{1}{2}}k^{-3}\delta_{ik}$, because $\bar{h}^{TT}{}_{ik}(\mathbf{k})k_k = 0$ in consequence of the transversality of $h^{TT}{}_{ik}$, and the contributions from δ functions are eliminated by the multiplicative factor k^4. In this way we get the gauge invariant form (20G) of the gravitational ground state functional.

The functional (20G) is in a sense simpler than the electromagnetic functional (19E), because it depends only on the direction of the vector joining the points $\mathbf{x}$ and $\mathbf{x}'$ and not on its magnitude. However, the difference between these expressions is only apparent, because the electromagnetic functional can be transformed into the form (20E), containing second derivatives of the vector potential A_i and closely analogous to (20G). To prove this statement, it is sufficient to introduce the Fourier transforms into (20E) and use (3) to return back to (18E), or we can reduce (20E) directly to (19E) through partial integrations, because

$$B_{i,k}(\mathbf{x})B_{i,l}(\mathbf{x}')n_k n_l = \partial_k(B_i(\mathbf{x})B_{i,l}(\mathbf{x}')n_k n_l) - \partial_i'(B_i(\mathbf{x})B_i(\mathbf{x}')(n_k n_l)_{,k}) - B_i(\mathbf{x})B_i(\mathbf{x}')(n_k n_l)_{,kl},$$
$$(n_k n_l)_{,kl} = 2\,|\mathbf{x}-\mathbf{x}'|^{-2}.$$

We see that, using nonradial kernels, we can express the state functionals in a number of equivalent ways. For example, the electromagnetic ground state functional can be written as

$$\Psi = \mathcal{N} \exp\left(-(8\pi^2)^{-1}\int d^3x \int d^3x' \times [\operatorname{curl}\mathbf{B}(\mathbf{x})\cdot\mathbf{n}][\operatorname{curl}\mathbf{B}(\mathbf{x}')\cdot\mathbf{n}]\right),$$

and also in several other forms.

3. THE EINSTEIN-SCHRÖDINGER EQUATION AND ITS LINEAR APPROXIMATION

Initial Value Equations and their Quantum Counterparts

In the general theory of relativity, the dynamics of the gravitational field is hidden in the initial value equations

$$\mathcal{H}^i \equiv -2\pi^{ik}{}_{|k} = -2\pi^{ik}{}_{,k} - g^{il}(2g_{lm,k} - g_{km,l})\pi^{km} = 0, \quad (4)$$

$$\mathcal{H} \equiv g^{-\frac{1}{2}}(g_{il}g_{km} - \tfrac{1}{2}g_{ik}g_{lm})\pi^{ik}\pi^{lm} - g^{\frac{1}{2}}R = 0. \quad (5)$$

The momenta π^{ik} are determined by the extrinsic curvature tensor K_{ik} of a 3-dimensional spacelike hypersurface on which the initial data are given:

$$\pi^{ik} = -g^{\frac{1}{2}}(K^{ik} - Kg^{ik}). \quad (6)$$

The extrinsic curvature K_{ik} depends on the metric ${}^4g_{\iota\kappa}$ of the surrounding space–time through the lapse function N and shift functions N_i:

$$K_{il} = \tfrac{1}{2}N^{-1}(-g_{ik,0} + N_{i|k} + N_{k|i}), \quad (7)$$

$$N_i = {}^4g_{0i}, \quad N = (-{}^4g^{00})^{\frac{1}{2}}. \quad (8)$$

If the initial value equations (4) and (5) are satisfied on all possible spacelike sections through the space–time, it is ensured that the gravitational field satisfies all the other Einstein vacuum equations.

In quantum mechanics, the initial value equations (4) and (5) become constraints imposed on the state functional Ψ:

$$\mathcal{H}^i\Psi = 0, \tag{9}$$

$$\mathcal{H}\Psi = 0. \tag{10}$$

As in the classical theory, these constraints contain implicitly all information about the time evolution of state. This general approach was suggested by Dirac[12] and extensively studied by DeWitt.[8]

In the metric representation, the functional Ψ is chosen as a functional of the components g_{ik} of the metric tensor, and the momenta are replaced by the operators

$$\pi^{ik}(\mathbf{x}) = -i\frac{\delta}{\delta g_{ik}(\mathbf{x})}. \tag{11}$$

The constraint (9) then means that the state functional does not change under a transformation of coordinates x^i, so that Ψ depends only on the intrinsic geometry $\mathcal{G}$ of the 3-space. Equation (10) is then the only remaining condition that the state functional must satisfy. Because of its resemblance to the Schrödinger equation of ordinary quantum theory, Wheeler called it the Einstein–Schrödinger equation. This equation tells us how the state functional evolves on the set of all possible 3-dimensional geometries—i.e., on superspace.

How Do We Linearize the Quantum Constraints?

The state functional of the gravitational vacuum (20G) that we obtained in the last section was expressed as a function of the linearized metric tensor, via the linearized Ricci tensor, so that it was not affected by gauge transformations $h_{ik} \to h_{ik} + \xi_{i,k} + \xi_{k,i}$, induced by linearized transformations of coordinates. It seems, therefore, easy to interpret this functional as a state of the gravitational field on superspace and to show that it approximately satisfies the Einstein–Schrödinger equation (10).

One should find an approximation to the Einstein–Schrödinger equation corresponding to the linearization of the classical Einstein equations. This task is seemingly quite straightforward. It is tempting to define the "linearized superspace" as the set of all 3-geometries that differ only by a small amount from the Euclidean geometry, so that in a properly chosen coordinate system

$$g_{ik}(\mathbf{x}) = \delta_{ik} + h_{ik}(\mathbf{x}), \quad |h_{ik}(\mathbf{x})| \ll 1.$$

If we define the state functionals on the "linearized superspace" and require that they are unaffected by linearized transformations of coordinates, we find out that these functionals satisfy the "linearized" form of Eq. (9), viz.,

$$\left(\frac{\delta\Psi}{\delta h_{ik}}\right)_{,k} = 0. \tag{12}$$

It remains to decide how to linearize the Einstein–Schrödinger equation. Leutwyler[13] proposed to keep the lowest-order terms in the "kinetic" and "potential" parts of this equation separately and to write

$$\left((\delta_{il}\delta_{km} - \tfrac{1}{2}\delta_{ik}\delta_{lm})\frac{\delta^2}{\delta h_{ik}\delta h_{lm}} + h_{ii,kk} - h_{ik,ik}\right)\Psi = 0. \tag{13}$$

Unfortunately, this simple procedure does not reproduce the quantized form of the linearized theory of gravitation. However, before modifying the procedure, we find it instructive to analyze the reasons for its failure. This will be done in detail in the following four subsections.

Linear Approximation Inadequate to Handle Thin Sandwich Theorem

The metric representation of the state functional is based on the assumption that the components of the metric tensor can be taken as independent variables. This assumption has been formulated in the classical theory as the "thin sandwich theorem," according to which an arbitrary initial metric g_{ik}, together with its arbitrary rate of change $g_{ik,0}$, uniquely determine (under appropriate boundary conditions) the space–time in which the hypersurface carrying the initial geometry is embedded. The initial value equations (4) and (5) are interpreted, by means of (6), (7), and (8), as the equations for the lapse function N and the shift functions N_i. Supposing that the thin sandwich theorem is correct, we see that N and N_i are determined uniquely by these equations and, together with other data g_{ik} and $g_{ik,0}$, they fix the initial momentum π^{ik}. The lapse function characterizes the proper time that elapses between the neighboring hypersurfaces $x^0 = \text{const}$ and $x^0 + dx^0 = \text{const}$. The intrinsic geometry $\mathcal{G} \equiv \{g_{ik}\}$ therefore carries information about time[14]; roughly speaking, from the three independent components of the metric tensor remaining after dividing the 3-dimensional metrics into equivalence classes by the group of coordinate transformations, one component represents an intrinsic time, and the other two represent the proper dynamical degrees of freedom of the gravitational field.

These features of geometrodynamics in the general case are obscured by the usual linearization of Einstein's equations. There the quantities g_{ik}, π^{ik}, N_i, and N are expanded into power series in a small

parameter λ:

$$\begin{aligned} g_{ik} &= \delta_{ik} + \lambda h^{(1)}{}_{ik} + \lambda^2 h^{(2)}{}_{ik} + \cdots, \\ \pi^{ik} &= \lambda \pi^{(1)ik} + \lambda^2 \pi^{(2)ik} + \cdots, \\ N_i &= \lambda N^{(1)}{}_i + \lambda^2 N^{(2)}{}_i + \cdots, \\ N &= 1 + \lambda N^{(1)} + \lambda^2 N^{(2)} + \cdots. \end{aligned} \tag{14}$$

Substituting these expansions into the initial value equations (4) and (5) and collecting the terms with the same power of λ, we get the initial value equations in successive orders of approximation. Specifically, to the first order in λ, we get

$$\pi^{(1)ik}{}_{,k} = 0, \tag{15}$$

$$R^{(1)} = -h^{(1)}{}_{ii,kk} + h^{(1)}{}_{ik,ik} = 0. \tag{16}$$

However, Eq. (16) tells us that the initial linearized geometry $h^{(1)}{}_{ik}$ cannot be freely specified. In fact, not only $h^{(1)}{}_{ik}$, but also $h^{(1)}{}_{ik,0}$ is limited. Using the definitions (6), (7), and (8), we substitute the power expansions (14) into Eq. (15) and get

$$\pi^{(1)ik}{}_{,k} \equiv -\tfrac{1}{2}(h^{(1)}{}_{kk,0i} - h^{(1)}{}_{ik,0k}) - \tfrac{1}{2}(N^{(1)}{}_{i,k} - N^{(1)}{}_{k,i})_{,k} = 0.$$

Differentiating this equation with respect to x^i, we see that the term containing the shift functions $N^{(1)}{}_i$ drops out, and the rate of change of the initial geometry $h^{(1)}{}_{ik,0}$ is subject to the same restriction as the geometry itself, viz.,

$$-h^{(1)}{}_{ii,0kk} + h^{(1)}{}_{ik,0ik} = 0.$$

Moreover, the lapse function $N^{(1)}$ does not enter into the first-order approximation equations at all and is therefore left completely undetermined by them. We conclude that in the linear approximation the geometry is neither freely prescribed nor does it carry enough information to determine the proper time interval between neighboring spacelike hypersurfaces.

Extrinsic Curvature More Sensitive to Small Deformations of Spacelike Hypersurface than Intrinsic Geometry

The plausibility of this conclusion can be seen from yet another point of view. Let us study entirely flat space–time in which the Cartesian system of coordinates (x^i, t) is introduced. We pick out a space hypersurface by giving t as a function of x^i,

$$t = t(x^i), \tag{17}$$

and use x^i also as intrinsic coordinates on this hypersurface. The intrinsic geometry g_{ik} and the extrinsic curvature K_{ik} of the hypersurface are then given by

$$g_{ik} = \delta_{ik} - t_{,i}t_{,k}, \tag{18}$$

$$K_{ik} = t_{,ik}(1 - t_{,l}t_{,l})^{-\frac{1}{2}}. \tag{19}$$

The function $t(x^i)$ represents a generalized time parameter, characterizing a deformation of the hypersurface from the chosen hyperplane $t = 0$. We see that the intrinsic geometry of the hypersurface (17) is affected by this deformation only in the second order terms. Even if space geometry is a carrier of information about time, as the title of the Baierlein, Sharp, and Wheeler paper[14] suggests, we must conclude that it is a poor carrier of information about time in the immediate vicinity of flat space–time. Unfortunately, it is just the immediate vicinity of flat space–time in which we are interested in the linearized theory of gravitation.

On the other hand, the extrinsic curvature (19) is influenced by a small deformation $t(x^i)$ from the hyperplane already in the terms linear in $t(x^i)$. It is therefore easier to determine $t(x^i)$ by looking at the extrinsic curvature of a hypersurface, rather than at its intrinsic geometry. We shall use this insight later on, while adapting the representation to match the linear approximation.

The Two Independent Degrees of Freedom of Linearized Geometry

We have seen that the components of the linearized metric tensor are subject to the restriction (16). To exhibit the independent degrees of freedom explicitly, we can use the Arnowitt–Deser–Misner decomposition of a symmetric tensor f_{ik} into transversal traceless $f^{\mathrm{TT}}{}_{ik}$, transversal $f^{\mathrm{T}}{}_{ik}$, and longitudinal $f^{\mathrm{L}}{}_{ik}$ parts[11,12]:

$$f_{ik} = f^{\mathrm{TT}}{}_{ik} + f^{\mathrm{T}}{}_{ik} + f^{\mathrm{L}}{}_{ik}, \tag{20}$$

$$f^{\mathrm{TT}}{}_{ik,k} = f^{\mathrm{TT}}{}_{ii} = 0, \quad f^{\mathrm{T}}{}_{ik,k} = 0, \tag{21}$$

$$f^{\mathrm{T}}{}_{ik} = \tfrac{1}{2}(f^{\mathrm{T}}\delta_{ik} - \Delta^{-1}f^{\mathrm{T}}{}_{,ik}), \tag{22}$$

$$f^{\mathrm{L}}{}_{ik} = f_{i,k} + f_{k,i}, \tag{23}$$

$$f^{\mathrm{T}} = f_{ii} - \Delta^{-1}f_{ik,ik}, \tag{24}$$

$$f_i = \Delta^{-1}(f_{ik,k} - \tfrac{1}{2}\Delta^{-1}f_{kl,kli}), \tag{25}$$

$$f^{\mathrm{TT}}{}_{ik} = f_{ik} - f^{\mathrm{T}}{}_{ik}(f_{lm}) - f^{\mathrm{L}}{}_{ik}(f_{lm}). \tag{26}$$

The restriction (16) then means (under the usual boundary conditions at spatial infinity) that $h^{(1)\mathrm{T}} = 0$. It "freezes out" the freedom one has in the general case to move the spacelike hypersurface forward by one amount at one place and by another amount at another place. Thus linearization effectively takes away one of the "three degrees of freedom per space

point" in the choice of $\mathfrak{G}$. What are then the two remaining independent degrees of freedom? They are not contained in the three functions $h^{(1)}{}_i$. These describe merely the freedom in the choice of space coordinates and can be annulled by a gauge transformation $h^{(1)}{}_{ik} \to h^{(1)}{}_{ik} + \xi^{(1)}{}_{i,k} + \xi^{(1)}{}_{k,i}$. What is left are the transverse traceless components $h^{(1)\mathrm{TT}}{}_{ik}$. There are only two free functions in a transverse traceless tensor. They represent exactly the two independent dynamical degrees of freedom per space point that we have in the choice of $\mathfrak{G}$.

Why Does the Straightforward Linearization of the Einstein–Schrödinger Equation Fail?

Let us summarize the reasons why the linearization of the Einstein–Schrödinger equation proposed by Leutwyler does not lead us to the quantized form of the linearized theory of gravitation. We shall then be in a position to propose the necessary modifications.

We have seen that in the classical geometrodynamics the linearized geometry cannot be freely prescribed and the linearized geometry limited by the condition $h^{\mathrm{T}} = 0$ cannot be used to characterize time. Therefore, we can hardly expect that in the quantum geometrodynamics the state functional $\Psi(h_{ik})$ ranges over a set of freely specifiable linearized 3-geometries and describes thereby how the probability amplitude of the gravitational field changes with time. This undermines the basic assumption that led us to Eq. (12).

Not even Eq. (13) is consistent from the point of view of the usual linearization procedure, because it leaves quadratic terms in the momenta, but neglects quadratic terms in the first derivatives of the metric tensor. We know, however, that these quadratic terms enter the Hamiltonian (14G) of the linearized theory. It therefore appears inescapable to take the equations (9) and (10) at least to the quadratic terms, if we want to get from them the quantized form of the linearized theory of gravitation.

Returning to Leutwyler's linearization, we can find the expansion method on which it is based. Comparing the linearized quantum constraints (12) and (13) with the initial value equations (4) and (5), we are led to the expansions

$$g_{ik} = \delta_{ik} + \lambda^2 h^{(2)}{}_{ik} + \lambda^4 h^{(4)}{}_{ik} + \cdots,$$
$$\pi^{ik} = \lambda \pi^{(1)ik} + \lambda^3 \pi^{(3)ik} + \cdots.$$

They enable us to treat Eq. (4) in the first order and, at the same time, Eq. (5) in the second order of a small parameter λ, while the quadratic combination of momenta $\pi^{(1)ik}$ is just balanced by the linear combination of coordinates $h^{(2)}{}_{ik}$. This procedure is, of course, different from the conventional linear approximation method, and we cannot expect it will simply repeat its results.

The Extrinsic Time Representation

When we are forced to treat the quantum constraints (9) and (10) up to the quadratic terms, the metric representation is not really advantageous. The state functionals of the linearized theory are defined on hyperplanes $t = \text{const}$ with respect to the flat space–time background. We have seen that the intrinsic geometry of these hyperplanes is insensitive to their small deformations. It follows that we must take the "square root" of the deformed metric (18) to recover the time $t(\mathbf{x})$. This is possible in principle, but inconvenient in practice. On the other hand, the extrinsic curvature (19) is affected linearly by a small deformation $t(\mathbf{x})$. The time $t(\mathbf{x})$ can, therefore, more easily be extracted from the momenta π^{ik} than from the coordinates g_{ik}. For small deformations $t(\mathbf{x})$ there is a straightforward way to do it, viz.,

$$t(\mathbf{x}) = -\tfrac{1}{2}\Delta^{-1}\pi^{\mathrm{T}}. \tag{27}$$

This is just an inverse formula to (19), as we can check using the definitions (6) of the momenta and the prescription (24) which tells how to find a transversal part of a tensor. However, whereas to a given $t(\mathbf{x})$ there corresponds the unique $\pi^{\mathrm{T}}(\mathbf{x})$, to a given $\pi^{\mathrm{T}}(\mathbf{x})$ there corresponds a whole family of $t(\mathbf{x})$'s. Indeed, we can add an arbitrary linear combination $(t_0 - \mathbf{v}\cdot\mathbf{x})(1 - v^2)^{-\frac{1}{2}}$ of x^i to the right-hand side of (27) and still have a solution of (19). The hypersurfaces $t(\mathbf{x})$ of this family differ from each other by a constant time translation and/or Lorentz transformation at spatial infinity. This corresponds to different boundary conditions for the Δ^{-1} operator.

The basic idea expressed by Eq. (27) is that time can be better treated as a momentum variable π^{T} than as a superspace coordinate. It is a representation by $h^{\mathrm{TT}}{}_{ik}(\mathbf{x})$, $h_i(\mathbf{x})$, $t(\mathbf{x}) = -\tfrac{1}{2}\Delta^{-1}\pi^{\mathrm{T}}$, rather than the metric representation $h_{ik}(\mathbf{x})$ or $h^{\mathrm{TT}}{}_{ik}(\mathbf{x})$, $h_i(\mathbf{x})$, $h^{\mathrm{T}}(\mathbf{x})$, that makes the interpretation of linearized quantum theory of gravitation simple. In the metric representation, time is implicitly characterized by the intrinsic geometry of a spacelike hypersurface. In the new representation, it is reconstructed from the extrinsic curvature. Therefore, wanting a descriptive name for the new representation, we shall call it the "extrinsic time representation."

Just as π^{T} characterizes the deformation $t(\mathbf{x})$ of a spacelike hypersurface from a hyperplane on the background of flat space–time, $h_i(\mathbf{x})$ characterizes a deformation of a system of space coordinates on this

hypersurface from a Cartesian system of coordinates. Deforming our system of coordinates from the Cartesian system, $x^i \to x^i + \xi^i$, we get in the first order the metric tensor, the longitudinal part of which gives directly the deformation $h_i = \xi_i$.

The meaning of the quantities π^{T} and h_i as parameters specifying the choice of space–time coordinates was discovered by Arnowitt, Deser, and Misner.[11] In an effort to exclude the surplus variables from the initial value equations (4) and (5) and thereby to identify the proper dynamical degrees of freedom of the gravitational field, they fixed π^{T} and h_i by coordinate conditions

$$t = -\tfrac{1}{2}\Delta^{-1}\pi^{\mathrm{T}}, \quad h_i = 0, \tag{28}$$

that selected (at least for weak fields) from all systems of coordinates a system that was as close as possible to a Cartesian system of coordinates in the flat space–time.

We took a slightly different attitude in this paper, corresponding to the philosophy that the basic variables of representation have a twofold task: to describe the changing gravitational field and simultaneously the time in which it changes. We shall therefore leave $t(\mathbf{x}) = -\tfrac{1}{2}\Delta^{-1}\pi^{\mathrm{T}}$ and $h_i(\mathbf{x})$ undetermined as arbitrary functions of $\mathbf{x}$. They specify the choice of hypersurface and the choice of coordinates on this hypersurface as deformations from the privileged system of coordinates in which conditions (28) hold. Intuitively, the four quantities $t(\mathbf{x})$ and $h_i(\mathbf{x})$ represent a "many-fingered time." This particular mixture of Dirac's approach to quantization with the Arnowitt–Deser–Misner choice of variables does not appear to have been treated in the literature, though it has a number of interesting features.

Successive Approximations to Initial Value Equations in the Arnowitt–Deser–Misner Formalism

The general method of successive approximations to the initial value equations (4) and (5) based on the power series expansion (14) was already mentioned. It assumes a very convenient form in the Arnowitt–Deser–Misner formalism. In the first approximation, we choose $\pi^{(1)\mathrm{TT}ik}$, $\pi^{(1)\mathrm{T}}$ and $h^{(1)\mathrm{TT}}{}_{ik}$, $h^{(1)}{}_i$ (i.e., $2 + 1 + 2 + 3 = 8$ numbers per space point) freely. Equation (4) then determines $\pi^{(1)i} = 0$ (3 numbers per space point), and Eq. (5) determines $h^{(1)\mathrm{T}} = 0$ (1 number per space point). Substituting these values to the second approximation, we get

$$\mathcal{H}^{(2)i} = p^{(2)i} + h^{(1)\mathrm{TT}}{}_{rs,i}\pi^{(1)\mathrm{TT}rs} + H_{ik,k} = 0, \tag{29}$$

$$\mathcal{H}^{(2)} = -E^{(2)} + \pi^{(1)\mathrm{TT}ik}\pi^{(1)\mathrm{TT}ik} + \tfrac{1}{4}h^{(1)\mathrm{TT}}{}_{ik,l}h^{(1)\mathrm{TT}}{}_{ik,l} + H_{k,k} = 0. \tag{30}$$

The quantities

$$p^i(\mathbf{x}) = -2(\Delta\pi^i + \pi^k{}_{,ki})$$

and

$$E(\mathbf{x}) = -\Delta h^{\mathrm{T}}$$

are canonically conjugate to $h_i(\mathbf{x})$ and $-t(\mathbf{x})$. They can be interpreted as effective momentum density and effective energy density of the gravitational field, respectively. The expressions H_{ik} and H_i are quadratic functions of $\pi^{(1)\mathrm{TT}ik}$, $\pi^{(1)\mathrm{T}}$, $h^{(1)\mathrm{TT}}{}_{ik}$, and $h^{(1)}{}_i$, the form of which we need not know in the following. They are long and complicated and therefore will not be written out here. We see that we can again freely prescribe the eight second-order quantities $\pi^{(2)\mathrm{TT}ik}$, $\pi^{(2)\mathrm{T}}$, and $h^{(2)\mathrm{TT}}{}_{ik}$, $h^{(2)}{}_i$, and determine thereby the three second-order quantities $\pi^{(2)i}$ and one second-order quantity $h^{(2)\mathrm{T}}$ through Eqs. (29) and (30). In general, $\pi^{\mathrm{TT}ik}$, π^{T}, and $h^{\mathrm{TT}}{}_{ik}$, h_i can be prescribed freely at any stage of the approximation, and the initial value equations then determine π^i and h^{T}.

Linearized Quantum Constraints in the Extrinsic Time Representation

In the extrinsic time representation, the quantities canonically conjugate to $h^{\mathrm{TT}}{}_{ik}$, h_i, and t are replaced by the operators

$$\pi^{\mathrm{TT}ik}(\mathbf{x}) = -i\frac{\delta}{\delta h^{\mathrm{TT}}{}_{ik}(\mathbf{x})}, \quad p^i(\mathbf{x}) = -i\frac{\delta}{\delta h_i(\mathbf{x})},$$

$$E(\mathbf{x}) = i\frac{\delta}{\delta t(\mathbf{x})}. \tag{31}$$

The constraints (9) and (10) then determine how the state functional $\Psi(h^{\mathrm{TT}}{}_{ik}, h_i, t)$ evolves in the variables h_i and t. Specifically, if we prescribe $\Psi(h^{\mathrm{TT}}{}_{ik}, 0, 0) = \Psi_0(h^{\mathrm{TT}}{}_{ik})$, the state functional $\Psi(h^{\mathrm{TT}}{}_{ik}, h_i, t)$ is given for an arbitrary $h_i(\mathbf{x})$ and $t(\mathbf{x})$ by means of (9) and (10). We can return from the extrinsic time representation to the metric representation by a formal Fourier transformation in the function variable $t(\mathbf{x})$:

$$\Psi(h^{\mathrm{TT}}{}_{ik}, h_i, h^{\mathrm{T}}) = \int Dt\Psi(h^{\mathrm{TT}}{}_{ik}, h_i, t)\exp\left(i\int d^3x\, t(\mathbf{x})\Delta h^{\mathrm{T}}(\mathbf{x})\right).$$

A perturbation method of solving Eqs. (9) and (10) corresponds to the approximation method of solving the classical initial value equations. If we replace the super-Hamiltonian $\mathcal{H}$ and supermomentum $\mathcal{H}^i$ by perturbed operators,

$$\mathcal{H}(\lambda) \equiv \mathcal{H}(\lambda h^{\mathrm{TT}}{}_{ik}, \lambda h_i, \lambda^2 h^{\mathrm{T}}, \lambda\pi^{\mathrm{TT}ik}, \lambda^2\pi^i, \lambda\pi^{\mathrm{T}}) = \lambda^2\mathcal{H}^{(0)} + \lambda^3\mathcal{H}^{(1)} + \cdots,$$

$$\mathcal{H}^i(\lambda) \equiv \mathcal{H}^i(\lambda h^{\mathrm{TT}}{}_{ik}, \lambda h_i, \lambda^2 h^{\mathrm{T}}, \lambda\pi^{\mathrm{TT}ik}, \lambda^2\pi^i, \lambda\pi^{\mathrm{T}}) = \lambda^2\mathcal{H}^{(0)i} + \lambda^3\mathcal{H}^{(1)i} + \cdots,$$

and assume that the state functional is expanded in a power series in the small parameter λ,

$$\Psi = \Psi^{(0)} + \lambda\Psi^{(1)} + \lambda^2\Psi^{(2)} + \cdots,$$

we get $\Psi^{(0)}$ as a solution of the equations

$$\mathcal{H}^{(0)}\Psi^{(0)} = 0, \quad \mathcal{H}^{(0)i}\Psi^{(0)} = 0. \tag{32}$$

We get the operators $\mathcal{H}^{(0)}$ and $\mathcal{H}^{(0)i}$ if we cease to distinguish the first- and second-order quantities in Eqs. (29) and (30), and substitute there the operators (31) for $\pi^{\mathrm{TT}ik}$, p^i, and E. Equations (32) then assume the form

$$i\frac{\delta\Psi^{(0)}}{\delta t(\mathbf{x})} = -\frac{\delta^2\Psi^{(0)}}{\delta h^{\mathrm{TT}}{}_{ik}(\mathbf{x})\delta h^{\mathrm{TT}}{}_{ik}(\mathbf{x})} + \tfrac{1}{4}h^{\mathrm{TT}}{}_{ik,l}(\mathbf{x})h^{\mathrm{TT}}{}_{ik,l}(\mathbf{x})\Psi^{(0)} + H_{k,k}\Psi^{(0)}, \tag{33}$$

$$i\frac{\delta\Psi^{(0)}}{\delta h_i(\mathbf{x})} = -ih^{\mathrm{TT}}{}_{rs,i}(\mathbf{x})\frac{\delta\Psi^{(0)}}{\delta h^{\mathrm{TT}}{}_{rs}(\mathbf{x})} + H_{ik,k}\Psi^{(0)}. \tag{34}$$

These equations resemble the Schrödinger equation even more closely then the Einstein–Schrödinger equation in the metric representation (which is, in fact, an equation of the Klein–Gordon type). The differentiation with respect to a single time parameter is, of course, replaced by functional differentiations with respect to $t(\mathbf{x})$ and $h_i(\mathbf{x})$, as is appropriate in a many-fingered time formalism. However, as we shall see in the next section, we can further specialize the variations of the spacelike hypersurface and pass into the single time formalism. In this process, the complicated operator terms $H_{k,k}$ and $H_{ik,k}$, being perfect divergences, will allow themselves to be eliminated.

Writing down the functional differential equations (33) and (34), we have reached the final goal of this section. These equations are the natural starting point for treating weak gravitational fields within the Dirac formalism. We shall illustrate the analysis by applying these equations to the particular problem of gravitational vacuum in the next section.

4. GROUND STATE FUNCTIONAL IN THE EXTRINSIC TIME REPRESENTATION

Time Translations

Equations (33) and (34) determine the functional $\Psi^{(0)}$ for every $h_i(\mathbf{x})$ and $t(\mathbf{x})$, if we know its value $\Psi_0^{(0)}$ for $h_i(\mathbf{x}) = 0$ and $t(\mathbf{x}) = 0$, i.e., in a privileged system of coordinates on a privileged initial spacelike hypersurface $t(\mathbf{x}) = 0$, $\pi^{\mathrm{T}}(\mathbf{x}) = 0$. However, we can also ask how $\Psi^{(0)}$ changes if we pass from one privileged hypersurface $\pi^{\mathrm{T}}(\mathbf{x}) = 0$, $h_i(\mathbf{x}) = 0$, corresponding to $t(\mathbf{x}) = t_0 = \text{const}$, to another privileged hypersurface $\pi^{\mathrm{T}}(\mathbf{x}) = 0$, $h_i(\mathbf{x}) = 0$, corresponding to a slightly changed time parameter $t(\mathbf{x}) = t_0 + \delta t_0 = \text{const}$. The functional $\Psi^{(0)}$ depends in this case on one real parameter t_0 and

$$\frac{\partial\Psi^{(0)}}{\delta t_0}\delta t_0 = \int d^3x\,\frac{\delta\Psi^{(0)}}{\delta t(\mathbf{x})}\delta t_0. \tag{35}$$

In order that $\Psi^{(0)}$ describe a stationary state with energy $E_0 = \text{const}$, we must have $i\partial\Psi^{(0)}/\partial t_0 = E_0\Psi^{(0)}$. Equations (33) and (35) then yield

$$\int d^3x\Big(-\frac{\delta^2\Psi^{(0)}}{\delta h^{\mathrm{TT}}{}_{ik}(\mathbf{x})\delta h^{\mathrm{TT}}{}_{ik}(\mathbf{x})} + \tfrac{1}{4}h^{\mathrm{TT}}{}_{ik,l}(\mathbf{x})h^{\mathrm{TT}}{}_{ik,l}(\mathbf{x})\Psi^{(0)}\Big) = E_0\Psi^{(0)}. \tag{36}$$

The terms containing the operator $H_{k,k}$ drop out because they can be transformed into an integral over a 2-dimensional surface at spatial infinity, which vanishes because of the boundary conditions.

We can easily verify that the ground state functional (19G) satisfies Eq. (36). We get

$$\frac{\delta\Psi^{(0)}}{\delta h^{\mathrm{TT}}{}_{ik}(\mathbf{x})} = \frac{1}{4\pi^2}\int d^3x'\,\frac{h^{\mathrm{TT}}{}_{ik,ll}(\mathbf{x}')}{|\mathbf{x}-\mathbf{x}'|^2}\Psi^{(0)}, \tag{37}$$

$$-\frac{\delta^2\Psi^{(0)}}{\delta h^{\mathrm{TT}}{}_{ik}(\mathbf{x})\delta h^{\mathrm{TT}}{}_{ik}(\mathbf{x})} = \frac{1}{4\pi^2}\Big[\Delta\frac{1}{|\mathbf{x}-\mathbf{x}'|^2}\Big]_{\mathbf{x}=\mathbf{x}'}\Psi^{(0)} - \frac{1}{(4\pi^2)^2}\int d^3x'\int d^3x''\,\frac{\Delta' h^{\mathrm{TT}}{}_{ik}(\mathbf{x}')\Delta'' h^{\mathrm{TT}}{}_{ik}(\mathbf{x}'')}{|\mathbf{x}-\mathbf{x}'|^2\,|\mathbf{x}-\mathbf{x}''|^2}\Psi^{(0)}. \tag{38}$$

The first term on the right-hand side of Eq. (38) diverges. We see from (36) that it represents the infinite energy density of the gravitational vacuum. To get rid of it, we should introduce a corresponding renormalization counterterm into the Einstein–Schrödinger equation (10). If we use the relations

$$\int d^3x\,|\mathbf{x}-\mathbf{x}'|^{-2}\,|\mathbf{x}-\mathbf{x}''|^{-2} = \pi^3\,|\mathbf{x}'-\mathbf{x}''|^{-1},$$

$$\Delta'\,|\mathbf{x}'-\mathbf{x}''|^{-1} = -4\pi\delta(\mathbf{x}'-\mathbf{x}'')$$

and integrate the second term on the right-hand side of Eq. (38) with respect to $\mathbf{x}$, as required by (36), we get

$$-\frac{1}{(4\pi^2)^2}\int d^3x\int d^3x'\int d^3x''\,\frac{\Delta' h^{\mathrm{TT}}{}_{ik}(\mathbf{x}')\Delta'' h^{\mathrm{TT}}{}_{ik}(\mathbf{x}'')}{|\mathbf{x}-\mathbf{x}'|^2\,|\mathbf{x}-\mathbf{x}''|^2} = -\frac{1}{16\pi}\int d^3x'\int d^3x''\,\frac{\Delta' h^{\mathrm{TT}}{}_{ik}(\mathbf{x}')\Delta'' h^{\mathrm{TT}}{}_{ik}(\mathbf{x}'')}{|\mathbf{x}'-\mathbf{x}''|} = \frac{1}{16\pi}\int d^3x'\int d^3x''\,h^{\mathrm{TT}}{}_{ik,l}(\mathbf{x}')h^{\mathrm{TT}}{}_{ik,l}(\mathbf{x}'')\Delta'\frac{1}{|\mathbf{x}'-\mathbf{x}''|} = -\tfrac{1}{4}\int d^3x'\,h^{\mathrm{TT}}{}_{ik,l}(\mathbf{x}')h^{\mathrm{TT}}{}_{ik,l}(\mathbf{x}').$$

This term just cancels the second term on the left-hand side of Eq. (36), so that Eq. (36) is satisfied.

Spatial Translations

Equation (36) follows from Eq. (33) when we restrict ourselves to constant translations in time. A similar procedure can be used for Eq. (34), when we restrict ourselves to constant translations in space. The ground state functional should be invariant under such translations, the vacuum being the state of zero momentum. This means that

$$\int d^3x \frac{\delta\Psi^{(0)}}{\delta h_i(\mathbf{x})} = -\int d^3x\, h^{\mathrm{TT}}{}_{rs,i}(\mathbf{x}) \frac{\delta\Psi^{(0)}}{\delta h^{\mathrm{TT}}{}_{rs}(\mathbf{x})} = 0. \quad (39)$$

Let us check that the functional (19G) satisfies this equation as well as Eq. (36). Using (37), we see that we must verify

$$\int d^3x \int d^3x' \frac{h^{\mathrm{TT}}{}_{rs,ll}(\mathbf{x}')h^{\mathrm{TT}}{}_{rs,i}(\mathbf{x})}{|\mathbf{x}-\mathbf{x}'|^2} = 0.$$

The last integral can be transformed by a sequence of integrations by parts to the form

$$\int d^3x \int d^3x'\, h^{\mathrm{TT}}{}_{rs,l}(\mathbf{x}')h^{\mathrm{TT}}{}_{rs,l}(\mathbf{x})\partial_i |\mathbf{x}-\mathbf{x}'|^{-2}.$$

Because $h^{\mathrm{TT}}{}_{rs,l}(\mathbf{x}')h^{\mathrm{TT}}{}_{rs,l}(\mathbf{x})$ is symmetric under the interchange $\mathbf{x} \leftrightarrow \mathbf{x}'$ whereas $\partial_i |\mathbf{x}-\mathbf{x}'|^{-2}$ is antisymmetric, this integral vanishes as required.

How to Get off the Hypersurface of Constant $t(\mathbf{x})$

As already remarked, Eqs. (33) and (34) determine $\Psi^{(0)}(h^{\mathrm{TT}}{}_{ik}, h_i, t)$, if $\Psi^{(0)}(h^{\mathrm{TT}}{}_{ik}, 0, 0)$ is known. We expand the functional in a power series in the function variables $h_i(\mathbf{x})$ and $t(\mathbf{x})$:

$$\begin{aligned}\Psi^{(0)} = \Psi_0^{(0)}(h^{\mathrm{TT}}{}_{ik}) &+ \int d^3x\, a_i(\mathbf{x})h_i(\mathbf{x}) + \int d^3x\, a(\mathbf{x})t(\mathbf{x}) \\ &+ \int d^3x \int d^3x'\, b_{ik}(\mathbf{x},\mathbf{x}')h_i(\mathbf{x})h_i(\mathbf{x}') \\ &+ \int d^3x \int d^3x'\, b_i(\mathbf{x},\mathbf{x}')(h_i(\mathbf{x})t(\mathbf{x}') + h_i(\mathbf{x}')t(\mathbf{x})) \\ &+ \int d^3x \int d^3x'\, b(\mathbf{x},\mathbf{x}')t(\mathbf{x})t(\mathbf{x}') + \cdots. \quad (40)\end{aligned}$$

The coefficients of this power series can be determined by repeated variations of Eqs. (33) and (34) with respect to $t(\mathbf{x})$ and $h_i(\mathbf{x})$ in the point $t(\mathbf{x}) = h_i(\mathbf{x}) = 0$. For example,

$$a(\mathbf{x}) = \left[\frac{\delta\Psi^{(0)}}{\delta t(\mathbf{x})}\right]_{t(\mathbf{x})=h_i(\mathbf{x})=0} \quad (41)$$

can be found by substituting $t(\mathbf{x}) = h_i(\mathbf{x}) = 0$ into the operator $H_{k,k}$ on the right-hand side of Eq. (33) and taking there $\Psi_0^{(0)}$ instead of $\Psi^{(0)}$. However, it is necessary to take into account that Eqs. (33) and (34) were obtained when we neglect the cubic and higher-order terms in the exact super-Hamiltonian and supermomentum. Further terms $\Psi^{(1)}, \Psi^{(2)}, \cdots$ of the perturbation expansion of the state functional modify, therefore, those terms of the power series expansion of $\Psi^{(0)}$ that depend on cubic and higher-order combinations of the variables $h^{\mathrm{TT}}{}_{ik}$, h_i, and π^{T}. It is therefore illusory to retain terms in the power series expansion of $\Psi^{(0)}$ of higher order than those we have explicitly written down in (40).

The first two integrals in expansion (40) indicate what happens if we slightly deform the original hypersurface and the system of coordinates on this hypersurface and observe the change in the state functional, linear in deformation. Determining the coefficient $a(\mathbf{x})$ as in (41), we get

$$a(\mathbf{x}) = -\frac{i}{4\pi^2}\left(\Delta \frac{1}{|\mathbf{x}-\mathbf{x}'|^2}\right)_{\mathbf{x}=\mathbf{x}'} + \text{second order terms in } h^{\mathrm{TT}}{}_{ik}.$$

The first term drops out when we renormalize the energy, so that the integral $\int d^3x\, a(\mathbf{x})t(\mathbf{x})$ is cubic in the variables $h^{\mathrm{TT}}{}_{ik}(\mathbf{x})$ and $t(\mathbf{x})$ and may be neglected. The same conclusion is reached about the term $\int d^3x\, a_i(\mathbf{x})h_i(\mathbf{x})$ if we use (34). A deformation of the hypersurface therefore leaves the ground state functional practically unchanged in the terms linear in the deformation parameters $t(\mathbf{x})$ and $h_i(\mathbf{x})$. This is just what can be expected of the state with zero energy.

ACKNOWLEDGMENTS

The author would like to thank Professor J. A. Wheeler for constant encouragement and invaluable advice. His enlightening suggestions helped to clarify the problems and paved the way for their solution. Useful comments from J. R. Klauder, C. W. Misner, and J. W. York are gratefully acknowledged. The last one also painfully corrected the wildest excesses of my English.

APPENDIX

To prove the relation (3) in the text, we study the Fourier transform of

$$f_{ik}(\mathbf{x}) = n_i n_k e^{-\epsilon|\mathbf{x}|}, \quad n_i = x_i/|\mathbf{x}|, \quad \epsilon > 0:$$

$$\bar{f}_{ik}(\mathbf{k}) = (2\pi)^{-\frac{3}{2}} \int d^3x\, n_i n_k e^{-i\mathbf{k}\cdot\mathbf{x}-\epsilon|\mathbf{x}|}. \quad (A1)$$

Because $\bar{f}_{ik}(\mathbf{k})$ must be form invariant as a function of $\mathbf{k}$ under rotations of a Cartesian system of coordinates,

$\tilde{f}_{ik}(\mathbf{k})$ must be of the form

$$\tilde{f}_{ik}(\mathbf{k}) = A(k)\delta_{ik} + B(k)k_ik_k. \tag{A2}$$

To determine the scalar functions $A(k)$ and $B(k)$ of $k = |\mathbf{k}|$, we contract Eq. (A1) in i and k and multiply Eq. (A1) by the tensor k_ik_k/k^2. Because of (A2) we get

$$3A + Bk^2 = (2\pi)^{-\frac{3}{2}}\int d^3x\, e^{-i\mathbf{k}\cdot\mathbf{x}-\epsilon|\mathbf{x}|}, \tag{A3}$$

$$\begin{aligned} A + Bk^2 &= (2\pi)^{-\frac{3}{2}}\int d^3x(n_ik_i/k)e^{-i\mathbf{k}\cdot\mathbf{x}-\epsilon|\mathbf{x}|} \\ &= (2\pi)^{-\frac{3}{2}}\int_0^\infty dr\left(r^2e^{-\epsilon r}\int_0^\pi d\vartheta\cos^2\vartheta\sin\vartheta\right. \\ &\qquad \left.\times e^{-ikr\cos\vartheta}\int_0^{2\pi}d\varphi\right). \end{aligned} \tag{A4}$$

The integral on the right-hand side of (A3) is simply a δ function:

$$3A + Bk^2 \underset{\epsilon\to 0}{\longrightarrow} (2\pi)^{\frac{3}{2}}\delta(\mathbf{k}). \tag{A5}$$

On the right-hand side of (A4) we can easily integrate over the angles ϑ and φ and get

$$\begin{aligned} &A + Bk^2 \\ &= (2\pi)^{-\frac{1}{2}}\int_0^\infty dr\, r^2e^{-\epsilon r}\left(\frac{2\sin kr}{kr} + \frac{4\cos kr}{k^2r^2} - \frac{4\sin kr}{k^3r^3}\right) \\ &\equiv \Lambda_1 + \Lambda_2 + \Lambda_3. \end{aligned}$$

The individual integrals over r give

$$\begin{aligned} \Lambda_1 &= (2\pi)^{\frac{1}{2}}\frac{1}{k^2+\epsilon^2}\frac{1}{\pi}\frac{\epsilon}{\epsilon^2+k^2} \\ &\underset{\epsilon\to 0}{\longrightarrow} (2\pi)^{\frac{1}{2}}k^{-2}\delta(k) \\ &= (2\pi)^{\frac{3}{2}}\delta(\mathbf{k}), \end{aligned}$$

$$\begin{aligned} \Lambda_2 &= 2(2\pi)^{\frac{1}{2}}k^{-2}\frac{1}{\pi}\frac{\epsilon}{\epsilon^2+k^2} \\ &\underset{\epsilon\to 0}{\longrightarrow} 2(2\pi)^{\frac{1}{2}}k^{-2}\delta(k) \\ &= 2(2\pi)^{\frac{3}{2}}\delta(\mathbf{k}), \end{aligned}$$

$$\begin{aligned} \Lambda_3 &= -4(2\pi)^{-\frac{1}{2}}k^{-3}\int_0^\infty d(kr)\frac{\sin kr}{kr}e^{-\epsilon r} \\ &\underset{\epsilon\to 0}{\longrightarrow} -(2\pi)^{\frac{1}{2}}k^{-3}, \end{aligned}$$

so that

$$A + Bk^2 \underset{\epsilon\to 0}{\longrightarrow} 3(2\pi)^{\frac{3}{2}}\delta(\mathbf{k}) - (2\pi)^{\frac{1}{2}}k^{-3}. \tag{A6}$$

From (A5) and (A6) we find the coefficients A and B and thus obtain the relation (3) in the main text.

* Partly supported by National Science Foundation Grant No. GP 7669.

† On leave of absence from Faculty of Mathematics and Physics, The Charles University, Prague, Czechoslovakia.

[1] J. A. Wheeler, Ann. Phys. (N.Y.) **2**, 604 (1957).
[2] J. A. Wheeler, *Geometrodynamics* (Academic, New York, 1962).
[3] J. A. Wheeler, "Geometrodynamics and the Issue of the Final State," in *Relativity, Groups and Topology*, C. DeWitt and B. DeWitt, Eds. (Gordon and Breach, New York, 1964).
[4] J. A. Wheeler, "Superspace and the Nature of Quantum Geometrodynamics," in *Battelle Rencontres*, C. DeWitt and J. A. Wheeler, Eds. (Benjamin, New York, 1968).
[5] V. Fock, Zh. Fiz. **6**, 425 (1934).
[6] K. Symanzig, Z. Naturforsch **9a**, 809 (1954).
[7] J. V. Novozilov and A. V. Tolub, Fortschr. Physik **6**, 50 (1958).
[8] B. DeWitt, Phys. Rev. **160**, 1113 (1967).
[9] See Ref. 2, p. 74.
[10] C. W. Misner, K. S. Thorne, and J. A. Wheeler, "An Open Letter to Relativity Theorists," August 19, 1968.
[11] R. Arnowitt, S. Deser, and C. W. Misner, "The Dynamics of General Relativity," in *Gravitation: An Introduction to Current Research*, L. Witten, Ed. (Wiley, New York, 1962). There are references to a series of papers on this problem by the authors. See also S. Deser, Ann. Inst. Henri Poincaré **A7**, 149 (1967).
[12] P. A. M. Dirac, Can. J. Math. **2**, 129 (1950).
[13] H. Leutwyler, "Boundary Conditions for the State Functional in Quantum Theory of Gravity," in Ref. 4.
[14] R. F. Baierlein, D. H. Sharp, and J. A. Wheeler, Phys. Rev. **126**, 1864 (1962).

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 11, NUMBER 12 DECEMBER 1970

Energy–Momentum Spectrum and Vacuum Expectation Values in Quantum Field Theory

JAMES GLIMM*
Courant Institute of Mathematical Sciences, New York University, New York, New York 10012
AND
ARTHUR JAFFE†
Lyman Laboratory of Physics, Harvard University, Cambridge, Massachusetts 02138

(Received 28 May 1970)

We consider nonlinear boson self-interactions with a periodic spatial cutoff. We prove that the energy–momentum spectrum lies in the forward light cone. A momentum cutoff does not influence this result. For theories with finite-field strength renormalization, we obtain bounds on the vacuum expectation values of products of the ϕ_t's and $\nabla\phi$'s. These bounds are uniform in the volume (and possible momentum) cutoff.

1. INTRODUCTION

In this paper we prove results about the spectrum and the vacuum expectation values for models with boson interactions. We use a periodic space cutoff and a translation invariant Hamiltonian, so that the energy and momentum operators commute. Our results are independent of the possible presence of a momentum cutoff, and rely on the fact that the ground state of the (cutoff) theory is nondegenerate. Our results also hold for models with fermions (such as the Yukawa coupling) once the uniqueness of the ground state (vacuum) is established. In the Yukawa coupling, for instance, the vacuum is known to be nondegenerate for small values of the coupling constant λ, $|\lambda| \leq \lambda_0$, but is is not known whether λ_0 approaches 0 as the volume of the periodic box tends to infinity.[1]

In Sec. 2, we prove that the energy–momentum spectrum lies in the forward light cone. In the remainder of the paper we derive bounds on the vacuum expectation values of certain products of field operators. These bounds are uniform in the space cutoff and the momentum cutoff.

It is known that the vacuum expectation values

$$\langle \Omega, \phi(x_1, t_1) \cdots \phi(x_n, t_n)\Omega \rangle \tag{1.1}$$

of a quantum field ϕ in its vacuum state Ω uniquely determine the field ϕ. The Wightman axioms for field theory can be stated in terms of the vacuum expectation values,[2] and given such expectation values one can construct a Hilbert space $\mathcal{H}$ of states and the field ϕ acting on $\mathcal{H}$.[2] Furthermore, the scattering of particles described by the theory is given in terms of the vacuum expectation values by the reduction formulas of Lehmann, Symanzik, and Zimmermann.[3] Wightman's program[4] for constructing the field ϕ is to find a sequence of approximating fields ϕ_j with vacuums Ω_j and to take the limit as $j \to \infty$ in the approximate vacuum expectation values

$$\langle \Omega_j, \phi_j(x_1, t_1) \cdots \phi_j(x_n, t_n)\Omega_j \rangle. \tag{1.2}$$

The limiting expectation values (1.1) are then used to construct ϕ and the Hilbert space $\mathcal{H}$.

We derive bounds, uniform in j, for the approximate vacuum expectation values (1.2) in which ϕ is replaced by $\nabla\phi = \phi_x$ or by $\pi = \phi_t$. These bounds are valid for boson interactions which are superrenormalizable and which have a finite-field strength renormalization constant.

2. THE ENERGY MOMENTUM SPECTRUM

We study boson self-interactions in a periodic box of volume $|V|$. We assume translation invariance and positivity of the Hamiltonian. We then prove that the energy–momentum spectrum lies in the forward light cone. In 2-dimensional space–time, these results apply to theories with no momentum cutoff. Our results also apply to models with a momentum cutoff, although the momentum or periodic space cutoffs presumably destroy Lorentz covariance.

We work in $(s + 1)$-dimensional space–time. Let V be a periodic box in s space having volume $|V|$, and let Γ_V be the lattice of allowed momenta. Then

$$\phi_{V,\kappa}(x) = (2\pi)^{-\frac{1}{2}s} 2^{-\frac{1}{2}} \int_{|k| \leq \kappa} \exp(-ik_V x) \times [a(k)^* + a(-k)]\mu(k_V)^{-\frac{1}{2}}\, dk \tag{2.1}$$

is a cutoff scalar boson field in the box V. Here $k_V \in \Gamma_V$ is the lattice point nearest k. The field $\phi_V(x) = \phi_{V,\infty}(x)$ has no momentum cutoff. The conjugate time zero field is

$$\pi_{V,\kappa}(x) = i(2\pi)^{-\frac{1}{2}s} 2^{-\frac{1}{2}} \int_{|k| \leq \kappa} \exp(-ik_V x) \times [a(k)^* - a(-k)]\mu(k_V)^{\frac{1}{2}}\, dk.$$

Let $\mathfrak{F}$ be a polynomial bounded from below. We set

$$H_0 = \int a(k)^* \mu(k_V) a(k)\, dk, \tag{2.2}$$

$$H_I = \int_V :\mathfrak{F}(\phi_{V,\kappa}(x)):\, dx, \tag{2.3}$$

$$E = \inf \text{spectrum}\,(H_0 + H_I), \tag{2.4}$$

$$H = H_0 + H_I - E, \tag{2.5}$$

$$P = \int a(k)^* k_V a(k)\, dk. \tag{2.6}$$

For $\kappa < \infty$, the operators (2.1)–(2.6) are essentially self-adjoint on the domain $\mathfrak{D}$ consisting of Fock space vectors with a finite number of particles and wave-functions in $\mathcal{S}$.[5] For $\kappa = \infty$, $s + 1 = 2$, the operators H_0, H_I, and P are essentially self-adjoint on $\mathfrak{D}$, and H is essentially self-adjoint on $\mathfrak{D}_1 = \mathfrak{D}(H_0) \cap \mathfrak{D}(H_I)$.[6,7] The operator H commutes with the operator P. The operator H has a unique ground-state vector $\Omega(V, \kappa) = \Omega$, satisfying $H\Omega = 0$, and zero is an isolated point of the spectrum of H.[8,7]

The self-adjointness of H and P and the existence of a ground state for H extend to the operators

$$H(\beta) = H \cosh \beta + P \sinh \beta,$$
$$P(\beta) = H \sinh \beta + P \cosh \beta.$$

Since there is no free constant in $H(\beta)$, the ground state $\Omega(V, \kappa, \beta) = \Omega(\beta)$ of $H(\beta)$ is the eigenvector of a possibly nonzero eigenvalue $E(\beta)$. This eigenvalue is simple and isolated, by the same proof that works for H. Clearly $H(\beta)^2 - P(\beta)^2 = H^2 - P^2$ on $\mathfrak{D}_1 \times \mathfrak{D}_1$.

Theorem 2.1: The joint spectrum of H and P above lies in the forward light cone. That is,

$$0 \leq H(\beta),$$
$$0 \leq H^2 - P^2,$$
$$E(\beta) = 0, \quad \Omega(\beta) = \Omega.$$

Proof: For $s + 1 = 2$, the resolvents

$$[H(V, \kappa, \beta) - \zeta]^{-1}$$

converge in norm as $\kappa \to \infty$ to the resolvent of the limiting Hamiltonian $H(V, \beta) = H(V, \infty, \beta)$. The operator $H(V, \beta)$ commutes with P, and their joint spectrum is contained in the limit of the joint spectrum of $H(V, \kappa, \beta)$ and P, as $\kappa \to \infty$ (Ref. 9, p. 432). Thus it is sufficient to prove the theorem in the case $\kappa < \infty$.

On the core $\mathfrak{D}$, the operators $H(\beta)$ are strongly continuous in β and in the coefficients of $\mathfrak{F}$. Thus the resolvents are also strongly continuous in β and $\mathfrak{F}$. Because $E(\beta)$ is an isolated eigenvalue, the projection $Q(\beta)$ on the vector $\Omega(\beta)$ depends continuously on β and $\mathfrak{F}$ (Ref. 9, pp. 437–38). Since H and P commute, so do $H(\beta)$ and P, and so do $Q(\beta)$ and P. The range $Q(\beta)$ is one dimensional, and so it is contained in an eigenspace of P. The corresponding eigenvalue belongs to the discrete lattice Γ_V and depends continuously on β and $\mathfrak{F}$. Thus the eigenvalue must be independent of β and P. From the case $\beta = 0$, $\mathfrak{F} = 0$, $H = H_0$, $\Omega(\beta) = \Omega_0 =$ Fock vacuum, we see that

$$P\Omega(\beta) = 0. \tag{2.7}$$

The ground state energy of $H(\beta)$ is nonnegative because, by (2.7) and the fact that $0 \leq H$,

$$0 \leq \cosh \beta\, \langle \Omega(\beta), H\Omega(\beta) \rangle = \langle \Omega(\beta), H(\beta)\Omega(\beta) \rangle$$
$$= E(\beta).$$

Thus $0 \leq H(\beta)$ and, on $\mathfrak{D} \times \mathfrak{D}$,

$$0 \leq H(\beta)H(-\beta)(\cosh \beta)^{-2} = H^2 - (\tanh \beta)^2 P^2,$$

and so, by limits as $\beta \to \infty$, $0 \leq H^2 - P^2$.

The ground state energy of $H(\beta)$ is nonpositive because $0 = \langle \Omega, H(\beta)\Omega \rangle$. Thus $E(\beta) = 0$ and

$$H(\beta)\Omega(\beta) = 0. \tag{2.8}$$

By (2.7) and (2.8), $H\Omega(\beta) = 0$, and by the uniqueness of the vacuum for H,

$$\Omega(\beta) = \Omega$$

is independent of β. This completes the proof of the theorem.

We remark that the result $\Omega(\beta) = \Omega$ agrees with the prediction of perturbation theory. To first-order perturbation theory, $\Omega^{(1)} = \Omega_0 - H_0^{-1} H_I \Omega_0$. The expression for $\Omega(\beta)$ to first order is

$$\Omega(\beta)^{(1)} = \Omega_0 - (H_0 \cosh \beta + P \sinh \beta)^{-1} \cosh \beta H_I \Omega_0$$
$$= \Omega_0 - H_0^{-1} H_I \Omega_0 = \Omega^{(1)},$$

where we have used $[P, H_I] = 0$ and $P\Omega_0 = 0$. Similar calculations yield $\Omega(\beta)^{(n)} = \Omega^{(n)}$ to all orders n.

3. TRANSFORMATIONS OF THE HAMILTONIAN

In this section we derive a uniform bound on $\|(H + I)^{-\frac{1}{2}} \pi(f) (H + I)^{-\frac{1}{2}}\|$. Let f be a smooth function on V with periodic boundary conditions. The operators

$$\phi_V(x) = \int \phi_V(x) f(x)\, dx$$

and

$$\pi_V(f) = \int \pi_V(x) f(x)\, dx$$

are self-adjoint and essentially self-adjoint on the domain $\mathfrak{D}$ of Sec. 2. We use the unitary operator

$$U = U(V,f) = \exp[i\phi_V(f)] \tag{3.1}$$

to transform the Hamiltonians H and $H(\beta)$ of Sec. 2.

Theorem 3.1: For $\kappa < \infty$ and any s, or for $s+1 = 2$, $\kappa \leq \infty$, we have

$$0 \leq H \pm \pi_V(f) + \tfrac{1}{2}\|f\|_2^2, \tag{3.2}$$

$$0 \leq H + P\tanh\beta \pm (\pi_V(f) - (\nabla\phi_V)(f)\tanh\beta) + \tfrac{1}{2}\|f\|_2^2. \tag{3.3}$$

Remark: In a theory with a spatially cutoff Hamiltonian $H(g)$, but no periodic box,[1,6-8] our proof shows that the corresponding estimate (3.2) for $\pi(f)$ is valid.

Proof: By Ref. 10, there is a cutoff-dependent and β-dependent bound

$$\pm\pi_V(f), \quad \pm\nabla\phi_V(f) \leq \epsilon N + \text{const} \leq H + \text{const}, \quad H + P\tanh\beta + \text{const}.$$

Thus the point to be established is the exact value $\|f\|_2^2$ of the change in the vacuum energy caused by the perturbations in (3.2) and (3.3). We prove that (3.2) and (3.3) are unitarily equivalent, via the unitary (3.1), to H and $H + P\tanh\beta$, respectively. On the domain $\mathfrak{D}$, we have the convergent power series expansions

$$H_0U = UH_0 + \sum_{n=0}^{\infty}\sum_{j=0}^{n}(n!)^{-1} \times [i\phi_V(f)]^{n-j-1}\pi_V(f)[i\phi_V(f)]^j,$$

$$H_0U = U[H_0 + \pi_V(f) + \tfrac{1}{2}\|f\|_2^2]. \tag{3.4}$$

Since $\mathfrak{D}$ is a core for H_0, we conclude that $U:\mathfrak{D}(H_0) \to D(H_0)$ and (3.4) holds as an operator identity on $D(H_0)$. Hence

$$U^*H_0U = H_0 + \pi_V(f) + \tfrac{1}{2}\|f\|_2^2.$$

Similarly, as an operator identity on $\mathfrak{D}(H_0) \subset \mathfrak{D}(P)$,

$$U^*PU = P - (\nabla\phi_V)(f).$$

Since U commutes with H_I, $U:\mathfrak{D}(H_I) \to \mathfrak{D}(H_I)$ and $U^*H_IU = H_I$.

The operators H and $H + P\tanh\beta$ are essentially self-adjoint on the domain $\mathfrak{D}_1 = \mathfrak{D}(H_0) \cap \mathfrak{D}(H_I)$. For H, this statement is proved in Refs. 6 and 7, and the same proof is valid for $H + P\tanh\beta$. Thus U^*HU is a positive self-adjoint operator that is essentially self-adjoint on $U^*\mathfrak{D}_1 = \mathfrak{D}_1$ and on $\mathfrak{D}_1$:

$$0 \leq U^*HU = U^*(H_0 + H_I - E)U = H_0 + \pi_V(f) + \tfrac{1}{2}\|f\|_2^2 + H_I - E = H + \pi_V(f) + \tfrac{1}{2}\|f\|_2^2.$$

The same equation holds with f replaced by $-f$, and so (3.2) holds. The proof of (3.3) is identical.

Transformation of H by $\exp[i\pi_V(f)]$ gives us

$$0 \leq H + \int :\mathfrak{F}(\phi_V(x) + f(x)):dx + m^2\phi(f) - \phi(\Delta f) + \tfrac{1}{2}m^2\|f\|^2 + \tfrac{1}{2}\|\nabla f\|_2^2. \tag{3.5}$$

We have not made significant use of this inequality, but we remark for quadratic interactions, $\mathfrak{F}(\xi) = \xi^2 + b\xi + c$, the bound (3.5) leads to an estimate on the vacuum expectation values (1.2) which is uniform in the cutoffs.

We observe that, for $s + 1 = 3$ and $\mathfrak{F}(\xi) = \xi^4 +$ lower-order terms, the coefficient of ξ^2 contains a mass renormalization term which is divergent as $\kappa \to \infty$. Neither this divergence nor the divergence of E (as $\kappa \to \infty$ or $|V| \to \infty$) interferes with our estimates. For $s + 1 = 4$, however, there are divergent counterterms containing $:\pi^2(x):$ coming from the field strength renormalization, and, if these terms are included in H, our estimates are no longer uniform in the cutoffs.

4. ESTIMATES ON THE VACUUM EXPECTATION VALUES

We use the time-dependent Heisenberg picture fields

$$\phi_\nu(x,t) = e^{itH}\phi_\nu(x)e^{-itH}, \tag{4.1}$$

where ϕ_ν equals ϕ, ϕ_t, or $\nabla\phi$ and $\phi(x)$, etc., is a time-zero free field. Recall that the Hamiltonian H contains a periodic space cutoff, and possibly a momentum cutoff. We use the uniform estimates (3.2) and (3.3) to derive bounds on the vacuum expectation values of the ϕ_ν.

For $f \in \mathfrak{C}_0^\infty$, the space–time averaged field is

$$\phi(f) = \int \phi(x,t)f(x,t)\,dx\,dt; \tag{4.2}$$

the fields $\phi_t(f)$ and $(\nabla\phi)(f)$ are similarly defined.

Lemma 4.1: Let $\kappa < \infty$ or $s + 1 = 2$ and $\kappa \leq \infty$. Then for $f \in \mathfrak{C}_0^\infty$, $\phi(f)$, $\phi_t(f)$, and $(\nabla\phi)(f)$ are operators on $\mathfrak{D}(H^{\frac{1}{2}})$.

For a proof of this lemma, see Ref. 8, Lemma 3.2.2. We have the cutoff-dependent estimate

$$\|\phi_\nu(f)\theta\| \leq |f|\,\|(H+I)^{\frac{1}{2}}\theta\|, \tag{4.3}$$

where $|f|$ is some Schwartz space norm on f, and $\phi_\nu = \phi$, ϕ_t or $\nabla\phi$, as defined by (4.1). The integral

(4.2) is by definition a weak integral of bilinear forms, and for $\nabla\phi$ we can integrate by parts in (4.2) to obtain $(\nabla\phi)(f) = -\phi(\nabla f)$. One can also show that $\phi_t(f)$ equals

$$\int \exp(itH)\pi(x)\exp(-itH)f(x,t)\,dx\,dt = \pi(f) = [iH, \phi(f)]$$

by evaluating $[iH, \phi(f)]$. This was proved for the case $\kappa = \infty$, $s+1=2$, $\mathfrak{P}$ having degree four.[8] The restriction to $\mathfrak{P}$ of degree four comes into the proof at one point only, and is easily avoided. Namely for such $\mathfrak{P}$, $H_0 + H_I$ is closed and self-adjoint on the domain $\mathfrak{D}(H_0) \cap D(H_I) = \mathfrak{D}_1$, while for general $\mathfrak{P}$, $H_0 + H_I$ is essentially self-adjoint on this domain.

As in Ref. 8, we conclude from (4.3) that $\mathcal{C}^\infty$ vectors for H are C^∞ vectors for $\phi_\nu(f)$ and that

$$\|H^m\phi_\nu(f)\theta\| \leq |f|_m \|(H+I)^{m+\frac{1}{2}}\theta\|, \tag{4.4}$$

where $|f|_m$ is a Schwartz space norm depending on the cutoffs. Thus the vacuum expectation values $\langle\Omega, \phi_{\nu_1}(x_1,t_1)\cdots\phi_{\nu_n}(x_n,t_n)\Omega\rangle$ are defined in the cutoff theory. Here Ω is the vacuum for the cutoff Hamiltonian H. To obtain estimates independent of the cutoff, we must replace (4.3) and (4.4) by bounds derived from Theorem 3.1. From Theorem 3.1 and the fact that H commutes with P, we have as bilinear forms on $\mathfrak{D}(H) \times \mathfrak{D}(H)$

$$0 \leq H \pm \pi(f) + \tfrac{1}{2}\|f\|_2^2, \tag{4.5}$$

$$0 \leq H + P\tanh\beta \pm (\pi(f) - (\nabla\phi)(f)\tanh\beta) + \tfrac{1}{2}\|f\|_2^2, \tag{4.6}$$

where $f \in \mathcal{C}_0^\infty(R^2)$ and $\pi(f)$ and $(\nabla\phi)(f)$ are the space–time averaged fields (4.2). These inequalities extend by continuity to $\mathfrak{D}(H^{\frac{1}{2}}) \times \mathfrak{D}(H^{\frac{1}{2}})$.

Theorem 4.1: Let $\kappa < \infty$ and s be arbitrary or let $\kappa \leq \infty$ and $s+1=2$. Then for $\theta \in \mathcal{C}^\infty(H)$, we have

$$\|H^j\pi(f)\theta\| + \|H^j\nabla\phi(f)\theta\| \leq |f|_j \|(H+I)^{j+4}\theta\|.$$

The Schwartz space norms $|f|_j$ are independent of the cutoffs κ and V, and they are independent of the polynomial $\mathfrak{P}$ and mass m which define H.

Proof: The case $j > 0$ follows from the case $j = 0$ and the commutator formulas

$$(\nabla\phi)(D_t f) = -[iH, (\nabla\phi)(f)]$$

and $\pi(D_t f) = -[iH, \pi(f)]$. From (4.5), we have

$$\pm\pi\left(\frac{f}{\|f\|_2}\right) \leq H + \tfrac{1}{2} \leq H + I,$$

$$\pm\pi(f) \leq \|f\|_2 (H+I). \tag{4.7}$$

Similarly, from (4.6) and (4.7) and Theorem 2.1, we have

$$\pm(\nabla\phi)(f) \leq 3\|f\|_2 (H+I). \tag{4.8}$$

We set $\psi = (H+I)^2\theta$, $R = (H+I)^{-1}$. Then we have

$$\begin{aligned}
\|\pi(f)\theta\|^2 &= \langle\psi, R^2\pi(f)^2R^2\psi\rangle \\
&\leq |\langle\psi, R\pi(f)R\pi(f)R^2\psi\rangle| \\
&\quad + |\langle\psi, R[R,\pi(f)]\pi(f)R^2\psi\rangle| \\
&\leq \|R^{\frac{1}{2}}\pi(f)R^{\frac{1}{2}}\|^2 \|\psi\|^2 \\
&\quad + |\langle\psi, R^2[\pi(f), H]R\pi(f)R^2\psi\rangle| \\
&\leq \|f\|_2^2 \|(H+I)^2\theta\|^2 \\
&\quad + |\langle\psi, R^2\pi(D_t f)R\pi(f)R^2\psi\rangle| \\
&\leq (\|f\|_2^2 + \|f\|_2 \|D_t f\|_2) \|(H+I)^2\theta\|^2.
\end{aligned} \tag{4.9}$$

In deriving (4.9) we use the bound (4.7). The proof of the bound on $(\nabla\phi)(f)$ is similar, and the proof of the theorem is complete.

Theorem 4.2: Let $\kappa < \infty$ for any s or let $\kappa \leq \infty$ for $s+1=2$. Then there is a Schwartz space norm $|\cdot|_n$ defined on $\mathcal{S}(R^{(s+1)n})$ such that, for $f \in \mathcal{S}(R^{(s+1)n})$, we have

$$\left|\int f(x,t)\langle\Omega, \phi_{\nu_1}(x_1,t_1)\cdots\phi_{\nu_n}(x_n,t_n)\Omega\rangle\,dx\,dt\right| \leq |f|_n.$$

If each ϕ_ν is a π or a $\nabla\phi$, then the norm $|\cdot|_n$ is independent of κ, V, m, and $\mathfrak{P}$.

Remark: In a theory with a spatially cutoff Hamiltonian $H(g)$, the theorem holds if each φ_ν is a π.

* Supported in part by the U.S. Air Force Office of Scientific Research, Contract AF 49(638)-1719.

† Alfred P. Sloan Foundation Fellow. Supported in part by the U.S. Air Force Office of Scientific Research, Contract F 44620-70-C-0030.

[1] J. Glimm and A. Jaffe "Self-Adjointness of the Yukawa$_2$ Hamiltonian," Ann. Phys. (N.Y.) (to be published).

[2] R. Streater and A. Wightman, *PCT, Spin and Statistics, and All That* (Benjamin, New York, 1964).

[3] H. Lehmann, K. Symanzik, and W. Zimmermann, Nuovo Cimento **1**, 205 (1955); K. Hepp, in *Brandeis University Summer Institute in Theoretical Physics, 1965*, M. Chretien and S. Deser, Eds. (Gordon and Breach, New York, 1966).

[4] A. Wightman, in *1964 Cargèse Summer School Lectures*, M. Lévy, Ed. (Gordon and Breach, New York, 1967).

[5] A. Jaffe, thesis, Princeton University, 1965.

[6] J. Glimm and A. Jaffe, Phys. Rev. **176**, 1945 (1968).

[7] L. Rosen, Commun. Math. Phys. **16**, 157 (1970); L. Rosen, "The ϕ_2^{2n} Quantum Field Theory: Higher Order Estimates," Preprint.

[8] J. Glimm and A. Jaffe, Ann. Math. **91**, 362 (1970).

[9] T. Kato, *Perturbation Theory for Linear Operators* (Springer, New York, 1966).

[10] E. Nelson, "A Quartic Interaction in Two Dimensions," in *Mathematical Theory of Elementary Particles*, R. Goodman and I. Segal, Eds, (M.I.T. Press, Cambridge, Mass., 1966), and J. Glimm, Commun. Math. Phys. **8**, 12 (1968).

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 11, NUMBER 12 DECEMBER 1970

Semiclassical Perturbation Theory of Molecular Collisions. III. Graphical Angular Momentum Methods and the *n*th Order*

H. A. RABITZ† AND R. GORDON
Department of Chemistry, Harvard University, Cambridge, Massachusetts 02138

(Received 2 March 1970)

A review of graphical angular momentum methods is presented. As an example, the graphical methods are applied to the coupling problem in the semiclassical first and second Born approximations with an electric multipole potential. The general *n*th-order Born term is then considered. A simple expression is derived for the matrix elements of the anisotropic terms in the potential. The evaluation of the remaining integrals over time is discussed.

INTRODUCTION

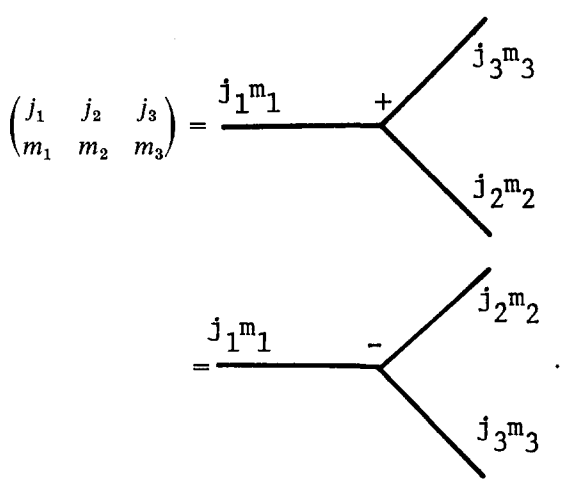
In Paper I we derived explicit expressions for the semiclassical first and second Born approximations, using time-dependent perturbation theory with an electric multipole potential.[1] We performed the necessary angular momentum algebra by finding the correct formulas in the standard books on angular momentum theory. This is a rather haphazard procedure. It would be much better to have a systematic method of handling such problems which could be extended to cases where we cannot find handy formulas in the standard books. This can be done without the use of diagrams, but it is much easier to use graphical methods. We shall essentially follow the diagrammatic methods outlined by Brink and Satchler[2]; proofs of certain graphical procedures will also be presented. Our diagrams will differ from those defined by Yutsis *et al.*[3] in certain rules and phases. This must be kept in mind when referring to their book. Recently Massot *et al.*[4] developed diagrammatic methods which include the graphical representation of tensor operators. The convenience of this is somewhat outweighed by the added complication of the diagrams; we will carry the tensor operators along with our diagrams. Only the essentials of the graphical methods will be presented in Sec. I. For more details see Brink and Satchler[2] and Yutsis *et al.*[3]

As an example, the graphical methods will be applied in Sec. II to the coupling problems dealt with in Paper I.[1] The technique will then be applied to the general *n*th-order Born term in Sec. III to yield a simple expression for the product of matrix elements involved. Possible methods of performing the remaining integrals over time will also be discussed.

The graphical method is not restricted to scattering problems[5]; any problem which involves the coupling of angular momenta can be treated by the graphical approach. The graphical angular momentum methods can also be made compatible with some many-body diagrammatic methods.[6]

I. GRAPHICAL ANGULAR MOMENTUM METHODS

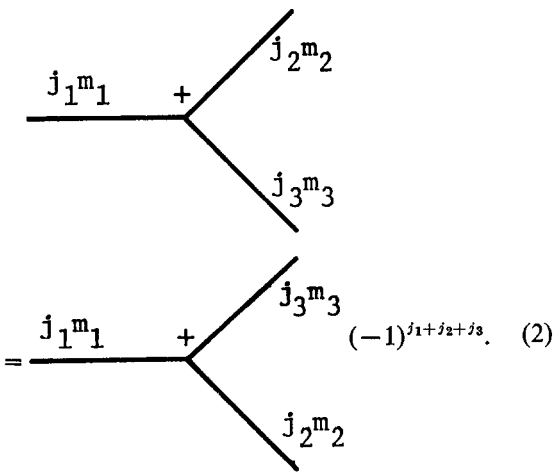
The basis of the graphical methods lies in the diagrammatic representation of the Wigner 3-*j* symbol[7]



(1)

The node represents the coupling of the three angular momenta. This does not correspond with the vector diagrams drawn for the classical coupling of three angular momenta. Ponzano and Regge[8] have recently used such a graphical representation; however, it has limited applicability in our work. The + sign corresponds to an anticlockwise orientation of the node; the − sign corresponds to a clockwise orientation of the node. From the permutation properties of the 3-*j*

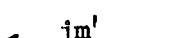
symbols, we arrive at the relations



(2)

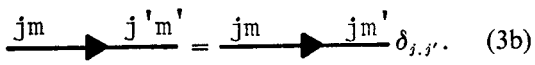
A rotation of the diagram does not change its meaning since it corresponds to an even permutation of the columns in the 3-j symbol. Likewise, a distortion of the symbol does not change its meaning provided that two lines are not crossed. The metric tensor[7] is defined as

$$\begin{pmatrix} j \\ m \quad m' \end{pmatrix} = (-1)^{j+m}\delta_{m,-m'} =$$



. (3a)

Also note that



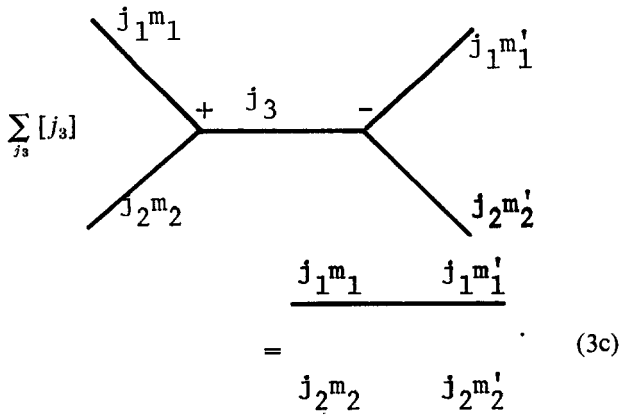
(3b)

The use of arrows will be discussed shortly.

The basic rule in using these graphical methods is that two lines may be joined if they have the same J and M labels on them. The resulting line is just labeled by J if M is summed over; the summation is implicitly understood in such cases. We illustrate this with the unitarity relation of 3-j symbols,

$$\sum_{j_3 m_3} [j_3] \begin{pmatrix} j_1 & j_2 & j_3 \\ m_1 & m_2 & m_3 \end{pmatrix} \begin{pmatrix} j_1 & j_2 & j_3 \\ m_1' & m_2' & m_3 \end{pmatrix} = \delta_{m_1,m'_1}\delta_{m_2,m'_2},$$

where $[J] \equiv 2J + 1$. Graphically this is written as



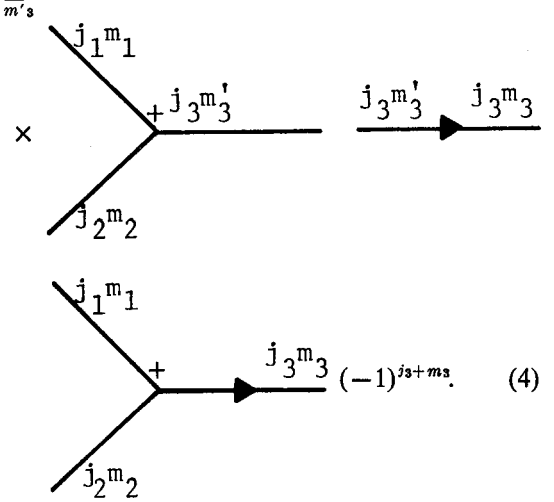
(3c)

It is clear from this that a line without an arrow just represents a δ function. We can derive a useful relation if we contract a 3-j symbol with a metric tensor:

$$\begin{pmatrix} j_1 & j_2 & j_3 \\ m_1 & m_2 & -m_3 \end{pmatrix}$$

$$= \sum_{m'_3} (-1)^{j_3+m_3} \begin{pmatrix} j_1 & j_2 & j_3 \\ m_1 & m_2 & m_3' \end{pmatrix} \begin{pmatrix} j_3 & & \\ m_3 & & m_3' \end{pmatrix}$$

$$= \sum_{m'_3} (-1)^{j_3+m_3}$$



(4)

We see that a negative m in a 3-j symbol will produce an arrow in its resulting diagram. The above relation can be used as a definition of the arrow symbol. Note that it is not possible to remove a single arrow from a line, since this would correspond to changing the sign on one m value in the 3-j symbol. Using the metric tensors, we can derive some simple rules for using the arrows:

$$\begin{aligned} &\text{j1m1} \longrightarrow \text{j1m'1} = \sum_{m_2} \begin{pmatrix} j_1 \\ m_1' \quad m_2 \end{pmatrix} \begin{pmatrix} j_1 \\ m_2 \quad m_1 \end{pmatrix} \\ &= \sum_{m_2} (-1)^{2j_1+m'_1+m_2} \delta_{m_1,-m_2}\delta_{m_2,-m'_1} \\ &= \underline{\text{j1m1} \qquad \text{j1m'1}}\ (-1)^{2j_1}, \end{aligned} \tag{5a}$$

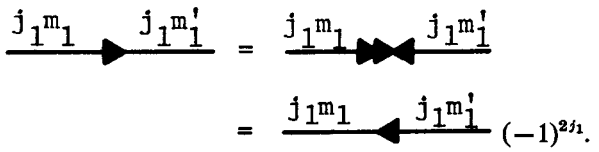
$$\begin{aligned} &\text{j1m1} \longleftarrow \text{j1m'1} = \sum_{m_2} \begin{pmatrix} j_1 \\ m_1' \quad m_2 \end{pmatrix} \begin{pmatrix} j_1 \\ m_1 \quad m_2 \end{pmatrix} \\ &= \sum_{m_2} (-1)^{2j_1+m_1+m'_1} \delta_{m_1,-m_2}\delta_{m'_1,-m} \\ &= \underline{\text{j1m1} \qquad \text{j1m'1}}\ . \end{aligned} \tag{5b}$$

Note that even if j is half-integer, $(j + m)$ is an integer. Hence,

$$(-1)^{2(j_1+m_1)} = 1.$$

These results are useful since we can change the direction of an arrow, provided that we are careful

with phases. That is,



(6)

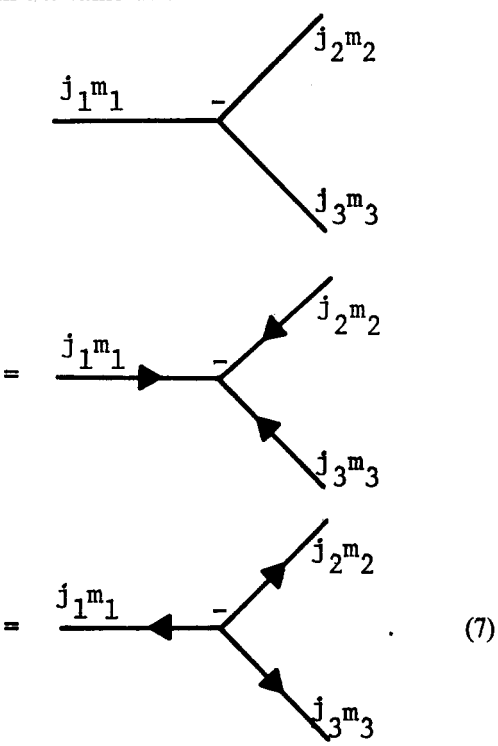
Another useful relation is obtained from

$$\begin{pmatrix} j_1 & j_2 & j_3 \\ m_1 & m_2 & m_3 \end{pmatrix} = (-1)^{j_1+j_2+j_3}\begin{pmatrix} j_1 & j_2 & j_3 \\ -m_1 & -m_2 & -m_3 \end{pmatrix}$$

and the fact that

$$m_1 + m_2 + m_3 = 0.$$

Writing this out graphically with the aid of Eq. (4), we arrive at the rule that an arrow can simultaneously be added to each line around a node, provided that they all point in the same direction:



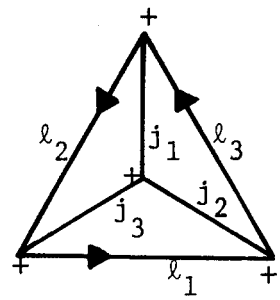
(7)

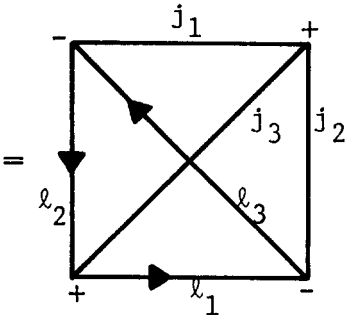
The last step follows from applying the rules for changing the direction of arrows in Eq. (6).

Using these rules, we can write out the diagrams for any complex summation of 3-*j* symbols. If the resulting diagram has no open lines (noncontracted lines), then it will represent a 3*n*-*j* coefficient. As with most diagrammatic methods, their topology is very interesting. In fact, the symmetry properties of the 3*n*-*j* coefficients can be obtained graphically by identifying the rotation group that describes the geometric properties of the diagram. It can be helpful to construct 3-dimensional models of the diagrams. However, the definition of the + or − sign on a node must be modified for the 3-dimensional case.

The equivalent diagrams for the 6-*j* and 9-*j* symbols are

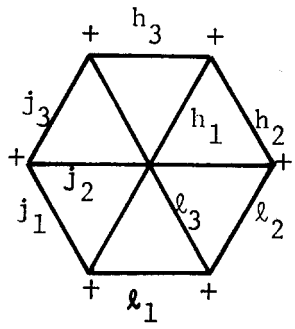
$$\begin{Bmatrix} j_1 & j_2 & j_3 \\ l_1 & l_2 & l_3 \end{Bmatrix} =$$





(8a)

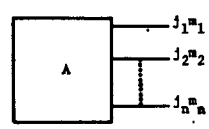
$$\begin{pmatrix} j_1 & j_2 & j_3 \\ l_1 & l_2 & l_3 \\ h_1 & h_2 & h_3 \end{pmatrix} =$$



(8b)

There are other ways of writing out these diagrams, but they are all topologically equivalent. In fact, it is a general rule that all topologically equivalent diagrams will represent the same algebraic expression or symbol (perhaps with the addition of a phase factor). The general study of the higher 3*n*-*j* symbols is very interesting in itself. It is sufficient to say that it is possible to generate all topologically nonequivalent 3*n*-*j* symbols from all the 3(n − 1)-*j* symbols. For more details see Yutsis *et al.*[3]

In applications, we frequently have diagrams with open lines, i.e., lines with noncontracted projection quantum numbers. Such diagrams can be represented by a general block of contracted lines with open lines coming out as follows:



$$= F_A\begin{pmatrix} j_1 & j_2 & \cdots & j_n \\ m_1 & m_2 & \cdots & m_n \end{pmatrix}. \quad (9a)$$

The label A represents the particular arrangement of the contracted lines. Any $F_A(\ )$ can be written out as

a summation over a product of 3-j symbols. All internal (contracted) lines must have arrows, and all external lines must not have arrows. Such diagrams are called *normal* diagrams. Only normal diagrams arise in physical problems. The $F_A(\ )$ are frequently called JM coefficients. The simplest nontrivial case is just the usual 3-j symbol. In general we can write

$$F_A\begin{pmatrix} j_1 & \cdots & j_n \\ m_1 & \cdots & m_n \end{pmatrix} = \sum_{\substack{\{n_i\} \\ i=1}}^{r} (-1)^{\sum_h (l_h + n_h)} \begin{pmatrix} l_1 & l_2 & j_1 \\ m_1 & n_2 & m_1 \end{pmatrix} \begin{pmatrix} l_1 & l_3 & l_4 \\ -n_1 & n_3 & n_4 \end{pmatrix} \times \begin{pmatrix} l_3 & j_2 & j_3 \\ -n_3 & m_2 & m_3 \end{pmatrix} \cdots \begin{pmatrix} l_r & l_{r-1} & j_n \\ n_r & n_{r-1} & m_n \end{pmatrix}, \quad (9b)$$

where the phase factor guarantees that the diagram is in normal form. The JM coefficients can be looked at as the coupling coefficients for the states $\phi(j_1 m_1)$, $\phi(j_2 m_2), \cdots, \phi(j_n m_n)$ up to a state with zero angular momentum. That is,

$$|j_1 j_2 \cdots j_n 00\rangle_A = \sum_{\{m_i\}} F_A\begin{pmatrix} j_1 & \cdots & j_n \\ m_1 & \cdots & m_n \end{pmatrix} \phi(j_1 m_1)\phi(j_2 m_2) \cdots \phi(j_n m_n).$$

The states are in reference to some space fixed coordinate system. This definition of $F_A(\ )$ could also be extended to the coupling up of tensor operators $T^{j_i}_{m_i}$ to a resultant $T^0_{A^0}$. We shall later use them for this purpose. With these definitions the 3-j symbol couples up three angular momenta to a zero resultant state. To see this, we note that

$$|j_1 j_2 j_3 m_3\rangle = \sum_{m_1 m_2} |j_1 m_1\rangle\, |j_2 m_2\rangle \langle j_1 m_1 j_2 m_2 \,|\, j_1 j_2 j_3 m_3\rangle$$

and

$$|j_1 j_2 j_3 00\rangle = \sum_{m_3 m_3'} |j_3 m_3'\rangle\, |j_1 j_2 j_3 m_3\rangle \times \langle (j_1 j_2 j_3 m_3) j_3 m_3' \,|\, (j_1 j_2 j_3) j_3 00\rangle.$$

Writing the coupling coefficient in terms of a 3-j symbol,[7]

$$\langle (j_1 j_2 j_3 m_3) j_3 m_3' \,|\, (j_1 j_2 j_3) j_3 00\rangle = \begin{pmatrix} j_3 & j_3 & 0 \\ m_3 & m_3' & 0 \end{pmatrix} = \frac{(-1)^{j_3 - m_3}}{([j_3])^{\frac{1}{2}}} \delta_{m_3, -m_3'},$$

we get

$$|j_1 j_2 j_3 00\rangle = \sum_{m_3} \frac{(-1)^{j_3 - m_3}}{([j_3])^{\frac{1}{2}}} |j_3 m_3\rangle\, |j_1 j_2 j_3 - m_3\rangle = (-1)^{j_2 - j_1 + j_3} \sum_{m_1 m_2} \begin{pmatrix} j_1 & j_2 & j_3 \\ m_1 & m_2 & m_3 \end{pmatrix} \times |j_1 m_1\rangle\, |j_2 m_2\rangle\, |j_3 m_3\rangle.$$

So we see that $F_A(\ )$ is essentially a 3-j symbol if A represents a single note. We could extend this to a generalized Wigner coefficient

$$\begin{pmatrix} j_1 & j_2 & \cdots & j_n \\ m_1 & m_2 & \cdots & m_n \end{pmatrix},$$

where the diagram for such a coefficient has a minimum number of contracted lines. See Yutsis *et al.*[3] for details.

From the definition of $F_A(\ )$ as the coupling coefficient up to the state $|00\rangle_A$, it is clear that

$$\sum_{i=1}^{n} m_i = 0. \quad (10a)$$

Now we shall see that

$$\alpha = \sum_{i=1}^{n} j_i = \text{integer} \quad (10b)$$

for the j_i in the open lines of $F_A(\ )$. Consider a rotation by π of the state $|j_1 \cdots j_n 00\rangle_A$,

$$R(\pi) = R_{j_1}(\pi) \times R_{j_2}(\pi) \cdots \times R_{j_n}(\pi),$$

where R_{j_i} acts on $\phi(j_i m_i)$ in the fashion

$$R_{j_i}(\omega)\phi(j_i m_i) = \sum_{m_i'} D^{j_i}_{m_i', m_i}(\omega)\phi(j_i m_i').$$

$D^j_{m',m}(\omega)$ is the rotation matrix as defined by Edmonds.[7] We now use the relation

$$D^j_{m',m}(\pi) = (-1)^{j+m}\delta_{m,-m'}$$

to obtain

$$\begin{aligned} R(\pi)\,|j_1 \cdots j_n 00\rangle_A &= \sum_{m_i} F_A\begin{pmatrix} j_1 & \cdots & j_n \\ m_1 & \cdots & m_n \end{pmatrix} \times R_{j_1}(\pi)\phi(j_1 m_1) \cdots R_{j_n}(\pi)\phi(j_n m_n) \\ &= \sum_{m_i, m_i'} F_A\begin{pmatrix} j_1 & \cdots & j_n \\ m_1 & \cdots & m_n \end{pmatrix} \times D^{j_1}_{m_1', m_1}(\pi)\phi(j_1 m_1') \cdots D^{j_n}_{m_n', m_n}(\pi)\phi(j_n m_n') \\ &= \sum_{m_i} F_A\begin{pmatrix} j_1 & \cdots & j_n \\ m_1 & \cdots & m_n \end{pmatrix} (-1)^{\sum_h (j_h + m_h)} \times \phi(j_1 - m_1) \cdots \phi(j_n - m_n). \end{aligned}$$

Note that $\sum_{i=1}^{n} m_i = 0$ to eliminate it from the phase. Now apply the reverse rotation $R^{-1}(\pi) = R(-\pi)$. Again we use the relation

$$D^j_{m',m}(-\pi) = (-1)^{j-m}\delta_{m',-m}$$

to arrive at

$$|j_1 \cdots j_n 00\rangle_A = (-1)^{2\alpha} \sum_{m_i} F_A\begin{pmatrix} j_1 & \cdots & j_2 \\ m_1 & \cdots & m_n \end{pmatrix} \times \phi(j_1 m_1) \cdots \phi(j_n m_n).$$

We conclude that

$$\alpha = \sum_{i=1}^{n} j_i = \text{integer}$$

by comparison with the original definition of $|j_1 \cdots j_n 00\rangle_A$. Another useful relation is

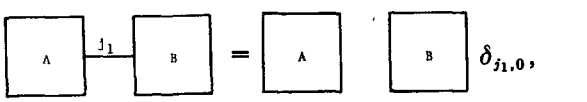
$$F_A\begin{pmatrix} j_1 & \cdots & m_n \\ -m_1 & \cdots & -m_n \end{pmatrix} = (-1)^{\alpha} F_A\begin{pmatrix} j_1 & \cdots & j_n \\ m_1 & \cdots & m_n \end{pmatrix}. \quad (10c)$$

This relation is evidently true from the last derivation if we note that

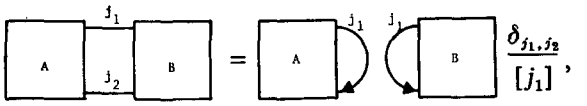
$$R(\omega)\,|j_1 \cdots j_n 00\rangle_A = |j_1 \cdots j_n 00\rangle_A .$$

This follows because the state $|00\rangle_A$ has no spatial dependence.

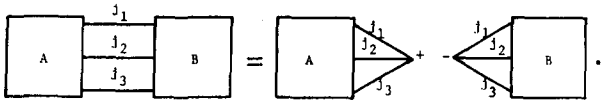
We are now ready to derive some powerful rules for manipulating diagrams. These simple rules essentially justify the use of the graphical methods. We want to show that the following operations are valid if the left or right block is in normal form:



(11a)

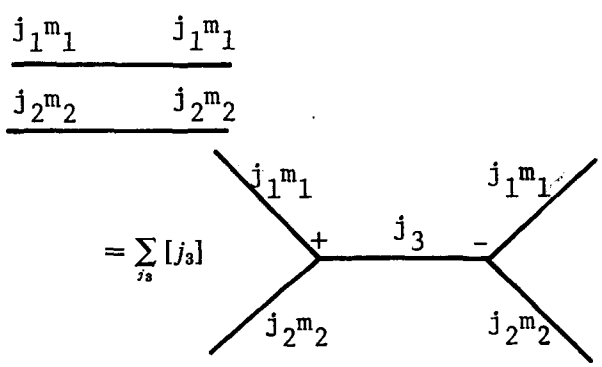


(11b)

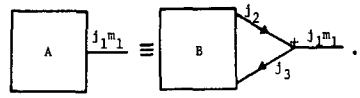


(11c)

For four or more connecting lines we cannot simply cut the diagrams; but we can utilize the relation



to reduce the number of connecting lines to three, and then cut the diagram. Appendix A discusses the general n-line cutting rules. Note that a line from a block really means



To prove these rules, we shall use $F_A(\ )$ notation. Consider

$$F_A\begin{pmatrix} j_1 & j_2 & j_3 \\ m_1 & m_2 & m_3 \end{pmatrix} = \sum_{\{n_i\}} (-1)^{\sum_h (l_h + n_h)} \begin{pmatrix} l_1 & l_2 & j_1 \\ n_1 & n_2 & m_1 \end{pmatrix} \times \begin{pmatrix} l_1 & l_3 & l_4 \\ -n_1 & n_3 & n_4 \end{pmatrix} \cdots \begin{pmatrix} j_3 & l_r & l_{r-1} \\ m_3 & -n_r & n_{r-1} \end{pmatrix},$$

where A can be any arrangement of contracted angular momenta and $i = 1, 2, \cdots, r$. We now use the following relation between 3-j coefficients and D matrices[7]:

$$\begin{pmatrix} l_1 & l_2 & l_3 \\ m_1 & m_2 & m_3 \end{pmatrix} = \sum_{m'_1 m'_2 m'_3} D^{l_1}_{m'_1, m_1}(\omega) D^{l_2}_{m'_2, m_2}(\omega) \times D^{l_3}_{m'_3, m_3}(\omega) \begin{pmatrix} l_1 & l_2 & l_3 \\ m'_1 & m'_2 & m'_3 \end{pmatrix}.$$

Note that ω is purely arbitrary. We want to expand each 3-j symbol in $F_A(\ )$ using the same ω:

$$F_A\begin{pmatrix} j_1 & j_2 & j_3 \\ m_1 & m_2 & m_3 \end{pmatrix} = \sum_{\{n_i\}} \sum_{\{n'_i\}} \sum_{\{n''_i\}} \sum_{m'_1 m'_2 m'_3} (-1)^{\sum_h (l_h + n_h)} \begin{pmatrix} l_1 & l_2 & j_1 \\ n'_1 & n'_2 & m'_1 \end{pmatrix} \times \begin{pmatrix} l_1 & l_3 & l_4 \\ -n''_1 & n'_3 & n'_4 \end{pmatrix} \cdots \begin{pmatrix} j_3 & l_r & l_{r-1} \\ m_3 & -n''_r & n'_{r-1} \end{pmatrix} D^{l_1}_{n'_1, n_1}(\omega) \times D^{l_1}_{-n''_1, -n_1}(\omega) \cdots D^{j_1}_{m'_1, m_1}(\omega) D^{j_1}_{m'_2, m_2}(\omega) D^{j_3}_{m'_3, m_3}(\omega),$$

where $\{n_i\} \to \{n'_i\}$ and $\{-n_i\} \to \{-n''_i\}$. Notice that the projection quantum number for each internal line occurs on two D matrices. Now we use the orthogonality property of the D matrices:

$$\sum_{n_i} (-1)^{-l_i - n_i} D^{l_i}_{n'_i, n_i}(\omega) D^{l_i}_{-n''_i, -n_i}(\omega) = \sum_{n_i} (-1)^{-l_i - n_i} D^{l_i}_{n'_i, n_i}(\omega) D^{l_i *}_{n''_i, n_i}(\omega) (-1)^{n_i - n''_i} = (-1)^{l_i + n'_i} \delta_{n'_i, n''_i}.$$

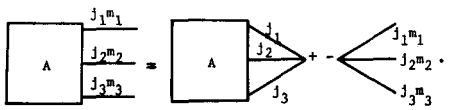
If $F_A(\ )$ was not in normal form, we could not use this relation. This is why one of the blocks in the coupling rules must be in normal form. We apply this to get

$$F_A\begin{pmatrix} j_1 & j_2 & j_3 \\ m_1 & m_2 & m_3 \end{pmatrix} = \sum_{m'_1, m'_2, m'_3} F_A\begin{pmatrix} j_1 & j_2 & j_3 \\ m'_1 & m'_2 & m'_3 \end{pmatrix} \times D^{j_1}_{m'_1, m_1}(\omega) D^{j_2}_{m'_2, m_2}(\omega) D^{j_3}_{m'_3, m_3}(\omega).$$

Now we multiply both sides by $(8\pi^2)^{-1}$ and integrate over ω. The right-hand side integrates to a product of two 3-j symbols[7]:

$$F_A\begin{pmatrix} j_1 & j_2 & j_3 \\ m_1 & m_2 & m_3 \end{pmatrix} = \begin{pmatrix} j_1 & j_2 & j_3 \\ m_1 & m_2 & m_3 \end{pmatrix} \times \sum_{m'_1 m'_2 m'_3} F_A\begin{pmatrix} j_1 & j_2 & j_3 \\ m'_1 & m'_2 & m'_3 \end{pmatrix}\begin{pmatrix} j_1 & j_2 & j_3 \\ m'_1 & m'_2 & m'_3 \end{pmatrix} = (3n\text{-}j \text{ coef.})_A \times \begin{pmatrix} j_1 & j_2 & j_3 \\ m_1 & m_2 & m_3 \end{pmatrix}.$$

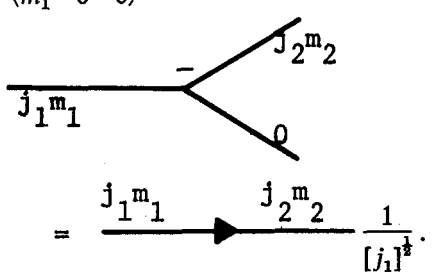
The sum over m'_1, m'_2, and m'_3 is a closed diagram in normal form or a $3n$-j coefficient. We now have essentially proved what we wanted if we transcribe it to graphical form:



If we had attached a block B to the right side, we would get the three line cutting rule. It is not necessary that B be in normal form since only A and its three external lines enter into the derivation. The other cutting rules can be obtained with the help of the relations

$$\begin{pmatrix} j_1 & j_2 & 0 \\ m_1 & m_2 & 0 \end{pmatrix} = \frac{\delta_{j_1 j_2}\delta_{m_1,-m_2}}{[j_1]^{\frac{1}{2}}}(-1)^{j_1-m_1},$$

$$\begin{pmatrix} j_1 & 0 & 0 \\ m_1 & 0 & 0 \end{pmatrix} = \delta_{j_1,0}\delta_{m_1,0},$$



As a last note, we must remember that all three angular momenta coming from a node must satisfy the triangle relations for angular momentum coupling.

We now have all the basic rules for handling open and closed diagrams. Their usefulness will be illustrated with a few examples in the next part.

II. GRAPHICAL METHODS AND THE FIRST AND SECOND BORN APPROXIMATION

We shall rederive the angular momentum coupling in the first and second Born approximations with multipolar potentials as examples of how the graphical methods can be used.[1] As we shall see, it is no longer necessary to fall back on "formula X in book Y" to calculate the necessary sums over the projection quantum numbers. Using the graphical approach, we shall derive all the necessary relations we need.

It is useful to start at the beginning and write out the diagrammatic representation of the basic matrix element,[1]

$$\langle n_i|\, T_{l-\gamma}\,|n_j\rangle,$$

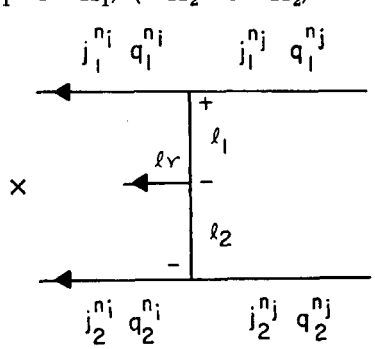
$$T_{l-\gamma} = \sum_{m_1 m_2} \langle l_1 m_1 l_2 m_2 \mid l_1 l_2 l\gamma\rangle Y_{l_1 m_1}(\hat{r}_1) Y_{l_2 m_2}(\hat{r}_2),$$

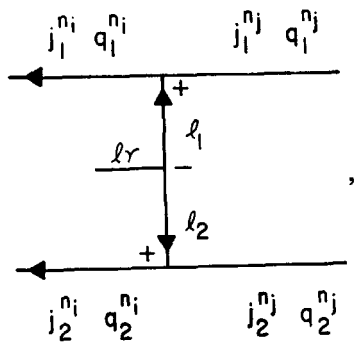
where $|n_i\rangle = |J_1^{n_i} q_1^{n_i} K_1\rangle\,|J_2^{n_i} q_2^{n_i} K_2\rangle$ are symmetric top wavefunctions and i is a dummy index. J is the rotational quantum number, q its projection on the space fixed axis, and K its projection on the molecule fixed axis (note that J, q, and K are integers). The subscripts 1 and 2 refer to the two molecules. Taking the indicated matrix elements over the spherical harmonics, we get

$$\langle n_i|\, T_{l-\gamma}\,|n_j\rangle = (-1)^{K_1+K_2+j_1^{n_i}+j_2^{n_i}}\left(\frac{[l_1][l_2][l][j_1^{n_i}][j_2^{n_i}][j_1^{n_j}][j_2^{n_j}]}{(4\pi)^2}\right)^{\frac{1}{2}} \times \begin{pmatrix} j_1^{n_i} & l_1 & j_1^{n_j} \\ -K_1 & 0 & K_1 \end{pmatrix}\begin{pmatrix} j_2^{n_i} & l_2 & j_2^{n_j} \\ -K_2 & 0 & K_2 \end{pmatrix} \times \sum_{m_1 m_2}\begin{pmatrix} l_1 & l_2 & l \\ m_1 & m_2 & -\gamma \end{pmatrix}\begin{pmatrix} j_1^{n_i} & l_1 & j_1^{n_j} \\ -q_1^{n_i} & m_1 & q_1^{n_j} \end{pmatrix} \times \begin{pmatrix} j_1^{n_i} & l_2 & j_2^{n_j} \\ -q_2^{n_i} & m_2 & q_2^{n_j} \end{pmatrix}(-1)^{j_1^{n_i}+q_1^{n_i}+j_2^{n_i}+q_2^{n_i}+l+\gamma}.$$

We now want to represent the sum over m_1 and m_2 graphically:

$$\langle n_i|\, T_{l-\gamma}\,|n_j\rangle = (-1)^{j_1^{n_i}+j_2^{n_i}+K_1+K_2}\left(\frac{[l_1][l_2][l][j_1^{n_i}][j_2^{n_i}][j_1^{n_j}][j_2^{n_j}]}{(4\pi)^2}\right)^{\frac{1}{2}} \times \begin{pmatrix} j_1^{n_i} & l_1 & j_1^{n_j} \\ -K_1 & 0 & K_1 \end{pmatrix}\begin{pmatrix} j_2^{n_i} & l_2 & j_2^{n_j} \\ -K_2 & 0 & K_2 \end{pmatrix}$$



$= h(l, n_i, n_j)$  , (12a)

where

$$h(l, n_i, n_j) = (-1)^{j_1^{n_i}+j_2^{n_j}+l_2+K_1+K_2}\left(\frac{[l_1][l_2][l][j_1^{n_i}][j_2^{n_i}][j_1^{n_j}][j_2^{n_j}]}{(4\pi)^2}\right)^{\frac{1}{2}} \times \begin{pmatrix} j_1^{n_i} & l_1 & j_1^{n_j} \\ -K_1 & 0 & K_1 \end{pmatrix}\begin{pmatrix} j_2^{n_i} & l_2 & j_2^{n_j} \\ -K_2 & 0 & K_2 \end{pmatrix}. \quad (12b)$$

We will treat $h(l, n_i, n_j)$ as a numerical coefficient. With this graphical representation of the matrix element, we now have the basic building block for the diagrams in the first and second Born approximations.

The first Born approximation for the transition probability can be written as[1]

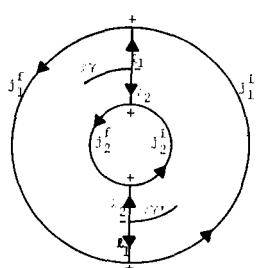
$$P_1(i \to f) = \frac{1}{\hbar^2[j_1^i][j_2^i]} \sum_{q_1^i q_1^f q_2^i q_2^f} \sum_{\gamma,\gamma'} \langle f|\, T_{l-\gamma}\,|i\rangle \times \langle i|\, T_{l-\gamma'}\,|f\rangle\, I_1(l,\gamma) I_1^*(l,\gamma'),$$

where $I_1(l\gamma)$ is an integral over time defined in Eq. (20b).

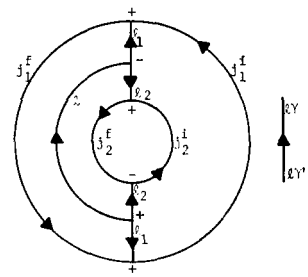
We are summing over initial and final projection quantum numbers. It is not necessary to do this; however, the resulting nonsummed expression would be more complicated. We want to focus on the summation

$$A_2 = \sum_{q_1^i q_1^f} \sum_{q_2^i q_2^f} \frac{\langle f|\, T_{l-\gamma}\,|i\rangle\langle i|\, T_{l-\gamma'}\,|f\rangle}{([j_1^i][j_2^i])}.$$

This can be written out in graphical form by using Eqs. (12):

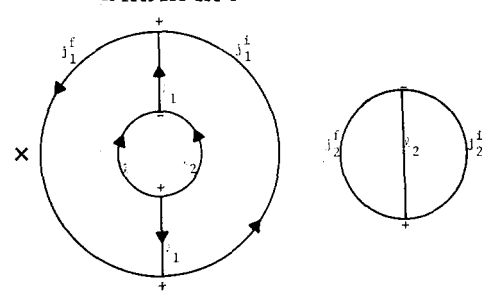
$A_2 =$  . (13)

Note that the diagram is in normal form, as it should be. We now apply the two line coupling rule in Eq. (11b) to the open lines:

$A_2 = \frac{h(l,f,i)h(l,i,f)}{[j_1^i][j_2^i][l]}$  $\times (-1)^{j_2^f+j_2^i+l_2}$.

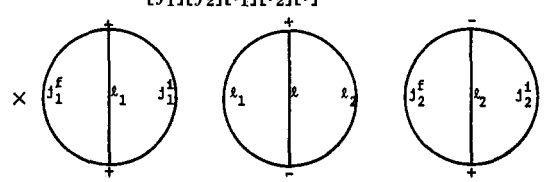
The added phase factor comes from changing the sign at one node. We can now split out the "bubble" in the middle and use Eq. (3a) for the line on the right:

$$A_2 = \frac{h(l,f,i)h(l,i,f)\delta_{\gamma,-\gamma'}(-1)^{l+\gamma+j_2^f+j_2^i+l_2}}{[j_1^i][j_2^i][l_2][l]}$$

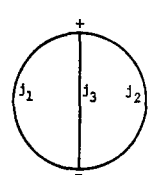


Splitting out the bubble again, we get

$$A_2 = \frac{h(l,f,i)h(l,i,f)\delta_{\gamma,-\gamma'}(-1)^{l+\gamma+j_2^f+j_2^i+l_2}}{[j_1^i][j_2^i][l_1][l_2][l]}$$



Now we note that



$$= \sum_{m_1 m_2 m_3} \begin{pmatrix} j_1 & j_2 & j_3 \\ m_1 & m_2 & m_3 \end{pmatrix}\begin{pmatrix} j_1 & j_2 & j_3 \\ m_1 & m_2 & m_3 \end{pmatrix} = 1.$$

This finally results in the expression (after changing a few signs at nodes and removing a few arrows)

$$A_2 = \frac{h(l,f,i)h(l,i,f)\delta_{\gamma,-\gamma'}(-1)^{\gamma+j_1^i+j_1^f+j_2^i+j_2^f}}{[j_1^i][j_2^i][l_1][l_2][l]}.$$

This now yields for the transition probability

$$P_1(i\to f) = \frac{h(l,f,i)h(l,i,f)(-1)^{j_1^i+j_1^f+j_2^i+j_2^f}}{[j_1^i][j_2^i][l_1][l_2][l]} \times \sum_\gamma (-1)^\gamma I_1(l,\gamma) I_1^*(l,-\gamma). \quad (14)$$

If the h's are now replaced by their algebraic expressions, we will arrive at the same result that was originally obtained algebraically in Paper I.

As a final example we will do the angular momentum sums for the second Born approximation. The second-order transition probability can be written as[1]

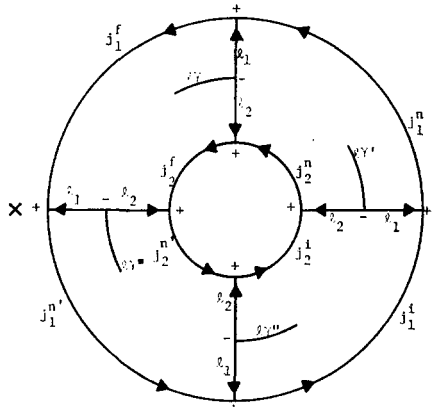
$$P_2(i\to f) = \sum_{\gamma' s} A_4 I_2(l,\gamma,\gamma') I_2^*(l,\gamma'',\gamma'''), \quad (15a)$$

where

$$A_4 = ([j_1^i][j_2^i])^{-1} \sum \langle f|\, T_{l-\gamma}\,|n\rangle\langle n|\, T_{l-\gamma'}\,|i\rangle \times \langle i|\, T_{l-\gamma''}\,|n'\rangle\langle n'|\, T_{l-\gamma'''}\,|f\rangle. \quad (15b)$$

We will concern ourselves with the calculation of A_4. $I_2(l,\gamma,\gamma')$ is a second-order translation integral defined in Eq. (20b). As before, we write out the graphical representation of the matrix elements:

$$A_4 = ([j_1^i][j_2^i])^{-1} h(l,f,n)h(l,n,i)h(l,i,n')h(l,n',f)$$

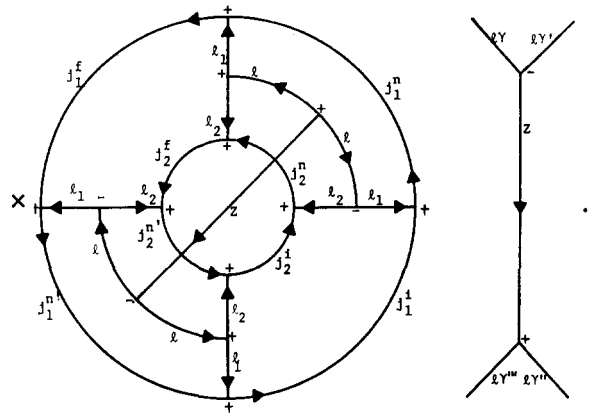


(15c)

It is worth noting the similarity of this diagram to the one for the first Born approximation. As one might expect, this circular ring character can be generalized to the nth-order Born approximation. We will show this to be true later. It really stems from the rotational invariance of the transition probability. If the initial and final projection quantum numbers were not summed over, the circle would be broken. This ring character would also arise in any perturbation theory (whether in scattering or not), if the potential is expressed as a sum of spherical tensor operators and the basis is in the angular momentum representation.

There is no unique way to proceed with manipulating this diagram. We want to be guided to a final result that can be easily evaluated numerically. Certainly, whatever outward form the final results take, they are all mathematically equivalent. Since there are four open lines we can no longer split out the inner circle. For simplicity we shall only be concerned with $\bar{A}_4$, the diagram in A_4. It is instructive to note that $\bar{A}_4$ can be written as

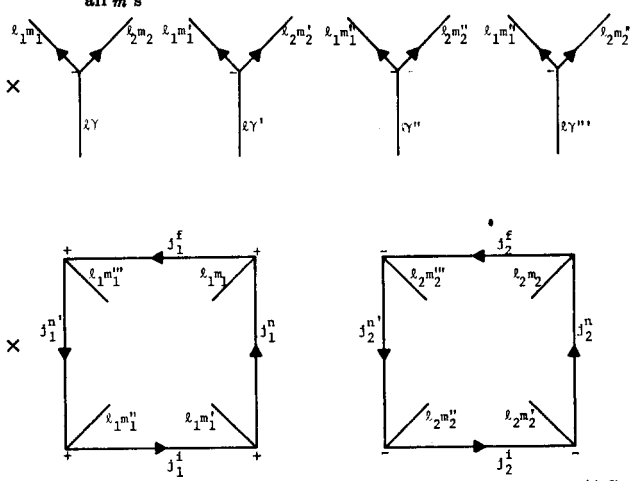
$$\bar{A}_4 = \sum_Z [Z]$$



This is of little practical importance since the closed diagram is a 21-j symbol. However, this is what one would expect for the second Born approximation. The potential is a sum of products of two tensor operators $[Y_{l_1 m_1}(1) Y_{l_1 m_1}(2)]$ which are coupled together. Since the potential enters four times in the second Born approximation, we are coupling together eight tensor operators to form a scalar. This coupling of eight angular momenta is represented by a 21-j symbol.

We shall proceed with $\bar{A}_4$ by first breaking it into a product of smaller pieces and then putting the pieces back together in a different fashion. $\bar{A}_4$ can be written as

$$\bar{A}_4 = \sum_{\text{all } m\text{'s}}$$

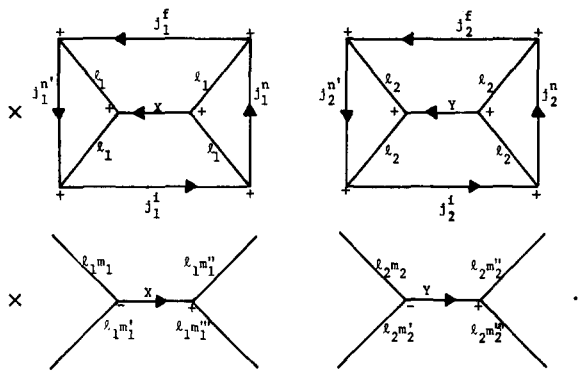


(16)

We shall now just focus on the rectangular diagrams; they can be written as a double sum over dummy

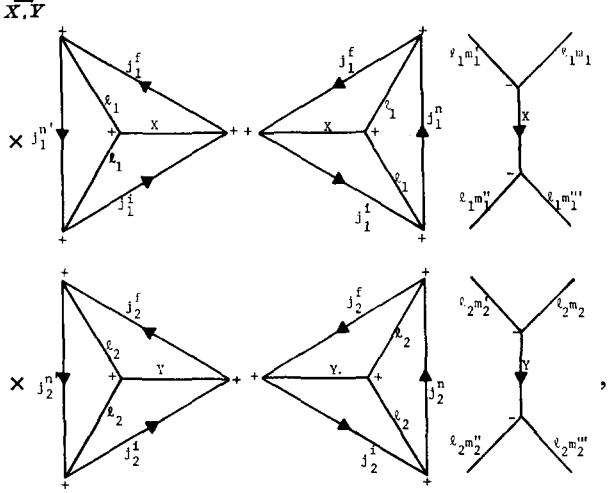
angular momenta X and Y:

$$\sum_{X,Y}[X][Y]$$



We have changed the signs on the corners of the second rectangle, but this does not give a phase factor. Each of the rectangles can now be split on the three lines connecting the right and left sides to give

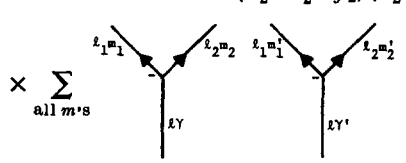
$$\sum_{X,Y}(-1)^{\delta}[X][Y]$$

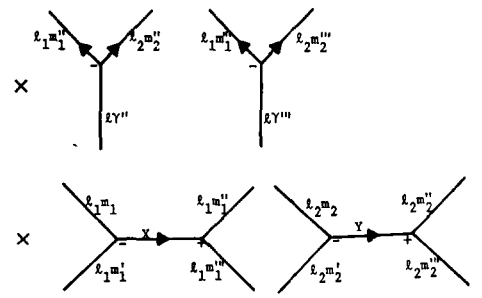


where $\delta = j_1^i + j_1^f + j_2^i + j_2^f + X + Y$. The phase factor comes from changing the signs at two nodes. Since these triangular diagrams are 6-j symbols, we have [by replacing the rectangular diagrams in Eq. (16) with the last expression]

$$\bar{A}_4 = \sum_{X,Y}[X][Y](-1)^{\delta}\begin{Bmatrix} j_1^i & j_1^f & X \\ l_1 & l_1 & j_1^n \end{Bmatrix}\begin{Bmatrix} j_1^i & j_1^f & X \\ l_1 & l_1 & j_1^{n'} \end{Bmatrix} \times \begin{Bmatrix} j_2^i & j_2^f & Y \\ l_2 & l_2 & j_2^n \end{Bmatrix}\begin{Bmatrix} j_2^i & j_2^f & Y \\ l_2 & l_2 & j_2^{n'} \end{Bmatrix}$$

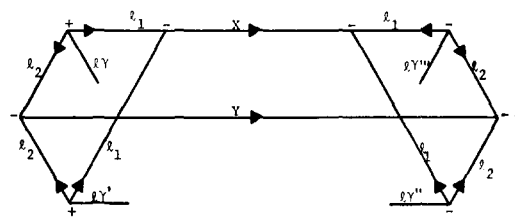
$$\times \sum_{\text{all } m\text{'s}}$$





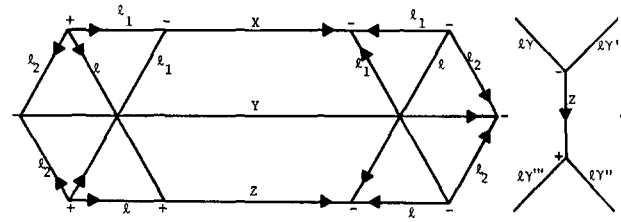
(17)

We now want to evaluate the sum over the m's. This sum can be written in the following form:



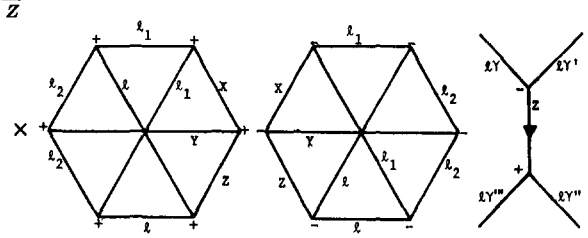
This can be written as

$$\sum_{Z}[Z] \times$$



This diagram can now be cut along the three lines labeled X, Y, Z to get

$$\sum_{Z}[Z]$$



$$\times (-1)^{X+Y}.$$

The phase factor comes from changing the signs on two nodes at the left hexagon. These hexagonal diagrams are 9-j symbols. Writing these out, we see that the above expression becomes

$$\sum_{Z,m_Z}[Z](-1)^{X+Y+Z+m_Z}\begin{Bmatrix} l_2 & Y & l_2 \\ l & Z & l \\ l_1 & X & l_1 \end{Bmatrix}\begin{Bmatrix} Z & Y & X \\ l & l_2 & l_1 \\ l & l_2 & l_1 \end{Bmatrix} \times \begin{pmatrix} Z & l & l \\ m_Z & \gamma & \gamma' \end{pmatrix}\begin{pmatrix} Z & l & l \\ -m_Z & \gamma''' & \gamma'' \end{pmatrix}.$$

We can now permute the columns of the 9-j symbols and reflect along a diagonal to give another factor of $(-1)^{X+Y+Z}$, finally getting

$$\sum_{Z,m_Z}[Z](-1)^{m_Z}\begin{Bmatrix} X & Y & Z \\ l_1 & l_2 & l \\ l_1 & l_2 & l \end{Bmatrix}^2 \times \begin{pmatrix} Z & l & l \\ m_Z & \gamma & \gamma' \end{pmatrix}\begin{pmatrix} Z & l & l \\ -m_Z & \gamma''' & \gamma'' \end{pmatrix}.$$

Plugging this expression back into Eq. (17) gives

$$\begin{aligned}\bar{A}_4 = \sum_{X,Y,Z}[X][Y][Z](-1)^{X+Y+j_1{}^i+j_1{}^f+j_2{}^i+j_2{}^f} \\ \times \begin{Bmatrix} j_1{}^i & j_1{}^f & X \\ l_1 & l_1 & j_1{}^{n'} \end{Bmatrix}\begin{Bmatrix} j_1{}^i & f_1{}^f & X \\ l_1 & l_1 & j_1{}^n \end{Bmatrix} \\ \times \begin{Bmatrix} j_2{}^i & j_2{}^f & Y \\ l_2 & l_2 & j_2{}^n \end{Bmatrix}\begin{Bmatrix} j_2{}^i & j_2{}^f & Y \\ l_2 & l_2 & j_2{}^{n'} \end{Bmatrix}\begin{Bmatrix} X & Y & Z \\ l_1 & l_2 & l \\ l_1 & l_2 & l \end{Bmatrix}^2 \\ \times \sum_{m_Z}(-1)^{m_Z}\begin{pmatrix} Z & l & l \\ m_Z & \gamma & \gamma' \end{pmatrix}\begin{pmatrix} Z & l & l \\ -m_Z & \gamma''' & \gamma'' \end{pmatrix}.\end{aligned} \tag{18}$$

When we put in the factors of $h(\ )$ from Eq. (12b) and $\bar{A}_4$ into A_4, we will arrive at the same expression for $P_2(i \to f)$ that we originally obtained algebraically in Paper I. We will not write this out in detail, but we can check the phase. Algebraically we arrived at the over-all phase $(-1)^{j_1{}^n+j_2{}^n+j_1{}^{n'}+j_2{}^{n'}+X+Y}$. Now graphically (with the phases from the h's) we get

$$\begin{aligned}(-1)^{X+Y+j_1{}^i+j_1{}^f+j_2{}^i+j_2{}^f} \\ \times (-1)^{j_1{}^f+j_2{}^n+j_1{}^n+j_2{}^i+j_1{}^i+j_2{}^{n'}+j_1{}^{n'}+j_2{}^f} \\ = (-1)^{X+Y+j_1{}^n+j_2{}^n+j_1{}^{n'}+j_2{}^{n'}}.\end{aligned}$$

So the phases check.

We have derived the same expressions for the first and second Born approximations without pulling any punches. It is always possible to do the necessary manipulations algebraically, but it can be very complicated and difficult at times. Graphical methods provide a rapid and efficient way of handling angular momentum coupling of any complexity. It is worth noting that the phases can also be derived graphically, as we did above.

III. GRAPHICAL METHODS AND THE nth-ORDER BORN APPROXIMATION

We can at least formally write out an expression for the general nth-order Born approximation with a multipole potential in the impact parameter framework. As we shall see, it *may* be possible to compute efficiently some terms of order higher than second. The problems involved are essentially twofold: the angular momentum algebra and the Born integrals. One must also be able to treat the mixed terms or interference terms between different orders in the Born series expansion. We shall show how a general formalism can be set up and how one may compute these terms up to perhaps third or fourth order. The angular momentum algebra will be treated first, followed by methods of evaluating the Born integrals.

First we write out an expression for the Born expansion of the S matrix[9]:

$$S = \sum_{p=0}^{\infty}\left(\frac{-i}{\hbar}\right)^p S_p, \tag{19a}$$

where

$$S_p = \int_{-\infty}^{+\infty}\bar{V}(t_1)\,dt_1\int_{-\infty}^{t_1}\bar{V}(t_2)\,dt_2\cdots\int_{-\infty}^{t_{p-1}}\bar{V}(t_p)\,dt_p. \tag{19b}$$

We are in the interaction representation, hence

$$\bar{V}(t) = e^{iH_0t/\hbar}V(t)^{-iH_0t/\hbar}, \tag{19c}$$

where $V(t)$ is the intermolecular potential and H_0 is the unperturbed Hamiltonian.[1] S_p is an operator above. If S_p is put into matrix form, the initial and final states will be connected through $p-1$ intermediate states. The states are $|n_1\rangle, |n_2\rangle, \cdots, |n_{p-1}\rangle$, where $|n\rangle$ was defined in Sec. II. The matrix elements of S_p are

$$\begin{aligned}\langle f|\,S_0\,|i\rangle &= \langle f|\,1\,|i\rangle = \delta_{if}, \\ \langle f|\,S_1\,|i\rangle &= \sum_l\sum_\gamma\langle f|\,T_{l-\gamma}\,|i\rangle\,I_1(l,\gamma), \\ \langle f|\,S_2\,|i\rangle &= \sum_{l,l'}\sum_{\gamma,\gamma'}\sum_n\langle f|\,T_{l-\gamma}\,|n_1\rangle\langle n_1|\,T_{l^1-\gamma^1}\,|i\rangle \\ &\quad\times I_2(l,l^1,\gamma,\gamma^1), \\ &\cdots \\ \langle f|\,S_p\,|i\rangle &= \sum_{l's}\sum_{\gamma's}\sum_{n's}\langle f|\,T_{l-\gamma}\,|n_1\rangle\langle n_1|\,T_{l1-\gamma1}\,|n_2\rangle \\ &\quad\times\langle n_2|\,T_{l2-\gamma2}\,|n_3\rangle \\ &\quad\times\cdots\times\langle n_{p-1}|\,T_{l^{p-1}-\gamma^{p-1}}\,|i\rangle\,I_p(\{l,\gamma\}),\end{aligned} \tag{20a}$$

where

$$\begin{aligned}I_1(l,\gamma) &= \alpha_l\int_{-\infty}^{+\infty}f^l_{\gamma,f,i}(t_1)\,dt_1, \\ I_2(l,l^1,\gamma,\gamma^1) &= \alpha_l\alpha_{l^1}\int_{-\infty}^{+\infty}f^l_{\gamma,f,n_1}(t_1)\int_{-\infty}^{t_1}f^{l^1}_{\gamma^1,n_1,i}(t_2)\,dt_1\,dt_2, \\ &\cdots \\ I_p(\{l,\gamma\}) &= \prod_{k=0}^{p-1}\alpha_{lk}\int_{-\infty}^{+\infty}f^l_{\gamma,f,n_1}(t_1) \\ &\quad\times\int_{-\infty}^{t_1}f^{l^1}_{\gamma^1,n_1,n_2}(t_2) \\ &\quad\times\cdots\times\int_{-\infty}^{t_{p-1}}f^{l^{p-1}}_{\gamma^{p-1},n_{p-1},i}(t_p)\,dt_1\,dt_2\cdots dt_p,\end{aligned} \tag{20b}$$

and

$$f^{l}_{\gamma,n,n'}(t) = \frac{e^{i\omega_{n,n'}t}Y_{l\gamma}(\Omega)}{(b^2 + v^2t^2)^{\frac{1}{2}(l+1)}},$$

where

$$h'\omega_{n,n'}(|n\rangle - |n'\rangle) = H_0(|n\rangle - |n'\rangle),$$

$$\alpha_l = \frac{4\pi(-1)^{l_2}}{[l]}\left(\frac{4\pi(2l+1)!}{(2l_1+1)!\,(2l_2+1)!}\right)^{\frac{1}{2}} Q_{l_1}Q_{l_2}.$$

H_0, Q_{l_1}, and Q_{l_2} are, respectively, the unperturbed Hamiltonian and the electric quadrupole moments of order l_1 and l_2. Papers I and II describe I_1 and I_2 in detail.[1,10] When the sum over the l's is performed, we mean to imply a summation over l_1 and l_2 as well. This is needed in order to consider the general multipole potential

$$V(t) = \sum_l V_l(t), \quad l = l_1 + l_2.$$

In practice, this sum will be taken up to some value $l_{\max}$.

The transition probability is written as

$$P(i \to f) = \frac{1}{[j_1^i][j_2^i]} \sum_{\mu_1^i\mu_1^f} \sum_{\mu_2^i\mu_2^f} |\langle f|\, S\, |i\rangle|^2, \quad i \neq f. \tag{21}$$

The summation is over initial and final projection quantum numbers. We want to examine this sum in detail. First note that, since $T_{l\gamma}$ is a spherical tensor operator,[3,7]

$$\langle n|\, T_{l\gamma}\, |n'\rangle^* = (-1)^{\gamma}\langle n'|\, T_{l-\gamma}\, |n\rangle.$$

This means that the complex conjugate matrix element can be written as

$$\langle f|\, S_{p'}\, |i\rangle^* = \sum_{l\text{'s}} \sum_{\bar\gamma\text{'s}} \sum_{\bar n\text{'s}} \langle i|\, T_{l-\bar\gamma}\, |\bar n_1\rangle\langle \bar n_1|\, T_{l^1-\bar\gamma^1}\, |\bar n_2\rangle$$
$$\times \cdots \times \langle \bar n_{p'-1}|\, T_{l^{p'-1}-\bar\gamma^{p'-1}}\, |f\rangle\, I^*_{p'}(\{\bar l, \bar\gamma, \bar n\}). \tag{22}$$

We have made use of the fact that[1] $Y_{l\gamma}(\Omega) \equiv Y_{l\gamma}(\theta, 0)$ is real. This implies[7] $Y^*_{l\gamma}(\Omega) = Y_{l-\gamma}(\Omega)(-1)^{\gamma} = Y_{l\gamma}(\Omega)$. Substituting Eqs. (20a) and (22) into (21), we arrive at the expression for the transition probability,

$$P(i \to f) = \sum_{p,p'} \sum_{l,\bar l} \sum_{\gamma,\bar\gamma} \sum_{n,\bar n} \sum_{\mu_1^i\mu_1^f} \sum_{\mu_2^i\mu_2^f} \left(\frac{-i}{\hbar}\right)^{p+p'} \langle f|\, T_{l-\gamma}\, |n_1\rangle$$
$$\times \cdots \times \langle n_{p-1}|\, T_{l^{p-1}-\gamma^{p-1}}\, |i\rangle\langle i|\, T_{\bar l-\bar\gamma}\, |\bar n_1\rangle$$
$$\times \langle \bar n_1|\, T_{l^1-\bar\gamma^1}\, |\bar n_2\rangle \cdots \langle \bar n_{p'-1}|\, T_{l^{p'-1}-\bar\gamma^{p'-1}}\, |f\rangle$$
$$\times \frac{I_p(\{l, \gamma, n\})I^*_{p'}(\{\bar l, \bar\gamma, \bar n\})}{[j_1^i][j_2^i]}. \tag{23}$$

This expression includes all the cross terms between the different orders of perturbation theory. Now we focus on the summation over all the projection quantum numbers in Eq. (23):

$$([j_1^i][j_2^i])^{-1} \sum_{q\bar q} \sum_{\mu_1^i\mu_1^f} \sum_{\mu_2^i\mu_2^f} \langle f|\, T_{\bar l-\gamma}\, |n_1\rangle\langle n_1|\, T_{l^1-\gamma^1}\, |n_2\rangle \times \cdots$$
$$\times \langle n_{p-1}|\, T_{l^{p-1}-\gamma^{p-1}}\, |i\rangle\langle i|\, T_{\bar l-\bar\gamma}\, |\bar n_1\rangle \times \cdots$$
$$\times \langle \bar n_{p'-1}|\, T_{l^{p'-1}-\bar\gamma^{p'-1}}\, |f\rangle. \tag{24}$$

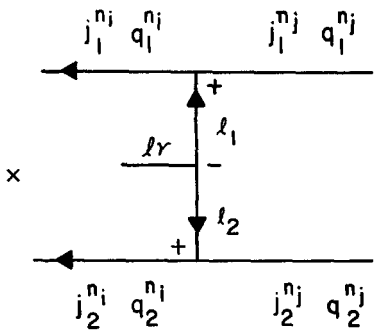
If each matrix element is considered a link in a chain, then this sum is a closed chain (note that the state $|f\rangle$ is on the right and left ends in the sum). This observation leads to the basic structure of the graphical representation in Eq. (24). Without any loss of generality, we will change notation a little and consider the more symmetric sum

$$A_r = \sum \frac{\langle n_1|\, T_{l-\gamma}\, |n_2\rangle\langle n_2|\, T_{l^1\gamma^1}\, |n_3\rangle \cdots \langle n_r|\, T_{l^{r-1}-\gamma^{r-1}}\, |n_1\rangle}{[j_1^i][j_2^i]} \tag{25}$$

There are r matrix elements in the sum. In Sec. II A_2 and A_4 were considered for the first and second Born approximations, respectively. Again note that the sum in Eq. (25) is over all projection quantum numbers.

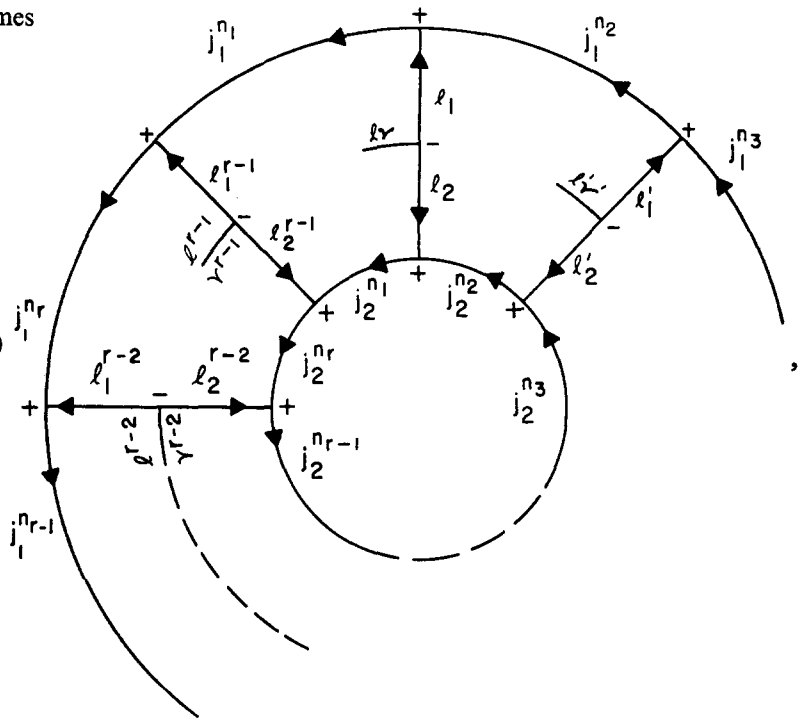
Now the task is to express Eq. (25) graphically and evaluate it in some convenient form. This is done by building up the diagram from the basic matrix element in Eq. (12):

$\langle n_i|\, T_{l-\gamma}\, |n_j\rangle = h(l, n_i, n_j)$



Using this, we see that A_r becomes

$$A_r = \prod_{k=0}^{r-1} \frac{h(l^k, n_{k+1}, n_{k+2})}{[j_1^k][j_2^k]}$$

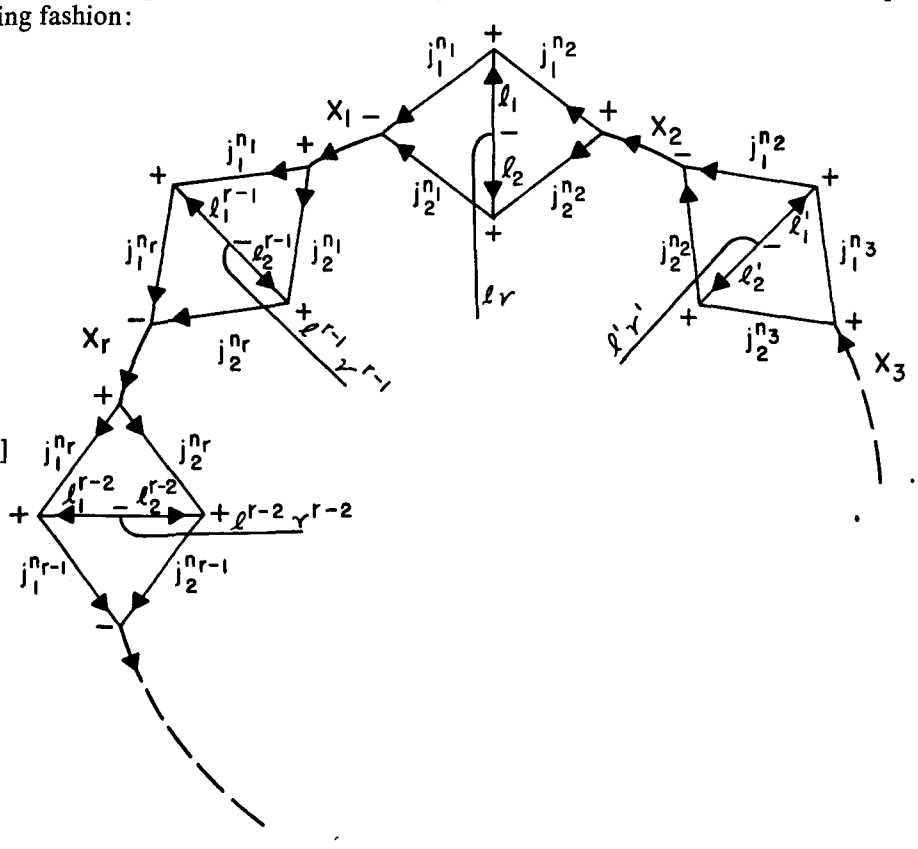


, (26)

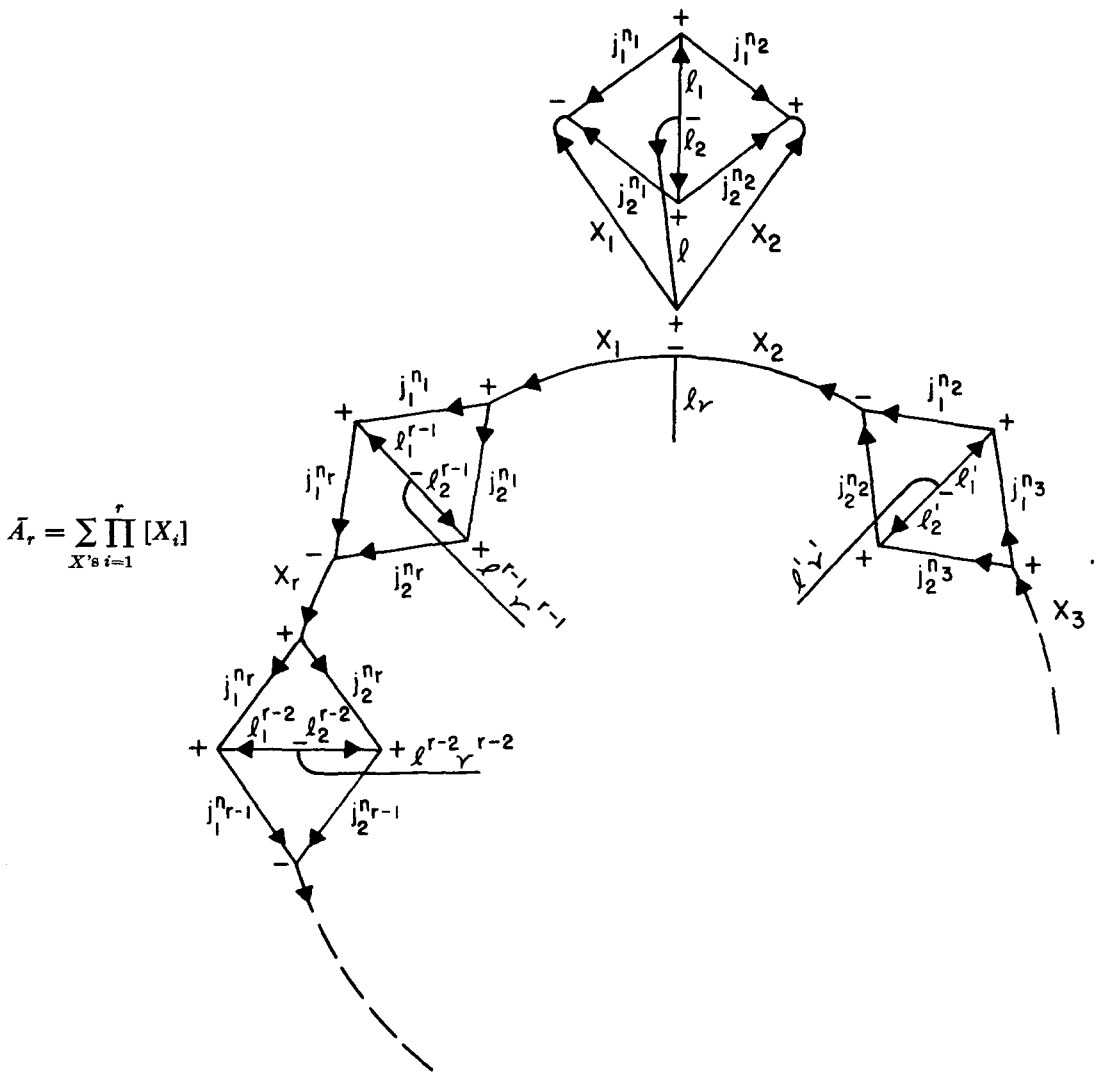
where $n_{r+1} \equiv n_1$.

Equations (13) and (15c) are special cases of Eq. (26). Each of the open lines couples with a tensor operator $Y_{l^j,\gamma^j}(\Omega)$, $j = 0, 1, \cdots, r-1$, in the Born integrals. We want to concern ourselves with the circular diagrams $\bar{A}_r$. We shall follow the procedure outlined in Appendix A. The two outer circles can be "pinched" together in the following fashion:

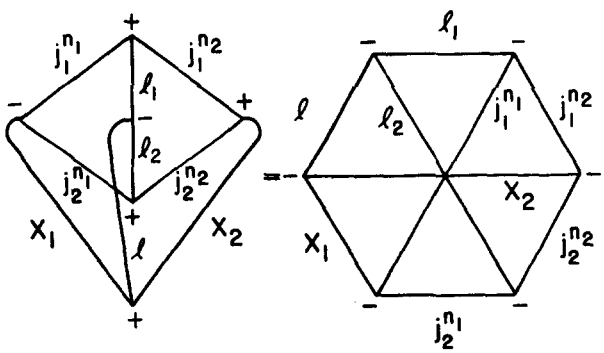
$$\bar{A}_r = \sum_{X\text{'s}} \prod_{i=1}^{r} [X_i]$$



In this form the diagram can be split into pieces. Note that each diamond-shaped piece can be considered to have three open lines connecting it to the remaining portion of the circle (including lines like l_γ). Since each diamond figure is in normal form, it can be split out:



This can be continued around the circle until all the"diamonds" are split out. Each of the split out groups can be deformed to a symmetric diagram, a 9-j symbol:



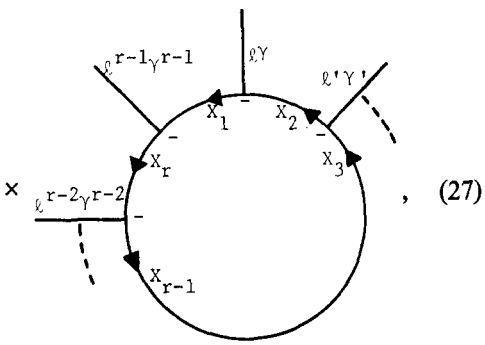
$$\times (-1)^{l_2+j_2^{n_1}+j_2^{n_2}+l+X_1+X_2}$$

$$= \begin{Bmatrix} X_1 & X_2 & l \\ j_2^{n_1} & j_2^{n_2} & l_2 \\ j_1^{n_1} & j_1^{n_2} & l_1 \end{Bmatrix} (-1)^{X_1+X_2+l_1+j_2^{n_1}+j_2^{n_2}}$$

$$= \begin{Bmatrix} X_1 & X_2 & l \\ j_1^{n_1} & j_1^{n_2} & l_1 \\ j_2^{n_1} & j_2^{n_2} & l_2 \end{Bmatrix} (-1)^{j_1^{n_1}+j_1^{n_2}+l_1}.$$

Using this result, we see that $\bar{A}_r$ becomes

$$\bar{A}_r = \sum_{X's} \prod_{i=1}^{r} [X_i](-1)^{j_1^{n_i}+j_1^{n_{i+1}}+l_2^{i-1}} \begin{Bmatrix} X_i & X_{i+1} & l^{i-1} \\ j_1^{n_i} & j_1^{n_{i+1}} & l_1^{i-1} \\ j_2^{n_i} & j_2^{n_{i+1}} & l_2^{i-1} \end{Bmatrix}$$

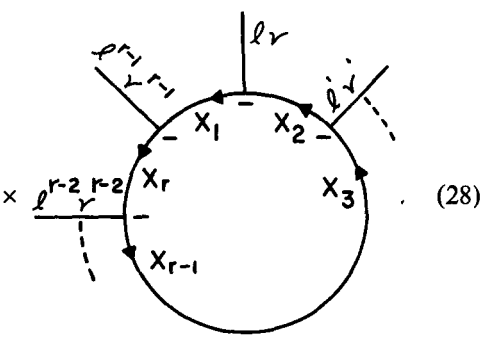


(27)

where $n_{r+1} \equiv n_1$ and $X_{r+1} \equiv X_1$ by definition. The circular diagram that remains is a product of 3-j symbols, each of which has one free index $l^j\gamma^j$, $j = 1, 2, \cdots, r-1$. This diagram is not a generalized Wigner coefficient[3] as defined by Yutsis *et al.* This is true since the number of contracted angular momenta is not minimal. By using Eq. (35) in Appendix A, expressions can be derived for $\bar{A}_r$ which have a generalized Wigner coefficient for the remaining open diagram. This would have three fewer dummy angular momenta $(X_1 \cdots X_{r-3})$, but it would be less symmetrical.

We have now reduced $\bar{A}_r$ to a useful expression. The 9-j symbols are not too difficult to handle. The triangle condition on the arguments in the 9-j symbols reduces the number of terms in the summation over $\{X_i\}$. Now we substitute Eq. (27) into Eq. (26) to get

$$A_r = \frac{1}{[j_1^i][j_2^i]} \sum_{x's} \prod_{i=1}^{r} \prod_{k=0}^{r-1} h(l^k, n_{k+1}, n_{k+2}) \times [X_i](-1)^{j_1^{n_i}+j_1^{n_{i+1}}+l_2^{i-1}} \begin{Bmatrix} X_i & X_{i+1} & l^{i-1} \\ j_1^{n_i} & j_1^{n_{i+1}} & l_1^{i-1} \\ j_2^{n_i} & j_2^{n_{i+1}} & l_2^{i-1} \end{Bmatrix}$$



(28)

If Eq. (28) is applied to the first or second Born approximation, the results will not appear like the ones derived in Sec. I or Paper I.[1] However, both results are equivalent; appropriate manipulation of the diagrams or symbols will show their equality. For the general case we have derived a formula which applies to all orders of perturbation theory. It may well happen that, for, say, the third Born approximation, a more convenient expression exists than our general result. One can endlessly manipulate the diagrams to any desired form in a particular case.

The remaining problem is to calculate the integrals $I_p(\{l, \gamma\})$ in Eq. (20b). It is possible to obtain asymptotic forms for these integrals in the limits $\omega_{n,n'} \to 0$ and $\omega_{n,n'} \to \infty$. This is done by looking at the asymptotic form of the integral

$$\int_{-\infty}^{t} f^l_{\gamma,n,n'}(t')\, dt'.$$

However, we would like a more general result for all $\omega_{n,n'}$. The basic problem with these integrals is their rapidly oscillating integrands. Ordinary quadrature procedures will not work with such integrands. However, we have found two possible approaches: the one considered in Paper I[1] for I_2 and an extension of methods used by Griem *et al.*[11]

In Paper I we approximated the *smooth* portion of the integrand by a sum of exponentials in time. The smooth portion of $f^l_{\gamma,n,n'}(t)$ can be expressed in terms of $(1 + z^2)^{-k-\frac{1}{2}}$, $z = vt/b$. v and b are the relative velocity and impact parameter, respectively. This function is replaced by

$$(1 + z^2)^{-k-\frac{1}{2}} \simeq \sum_{\beta=1}^{\beta^*} C_\beta(k) e^{-\beta|z|}.$$

$\beta^* \simeq 10$ for a good fit.[10] The coefficients $C_\beta(k)$ are determined by requiring the functions to be equal at several points. The remaining integrals can be done analytically, and the results are quite accurate[10] (to $\lesssim 1\%$ or so) for $0 \leq x \leq 10$ or 15 with I_2.

Griem *et al.*[11] evaluated integrals similar to I_2 by obtaining the real part of I_2 and expressing the imaginary part in terms of a principal value integral over the real part. This procedure can be extended to obtain the integrals I_p by a recursion procedure. The method is based on the fact that we can get the real or imaginary part (not both) of the pth-order integral in terms of all $p - 1$ lower-order integrals. Cauchy's integral theorem[12] can then be applied to get the remaining real or imaginary part. The complex conjugate of I_p will be needed to show that the real or imaginary part can be obtained from lower-order

integrals. First consider that

$$f^*_{\gamma,n,n'}(t) = \frac{e^{-i\omega_{nn'}t}Y_{l\gamma}(\Omega)}{(b^2 + v^2t^2)^{\frac{1}{2}(l+1)}}$$

since $Y_{l\gamma}(\Omega)$ is real (the azimuthal angle is taken as zero). Also note that

$$f^*_{\gamma,n,n'}(-t) = (-1)^{l+\gamma}f_{\gamma,n,n'}(t).$$

This follows from the properties of the associated Legendre polynomials. Now let $\{t_i\} \to \{-t_i\}$ in $I^*_p(\{l, \gamma, n\})$. This gives

$$I^*_p(\{l, \gamma, n\}) = (-1)^p \prod_{j=0}^{p-1} \alpha_{lj}(-1)^{(l^j+\gamma^j)} \int_{+\infty}^{-\infty} f^l_{\gamma,f,n_1}(t_1)\,dt_1 \times \int_{+\infty}^{t_1} f^{l^1}_{\gamma^1,n_1,n_2}(t_2)\,dt_2 \times \cdots \times \int_{+\infty}^{t_{p-1}} f^{l^{p-1}}_{\gamma^{p-1},n_{p-1},i}(t_p)\,dt_p. \quad (29)$$

Note the change in sign on the lower limits of the integrals. This integral can be rewritten by using the relation

$$\int_{+\infty}^{t} = -\int_{-\infty}^{+\infty} + \int_{-\infty}^{t}.$$

Equation (29) now becomes

$$I^*_p(\{l, \gamma, n\}) = (-1)^{p-1} \prod_{j=0}^{p-1} \alpha_{lj}(-1)^{(l^j+\gamma^j)} \int_{-\infty}^{+\infty} f^l_{\gamma,f,n}(t_1)\,dt_1 \times \left[\left(\int_{-\infty}^{t_1} - \int_{-\infty}^{+\infty}\right) f^{l^1}_{\gamma^1,n_1,n_2}(t_2)\,dt_2\right] \times \cdots \times \left[\left(\int_{-\infty}^{t_{p-1}} - \int_{-\infty}^{+\infty}\right) f^{l^{p-1}}_{\gamma^{p-1},n_{p-1},i}(t_p)\,dt_p\right]. \quad (30)$$

It is useful to define two quantities to see the structure of Eq. (30):

$$A_i = \int_{-\infty}^{t_{i-1}} f^{l^{i-1}}_{\gamma^{i-1},n_{i-1},n_i}(t_i)\,dt_i,$$

$$B_i = \int_{-\infty}^{+\infty} f^{l^{i-1}}_{\gamma^{i-1},n_{i-1},n_i}(t_i)\,dt_i.$$

A_i and B_i are to be interpreted as integral operators. With these definitions, Eq. (30) becomes

$$I^*_p(\{l, r, n\}) = (-1)^{p-1} \prod_{j=0}^{p-1} \alpha_{l^j}(-1)^{(l^j+\gamma^j)} \times \int_{-\infty}^{+\infty} f^l_{\gamma,f,n_1}(t_1)\,dt_1 \prod_{i=2}^{p} (A_i - B_i).$$

There is one term in the product of i that can be written as $\prod_{i=2}^{p} A_i$. This term gives rise to $I_p(\{l, r, n\})$ again. Hence,

$$I^*_p(\{l, \gamma, n\}) = (-1)^{\delta_p+1} I_p(\{l, \gamma, n\}) + (-1)^{\delta_p+1} \prod_{j=0}^{p-1} \alpha_{lj} \int_{-\infty}^{+\infty} f^l_{\gamma,f,n_1}(t_1)\,dt_1 \prod_{i=2}^{p}{}' (A_i - B_i),$$

where

$$\prod_{i=2}^{p}{}' (A_i - B_i) = \prod_{i=2}^{p} (A_i - B_i) - \prod_{i=2}^{p} A_i$$

and

$$\delta_p = \sum_{j=0}^{p-1} (l^j + \gamma^j + 1).$$

Rewriting again, we arrive at

$$\begin{Bmatrix}\mathrm{Re}\\ \mathrm{Im}\end{Bmatrix}_{\delta_p} I_p(\{l, \gamma, n\}) = -\tfrac{1}{2} \prod_{k=0}^{p-1} \alpha_{l^k} \int_{-\infty}^{+\infty} f^l_{\gamma,f,n_1}(t_1)\,dt_1 \prod_{i=2}^{p}{}' (A_i - B_i), \quad (31)$$

where

$$\begin{Bmatrix}\mathrm{Re}\\ \mathrm{Im}\end{Bmatrix}_n = \begin{cases}\mathrm{Re}, & n \text{ even}\\ \mathrm{Im}, & n \text{ odd.}\end{cases}$$

This is our desired result; the real or imaginary part of $I_p(\{l, \gamma, n\})$ can be obtained depending on whether δ_p is an even or odd integer, respectively. The right-hand side is a sum of products if I_s, where $s = 1, 2, \cdots, p-1$. So, if we have the integrals up to order $p-1$, then the real or imaginary part of I_p is determined. This is the key that allows us to get the Born integrals by a recursion procedure. There still remains the problem of getting the remaining real or imaginary part of I_p.

The integral I_p can be considered to be a function of the p-independent frequencies $\omega_{fn_1}, \omega_{n_1n_2}, \cdots, \omega_{n_{p-1}i}$ and the impact parameter and velocity. By an appropriate change of variables, the integrand can be made a function of the dimensionless independent variables $z_1 = \omega_{fn_1}v/b, z_2 = \omega_{n_1n_2}b/v, \cdots, z_p = \omega_{n_{p-1}i}b/v$. Suppose we remove the independence of all the $\{z_i\}$ and define the dependent variables

$$z_1 = z_1, \quad z_i = -z_1/(p-1) + \beta_i, \quad i = 2, 3, \cdots, p,$$

where the β_i are treated as constants. There is a reason for doing this. Using the above definitions and the new dimensionless time variables $x_i = t_iv/b, i = 1, \cdots, p$, we see that the exponential in the integrand of I_p has the form

$$\exp\left(i\sum_{j=1}^{p} x_jz_j\right) = \exp\left(\frac{iz_1[(x_1 - x_2) + (x_1 - x_3) + \cdots + (x_1 - x_p)]}{(p-1)}\right) \times \exp\left(i\sum_{j=2}^{p} x_j\beta_j\right).$$

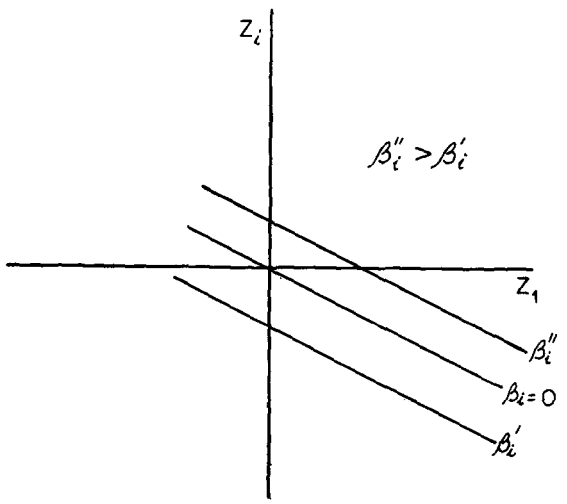


FIG. 1. z_i is plotted as a function of z_1 for various β_i values.

Now consider z_1 as a complex variable with Im $(z_1) > 0$. Then $e^{iz_1(x_1-x_i)/(p-1)} \to 0$ as $|z_1| \to \infty$, Im $(z_1) > 0$. This follows since $x_1 \geq x_i$ or $(x_1 - x_i) \geq 0$. We had to sacrifice the independence of the $\{z_i\}$ to get this result. This is not a serious problem since the $\{\beta_i\}$ can be varied independently. Figure 1 shows how z_i and z_1 are related for different values of β_i (z_1 and z_i are real here).

In Appendix A we show that $I_p(z_1)$ is analytic in the upper half-plane. Now the results in Appendix C can be applied to $I_p(z_1)$. The two Hilbert transform relations in Eq. (C5) can be combined to form

$$\begin{Bmatrix} \text{Re} \\ -\text{Im} \end{Bmatrix}_{\delta_{p+1}} I_p(z_1) = \frac{1}{\pi} P \int_{-\infty}^{+\infty} \frac{1}{z_1' - z_1} \begin{Bmatrix} \text{Re} \\ \text{Im} \end{Bmatrix}_{\delta_p} I_p(z_1')\, dz_1'. \tag{32}$$

So we now have the remaining part of $I_p(z_1)$. Note that $I_p(z_1)$ is still a function of $\{\beta_i\}$. The question is how accurately can the principal value integral be evaluated. The integral in Eq. (33) will show what is required:

$$P \int_{-a}^{+a} \frac{f(z)\, dz}{z}. \tag{33}$$

The singularity in a principal value integral can always be isolated, as in the above integral. The remaining integrals from $-\infty$ to $-a$ and a to ∞ have no singularities and are well behaved [assuming $f(z)$ is well behaved]. These integrals can be handled by appropriate quadrature methods. Now consider the integral in Eq. (33). First note that

$$P \int_{-a}^{+a} \frac{dz}{z} = 0.$$

Hence,

$$P \int_{-a}^{+a} \frac{f(z)\, dz}{z} = P \int_{-a}^{+a} \frac{f(z) - f(0)}{z}\, dz = \int_{-a}^{+a} \frac{f(z) - f(0)}{z}\, dz.$$

The last step follows since as, $z \to 0$,

$$\frac{f(z) - f(0)}{z} \to f'(0).$$

We assume $\lim f'(z)$ as $z \to 0$ is well behaved (i.e., finite). This procedure essentially removes the singularity at the origin. However, we now need the derivative at the origin. This could be a problem since $f(z)$ will be a numerical function in general. This implies that eventually the recursion procedure for I_p will break down. The inaccuracy of the numerical function will not allow the determination of its derivative. However, with the usual quadrature methods it should be possible to calculate the third- or fourth-order integrals before numerical instability sets in.

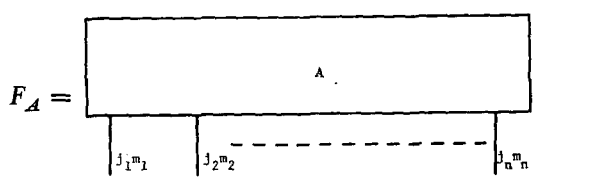
ACKNOWLEDGMENT

One of the authors (H. A. R.) would like to thank Professor P. C. H. Sandars for his excellent lectures on graphical methods.

APPENDIX A

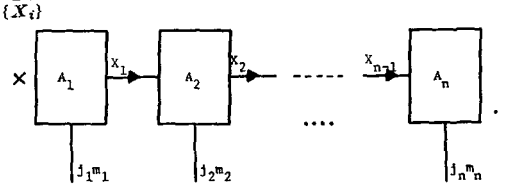
In Sec. I we derived the 1-, 2-, and 3-line cutting rules [Eq. (11)]. We would now like to consider the general n-line cutting rules. These rules are used in handling the general Born term. Yutsis *et al.*[3] derived an n-line cutting rule by using the properties of generalized Wigner coefficients. We shall use Eq. (11) to derive the general case.

With the help of Eq. (3c) the general block with n-lines can be written as



(A1a)

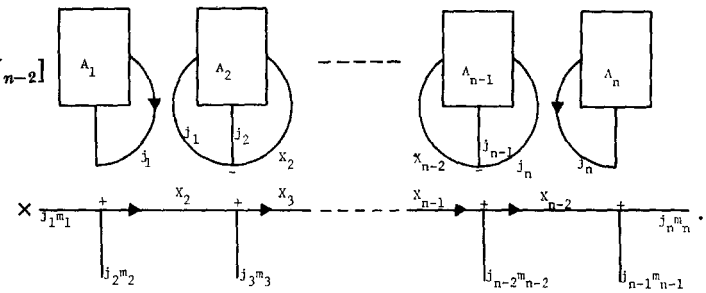
$$= \sum_{\{X_i\}} [X_1] \cdots [X_{n-1}]$$



(A1b)

This may require other "dummy" angular momenta besides $\{X_i\}$, but they will be hidden in the blocks $\{A_i\}$. Now repeatedly apply the 2- and 3-line cutting rules in Eq. (11b) and (11c) to obtain

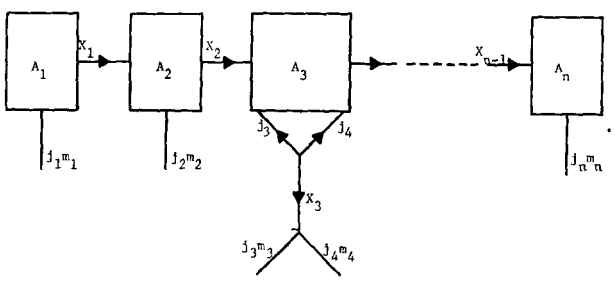
$$F_A = \sum_{\{X_i\}} [X_2] \cdots [X_{n-2}]$$



(A2)

Note that there are $(n-3)$ "dummy" angular momenta $\{X_i\}$. The remaining open diagram is a generalized Wigner coefficient.[3] It is possible to derive other cutting rules which end up with different generalized Wigner coefficients by breaking up F_A in a different fashion; i.e.,

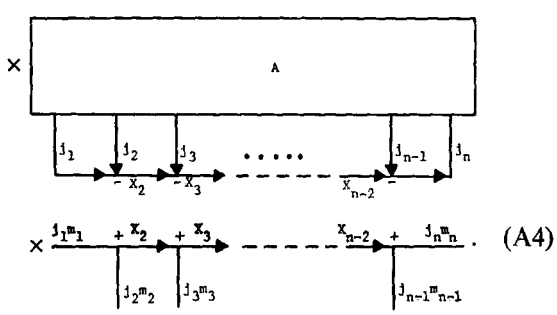
$$F_A = \sum_{\{X_i\}} [X_1] \cdots [X_{n-1}] \times$$



(A3)

It is interesting to note that repeated application of Eq. (11b) on the closed diagrams in Eq. (A2) will yield

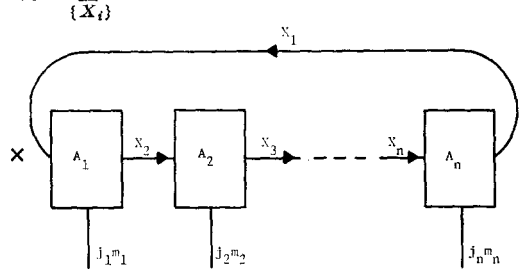
$$F_A = \sum_{\{X_i\}} [X_2] \cdots [X_{n-2}]$$



(A4)

We would now like to consider a cutting rule that does not end up with generalized Wigner coefficients. Again, with the help of Eq. (3c), F_A can be written as

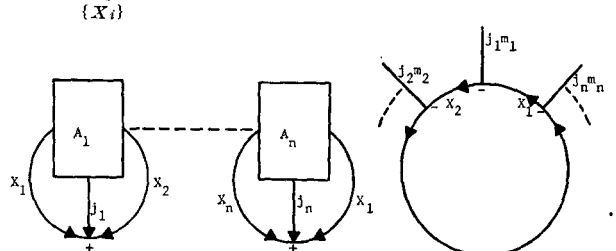
$$F_A = \sum_{\{X_i\}} [X_1] \cdots [X_n]$$



(A5)

This construction may seem artificial, but it arises quite naturally in treating the general Born terms. Repeated application of Eq. (11c) yields

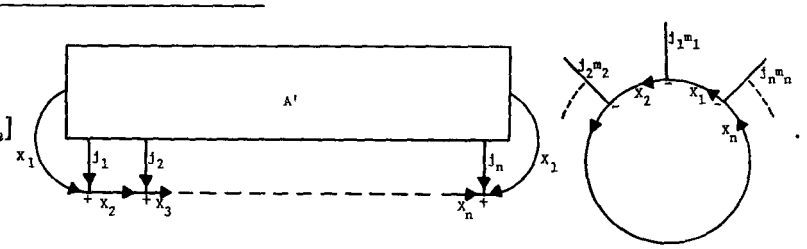
$$F_A = \sum_{\{X_i\}} [X_1] \cdots [X_n] \times$$



(A6)

Although there are n "dummy" angular momenta in Eq. (A6), the resulting open and closed diagrams are more symmetrical than the ones in Eq. (A2). The methods leading to Eq. (A6) were used in Sec. III to derive Eq. (28). Finally, note that by applying Eq. (11b) to Eq. (39) we obtain

$$F_A = \sum_{\{X_i\}} [X_1] \cdots [X_n]$$



(A7)

APPENDIX B

We want to look at the analytic properties of $I_p(z_1)$ in the upper half-plane. It is clear from its integrand that $I_p(z_1) \to 0$ as $|z_1| \to \infty$, Im $(z_1) \geq 0$. In fact, it goes to zero as z_1^{1-p-1} (this can be seen by integration by parts). We want to show that $I_p(z_1)$ is analytic in the upper half-plane. To do this, consider the Taylor expansion about $\tilde{z}_1 = \tilde{u}_1 + i\tilde{v}_1$, $\tilde{v}_1 \geq 0$:

$$I_p(z_1) = \sum_{n=0}^{\infty} \frac{I_p^n(\tilde{z}_1)}{n!} (z_1 - \tilde{z}_1)^n,$$

where

$$I_p^n(\tilde{z}_1) = \frac{d^n}{dz_1^n} I_p(z_1)\Big|_{z_1=\tilde{z}_1}.$$

We will show that this is a convergent series. $I_p^n(\tilde{z}_1)$ is

$$\begin{aligned} I_p^n(\tilde{z}_1) &= \left(\frac{-i}{(p-1)}\right)^n \int_{-\infty}^{+\infty}\int_{-\infty}^{x_1}\cdots\int_{-\infty}^{x_{p-1}} dx_1\cdots dx_p \\ &\times \exp\left(i\frac{\tilde{z}_1}{p-1}[(x_1-x_2)+\cdots+(x_1-x_p)]\right) \\ &\times [(x_1-x_2)+\cdots+(x_1-x_p)]^n \\ &\times \chi_\gamma^l(x_1)\cdots\chi_\gamma^{l^{p-1}}(x_p), \end{aligned}$$

where

$$\chi_\gamma^l(x) = \frac{Y_{l\gamma}(\theta(x), 0)}{(1+x^2)^{\frac{1}{2}(l+1)}}.$$

Now

$$\begin{aligned} |I_p^n(\tilde{z}_1)| \leq \frac{1}{(p-1)^n} &\int_{-\infty}^{+\infty}\int_{-\infty}^{x_1}\cdots\int_{-\infty}^{x_{p-1}} dx_1\cdots dx_p \\ &\times e^{-\tilde{v}_1[(x_1-x_2)+\cdots+(x_1-x_p)]/(p-1)} \\ &\times [(x_1-x_2)+\cdots+(x_1-x_p)]^n \\ &\times |\chi_\gamma^l(x_1)|\cdots|\chi_\gamma^{l^{p-1}}(x_p)|. \end{aligned}$$

It is clear that $\chi_\gamma^l(x)$ dies off as $x \to \infty$. At some value $\bar{x}$, the function $|\chi_\gamma^l(\bar{x})|$ will be a maximum. Hence

$$\begin{aligned} |I_p^n(\tilde{z}_1)| \leq \prod_{j=1}^{p-1} |\chi_\gamma^{l^j}(\bar{x}_{j+1})| &\frac{\partial^n}{\partial\tilde{v}_1^n}\int_{-\infty}^{+\infty} e^{-\tilde{v}_1 x_1}|\chi_\gamma^l(x_1)|\,dx_1 \\ &\times \int_{-\infty}^{x_1} e^{\tilde{v}_1 x_2/(p-1)}\,dx_2\cdots\int_{-\infty}^{x_{p-1}} e^{\tilde{v}_1 x_p/(p-1)}\,dx_p. \end{aligned}$$

Note that $|\chi_\gamma^l(x_1)|$ is still in the integrand. Performing the remaining integrals, we get

$$\begin{aligned} |I_p^n(\tilde{z}_1)| \leq \prod_{j=1}^{p-1} |\chi_\gamma^{l^j}(\bar{x}_{j+1})| &\frac{(p-1)^{p-1}}{(p-1)!}\frac{\partial^n}{\partial\tilde{v}_1^n}\frac{1}{\tilde{v}_1^{p-1}} \\ &\times \int_{-\infty}^{+\infty}|\chi_\gamma^l(x_1)|\,dx_1. \end{aligned}$$

This last integral is clearly convergent to some finite number. After taking the indicated derivatives, we finally arrive at

$$|I_p^n(\tilde{z}_1)| < \tau(p)(n+p-2)!\,(\tilde{v}_1^n)^{-1},$$

where

$$\begin{aligned} \tau(p) = \frac{1}{\tilde{v}_1^{p-1}}&\frac{(p-1)^{p-1}}{(p-1)!\,(p-2)!} \\ &\times \prod_{j=1}^{p-1}|\chi_\gamma^{l^j}(\bar{x}_{j+1})|\int_{-\infty}^{+\infty}|\chi_\gamma^l(x_1)|\,dx_1. \end{aligned}$$

It is clear that $|I_p^n(\tilde{z}_1)|$ exists and is finite for $\tilde{v}_1 > 0$. Now we get for the Taylor expansion of $|I_p(z_1)|$

$$\begin{aligned} |I_p(z_1)| &\leq \tau(p)\sum_{n=0}^{\infty}\frac{(n+p-2)!}{n!}\frac{|z_1-\tilde{z}_1|^n}{\tilde{v}_1^n} \\ &= \frac{\tau(p)(p-2)!\,\tilde{v}_1^{p-1}}{(\tilde{v}_1-|z_1-\tilde{z}_1|)^{p-1}}. \end{aligned}$$

This series is convergent for $p \geqslant 2$ and $|z_1 - \tilde{z}_1| < \tilde{v}_1$. We can vary $\tilde{z}_1$ in the upper half-plane and conclude that the Taylor series for $I_p(z_1)$ is convergent within a circle centered about $\tilde{z}_1$ and just touching the real axis. This implies that $I_p(z)$ is analytic[12] in the upper half-plane.

APPENDIX C

We would like to develop some integral relations known as Hilbert transforms or dispersion relations. These relations are discussed in detail in several sources.[13] Consider the function $f(z)$ to be analytic in the upper half-plane with the property $f(z) \to 0$ as Im $(z) \to \infty$. Also assume that $f(z)$ has no poles in the upper half-plane. This last assumption is not necessary, but the case of interest in this paper has this behavior.

Direct application of Cauchy's integral theorem[12] for a point z interior the closed curve C in Fig. 2 gives

$$f(z) = \frac{1}{2\pi i}\int_C \frac{f(z')\,dz'}{z'-z}.$$

Now let the radius of the semicircle in Fig. 2 go to ∞. The contribution to the integral along the circular arc goes to zero. Hence,

$$2\pi i f(z) = \int_{-\infty}^{+\infty}\frac{f(z')\,dz'}{z'-z}, \quad \text{Im } z > 0. \qquad \text{(C1)}$$

Suppose z approaches the real axis from above. Let

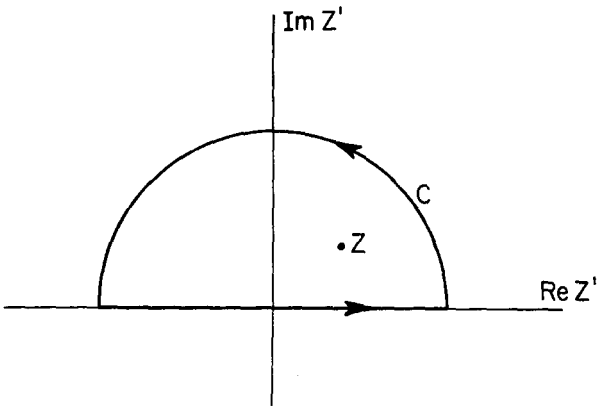


FIG. 2. The curve C is indicated in the complex z' plane.

$z \to z + i\epsilon$, where z is now real and $\epsilon \to 0$. Note that[13]

$$\lim_{\epsilon \to 0_+} \frac{1}{z' - z - i\epsilon} = P\frac{1}{z' - z} + i\pi\delta(z' - z), \quad \text{(C2)}$$

where P denotes the principal value when the appropriate integral is taken. The principal value integral is defined as

$$P\int_{-\infty}^{+\infty} \frac{f(z)}{z}\,dz = \lim_{\epsilon \to 0_+} \left(\int_{-\infty}^{-\epsilon} \frac{f(z)}{z}\,dz + \int_{\epsilon}^{\infty} \frac{f(z)}{z}\,dz\right).$$

Now we apply Eq. (C2) to the integral in Eq. (C1):

$$\lim_{\epsilon \to 0_+} \int_{-\infty}^{+\infty} \frac{f(z')\,dz'}{(z' - z - i\epsilon)} = P\int_{-\infty}^{+\infty} \frac{f(z')\,dz'}{(z' - z)} + i\pi f(z). \quad \text{(C3)}$$

Combining Eqs. (C1) and (C3), we get

$$f(z) = \frac{-i}{\pi} P\int_{-\infty}^{+\infty} \frac{f(z')\,dz'}{z' - z}. \quad \text{(C4)}$$

Now we take the real and imaginary part of this equation to get the Hilbert transform relations

$$\mathrm{Re}\, f(z) = \frac{1}{\pi} P\int_{-\infty}^{+\infty} \frac{\mathrm{Im}\, f(z')\,dz'}{(z' - z)},$$

$$\mathrm{Im}\, f(z) = -\frac{1}{\pi} P\int_{-\infty}^{+\infty} \frac{\mathrm{Re}\, f(z')\,dz'}{(z' - z)}, \quad \mathrm{Im}\, z = 0. \quad \text{(C5)}$$

These are the relations we set out to derive. The real and imaginary parts of an analytic function are not independent.

* Supported in part by a grant from the National Science Foundation.

† Present address: Theoretical Chemistry Institute, University of Wisconsin, Madison, Wisconsin 53706.

[1] Paper I, H. Rabitz and R. Gordon, J. Chem. Phys. **53**, 1815 (1970).

[2] D. M. Brink and G. R. Satchler, *Angular Momentum* (Oxford U.P., London, 1968), 2nd ed.

[3] A. P. Yutsis, I. B. Levinson, and V. V. Vavagas, *Theory of Angular Momentum* (National Science Foundation, Washington, D.C., 1962).

[4] J.-N. Massot, E. El-baz, and J. Lafoucriere, Rev. Mod. Phys. **39**, 288 (1967).

[5] C. F. Curtiss, J. Chem. Phys. **48**, 1725 (1968).

[6] B. R. Judd, *Second Quantization and Atomic Spectroscopy* (Johns Hopkins Press, Baltimore, Md., 1967), Chap. 5 and references cited therein.

[7] A. R. Edmonds, *Angular Momentum in Quantum Mechanics* (Princeton U.P., Princeton, N.J., 1957).

[8] G. Ponzano and T. Regge, "Semiclassical Limit of Racah Coefficients," in *Spectroscopic and Group Theoretical Methods in Physics* (Racah Memorial Volume), F. Bloch, S. G. Cohen, A. deShalit, S. Sambursky, and I. Talmi, Eds. (Interscience, New York, 1968).

[9] A. Messiah, *Quantum Mechanics* (Wiley, New York, 1962), Vol. II.

[10] Paper II, H. Rabitz and R. Gordon, J. Chem. Phys. **53**, 1831 (1970).

[11] H. R. Greim, M. Baranger, A. C. Kolb, and G. Oertel, Phys. Rev. **125**, 177 (1962).

[12] G. F. Carrier, M. Krook, and C. E. Pearson, *Functions of a Complex Variable* (McGraw-Hill, New York, 1966).

[13] T. Wu and T. Ohmura, *Quantum Theory of Scattering* (Prentice-Hall, Englewood Cliffs, N.J., 1962); P. Roman, *Advanced Quantum Theory* (Addison-Wesley, Reading, Mass., 1965).

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 11, NUMBER 12 DECEMBER 1970

Einstein Tensor and 3-Parameter Groups of Isometries with 2-Dimensional Orbits

HUBERT GOENNER*†
Department of Physics, Temple University, Philadelphia, Pennsylvania
AND
JOHN STACHEL‡
Department of Physics, Boston University, Boston, Massachusetts

(Received 10 April 1969)

The algebraic classification of the Weyl and Ricci tensors and the relation between them in a Riemann space with an isometry group possessing a nontrivial isotropy group are reviewed. All metrics with Minkowski signature, invariant under a 3-parameter isometry group with 2-dimensional orbits having nondegenerate metrics, are constructed from the group properties and are shown to have Ricci tensors with a double eigenvalue, and the orbits are shown to be surfaces of constant curvature. The null orbits are shown to have a triply degenerate eigenvalue of the Ricci tensor. The various additionally degenerate metrics are classified in further detail, extending the work of Plebański and Stachel.

1. INTRODUCTION

The algebraic classification (in the tangent space at a point of a Riemannian manifold) of the geometrical objects describing the gravitational field has become a convenient tool both in the search for exact solutions of Einstein's equations and in attempts at their physical interpretation.[1–3] The quantity best characterizing the gravitational field, the Riemann tensor, may be decomposed into objects belonging to various irreducible representations of the homogeneous Lorentz group: $D(2, 0) + D(0, 2)$ (Weyl conformal tensor), $D(1, 1)$ (traceless Ricci tensor), and $D(0, 0)$ (curvature scalar). Most attention has been devoted to the classification of the Weyl tensor, but, recently, earlier investigations into the algebraic structure of the Ricci (or Einstein) tensor have been resumed.

In general, one would expect neither a partial nor a complete coincidence of the algebraic structures (eigenelements) of the Weyl and Einstein tensors corresponding to a metric. However, if the metric allows a group of isometries, the situation changes. The nature of the subgroup acting in the tangent space at a point, the isotropy group, then plays a key role in connecting the classifications of the Weyl and Einstein tensor.[4]

After a review of the classification of the Weyl and Einstein tensors in Sec. 2, we combine the consideration of the *local* algebraic classification with that of the *global* action of the group of isometries for a discussion of the relation between the isometry group and eigenstructures of the Weyl and Einstein tensors in Sec. 3. In Sec. 4 we apply the classification scheme to all metrics allowing a 3-parameter Lie group with 2-dimensional orbits. Their Einstein tensors will be seen always to have one double eigenvalue for the orbits with nondegenerate metric and a triple eigenvalue for the null orbits. The various further degenerate subclasses of the Einstein tensor with two distinct double eigenvalues or with one quadruple eigenvalue are classified in detail in Sec. 5. A concluding section discusses the outlook for further applications of this classification method. In three appendices we describe the method of constructing the nondegenerate metric on the orbits directly from the Killing vectors, give the finite equations of the five different isometry groups which occur, and list the Einstein tensors for the canonical forms of various metrics appearing in the text.

2. REVIEW OF ALGEBRAIC CLASSIFICATION

A. Classification of the Weyl Tensor

Among various methods of classifying the Weyl conformal curvature tensor, one simple approach treats this tensor as a mapping of the space of bivectors onto itself. Because of the isomorphism between this Klein space and a 3-dimensional complex vector space with Euclidean metric, the classification task can be formulated as an eigenvalue problem for a complex, traceless 3×3 matrix (matrix method).[5] The decomposition into the various cases of distinct eigenvectors and eigenvalues leads to the Penrose diagram of Petrov types[3,6] (Table I). To a triplet of distinct eigenvectors in complex 3-space (nondegenerate Weyl tensor), there corresponds a tetrad of principal vectors in the tangent space of the Riemannian manifold. For the nontrivial degenerate types the eigenstructure of the Weyl tensor consists, in case D, of two orthogonal 2-flats (spacelike and timelike, respectively); and, for case N, of two orthogonal null 2-flats with a common null direction. Jordan, Ehlers, and Kundt[7] have given a refinement of the Penrose

TABLE I. Penrose diagram.

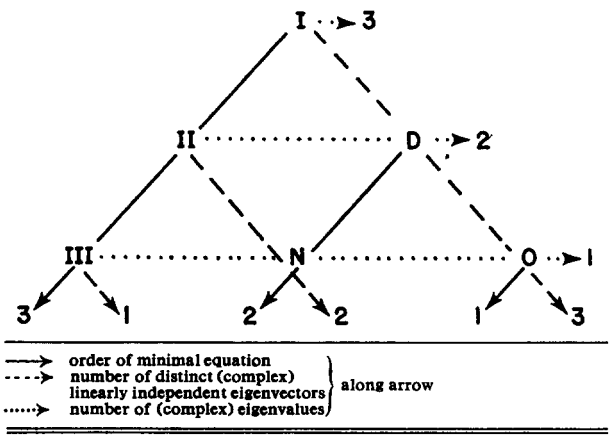


diagram by considering the degeneracies that may occur separately in the real and imaginary parts of the three complex eigenvalues, which correspond to four distinct real eigenvalues in the nondegenerate case.

B. Classification of Einstein (or Ricci) Tensor

Here one has to deal with the eigenvalue problem of simultaneously transforming two indefinite quadratic forms into Jordan canonical form.[2,8] A more geometrical approach to the same problem was followed by Churchill.[9] The results of such a classification may be put into a form analogous to the Penrose diagram (Table II). The line $I \to D \to DD$ DDD characterizes an Einstein tensor with two orthogonal eigen-2-flats (spacelike and timelike, respectively) *without null eigenvectors*. For the arrow $II \to N \to NN_{II}$, the timelike invariant plane contains a *double null eigenvector*; for the arrow $III \to NN_{III}$ there exists an invariant null plane with a *triple null eigenvector*. In contrast to the Penrose diagram of Table I, the more degenerated types may be reached only along the broken arrows (⇢). Plebański[10] and Petrov[11] have given a more detailed refinement of this classification, taking into account the various possibilities for nonnull eigenvectors. This refinement corresponds to the refinement of the Penrose diagram mentioned at the end of Sec. 2A. Table III connects this diagram with that given by Plebański.

TABLE II. Classification of Einstein tensor.

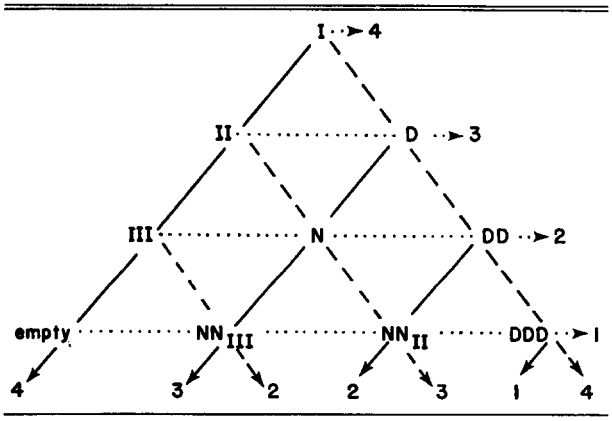


TABLE III. Comparison of Plebański types with types of Table II. Explanation of Plebański's symbols (see Ref. 10 for more details): The symbols T, N, and S are used if the eigenspace of an eigenvalue contains a timelike eigenvector, no timelike but a null eigenvector, or only spacelike eigenvectors, respectively. Z, $\bar{Z}$ refer to the eigenspace of a pair of complex conjugate eigenvalues. The subscripts refer to the power to which each factor is raised in the minimal equation of the matrix of this type.

This paper	Refs. 10, 14
I	$[Z - \bar{Z} - S_1 - S_2]_{[1-1-1-1]}$
	$[T - S_1 - S_2 - S_3]_{[1-1-1-1]}$
II	$[2N - S_1 - S_2]_{[2-1-1]}$
III	$[3N - S]_{[3-1]}$
D	$[Z - \bar{Z} - 2S]_{[1-1-1]}$
	$[2T - S_1 - S_2]_{[1-1-1]}$
	$[T - 2S_1 - S_2]_{[1-1-1]}$
DD	$[2T - 2S]_{[1-1]}$
	$[3T - S]_{[1-1]}$
	$[T - 3S]_{[1-1]}$
DDD	$[4T]_{[1]}$
N	$[2N - 2S]_{[2-1]}$
	$[3N - S]_{[2-1]}$
N_{III}	$[4N]_{[3]}$
N_{II}	$[4N]_{[2]}$

3. ISOTROPY GROUPS AND ALGEBRAIC CLASSIFICATION

A group of isometries leaves invariant the first fundamental form and all tensors constructed from this metric, such as the curvature tensor and its covariant derivatives, the Weyl conformal tensor, etc. We assume the group of isometries to be an r-parameter Lie group G_r. The *orbit* (or minimal invariant variety) of a point x on the manifold is the set of all points into which it is mapped by the group operations. That subgroup of G_r which maps point x into itself is called the *isotropy group* (or subgroup of stability), I_s of x, where s is the number of parameters of the isotropy group. We shall assume that all points in the region under consideration are regular so that s is the same for each point.[12] The orbits are isomorphic to the factor manifold G_r/I_s of G_r, modulo the isotropy group I_s. Consequently, the dimension q of the orbit is connected to r and s by[1,12,13]

$$q = r - s. \tag{3.1}$$

The isotropy group I_s of a point x induces a group of mappings of the tangent space of x into itself, which must thus form a subgroup of the proper homogeneous Lorentz group. As it leaves both the physical components of the conformal and the Einstein tensor invariant, the eigenelements of both tensors are likewise preserved.[13] By virtue of the nature of the eigenelements described above (Sec. 2), the following theorems hold.

Theorem 1[1,4,7,13]: A type I conformal tensor allows at most a discrete isotropy group I_0. D and N fields have $s \leq 2$, the 2-parameter groups I_2 acting within the eigen-2-flats. In the case of Petrov type O (vanishing conformal tensor), the isotropy group may be the full homogeneous (proper) Lorentz group.

Since the Einstein tensor has at least one invariant plane,[9] one sees similarly the following.

Theorem 2: A nonvanishing Einstein tensor allows at most a 2-parameter isotropy group. An Einstein tensor of type I allows at most an I_0, an Einstein tensor of type II or D allows at most an I_1.

Together, Theorems 1 and 2 lead to compatibility conditions between Weyl and Einstein tensor type. Theorems along such lines have recently been proven by Cahen, Debever, and Defrise[4] and Cahen and Defrise[14] for vacuum metrics.

Under the assumptions of constancy of the Petrov type throughout some region of space–time and the existence of a nondiscrete local isotropy group for each point of a region, Cahen, Debever, and Defrise give the following expressions for the admissible Ricci tensors[4].

Petrov type D:

$$R_{\alpha\beta} = \tfrac{1}{4}Rg_{\alpha\beta} + ak_\alpha k_\beta + bm_\alpha m_\beta + 2c[k_{(\alpha}m_{\beta)} + t_{(\alpha}\bar{t}_{\beta)}]; \quad (3.2)$$

Petrov type N:

$$R_{\alpha\beta} = \tfrac{1}{4}Rg_{\alpha\beta} + ak_\alpha k_\beta + 2b[k_{(\alpha}m_{\beta)} + t_{(\alpha}\bar{t}_{\beta)}], \quad (3.3)$$

where $(k, m, t, \bar{t})$ form a null tetrad (or Sachsbein) in the tangent space. By comparing Eqs. (3.2) and (3.3) with the normal forms of the traceless tensor $U_{\alpha\beta} = R_{\alpha\beta} - \frac{1}{4}Rg_{\alpha\beta}$ given by Plebański,[10] we conclude that, under the above assumptions, only the following Churchill–Plebański types may occur.

Petrov type D:

$$[Z - \bar{Z} - 2S]_{[1-1-1]},\ [T - 2S_1 - S_2]_{[1-1-1]},\ [2N - 2S]_{[2-1]},\ [4N]_{[2]},\ [4T]_{[1]},\ [2T - 2S]_{[1-1]}; \quad (3.4)$$

Petrov type N:

$$[3N - S]_{[2-1]},\ [2N - 2S]_{[2-1]},\ [2T - 2S]_{[1-1]},\ [4N]_{[2]},\ [4N]_{[3]},\ [4T]_{[1]}. \quad (3.5)$$

Of course, this list of possible types may be further restricted by additional symmetry assumptions. The types actually occurring among the metrics we discuss in the next section will be found by direct inspection.

4. METRICS ADMITTING A 3-PARAMETER GROUP OF ISOMETRIES WITH 2-DIMENSIONAL ORBITS

Plebański and Stachel[15] recently discussed the classification of all metrics with spherical symmetry, i.e., with a 3-parameter group of isometries having 2-dimensional spacelike orbits. We shall generalize this work by discussing the classification of metrics with 3-parameter isometry groups having 2-dimensional orbits, which may be spacelike, timelike, or null since the metric is of Minkowski signature. The isotropy group belonging to a G_3 with two-dimensional orbits must be a 1-parameter group, by Eq. (3.1), and it must induce a subgroup of the proper homogeneous Lorentz group in the tangent space at each point of the orbit. This means that the isotropy group must induce a 1-parameter group of pure rotations in the tangent space of the spacelike orbits, special Lorentz transformations in the tangent space of the timelike orbits, and of null rotations (singular Lorentz transformations) in the tangent space of the null orbits.

The existence of such an isometry group imposes restrictions on the algebraic structure of the conformal and Einstein tensors. Applying the results of the previous section, we see that the Petrov type can be only D, N, or O. (We shall treat O, in the following work, as a degeneracy of D or N.) The Churchill–Plebański type of the Einstein tensor is restricted to at most those types given in Eqs. (3.4) and (3.5).

The nonnull orbits have a 2-dimensional metric structure induced on them by the metric of the 4-space, definite for the spacelike orbits and indefinite for the timelike orbits. In either case, since the group of isometries induced on the orbit is maximal, the spaces must be of constant (positive, negative, or vanishing) curvature. (For definite metrics, this result goes back

TABLE IV. G_3's with 2-dimensional orbits.

Space-time character of orbit and sign of curvature		IX	VIII	VII ($q = 0$)	VI ($q = -1$)	II	Isotropy group
spacelike	+1	X					rotation
	−1		X				
	0			X			
timelike	+1		X				special Lorentz transformation
	−1		X				
	0				X		
null						X	null rotation

to Bianchi[16]; for indefinite metrics it is given by Eisenhart.[12])

Bianchi[17] has classified all real 3-parameter Lie groups into nine types, for each of which a canonical form for the commutation relations of its generators may be given (see Appendix A for details). One may now see which types are compatible with the character of the orbit (spacelike, timelike, or null) and with its curvature (positive, negative, or vanishing). The results of such a study, which is discussed in Appendix A, are given in Table IV.

Canonical forms of the metrics for the nonnull case may be obtained in the following way: We first solve the commutation relations of the Lie algebras to get the three Killing vectors. The metric on the orbit is then constructed from the Casimir operator quadratic in the infinitesimal operators. We then add two further generators commuting with the Killing vectors in order to build up the full 4-dimensional metric. The only additional requirement needed is that the Lie derivative of the 4-dimensional metric be zero with respect to the generators of the isometry group. (For details, see Appendix B.)

We thus obtain the following canonical forms for the metrics. Although somewhat different, the canonical forms given below (except for the null case) are equivalent to the corresponding ones listed by Petrov (Ref. 2, pp. 200–206). Other equivalent canonical forms for the metrics of Petrov type D are also to be found in Ref. 14.

1. Spacelike Orbit: $x^0 = const$, $x^1 = const$

We have

$$ds^2 = e^{2\alpha(x^0,x^1)}(dx^0)^2 - e^{2\beta(x^0,x^1)}(dx^1)^2 - e^{2\gamma(x^0,x^1)}[(dx^2)^2 + \Sigma^2(x^2)(dx^3)^2], \quad (4.1)$$

where

$$\Sigma(x^2) = \begin{Bmatrix} \sin x^2 \\ \sinh x^2 \\ 1 \end{Bmatrix} \text{ for } \begin{Bmatrix} \text{positive} \\ \text{negative} \\ \text{zero} \end{Bmatrix} \text{ 2-curvature.} \quad (4.2)$$

The Killing vectors take the form

curvature of orbit	ξ^α_1	ξ^α_2	ξ^α_3
+1	$\sin x^3\delta^\alpha_2 + \cos x^3 \cot x^2\delta^\alpha_3$	δ^α_3	$\cos x^3\delta^\alpha_2 - \sin x^3 \cot x^2\delta^\alpha_3$
−1	$\sin x^3\delta^\alpha_2 + \cos x^3 \coth x^2\delta^\alpha_3$	δ^α_3	$\cos x^3\delta^\alpha_2 - \sin x^3 \coth x^2\delta^\alpha_3$
0	δ^α_2	δ^α_3	$-x^3\delta^\alpha_2 + x^2\delta^\alpha_3$

(4.3)

2. Timelike Orbit: $x^1 = const$, $x^3 = const$

There are two alternate forms in this case, resulting from different identifications of the timelike coordinate. The first metric is static:

$$ds^2 = 2^{\gamma(x^1,x^3)}[\Sigma^2(x^2)(dx^0)^2 - (dx^0)^2] - e^{2\beta(x^1,x^3)}(dx^1)^2 - e^{2\alpha(x^1,x^3)}(dx^2)^2, \quad (4.4)$$

with Killing vectors

curvature of orbit	ξ^α_1	ξ^α_2	ξ^α_3
$+1$	$\sinh x^0\delta^\alpha_2 - \cosh x^0 \cot x^2\delta^\alpha_0$	δ^α_0	$\cosh x^0\delta^\alpha_2 - \sinh x^0 \cot x^2\delta^\alpha_0$
-1	$\sinh x^0\delta^\alpha_2 - \cosh x^0 \coth x^2\delta^\alpha_0$	δ^α_0	$\cosh x^0\delta^\alpha_2 - \sinh x^0 \coth x^2\delta^\alpha_0$

(4.5)

The second metric is

$$ds^2 = e^{2\gamma(x^1,x^3)}[(dx^0)^2 - \Sigma^2(x^0)(dx^2)^2] - e^{2\beta(x^1,x^3)}(dx^1)^2 - e^{2\alpha(x^1,x^3)}(dx^3)^2, \tag{4.6}$$

with Killing vectors

curvature of orbit	ξ^α_1	ξ^α_2	ξ^α_3
$+1$	$\sinh x^2\delta^\alpha_0 - \cosh x^2 \cot x^0\delta^\alpha_2$	δ^α_2	$\cosh x^2\delta^\alpha_0 - \sinh x^2 \cot x^0\delta^\alpha_2$
-1	$\sinh x^2\delta^\alpha_0 - \cosh x^2 \coth x^0\delta^\alpha_2$	δ^α_2	$\cosh x^2\delta^\alpha_0 - \sinh x^2 \coth x^0\delta^\alpha_2$

(4.7)

In the case of zero curvature on the orbits, the Killing vectors are, for both canonical forms (4.4) and (4.6) (apart from signs),

$$\underset{1}{\xi^\alpha} = 2^{-\frac{1}{2}}(-\delta^\alpha_0 + \delta^\alpha_2), \quad \underset{2}{\xi^\alpha} = 2^{-\frac{1}{2}}(\delta^\alpha_0 + \delta^\alpha_2),$$

$$\underset{3}{\xi^\alpha} = 2^{-\frac{1}{2}}[-x^0(\delta^\alpha_0 + \delta^\alpha_2) + x^2(-\delta^\alpha_0 + \delta^\alpha_2)] = x^2\underset{1}{\xi^\alpha} - x^0\underset{2}{\xi^\alpha}. \tag{4.8}$$

With the coordinate transformation

$$y^0 = 2^{-\frac{1}{2}}(x^0 - x^2), \quad y^1 = x^1,$$
$$y^2 = 2^{-\frac{1}{2}}(x^0 + x^2), \quad y^3 = x^3,$$

we obtain from Eqs. (4.4) and (4.6) the simpler canonical form

$$ds^2 = 2e^{2\gamma(y^1,y^3)}\,dy^0\,dy^2 - e^{2\beta(y^1,y^3)}(dy^1)^2 - e^{2\alpha(y^1,y^3)}(dy^3)^2, \tag{4.9}$$

with the Killing vectors

$$\underset{1}{\xi^\alpha} = \delta^\alpha_2, \quad \underset{2}{\xi^\alpha} = \delta^\alpha_0, \quad \underset{3}{\xi^\alpha} = x^2\delta^\alpha_2 - x^0\delta^\alpha_0. \tag{4.10}$$

3. Null Orbit (see Defrise[18]): $x^0 = const$, $x^1 = const$

We have

$$ds^2 = e^{2\alpha(x^0,x^1)}[2\,dx^0\,dx^3 - (dx^2)^2 + \sigma(x^0)(x^2)^2(dx^0)^2] - e^{2\beta(x^0,x^1)}(dx^1)^2, \tag{4.11}$$

with Killing vectors

$$\xi^\alpha_k = \delta^\alpha_3[b_k(x^0)x^2 + c_k(x^0)] + \delta^\alpha_2\int^{x^0} b_k(y)\,dy, \quad k = 2, 3, \tag{4.12}$$

where b_k has to satisfy the equation

$$\sigma(x^0)\int^{x^0} b_k(y)\,dy + \frac{db_k}{dx^0} = 0.$$

Only two of the three arbitrary functions α, β, and γ of Eqs. (4.1), (4.4), and (4.6) are essential. This stems from the fact that every 2-dimensional Riemannian space is conformally flat. One may thus put Eqs. (4.1), (4.4), and (4.6) into the forms

$$ds^2 = \exp 2\bar{\alpha}(x^0, x^1)[(dx^0)^2 - (dx^1)^2] - \exp 2\gamma(x^0, x^1)f(x^2, x^3)[(dx^2)^2 + (dx^3)^2] \tag{4.13}$$

and

$$ds^2 = \exp 2\gamma(x^1, x^3)g(x^0, x^2)[(dx^0)^2 - (dx^2)^2] - \exp 2\bar{\beta}(x^1, x^3)[(dx^1)^2 + (dx^3)^2]. \tag{4.14}$$

We shall, however, in our further work use the more flexible canonical forms (4.1), (4.4), and (4.6).

5. TWO DOUBLE EIGENVALUES AND QUADRUPLE EIGENVALUES

All the Einstein tensors belonging to metrics with a G_3 with 2-dimensional nonnull orbits have at least one *double* eigenvalue. The Einstein tensors of metrics with a G_3 with 2-dimensional null orbits have a triple eigenvalue. We shall now study the degenerate cases in more detail.

In order to obtain the condition for the occurrence of a second double eigenvalue for the nonnull orbits, we have to calculate the Einstein tensor corresponding to the canonical forms [Eqs. (4.1), (4.4), and (4.6)]. This is most easily done by means of the exterior differential calculus. The results are given in Appendix C.

A. Metrics With G_3 With Spacelike Orbits

For a metric with canonical form (4.1), a double root exists if and only if

$$-\ddot{\mu}e^{-2\alpha} - \mu''e^{-2\beta} \mp 2\dot{\mu}'e^{-(\alpha+\beta)} + \left(\frac{\partial}{\partial x^0}(e^{\alpha+\beta}) \pm \frac{\partial}{\partial x^1}e^{2\alpha}\right)\dot{\mu}e^{-3\alpha-\beta} + \mu'e^{-\alpha-3\beta}\left(\frac{\partial}{\partial x^1}(e^{\alpha+\beta}) \pm \frac{\partial}{\partial x^0}e^{2\beta}\right) = 0, \quad (5.1)$$

where $\gamma = \log \mu(x^0, x^1)$. A dot denotes differentiation with respect to x^0, and a prime denotes differentiation with respect to x^1.

Four cases arise here:

(1) the surfaces $\mu = \text{const}$ are spacelike;
(2) the surfaces $\mu = \text{const}$ are timelike;
(3) the surfaces $\mu = \text{const}$ are null;
(4) μ is constant throughout the region.

(1) If μ is a spacelike variable, we may set $\mu = x^1$, in which case (5.1) reduces to

$$\frac{\partial}{\partial x^1}e^{\alpha+\beta} = \mp\frac{\partial}{\partial x^0}e^{2\beta}. \quad (5.2)$$

This is the same condition as the one given, for spherically symmetric metrics, by Eq. (4.5) of Ref. 15. We note, especially, that Eq. (5.2) is independent of the curvature of the orbits. Thus, the introduction of a null coordinate u by

$$e^{\alpha+\beta} = \frac{\partial u}{\partial x^0}, \quad e^{2\beta} = \mp\frac{\partial u}{\partial x^1} \quad (5.3)$$

leads again to the canonical form

$$ds^2 = F(u, x^1)\,du^2 + 2\,du\,dx^1 - (x^1)^2\,d\omega^2, \quad F > 0, \quad (5.4)$$

now, however, with

$$d\omega^2 = (dx^2)^2 + \Sigma^2(x^2)(dx^3)^2, \quad (5.5)$$

with Σ given by Eq. (4.2).

(2) If μ is a timelike variable, we may proceed similarly or merely note that if $F < 0$ in metric (5.4), x^1 is now a timelike variable; so we may use the same canonical form with $F < 0$. If F were identically zero, this would make x^1 a null variable, and we would get case (3) below. If there is a surface $x^1 = \text{const}$ on which F makes a transition from positive to negative values, then x^1 is a null coordinate on this surface. This occurs in the case of the Schwarzschild and Vaidya metrics, for example.

(3) If μ is a null variable, we may always choose $\alpha = \beta$, since any 2-space is conformally flat, and rewrite the condition for μ to be null as $(\mu_{,0})^2 = (\mu_{,1})^2$, so that $\mu = \mu(x^0 - x^1)$ or $\mu = \mu(x^0 + x^1)$. Introducing $u = 2^{-\frac{1}{2}}(x^0 - x^1)$, $v = 2^{-\frac{1}{2}}(x^0 + x^1)$, we get $ds^2 = 2e^{2\alpha}\,du\,dv - \mu^2\,d\Omega$.

Takeno[19] has shown, for orbits of positive curvature, that this canonical form may be used in all cases by allowing μ to be a spacelike, timelike, or null variable, or a constant. However, we shall use the canonical form (5.4) for purposes of comparison with Ref. 15 in which the possibility of $\mu = \text{const}$ not being spacelike was not considered.

All the results of Sec. 4 of Ref. 15 can now be generalized immediately for metrics with a group of isometries G_3 acting transitively on *spacelike* 2-dimensional orbits of nonpositive curvature. For example, the Einstein tensor has the form

$$G_\alpha^\beta = \begin{pmatrix} \frac{f_{,1}}{(x^1)^2} & 0 & 0 & 0 \\ \frac{f_{,u}}{(x^1)^2} & \frac{f_{,1}}{(x^1)^2} & 0 & 0 \\ 0 & 0 & \frac{f_{,11}}{2x^1} & 0 \\ 0 & 0 & 0 & \frac{f_{,11}}{2x^1} \end{pmatrix}, \quad (5.6)$$

and the curvature scalar and conformal invariant are given by

$$R = G_\alpha^\alpha = (x^1)^{-2}(x^1 f)_{,11}\,, \quad (5.7)$$

$$C = [3C_{\alpha\beta\gamma\delta}C^{\alpha\beta\gamma\delta}]^{\frac{1}{2}} = \tfrac{1}{2}x^1[(x^1)^{-2}f]_{,11}. \quad (5.8)$$

The function f in (5.6), (5.7), and (5.8) is related to $F(u, x^1)$ of the canonical form (5.4) by

$$F = -\Sigma^{-1}\frac{d^2\Sigma}{(dx^2)^2} - \frac{f}{x^1}. \quad (5.9)$$

The further classification of the metrics (5.4) according to whether or not the conformal invariant and the Ricci scalar vanish may also be repeated; this leads, however, to a threefold subdivision of metrics according to the 2-curvature of the group orbits. One thus obtains [with the abbreviation $\lambda = \Sigma^{-1}d^2\Sigma/(dx^2)^2$] Table V.

In the case of spacelike orbits with positive curvature, many of these metrics have been named, e.g., the Schwarzschild metric, de Sitter metric, Vaidya metric, etc. (see Ref. 14 for the full listing). We may generalize these names and speak of positive, negative, and zero Schwarzschild, de Sitter, etc., solutions, depending on the curvature of the orbits. Cahen, Debever, and Defrise[4] have used the term "Schwarzschildian" to refer to this class of positive-, negative-, and zero-curvature Schwarzschild metrics, as well as to the

TABLE V. Metrics with spacelike orbits. (R, C) means $R \neq 0$ and $C \neq 0$. A zero in either place corresponds to the vanishing of the corresponding invariant.

Type	$F(u, x^1)$			
	$(0, 0)$	$(0, C)$	$(R, 0)$	(R, C)
$[4T]_{[1]}$	1	$-\lambda - \frac{2m}{x^1}$	$-\lambda - \frac{1}{3}\Lambda(x^1)^2$	$-\lambda - \frac{2m}{x^1} - \frac{1}{3}\Lambda(x^1)^2$
$[4N]_{[2]}$	empty	$-\lambda - \frac{2m(u)}{x^1}$	$-\lambda - \frac{1}{3}\Lambda(u)(x^1)^2$	$-\lambda - \frac{2m(u)}{x^1} - \frac{1}{3}\Lambda(u)(x^1)^2$
$[2T - 2S]_{[1-1]}$	empty	$-\lambda - \frac{2m}{x^1} + \frac{e^2}{(x^1)^2}$	$-\lambda + 2ax^1 + b(x^1)^2$	$-\lambda - \frac{f(x^1)}{x^1}$
$[2N - 2S]_{[1-1]}$	empty	$-\lambda - \frac{2m(u)}{x^1} + \frac{e^2(u)}{(x^1)^2}$	$-\lambda + 2a(u)x^1 + b(u)(x^1)^2$	$-\lambda - \frac{f(u, x^1)}{x^1}$

corresponding ones for timelike orbits, which we shall discuss in the next subsection. Birkhoff's theorem for empty space-times (and its generalization to some nonempty solutions) holds also for the groups of Bianchi type VIII and VII with spacelike 2-dimensional orbits, as Taub noted.[20] Cahen and Debever have also discussed the generalization of Birkhoff's theorem.[21] A fuller discussion of Birkhoff's theorem and its generalizations has been given by one of us (H. G.) elsewhere.[22] The static degenerate vacuum solutions of Table V have been discussed by Ehlers and Kundt[1] and are tabulated in Table 2-3.1 of Ref. 1 as fields of Class A.

B. Metrics with Timelike Orbits

The metrics with canonical forms (4.4) or (4.6) have an additional double root if and only if a relation like Eq. (5.1) holds. If $\gamma \equiv$ const, there is an additional double root, as may be seen from Eq. (C10). If γ is not a constant, it is always spacelike; in this case, we may take $\gamma = \log x^3$ and so obtain

$$\frac{\partial}{\partial x^3} e^{\alpha+\beta} = \pm \frac{\partial}{\partial x^1} e^{2\alpha}. \tag{5.10}$$

We can again satisfy Eq. (5.10) by putting

$$e^{\alpha+\beta} = \frac{\partial u}{\partial x^1}, \quad e^{2\alpha} = \frac{\partial u}{\partial x^3} \tag{5.11}$$

and so arrive at the canonical forms

$$ds^2 = -\left(\frac{\partial u}{\partial x^3}\right)^{-1}\left[\left(\frac{\partial u}{\partial x^3}\right)^2 (dx^3)^2 + \left(\frac{\partial u}{\partial x^1}\right)^2 (dx^1)^2\right] + (x^3)^2[(dx^0)^2\Sigma^2(x^2) - (dx^2)^2] \tag{5.12}$$

and

$$ds^2 = -\left(\frac{\partial u}{\partial x^3}\right)^{-1}[(u_{,3})^2(dx^3)^2 + (u_{,1})^2(dx^1)^2] + (x^3)^2[(dx^0)^2 - \Sigma^2(x^0)(dx^2)^2]. \tag{5.13}$$

However, u is no longer a null coordinate, and these forms do not appear to constitute a simplification.

In the special case where $\beta = \beta(x^3)$ and $\alpha = \alpha(x^3)$, one derives from Eq. (5.9) that

$$\frac{\partial}{\partial x^3}(\alpha + \beta) = 0. \tag{5.14}$$

Putting $e^{2\alpha} = f$ and $e^{2\beta} = g$, we find that Eq. (5.14) leads to

$$\dot{g}/g = -\dot{f}/f \quad \text{or} \quad g = f^{-1} + \text{const} \tag{5.15}$$

with $g = g(x^3)$ and $f = f(x^3)$. The components of the Einstein tensor (referred to the tetrad of differential forms of Appendix C) are in this case

$$G^0_0 = G^2_2 = -\frac{1}{2(x^3)^2}\frac{\partial}{\partial x^3}\left((x^3)^2\frac{\partial g}{\partial x^3}\right),$$
$$G^1_1 = G^3_3 = -\frac{1}{(x^3)^2}\left(\pm\lambda + \frac{\partial}{\partial x^3}(x^3 g)\right), \tag{5.16}$$
$$G^3_1 = 0,$$

with $+\lambda$ and $-\lambda$ corresponding to canonical forms (4.4) and (4.6), respectively.

The conformal invariant is

$$C = G^0_0 - G^1_1 - \frac{1}{x^3}\frac{\partial g}{\partial x^3}. \tag{5.17}$$

Again, we may discuss the case $R = 0$ and $C \neq 0$ which leads to

$$g = \mp\lambda + \frac{a}{x^3} + \frac{b}{(x^3)^2}$$

and

$$ds^2 = -\left(\mp\lambda + \frac{a}{x^3} + \frac{b}{(x^3)^2}\right)(dx^1)^2 - \frac{(dx^3)^2}{\mp\lambda + a/x^3 + b/(x^3)^2} + \begin{cases}(x^3)^2[(dx^0)^2\Sigma^2(x^2) - (dx^2)^2] \\ (x^3)^2[(dx^0)^2 - \Sigma^2(x^0)(dx^2)^2]\end{cases}. \tag{5.18}$$

For $b = 0$, the vacuum field equations are satisfied [$C = a/(x^3)^3$]. Further discussions of some of these metrics may be found in Refs. 14, 18, and 19. For $R \neq 0$, $C = 0$,

$$g = \mp\lambda + bx^3 + a(x^3)^2$$

so that

$$ds^2 = -[\mp\lambda + bx^3 + a(x^3)^2](dx^1)^2 - \frac{(dx^3)^2}{\mp\lambda + bx^3 + a(x^3)^2} + (x^3)^2\begin{cases}[(dx^0)^2\Sigma^2(x^2) - (dx^2)^2] \\ [(dx^0)^2 - \Sigma^2(x^0)(dx^2)^2]\end{cases}. \quad (5.19)$$

For $R = 0$ and $C = 0$, one obtains $f = -\lambda^{-1}$, $\lambda \neq 0$, and

$$ds^2 = \frac{1}{\lambda}[(dx^1)^2 + (dx^3)^2] + (x^3)^2\begin{cases}[(dx^0)^2\Sigma^2(x^2) - (dx^2)^2] \\ [(dx^0)^2 - \Sigma^2(x^0)(dx^2)^2]\end{cases}. \quad (5.20)$$

The metrics of Eqs. (5.18)–(5.20) are analogs of some of the metrics compiled in Table V. The static degenerate vacuum fields of this type have been given in Ref. 1, Table 2-3.1. However, Eq. (5.10) allows more general classes of solutions than does (5.2). For example, the most general metric with $\alpha = \alpha(x^1)$ and $\beta = \beta(x^3, x^1)$ and two double eigenvalues of the Einstein tensor is given by

$$ds^2 = -[kx^1 + l]^2(dx^3)^2 - [b(x^1) + 2kx^3]^2(dx^1)^2 + (x^3)^2[(dx^0)^2 - \Sigma^2(x^0)(dx^2)^2], \quad (5.21)$$

where $b(x^1)$ is an arbitrary function and k and l arbitrary constants. The metric (5.21) cannot satisfy the vacuum field equations.

C. Metric with Null Orbits

For a metric with a null group orbit, one has a triple eigenvalue in the most general case. We shall discuss the condition for a quadruple eigenvalue.

As one can see from Appendix C, the condition for the canonical form (4.11) to have a quadruple eigenvalue is

$$\alpha'' - \alpha'\beta' = 0. \quad (5.22)$$

Two cases must be distinguished:

(1) $\alpha' = 0, \rightarrow \alpha = \alpha(x^0), \quad \beta = \beta(x^0, x^1)$;

(2) $\alpha' \neq 0, \rightarrow \dfrac{\alpha''}{\alpha'} - \beta' = 0; \quad \text{or} \quad \beta = \log\alpha' - f(x^0).$

The canonical forms for a metric allowing a group of isometries with 2-dimensional null orbits and whose Einstein tensor has a quadruple eigenvalue thus are, respectively,

$$ds^2 = e^{2\alpha(x^0)}[2\,dx^0\,dx^2 - (dx^3)^2 + \sigma(x^0)(x^2)^2(dx^0)^2] - e^{2\beta(x^0,x^1)}(dx^1)^2 \quad (5.23)$$

and

$$ds^2 = e^{2\alpha(x^0,x^1)}[2\,dx^0\,dx^2 - (dx^3)^2 + \sigma(x^0)(x^2)^2(dx^0)^2] - (\alpha_{,1})^2e^{-2\beta(x^0)}(dx^1)^2. \quad (5.24)$$

Only the metric (5.23) can satisfy the vacuum field equations, which it does if α and β satisfy

$$\ddot{\beta} - 2\dot{\alpha}\dot{\beta} + \dot{\beta}^2 + \ddot{\alpha} - \dot{\alpha}^2 - \sigma = 0.$$

The Einstein tensor of metric (5.23) has three eigenvectors (one null and two spacelike belonging to the eigenvalue $\lambda = 0$); whereas the Einstein tensor belonging to (5.24) possesses only two eigenvectors (one null, one spacelike). Tables II and III show that the corresponding Churchill–Plebański types are $[4N]_{[2]}$ and $[4N]_{[3]}$. The first one may thus represent a null electromagnetic field.

6. CONCLUSIONS

In the preceding sections, we have seen that the combination of algebraic classification of the Weyl and Einstein tensors and of group-theoretic methods is a convenient tool in finding specific canonical forms for metrics with sufficient symmetry. The canonical forms belonging to metrics allowing a group of isometries may be found in the literature[2] or may be derived directly by coordinate-free differential geometric techniques. They may be further restricted by conditions inferred from a given matter distribution and imposed on the eigenstructure of the Einstein tensor.

While we have listed the general canonical forms for metrics with a G_3 and 2-dimensional orbits, only the more degenerate Einstein tensors have been classified in detail. Many of the metrics obtained are still to be investigated for possible physical significance. Some of these metrics, which may be interpreted as generated by pressure-free matter have been discussed by Ellis,[23] with references to earlier work. One may assume the orbits to constitute symmetry surfaces of the matter. However, it is not obvious what kind of matter distribution, if any, would generate, for example, a surface of constant negative curvature. The case of flat orbits seems to be a little easier, if one restricts himself to special global topologies (e.g., cylindrical orbits, etc.). We have not touched upon questions of global topology in this paper. The global topology of 2-dimensional surfaces of constant curvature has been completely classified, for definite or indefinite signature of the metric and

for positive, negative, and vanishing curvature. The results have been conveniently cataloged by Wolf,[24] who is responsible for the study of the timelike orbits. This should help in the study of the 4-dimensional global topology of the metrics with spacelike or timelike orbits.

As the theorems connecting the algebraic classification of conformal and Einstein tensors discussed in Sec. 3 are more general than the applications in this paper, it would be interesting to investigate Einstein tensors with more or less symmetry. For example, one might want to study metrics allowing a G_2 with 2-dimensional orbits (nondegenerate conformal tensor, rotational symmetry) or a G_4 with 3-dimensional orbits, this case containing among others the Taub–NUT space.[14,20,25] Cahen and Defrise[14] have begun a general study of metrics having a nontrivial isotropy group.

Furthermore, the method for construction of the metric on the orbits used in Appendix B might be exploited in the search for exact solutions of Einstein's equations. This method is rarely used in the relativity literature. The only reference to it we have found is in the work of Misner.[25]

ACKNOWLEDGMENTS

We wish to thank Dr. P. Havas for his hospitality at Temple University, where this work was started. One of us (J. S.) thanks Dr. M. Cahen for a copy of Ref. 14 prior to its publication and for discussion of its contents. One of us (H. G.) thanks Dr. L. Defrise for sending him a copy of her thesis, Ref. 18. We thank Dr. G. F. R. Ellis for his helpful comment on the preprint of this paper.

APPENDIX A: 3-PARAMETER ISOMETRY GROUPS WITH 2-DIMENSIONAL ORBITS

As we have discussed in Sec. 4, if the orbits of a 3-parameter isometry group are nonnull 2-dimensional surfaces, they must be surfaces of constant curvature, positive, negative, or vanishing. The isotropy group at each point of the orbit must be a 1-parameter group of real rotations in the case of a spacelike orbit (i.e., one containing only spacelike directions), a 1-parameter group of special Lorentz transformations in the case of a timelike orbit (i.e., one containing timelike as well as spacelike and null directions), and a 1-parameter group of null rotations or singular Lorentz transformations, in the case of a null orbit (i.e., one everywhere tangent to the null cone, containing one null and spacelike directions only).

Bianchi has shown[17] that there are nine nonisomorphic 3-dimensional Lie algebras over the field of the real numbers. We follow Petrov[2] in the numeration of these groups in Table VI, where they are listed with a canonical form for the commutation relations of their generators.

TABLE VI. Bianchi's classification of G_3's.

I	$[X_iX_j]=0$	$i,j=1,2,3$
II	$[X_1X_2]=0$	$[X_2X_3]=X_1,\ [X_3X_1]=0$
III	$[X_1X_2]=0$	$[X_2X_3]=0,\ [X_3X_1]=-X_1$
IV	$[X_1X_2]=0$	$[X_2X_3]=X_1+X_2,\ [X_3X_1]=-X_1$
V	$[X_1X_2]=0$	$[X_2X_3]=X_2,\ [X_3X_1]=-X_1$
VI	$[X_1X_2]=0$	$[X_2X_3]=qX_2[X_3X_1]=-X_1,\ (q\neq 0,1)$
VII	$[X_1X_2]=0$	$[X_2X_3]=-X_1+qX_2,$ $[X_3X_1]=-X_2,\ q^2<4$
VIII	$[X_1X_2]=X_1$	$[X_2X_3]=X_3,\ [X_3X_1]=-2X_2$
IX	$[X_1X_2]=X_3$	$[X_2X_3]=X_1,\ [X_3X_1]=X_2$

Over the field of complex numbers, types VI and VII are isomorphic, as are types VIII and IX (the only two nonsolvable groups), which are then isomorphic to $SL(2, C)$.

Now we shall discuss which of the above groups can be associated with 2-dimensional orbits of the various possible types. Our discussion will sometimes lean on and amplify that of Petrov[2] and sometimes differ from it somewhat. First of all, let us consider the nonnull orbits. If and only if the curvature of these orbits vanishes, the isometry group must induce a 2-parameter Abelian subgroup on the orbit (see Eisenhart,[12] for example). For the case of a spacelike orbit of vanishing curvature, the two commuting generators must have commutation relations with the third generator, which generates the 1-parameter family of rotations of the isotropy group, like those of the translations and rotations of a Euclidean plane. This leads us to Bianchi type VII, with $q=0$. Since VI and VII are isomorphic over the complex field and since a pure Lorentz transformation may be represented as a rotation through a complex angle, one can immediately guess that in the case of a timelike flat orbit we shall have Bianchi type VI. One may then check that the generators will have the commutation relations appropriate to a timelike 2-flat in Minkowski space, if we take $q=-1$ for Bianchi type VI.

If the orbits are nonflat, then the symmetry group cannot have an Abelian subgroup, so that we are led at once to the nonsolvable groups VIII and IX. For a spacelike orbit of positive curvature, the generators must have the commutation relations of the rotations on an ordinary 2-dimensional sphere and so must be of Bianchi type IX. Since the commutation relations of the other three cases must be different (i.e., spacelike orbit of negative curvature and timelike orbit of positive or negative curvature), they must be of Bianchi type VIII. As we shall see below, when we

discuss the finite equations for the groups, these commutation relations do admit the appropriate isotropy group in each case.

In order to discuss the Bianchi type associated with the null orbits, we consider an example of such an orbit. If we cut a null cone with a timelike hypersurface, the intersection forms such a 2-dimensional null orbit, which is generated locally by two translations, one in a spacelike and one in a null direction. The null direction is also the axis of the 1-parameter group of null rotations which constitutes the isotropy group at a point of the surface. The spacelike and null translations commute, as do the null translations and null rotations. Thus, only Bianchi types II and III need be considered. We have been unable to find an intuitive argument leading uniquely to type II, but the calculations of Defrise[18] show that only type II occurs.

Since the metric induced on the null orbit by the 4-dimensional metric of the space–time is degenerate, it cannot be constructed directly from the Killing vectors as we shall do in Appendix B for the other cases. The work of Defrise, discussed in Ref. 18, shows that the unique canonical form of the metric in this case is given by Eq. (4.11).

The simple and somewhat intuitive arguments given above can be formalized (except for the null case) using an extension of Cartan's method for the case of definite metrics,[26] based on the fact that the connection form on the orbits [see Eq. (C3)] is a 1-form in the Lie algebra of the isometry group. The results of this discussion of Bianchi types are summarized in Table IV.

We now give the finite equations for the groups which occur. As mentioned, the groups of Bianchi type IX and VIII are isomorphic to $SL(2, C)$ over the complex numbers. The finite equation of the group is therefore

$$z' = \frac{az + b}{cz + d}, \quad \begin{vmatrix} a & b \\ c & d \end{vmatrix} = 1. \tag{A1}$$

Since $SL(2, C)$ is isomorphic to the proper homogeneous Lorentz group, it contains as subgroups the group of rotations around an axis

$$w' = \frac{w \cos\varphi + \sin\varphi}{-w \sin\varphi + \cos\varphi}, \quad w = xy^{-1},$$

which is the isotropy group for the spacelike orbits, as well as the group of special Lorentz transformations

$$w' = \frac{w \cosh\psi + \sinh\psi}{w \sinh\psi + \cosh\psi}, \quad w = xt^{-1},$$

which is the isotropy group for timelike orbits.

For the group of Bianchi type VII, the finite equation is derived from Eq. (4.3):

$$\begin{aligned} (x^2)' &= x^2 \cos\varphi - x^3 \sin\varphi + a, \\ (x^3)' &= x^2 \sin\varphi + x^3 \cos\varphi + b, \\ (x^0)' &= x^0, \quad (x^1)' = x^1, \end{aligned} \tag{A2}$$

with group parameters a, b, and φ. The isotropy group (leaving the point $x^2 = x^3 = 0$ invariant) is again $O(1, R)$.

For Bianchi type VI we start with the canonical form (4.9), in which y^0 and y^2 are null coordinates. The finite equations of the group deriving from (4.10) are

$$\begin{aligned} &(y^0)' = ay^0 + b, \quad (y^2)' = a^{-1}y^2 + d, \\ &\qquad a, b, d \text{ group parameters}, \\ &(y^1)' = y^1, \quad (y^3)' = y^3. \end{aligned} \tag{A3}$$

Equation (A3) contains as a subgroup (isotropy group) the group of special Lorentz transformations ($b = d = 0$):

$$(y^0)' = ay^0, \quad (y^2)' = a^{-1}y^2.$$

This can be seen by rewriting the special Lorentz transformation

$$\begin{aligned} x' &= \gamma(x - vt), \\ ct' &= \gamma(ct - v/cx), \end{aligned}$$

with $\gamma = (1 - v^2/c^2)^{-\frac{1}{2}}$, in null coordinates $u = x - ct$ and $w = x + ct$:

$$u' = au, \quad w' = a^{-1}w,$$
$$a = \left(\frac{1 + v/c}{1 - v/c}\right)^{\frac{1}{2}}.$$

Finally, for Bianchi type II, the finite equations of the group, from Eq. (4.12), are

$$\begin{aligned} &(x^0)' = x^0, \\ &(x^1)' = x^1, \\ &(x^2)' = x^2 + \epsilon x^3 + a, \\ &(x^3)' = x^3 + b, \quad a, b, \epsilon \text{ group parameters}. \end{aligned} \tag{A4}$$

Again, x^0 and x^2 are null coordinates. The subgroup of isotropy ($a = b = 0$, leaving the origin fixed) now consists of the singular Lorentz transformations (null rotations)[27]

$$\begin{aligned} (x^0)' &= x^0, & (x^1)' &= x^1, \\ (x^2)' &= x^2 + \epsilon x^3, & (x^3)' &= x^3, \end{aligned} \tag{A5}$$

where we keep invariant either the null ray $x^1 = 0$, $x^2 = 0$, $x^3 = 0$ or the null ray $x^0 = x^1 = x^3 = 0$, but not both.

APPENDIX B: CONSTRUCTION OF THE METRIC ON THE NONNULL ORBITS FROM THE KILLING VECTORS

In order to obtain the canonical form of a metric allowing a group of isometries, we start by constructing the metric *on the orbits* from the Killing vectors.

Instead of solving Killing's equations, as Petrov[2] does, for example, we may solve the commutation relations of the corresponding Lie algebra

$$[X_a, X_b] = C^c_{ab} X_c, \quad X_a = \xi^\alpha_a \frac{\partial}{\partial x^\alpha}.$$

Then we calculate the quadratic invariants I of the Lie algebra. In the case of nonsolvable groups, this is done by help of the group metric g_{ab} and of the first Casimir operator $I = g^{ab} X_a X_b$; in the case of solvable groups, by direct inspection. These quadratic invariants, written as tensor products, provide the contravariant metric on the orbit (for the notation used, see Ref. 25):

$$\left(\frac{\partial}{\partial s}\right)^2 = g^{ab} X_a \otimes X_b$$

or, in the case of solvable groups,

$$\left(\frac{\partial}{\partial s}\right)^2 = \sum_{a,b} c_{ab} X_a \otimes X_b .$$

By construction, then, the Lie derivatives of the corresponding covariant metric with respect to Killing vectors vanish. We have summarized the relevant information in Table VII.

The structure constants in column 2, rows 2, 3, 4, and 6 do *not* correspond to the ones usually given [in Eq. (A1), for example]. By basis transformations of the Lie algebra one can convince himself, however, that the exhibited Bianchi types are correct. For Bianchi types VII and VI, quadratic operators commuting with all elements of the Lie algebra exist only if q takes the values $q = 0$ and $q = -1$, respectively.

In order to proceed from the metric on the orbits to the metric of the full Riemannian manifold, we find two vectors (operators) commuting with the Killing vectors. For Eqs. (4.3) these are $\partial/\partial x^0$ and $\partial/\partial x^1$; for Eqs. (4.5), (4.7), and (4.10), $\partial/\partial x^1$ and $\partial/\partial x^3$. Since the cross terms

$$\frac{\partial}{\partial x^0} \otimes \frac{\partial}{\partial x^1} + \frac{\partial}{\partial x^1} \otimes \frac{\partial}{\partial x^0}$$

and

$$\frac{\partial}{\partial x^1} \otimes \frac{\partial}{\partial x^3} + \frac{\partial}{\partial x^3} \otimes \frac{\partial}{\partial x^1},$$

respectively, can always be removed by a coordinate transformation and since the Lie derivative of the 4-dimensional metric with respect to the Killing vectors

TABLE VII. Metric on the group orbits.

Bianchi type of group	Lie algebra $[X_a X_b] = C^c_{ab} X_c$	Group metric	Casimir operator or quadratic invariant	Metric on the group orbit
IX	$C^3_{12} = C^1_{23} = C^2_{31} = -1$	$2\begin{pmatrix} -1 & & 0 \\ & -1 & \\ 0 & & -1 \end{pmatrix}$	$-(X_1^2 + X_2^2 + X_3^2)$	$-[(dx^2)^2 + \sin^2 x^2 (dx^3)^2]$
VIII	$C^2_{31} = C^1_{23} = 1$ $C^3_{12} = -1$	$2\begin{pmatrix} 1 & & 0 \\ & 1 & \\ 0 & & -1 \end{pmatrix}$	$X_1^2 + X_2^2 - X_3^2$	$-[(dx^0)^2 - \sin^2 x^0 (dx^2)^2]$
VIII	$C^3_{12} = C^3_{31} = -1$ $C^1_{23} = +1$	$2\begin{pmatrix} -1 & & 0 \\ & 1 & \\ 0 & & 1 \end{pmatrix}$	$-X_1^2 + X_2^2 + X_3^2$	$(dx^0)^2 - \sinh^2 x^0 (dx^2)^2$
VIII	$C^3_{12} = C^1_{23} = -1$ $C^2_{31} = +1$	$2\begin{pmatrix} 1 & & 0 \\ & -1 & \\ 0 & & 1 \end{pmatrix}$	$X_1^2 - X_2^2 + X_3^2$	$(dx^2)^2 + \sinh^2 x^2 (dx^3)^2$
VII ($q = 0$)	$C^2_{13} = -C^1_{23} = 1$	—	$X_1^2 + X_2^2$	$(dx^2)^2 + (dx^3)^2$
VI ($q = -1$)	$C^1_{13} = C^3_{13} = \frac{1}{\sqrt{2}}$ $C^1_{23} = C^2_{23} = \frac{1}{\sqrt{2}}$	—	$\frac{1}{2}(X_1 X_2 + X_2 X_1) = X_1 X_2$	$(dx^0)^2 - (dx^2)^2$
II	$C^1_{23} = 1$	—	X_1^2	singular

must vanish, we arrive at the canonical forms given in the text.

APPENDIX C: EINSTEIN TENSOR FOR CANONICAL FORMS (4.1), (4.4), (4.6), AND (4.11)

In the usual applications of the exterior differential calculus, one introduces a tetrad of differential 1-forms ω^α by help of which the metric may be written as $g_{\alpha\beta} = \eta_{\alpha\beta}\omega^\alpha\omega^\beta$ (where $\eta_{\alpha\beta}$ is the Minkowski metric). The components of the Riemann curvature tensor referred to this tetrad may be read off from the curvature form

$$\theta_{\alpha\beta} = \tfrac{1}{2}R_{\alpha\beta\gamma\delta}\omega^\gamma \wedge \omega^\delta, \tag{C1}$$

where

$$\theta^\alpha_\beta = d\omega^\alpha_\beta + \omega^\alpha_\sigma \wedge \omega^\sigma_\beta, \tag{C2}$$

with the connection form ω^α_σ defined by

$$d\omega^\alpha = \omega^\alpha_\sigma \wedge \omega^\sigma. \tag{C3}$$

It turns out to be convenient to use a nondiagonal tetrad of differential forms for the canonical form (4.11). The formalism given by Eqs. (C1)–(C3) will always work properly if one does not overlook that nonvanishing ω^σ_σ's (no summation over σ!) exist in place of some vanishing ω^α_β's ($\alpha \neq \beta$).

1. Metric with Spacelike Orbits

We have

$$ds^2 = \eta_{\alpha\beta}\omega^\alpha\omega^\beta$$

with

$$\omega^0 = e^\alpha\,dx^0, \quad \omega^1 = e^\beta\,dx^1, \quad \omega^2 = e^\gamma\,dx^2, \quad \omega^3 = \Sigma e^\gamma\,dx^3. \tag{C4}$$

The calculation leads to

$$G^\beta_\alpha = \begin{pmatrix} G_{00} & -G_{01} & 0 & 0 \\ -G_{01} & -G_{11} & 0 & 0 \\ 0 & 0 & -G_{22} & 0 \\ 0 & 0 & 0 & -G_{22} \end{pmatrix}, \tag{C5}$$

where

$$\begin{aligned}
G^0_0 &= (\dot\gamma^2 + 2\dot\beta\dot\gamma)e^{-2\alpha} + (-2\gamma'' + 2\beta'\gamma' - 3\gamma'^2)e^{-2\beta} - \Sigma^{-1}\Sigma_{,22}e^{-2\gamma},\\
G^1_1 &= (2\ddot\gamma - 2\dot\alpha\dot\gamma + 3\dot\gamma^2)e^{-2\alpha} - (\gamma'^2 + 2\alpha'\gamma')e^{-2\beta} - \Sigma^{-1}\Sigma_{,22}e^{-2\gamma},\\
G^2_2 &= G^3_3 \\
&= (\ddot\beta - \dot\alpha\dot\beta + \dot\beta^2 + \ddot\gamma - \dot\alpha\dot\gamma + \dot\beta\dot\gamma + \dot\gamma^2)e^{-2\alpha} \\
&\quad + (-\alpha'' + \alpha'\beta' - \alpha'^2 - \alpha'\gamma' - \gamma'' + \beta'\gamma' - \gamma'^2)e^{-2\beta},\\
G^1_0 &= -G^0_1 = 2(\dot\gamma' - \alpha'\dot\gamma + \dot\gamma\gamma' - \dot\beta\gamma')e^{-(\alpha+\beta)},
\end{aligned} \tag{C6}$$

where a dot denotes differentiation with respect to x^0 and a prime denotes differentiation with respect to x^1.

For the special choice of coordinates $\gamma = \log x^1$, one obtains

$$\Delta = \tfrac{1}{4}(G^0_0 - G^1_1)^2 - G^1_0G^0_1 = \frac{1}{(x^1)^2}e^{-2(\alpha+3\beta)}\left[\left(\frac{\partial}{\partial x^1}e^{\alpha+\beta}\right)^2 - \left(\frac{\partial}{\partial x^0}e^{2\beta}\right)^2\right]. \tag{C7}$$

The roots of the secular equation for (C5) are

$$\lambda_{1,2} = G^2_2, \quad \lambda_{3,4} = \tfrac{1}{2}(G^0_0 + G^1_1) \pm \Delta^{\frac{1}{2}};$$

thus, there is always one double root. There is an additional double root when $\Delta = 0$.

2. Metric with Timelike Orbits

1. Canonical Form (4.4)

We have

$$ds^2 = \eta_{\alpha\beta}\omega^\alpha\omega^\beta$$

with

$$\omega^0 = e^\gamma\Sigma\,dx^0, \quad \omega^1 = e^\beta\,dx^1, \quad \omega^2 = e^\gamma\,dx^2, \quad \omega^3 = e^\alpha\,dx^3. \tag{C8}$$

The components of the Einstein tensor, in this tetrad, are given by

$$G^\beta_\alpha = \begin{pmatrix} G_{00} & 0 & 0 & 0 \\ 0 & -G_{11} & 0 & -G_{13} \\ 0 & 0 & G_{00} & 0 \\ 0 & -G_{13} & 0 & -G_{33} \end{pmatrix}, \tag{C9}$$

with

$$\begin{aligned}
G^0_0 &= G^2_2 \\
&= (-\ddot\gamma - \dot\gamma\dot\beta + \dot\gamma\dot\alpha - \dot\gamma^2 - \ddot\beta + \dot\alpha\dot\beta - \dot\beta^2)e^{-2\alpha} \\
&\quad + (-\gamma'' + \gamma'\beta' - \gamma'\alpha' - \gamma'^2 + \alpha'\beta' - \alpha'^2 - \alpha'')e^{-2\beta},\\
G^1_1 &= (-2\ddot\gamma + 2\dot\gamma\dot\alpha - 3\dot\gamma^2)e^{-2\alpha} \\
&\quad + (-2\gamma'\alpha' - \gamma'^2)e^{-2\beta} - \Sigma^{-1}\Sigma_{,22}e^{-2\gamma},\\
G^3_3 &= (-2\dot\gamma\dot\beta - \dot\gamma^2)e^{-2\alpha} \\
&\quad + (-2\gamma'' + 2\gamma'\beta' - 3\gamma'^2)e^{-2\beta} - \Sigma^{-1}\Sigma_{,22}e^{-2\gamma},\\
G^3_1 &= 2(\dot\gamma' - \gamma'\dot\beta + \gamma'\dot\gamma - \alpha'\dot\gamma)e^{-(\alpha+\beta)},
\end{aligned} \tag{C10}$$

where a dot denotes differentiation with respect to x^3 and a prime denotes differentiation with respect to x^1.

The eigenvalues of the Einstein tensor are given by

$$\lambda_{1,2} = G^0_0, \quad \lambda_{3,4} = \tfrac{1}{2}(G^1_1 + G^3_3) \pm \Delta^{\frac{1}{2}}, \quad \Delta = \tfrac{1}{4}(G^1_1 - G^3_3)^2 - G^1_1G^3_1.$$

The specialization $\gamma = \log x^3$ leads to

$$\Delta = \frac{1}{(x^3)^2}e^{-2(\beta+3\alpha)}\left[\left(\frac{\partial}{\partial x^3}e^{\alpha+\beta}\right)^2 - \left(\frac{\partial}{\partial x^1}e^{2\alpha}\right)^2\right] \tag{C11}$$

and reduces Eq. (C10) to

$$
\begin{aligned}
G_0^0 = G_2^2 &= e^{-2\alpha}\left(\frac{\dot\alpha - \dot\beta}{x^3} - \ddot\beta + \dot\alpha\dot\beta - \dot\beta^2\right) \\
&\quad + e^{-2\beta}(\alpha'\beta' - \alpha'^2 - \alpha''), \\
G_1^1 &= e^{-2\alpha}\left(2\frac{\dot\alpha}{x^3} - \frac{1}{(x^3)^2}\right) - \frac{1}{(x^3)^2}\Sigma^{-1}\Sigma_{,22}, \\
G_3^3 &= e^{-2\alpha}\left(-2\frac{\dot\beta}{x^3} - \frac{1}{(x^3)^2}\right) - \frac{1}{(x^3)^2}\Sigma^{-1}\Sigma_{,22}, \\
G_3^1 = G_1^3 &= -2\frac{\alpha'}{x^3}e^{-(\alpha+\beta)}.
\end{aligned}
\tag{C12}
$$

With Eq. (5.14), this set of expressions goes over into Eq. (5.15).

2. Canonical Form (4.6)

Here one works with

$$
\begin{aligned}
ds^2 &= \eta_{\alpha\beta}\omega^\alpha\omega^\beta, \\
\omega^0 = e^\gamma\, dx^2, &\quad \omega^1 = e^\beta\, dx^1, \\
\omega^2 = e^\gamma\Sigma\, dx^0, &\quad \omega^3 = e^\alpha\, dx^3
\end{aligned}
\tag{C13}
$$

[exchange of ω^0 and ω^2 in (C8)].

The detailed calculation shows that the only change in Eqs. (C9)–(C12) is a change of sign of the terms proportional to $\Sigma^{-1}\Sigma_{,22}$.

3. Metric with Null Orbit

For this case, we work with the tetrad of external differential forms

$$
\begin{aligned}
&\omega^0 = e^\alpha\, dx^0, \quad \omega^1 = e^\beta\, dx^1, \\
&\omega^2 = e^\alpha\, dx^2, \quad \omega^3 = e^\alpha[dx^3 + \sigma(x^0)(x^2)^2\, dx^0]
\end{aligned}
\tag{C14}
$$

leading to $ds^2 = \gamma_{\alpha\beta}\omega^\alpha\omega^\beta$,

$$
\gamma_{\alpha\beta} = \begin{pmatrix} 0 & 0 & 0 & 1 \\ 0 & -1 & 0 & 0 \\ 0 & 0 & -1 & 0 \\ 1 & 0 & 0 & 0 \end{pmatrix}.
$$

The nonvanishing coefficients ω_β^α are

$$
\omega_0^0,\ \omega_1^0,\ \omega_0^1,\ \omega_2^0,\ \omega_0^2, \quad \text{and} \quad \omega_2^1.
$$

We have the further relations $\omega_1^3 = \omega_0^1$, $\omega_2^3 = \omega_0^2$, $\omega_1^2 = -\omega_2^1$, $\omega_3^1 = \omega_1^0$, $\omega_3^2 = \omega_2^0$, $\omega_3^0 = \omega_0^3 = \omega_1^1 = \omega_2^2 = 0$, and $\omega_3^3 = -\omega_0^0$. The detailed calculation leads to

$$
G_\alpha^\beta = \begin{pmatrix} G_{03} & -G_{01} & 0 & G_{00} \\ & -G_{11} & 0 & G_{01} \\ & & G_{03} & 0 \\ & \mathbf{O} & & G_{03} \end{pmatrix},
\tag{C15}
$$

with

$$
\begin{aligned}
G_{00} &= -(\ddot\beta - 2\dot\alpha\dot\beta + \dot\beta^2 + \ddot\alpha - \dot\alpha^2 - \sigma)e^{-2\alpha}, \\
G_{11} &= -3(2\alpha'' - 2\alpha'\beta' + 3\alpha'^2)e^{-2\alpha}, \\
G_{22} &= -(4\alpha'' - 4\alpha'\beta' + 9\alpha'^2)e^{-2\alpha}, \\
G_{03} &= -G_{22}, \quad G_{01} = -2(\dot\alpha' - \alpha'\dot\beta)e^{-(\alpha+\beta)},
\end{aligned}
\tag{C16}
$$

where a dot denotes differentiation with respect to x^0 and a prime denotes differentiation with respect to x^1. For a quadruple root we obtain the condition $-G_{11} = G_{03}$, which leads to Eq. (5.21).

* Work supported by the Aerospace Research Laboratories of the Office of Aerospace Research, U.S.A.F.

† Present address: Institut für Theoretische Physik, Bunsenstrasse 9, 34 Göttingen, Germany.

‡ Work supported in part by the National Science Foundation during a summer stay at Temple University.

[1] J. Ehlers and W. Kundt, "Exact Solutions of the Gravitational Fields Equations," *Gravitation*, L. Witten, Ed. (Wiley, New York, 1962), p. 49.

[2] A. Z. Petrov, *Einstein Spaces* (Pergamon, New York, 1969).

[3] F. A. E. Pirani, "Gravitational Radiation," *Gravitation*, L. Witten, Ed. (Wiley, New York, 1962), p. 199.

[4] M. Cahen, R. Debever, and L. Defrise, J. Math. Mech. **16**, 761 (1967).

[5] J. L. Synge, Commun. Inst. Advan . Stud. (Dublin), Ser. A, No. 15 (1960).

[6] R. Penrose, Ann. Phys. **10**, 179 (1960).

[7] P. Jordan, J. Ehlers, and W. Kundt, Akad. Wiss. Lit. (Mainz), Abhandl. Math. Nat. Kl.,**2**, (1960).

[8] L. P. Eisenhart, *Riemannian Geometry* (Princeton U.P., Princeton, N.J., 1964), p. 33.

[9] R. V. Churchill, Trans. Am. Math. Soc. **34**, 784 (1932).

[10] J. Plebański, Acta Phys. Polon. **26**, 963 (1964).

[11] A. Z. Petrov, "The Type of the Field and the Type of the Energy Momentum Tensor," in *Perspectives in Geometry and Relativity*, B. Hoffman, Ed. (Indiana U.P., Bloomington, 1966), p. 278.

[12] L. P. Eisenhart, *Continuous Groups of Transformation* (Princeton U.P., Princeton, N.J., 1933).

[13] R. P. Kerr, J. Math. Mech. **12**, 33 (1963).

[14] M. Cahen and L. Defrise, Commun. Math. Phys. **11**, 56 (1968).

[15] J. Plebański and J. Stachel, J. Math. Phys. **9**, 269 (1968).

[16] L. Bianchi, *Lezioni di geometria differenziale* (Spoerri, Pisa, 1902), 2nd ed., Vol. 1, p. 348.

[17] L. Bianchi, *Lezioni sulla teoria dei gruppi continui finiti di transformazioni* (Spoerri, Pisa, 1918).

[18] L. Defrise, "Groupes d'isotropie et groupes de stabilité conforme dans les espaces Lorentzians," Thèse, Université libre de Bruxelles, 1969, unpublished. Metric (4.11) generalizes the canonical form for the case of null orbits given by Petrov.[2]

[19] H. Takeno, Hiroshima University, Scientific Reports of the Research Institute for Theoretical Physics, No. 5, 1966.

[20] A. H. Taub, Ann. Math. **53**, 472 (1951).

[21] M. Cahen and R. Debever, Compt. Rend. Acad. Sci. (Paris) **260**, 815 (1968).

[22] H. Goenner, Commun. Math. Phys. **16**, 34 (1970).

[23] G. F. R. Ellis, J. Math. Phys. **8**, 1171 (1967).

[24] Joseph A. Wolf, *Spaces of Constant Curvature* (McGraw-Hill, New York, 1967).

[25] C. W. Misner, J. Math. Phys. **4**, 924 (1963); C. W. Misner and A. H. Taub, Zh. Eksp. Teor. Fiz. **55**, 233 (1968) [Sov. Phys.-JETP, **28**, 122 (1969)].

[26] H. Cartan, *Leçons sur la geometrie des espaces de Riemann* (Gauthier-Villars, Paris, 1946), 2nd ed., Chap. XII.

[27] J. L. Synge, *Relativity, The Special Theory* (North-Holland, Amsterdam, 1965), 2nd ed., p. 434.

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 11, NUMBER 12 DECEMBER 1970

Representations of Local I-Spin Charge Densities. I. Spaces Where Each I Occurs Only Once

RUI V. MENDES*
Center for Particle Theory, The University of Texas at Austin, Austin, Texas 78712
AND
YUVAL NE'EMAN†‡
Center for Particle Theory, The University of Texas at Austin, Austin, Texas 78712
and *Tel Aviv University, Tel Aviv, Israel*

(Received 29 December 1969)

An approach to the problem of representation of the algebra of currents that puts essential emphasis on the study of infinite-parameter Lie algebras is proposed. As an example, a class of irreducible Hermitian representations of the commutation relations $[V^i(\phi_1), V^j(\phi_2)] = i\epsilon^{ijk}V^k(\phi_1\phi_2)$, where the ϕ's are elements of a commutative algebra with identity, is derived. The dependence of the representations on the algebra $\{\phi\}$ is completely characterized by two functional equations that are explicitly solved, for $\{\phi\}$ an algebra of polynomials. States of well-defined momentum and rotational properties are constructed using translational and rotational invariance and forming direct integral spaces. The representations so constructed are seen to belong to two distinct subclasses, distinguished by the vanishing or nonvanishing of a length parameter $|\eta|$. The subclass with $|\eta| = 0$ is unbounded in isospin and has the trivial momentum-transfer structure characteristic of field-theoretical point particles. On the other hand, the spaces characterized by $|\eta| \neq 0$ are bounded in isospin and suited to describe particles with structure. A brief discussion on how to derive invariant form factors from the results here presented is included.

1. PHYSICAL MOTIVATION

Following Gell-Mann's suggestion,[1] the possibility that a theory of hadrons should involve the currents as basic coordinates has been discussed in many recent publications.[2] Two types of nontrivial examples have been investigated:

(a) the program of Dashen and Gell-Mann[3] using the representations of the chiral charges and other current densities (provided that they commute with α_z in a system of Dirac matrices, when using quark-field bilinears) between states with infinite momentum $p_z \to \infty$;

(b) the Sugawara[4] model, which provides a consistent description of the generators of the Poincaré group as integrals over quadratic functions of the currents.

In both programs, one is interested in the representations of the local charge and current densities, which are supposed to explain the actually observed spectrum of the hadrons. In the previous view[1,5] it was hoped that the spectrum of masses, spins, and unitary spins would reflect an approximate symmetry under a generalized system of global algebraic generators, the space integrals of time and space components of current densities. The energies would thus be related to some Casimir operators of this global Lie algebra; the order of operations was thus first a space integration of the densities, next a squaring of the integrals (e.g., for the quadratic Casimir operator), and then a summation over chiral, unitary, or $SU(6)$-spin indices.

In the alternative (local) approach, the procedure is inverted. Sugawara's energy spectrum will be fixed by the space integral of θ^{00}, the energy–momentum density, given as a quadratic function of the charge densities. First we square the densities, next we sum over chiral–unitary indices, and then we integrate over all space. The structure of the spectrum is already realized at the level of the infinite-parameter algebra of the densities. As a consequence, both the Dashen–Gell-Mann program and the Sugawara model lead us to shift our concern from the representations of integrated commutators (charge algebras) to the representations of the infinite-parameter Lie algebras associated with the algebra of currents.

In the approach of Dashen and Gell-Mann a solution is attempted in the following way: One takes matrix elements of the currents between physical states at infinite momentum; then one is left with the problem of satisfying the commutation relations and the kinematical constraints imposed by the transformation properties of states and currents. These kinematical constraints have been the main difficulty standing in the way of a solution to this problem.

In this paper an alternative approach is proposed: One writes the currents as functionals over an algebra of functions in 3-space; the unintegrated commutators written in functional form are then treated as an infinite-parameter Lie algebra and its representations constructed without any reference at this stage to the physical interpretation of the state vectors in the

representation spaces. The physical interpretation and construction of physical spaces would then be effected by the use of the relativistic transformation properties of the currents or in the case of the Sugawara model by operating with the energy–momentum tensor on the states. That this is a natural approach to solve the Sugawara model has already been pointed out by Sakita.[6] Here we suggest that it may also be a convenient approach even when no energy–momentum tensor is available, for, without having to worry about *ab initio* kinematical constraints, one has more freedom to explore the structural richness of the infinite-parameter Lie algebras. To show the feasibility of this approach, we study a simple example involving the equal-time commutation relations of the isospin charge densities:

$$[V_0^i(x), V_0^j(y)] = i\epsilon^{ijk}V_0^k(x)\delta^3(x-y), \quad x_0 = y_0. \tag{1}$$

Defining the currents as operator-valued functionals over an appropriate algebra of functions $\{\phi\}$, one has

$$V^i(\phi) = \int V_0^i(x)\phi(x)\, d^3x, \tag{2}$$

$$[V^i(\phi_1), V^j(\phi_2)] = i\epsilon^{ijk}V^k(\phi_1\phi_2). \tag{3}$$

To be able to use equal-time commutation relations, we here assume that a simple smearing over space at fixed time of the charge density operators is enough to ensure good behavior of $V^i(\phi)$. This is not so in some models; however, for the essentially model-free calculations done in this paper, no trouble is seen to arise from the above assumption. The commutation relations (3) define what will be called "local $SU2$."

The algebra $\{\phi\}$ is required to contain the identity. Thus the algebra defined by (3) contains the Lie algebra of $SU2$ as a finite parameter subalgebra.

In Sec. 2, conditions on the reduced matrix elements for a general Hermitian representation are derived. In Sec. 3, we restrict ourselves to representations where no two $SU2$ subspaces are equivalent (here called singleton representations), and $\{\phi\}$ is any commutative algebra with identity. The dependence of the matrix elements on the $SU2$ quantum numbers is completely factored out and the functional dependence is found to be characterized by two functional equations.

In Sec. 4 a general solution of the functional equations is found, for $\{\phi\}$ an algebra of polynomials. Finally, in Sec. 5 we use translational and rotational invariance to construct states with well-defined momentum and rotational properties.

2. GENERAL STRUCTURE

We define

$$V^{\pm} = V^1(\phi) + iV^2(\phi). \tag{4}$$

From (3) it follows that

$$[(V^{\pm})^3(\phi_1), (V^{\pm})^3(\phi_2)] = 0, \tag{5}$$

$$[V^+(\phi_1), V^-(\phi_2)] = 2V^3(\phi_1\phi_2), \tag{6}$$

$$[V^+(\phi_1), V^3(\phi_2)] = -V^+(\phi_1\phi_2), \tag{7}$$

$$[V^-(\phi_1), V^3(\phi_2)] = V^-(\phi_1\phi_2). \tag{8}$$

The representation spaces of $\{V^i(\phi)\}$ to be derived are assumed to be reduced in irreducible subspaces of the subalgebra $\{V^i(1)\} \sim SU2$.

The vectors in these subspaces are characterized by $|\rho II_3\rangle$, where $I(I+1)$ and I_3 are the eigenvalues of $V^i(1)V^i(1)$ and $V^3(1)$ and ρ is a multiplicity quantum number:

$$V^{\pm}(1)\,|\rho II_3\rangle = [(I \pm I_3 + 1)(I \mp I_3)]^{\frac{1}{2}}\,|\rho II_3 \pm 1\rangle, \tag{9}$$

$$V^3(1)\,|\rho II_3\rangle = I_3\,|\rho II_3\rangle. \tag{10}$$

From (3) one sees that $V^i(\phi)$ transforms like a $SU2$-vector operator. Using the Wigner–Eckart theorem with the appropriate Clebsch–Gordan coefficients, we see that the most general form of the representations of $\{V^i(\phi)\}$ is

$$\begin{aligned} V^{\pm}(\phi)\,|\rho II_3\rangle &= \sum_{\rho'}\{\pm[(I \mp I_3)(I \mp I_3 - 1)]^{\frac{1}{2}} \\ &\quad \times A^I{}_{\rho}{}^{\rho'}(\phi)\,|\rho' I - 1 I_3 \pm 1\rangle \\ &\quad - [(I \mp I_3)(I \pm I_3 + 1)]^{\frac{1}{2}}B^I{}_{\rho}{}^{\rho'}(\phi)\,|\rho' I I_3 \pm 1\rangle \\ &\quad \pm [(I \pm I_3 + 1)(I \pm I_3 + 2)]^{\frac{1}{2}} \\ &\quad \times C^I{}_{\rho}{}^{\rho'}(\phi)\,|\rho' I + 1 I_3 \pm 1\rangle\}, \end{aligned} \tag{11}$$

$$\begin{aligned} V^3(\phi)\,|\rho II_3\rangle &= \sum_{\rho'}\{[(I - I_3)(I + I_3)]^{\frac{1}{2}}A^I{}_{\rho}{}^{\rho'}(\phi)\,|\rho' I - 1 I_3\rangle \\ &\quad - I_3 B^I{}_{\rho}{}^{\rho'}(\phi)\,|\rho' I I_3\rangle \\ &\quad - [(I + I_3 + 1)(I - I_3 + 1)]^{\frac{1}{2}}C^I{}_{\rho}{}^{\rho'}(\phi)\,|\rho' I + 1 I_3\rangle\}. \end{aligned} \tag{12}$$

A scalar product is defined as

$$\langle \rho II_3 \,|\, \rho' I' I_3'\rangle = \delta_{\rho\rho'}\delta_{II'}\delta_{I_3 I_3'}, \tag{13}$$

and for Hermitian representations we require

$$\langle \rho' I' I_3 + 1|\, V^+(\phi)\,|\rho II_3\rangle = \langle \rho II_3|\, V^-(\phi)\,|\rho' I' I_3 + 1\rangle^*, \tag{14}$$

$$\langle \rho' I' I_3|\, V^3(\phi)\,|\rho II_3\rangle = \langle \rho II_3|\, V^3(\phi)\,|\rho' I' I_3\rangle^*. \tag{15}$$

(14) and (15) applied to (11) and (12) imply

$$B^I(\phi) = [B^I(\phi)]^+, \tag{16}$$

$$A^I(\phi) = -[C^{I-1}(\phi)]^+, \tag{17}$$

where $A^I(\phi)$, $B^I(\phi)$, and $C^I(\phi)$ are matrices on the multiplicity index ρ.

To determine the reduced matrix elements $A^I{}_\rho{}^{\rho'}$, $B^I{}_\rho{}^{\rho'}$, and $C^I{}_\rho{}^{\rho'}$, we substitute (11) and (12) in (5)–(8), and after some algebra one finds the following equations for the matrices B and C:

$$C^I(\phi_1)C^{I+1}(\phi_2) = C^I(\phi_2)C^{I+1}(\phi_1), \tag{18}$$

$$C^I(\phi_1)B^{I+1}(\phi_2) + B^I(\phi_1)C^I(\phi_2) = C^I(\phi_2)B^{I+1}(\phi_1) + B^I(\phi_2)C^I(\phi_1), \tag{19}$$

$$C^I(\phi_1)[C^I(\phi_2)]^+ - \text{h.c.} = \frac{I^2}{2I+1}[B^I(\phi_2), B^I(\phi_1)], \tag{20}$$

$$[C^{I-1}(\phi_2)]^+C^{I-1}(\phi_1) - \text{h.c.} = \frac{(I+1)^2}{2I+1}[B^I(\phi_2), B^I(\phi_1)], \tag{21}$$

$$-C^I(\phi_1)B^{I+1}(\phi_2)(I+2) + B^I(\phi_2)C^I(\phi_1)(I+1) - B^I(\phi_1)C^I(\phi_2) = C^I(\phi_1\phi_2), \tag{22}$$

$$[C^{I-1}(\phi_2)]^+C^{I-1}(\phi_1)(2I+1) - 2[C^{I-1}(\phi_1)]^+C^{I-1}(\phi_2) - C^I(\phi_1)[C^I(\phi_2)]^+(2I+3) + [B^I(\phi_1), B^I(\phi_2)](I+2) + B^I(\phi_2)B^I(\phi_1) = -B^I(\phi_1\phi_2). \tag{23}$$

(17) was used to eliminate the equations involving $A^I(\phi)$.

3. SINGLETON REPRESENTATIONS

In this section we find the Hermitian irreducible representations of the commutation relations (3) subject to the condition that any $SU2$ subspace of a given kind appears at most once (singleton representations). The algebra $\{\phi\}$ is any commutative algebra with identity.

The result is the following:

Theorem: The Hermitian irreducible singleton representations of "local $SU2$" are characterized by the reduced matrix elements

$$A^{I+1}(\phi) = C^I(\phi) = \epsilon_I i\left[\left(\frac{\gamma^2}{(I+1)^2} - \mu\right)\left(\frac{(I+1)^2 - I_0^2}{(2I+3)(2I+1)}\right)\right]^{\frac{1}{2}} C(\phi), \tag{24}$$

$$B^I(\phi) = \frac{\gamma I_0}{I(I+1)}C(\phi) - Q(\phi), \tag{25}$$

where

(i) a particular choice of phases was made to guarantee $A^{I+1}(\phi) = C^I(\phi)$,

(ii) I_0 is the smallest I in the representation space,

(iii) ϵ_I is an arbitrary sign, one for each I,

(iv) γ is an arbitrary real number,

(v) $C(\phi)$ and $Q(\phi)$ are real linear functionals obeying the equations and boundary conditions

$$C(\phi_1)Q(\phi_2) + C(\phi_2)Q(\phi_1) = C(\phi_1\phi_2), \tag{26}$$

$$Q(\phi_1\phi_2) - Q(\phi_1)Q(\phi_2) = \mu C(\phi_1)C(\phi_2), \tag{27}$$

$$C(1) = 0, \quad Q(1) = 1, \tag{28}$$

(vi) μ is a real number, not completely arbitrary, which is required to belong to the range of $Q(\phi^2) - Q(\phi)^2$ with the restriction $C(\phi) \neq 0$,

(vii) $\gamma^2/(I+1)^2 - \mu \geq 0$ for all I's.

Proof: In the case of singleton representations each I appears only once, $B^I(\phi)$ and $C^I(\phi)$ are just numbers, and the commutators of the B's in (20), (21), and (23) vanish.

(a) Suppose that the representation space contains only one I. Then all C's vanish, and of Eqs. (18)–(23) only (23) survives with the form

$$-B^I(\phi_2)B^I(\phi_1) = B^I(\phi_1\phi_2).$$

This is a particular case of the general result stated above.

(b) Suppose now that the representation space contains more than one I. Call I_0 the smallest I in the space and $I_{\max}$ the largest one when it exists. Then, for at least one ϕ_r, one has $C^{I_0}(\phi_r) \neq 0$, for otherwise the I_0 subspace would be invariant, thus contradicting the irreducibility hypothesis.

From (18) one writes

$$C^{I_0+1}(\phi) = C^{I_0}(\phi)C^{I_0+1}(\phi_r)/C^{I_0}(\phi_r)$$

and, if $I_0 + 1$ is not the highest I in the space, then $C^{I_0+1}(\phi_r) \neq 0$; for, otherwise, $C^{I_0+1}(\phi) = 0$ for every ϕ, and the irreducibility hypothesis would be contradicted.

Iterating this reasoning, one concludes that for an irreducible representation space there is at least one ϕ_r such that $C^I(\phi_r) \neq 0$ for all I's except the largest one ($I_{\max}$).

One such ϕ_r may then be used to write all the $C^I(\phi)$ in the form

$$C^I(\phi) = C^I(\phi_r)C^{I+1}(\phi)/C^{I+1}(\phi_r) = C^I(\phi_r)C(\phi), \tag{29}$$

where it follows from (18) that $C(\phi)$ is a linear functional independent of I.

From (20) or (21) it follows that, using (29), we obtain

$$C(\phi_1)C^*(\phi_2) = C(\phi_2)C^*(\phi_1).$$

Since this equation has to be verified for every pair ϕ_1, ϕ_2, one concludes, putting $\phi_1 = \phi_r$, that $C(\phi)$ is a real functional.

Substituting now (29) in (22) and cancelling the $C^I(\phi_r)$ that appears in both sides, one gets

$$-C(\phi_1)B^{I+1}(\phi_2)(I+2) + C(\phi_1)B^I(\phi_2)(I+1) - C(\phi_2)B^I(\phi_1) = C(\phi_1\phi_2). \quad (30)$$

Putting $\phi_1 = \phi_2 = \phi$ and multiplying both sides by $I+1$, one obtains a simple difference equation

$$C(\phi)B^{I+1}(\phi)(I+2)(I+1) - C(\phi)B^I(\phi)(I+1)I = -C(\phi^2)(I+1). \quad (31)$$

If $C(\phi) \neq 0$, iteration of (31) gives

$$B^I(\phi) = \frac{\gamma(\phi)I_0}{I(I+1)} - \frac{C(\phi^2)}{2C(\phi)},$$

where

$$\gamma(\phi) = \left(B^{I_0}(\phi) + \frac{C(\phi^2)}{2C(\phi)}\right)(I_0+1). \quad (32)$$

Substitution of (32) and (29) in (19) gives

$$[C(\phi_1)\gamma(\phi_2) - C(\phi_2)\gamma(\phi_1)]I_0 = 0.$$

Putting $\gamma(\phi_r) = \gamma$, the conclusion is that

$$\gamma(\phi) = \gamma C(\phi) \quad \text{unless} \quad I_0 = 0. \quad (33)$$

If $I_0 = 0$, $\gamma(\phi)$ is undetermined. In this case, however, the first term of the right-hand side of (32) vanishes identically, and (33) may be adopted for all cases. From the hermiticity condition (16) and the fact derived above that $C(\phi)$ is real, it follows that γ is an arbitrary real number. It is arbitrary because $B^{I_0}(\phi_r)$ is arbitrary.

If $C(\phi) = 0$, one obtains directly from (30)

$$B^I(\phi) = -C(\phi\phi_1)/C(\phi_1),$$

where ϕ_1 is arbitrary, subject only to the restriction $C(\phi_1) \neq 0$. The two cases $C(\phi) = 0$ and $C(\phi) \neq 0$ may then be described by one equation

$$B^I(\phi) = [\gamma I_0/I(I+1)]C(\phi) - Q(\phi), \quad (25)$$

where

$$Q(\phi) = C(\phi^2)/2C(\phi) \quad \text{for} \quad C(\phi) \neq 0$$
$$= C(\phi\phi_1)/C(\phi_1) \quad \text{for} \quad C(\phi) = 0 \quad \text{and} \quad C(\phi_1) \neq 0. \quad (34)$$

The definition of the current densities as operator-valued *linear* functionals requires that both $C(\phi)$ and $C(\phi^2)/C(\phi)$ [defined for $C(\phi) \neq 0$] be *linear* functionals. Another condition on the functional $C(\phi)$ may be derived from (34), namely that for $C(\phi) = 0$ the quotient $C(\phi\phi_1)/C(\phi_1)$, $C(\phi_1) \neq 0$, be independent of ϕ_1.

Now, substitution of (25) in (30) with $\phi_1 \neq \phi_2$ leads to

$$C(\phi_1)Q(\phi_2) + C(\phi_2)Q(\phi_1) = C(\phi_1\phi_2). \quad (26)$$

It is easily seen that (26) contains the definitions (34) and also the linearity condition of $C(\phi^2)/C(\phi)$ for $C(\phi) \neq 0$.

All of Eqs. (18)–(22) have been used and their consequences derived. We now use the remaining Eq. (23) to compute the explicit form of $C^I(\phi_r)$. Using (29), (21), and (25), we write Eq. (23) in the form

$$C(\phi_1)C(\phi_2)[|C^I(\phi_r)|^2(2I+3) - |C^{I-1}(\phi_r)|^2(2I-1)] = \frac{\gamma^2 I_0^2 C(\phi_1)C(\phi_2)}{I^2(I+1)^2} + [Q(\phi_1)Q(\phi_2) - Q(\phi_1\phi_2)].$$

(26) was used to simplify the right-hand side.

This equation implies that, for $C(\phi_1) = 0$,

$$Q(\phi_1\phi_2) = Q(\phi_1)Q(\phi_2) \quad \text{for any} \quad \phi_2$$

and, for $C(\phi_1)$ and $C(\phi_2) \neq 0$,

$$[Q(\phi_1\phi_2) - Q(\phi_1)Q(\phi_2)]/C(\phi_1)C(\phi_2) = \mu,$$

where μ is a real number independent of ϕ_1 and ϕ_2, in particular

$$\mu = Q(\phi_r^2) - Q(\phi_r)^2.$$

These two conditions may be summarized in the form

$$Q(\phi_1\phi_2) - Q(\phi_1)Q(\phi_2) = \mu\,|C(\phi_1)|\,|C(\phi_2)|. \quad (27')$$

$|C^I(\phi_r)|^2$ is now obtained from

$$|C^I(\phi_r)|^2(2I+3) - |C^{I-1}(\phi_r)|^2(2I-1) = \frac{|\gamma|^2 I_0^2}{I^2(I+1)^2} - \mu. \quad (35)$$

If I_0 is the smallest I in the space,

$$C^{I_0-1}(\phi_r) = -A^{I_0}(\phi_r)^* = 0.$$

Using this condition and iterating (35), one obtains

$$|C^I(\phi_r)|^2 = \left(\frac{\gamma^2}{(I+1)^2} - \mu\right)\frac{[(I+1)^2 - I_0^2]}{(2I+3)(2I+1)}. \quad (36)$$

We now prove that the phases of the vectors in the representation space may be chosen in such a way that $C^I(\phi_r)$ is purely imaginary, thus implying together

with the reality of $C(\phi)$ that

$$C^I(\phi) = A^{I+1}(\phi) \tag{37}$$

for any ϕ and any I.

Suppose that one representation has been found and that the phase difference of $C^I(\phi_r)$ and $A^{I+1}(\phi_r) = -C^I(\phi_r)^*$ is θ_I:

$$C^I(\phi_r) = e^{i\theta_I} A^{I+1}(\phi_r).$$

Now, we multiply all the vectors in each $SU2$ subspace by the phase factor $\exp(\frac{1}{2}i\sum_{\alpha=I_0}^{I-1}\theta_\alpha)$. One sees by inspection of (11) and (12) that the B's are unchanged, and only the A's and C's are changed in the following way:

$$A^I(\phi_r) \to [A^I(\phi_r)]' = A^I(\phi_r)e^{i\theta_{I-1}/2},$$
$$C^I(\phi_r) \to [C^I(\phi_r)]' = C^I(\phi_r)^{-i\theta_{I-1}/2}.$$

One obtains then

$$[C^I(\phi_r)]' = [A^{I+1}(\phi_r)]' \Rightarrow [C^I(\phi_r)]' \text{ purely imaginary.}$$

Assuming that this choice has been made, we obtain the result (24).

We have thus obtained most of the results stated in the beginning of this section. The I dependence of the reduced matrix elements has been completely isolated, and the functional dependence is specified by two functional equations.

The boundary conditions $C(1) = 0$ and $Q(1) = 1$ are required for the representations to reduce to the $SU2$ representations defined in (9) and (10) when $\phi = 1$. If $\{\phi\}$ is a topological algebra, it would be natural to require continuity of the functionals, and one should have, in addition, both $\lim C(\phi) = 0$ and $\lim B(\phi) = 1$ as $\phi \to 1$.

The representations contain only integer or only half-integer I's. From (36) it follows that γ, μ, and $I_{\max}$ have to be such that

$$[\gamma^2/(I+1)^2] - \mu \geq 0 \quad \text{for} \quad I_0 \leq I \leq I_{\max}.$$

Thus, if μ is positive, one has

$$\gamma^2/\mu = n^2,$$

where n is either an integer or a half-integer and the representation space is finite dimensional. The real number μ is not arbitrary. It depends on the particular algebra $\{\phi\}$ to be used. Finally, to prove the irreducibility, one notes that the representation space may be turned into a Hilbert space in a natural way, by using a scalar product derived from (13) and the usual completion procedure. Hence, Schur's lemma will apply whether the space is finite or infinite dimensional, and it will be enough to prove that the only operator commuting with all $V^i(\phi)$ is a multiple of the identity.

Assuming $[X, V^i(\phi)] = 0$ and writing

$$X\,|II_3\rangle = \sum_{I'I_3'} x(I'I_3', II_3)\,|I'I_3'\rangle,$$

we see that it follows from $[X, V^i(1)] = 0$ that

$$x(I'I_3', II_3) = x(I)\delta_{II'}\delta_{I_3I_3'}$$

and from $[X, V^i(\phi)] = 0$ and (11) and (12) that

$$C^I(\phi)x(I+1) = x(I)C^I(\phi).$$

(24) implies that, unless I is the highest I in the space, there always exists a ϕ such that $C^I(\phi) \neq 0$. Hence, $x(I)$ is independent of I, and X is a multiple of the identity operator. ∎

The solution of the functional equations will depend strongly on the nature of the algebra $\{\phi\}$.

In particular, if $\{\phi\}$ is a finite commutative algebra, one ends up with representations of finite-parameter Lie algebras.

Of more concern to us in the problem of representation of current densities is the case of function algebras with pointwise multiplication.

In the following section, the explicit solutions of (26) and (27) will be found, for $\{\phi\}$ an algebra of polynomials in three variables with pointwise multiplication.

4. SOLUTION OF THE FUNCTIONAL EQUATIONS FOR AN ALGEBRA OF POLYNOMIALS

In this section $\{\phi\}$ is an algebra of polynomials in three variables. A basis for this algebra may be written $\{x_1^p, x_2^q, x_3^r; p, q, r = 0, 1 \cdots\}$. The equations to be solved are (26) and (27). (26) may be written in the form

$$C(\phi_1\phi_2) = \frac{1}{2}\left(\frac{C(\phi_1^2)}{C(\phi_1)}C(\phi_2) + \frac{C(\phi_2^2)}{C(\phi_2)}C(\phi_1)\right) \quad \text{if } C(\phi_1) \text{ and } C(\phi_2) \neq 0, \tag{38a}$$

$$\frac{C(\phi\phi_1)}{C(\phi_1)} = \frac{C(\phi\phi_2)}{C(\phi_2)} \quad \text{if } C(\phi_1) \text{ and } C(\phi_2) \neq 0 \text{ and } C(\phi) = 0. \tag{38b}$$

First we note that the boundary condition $C(1) = 0$ satisfies (38b).

Next, we restrict ourselves to a subalgebra of polynomials in one variable only, $\{\phi_x\} = \{x^p; p = 0, 1, 2, \cdots\}$, and prove the following:

The complete specification of the functional $C(\phi)$, a solution of Eqs. (38), requires that at most $C(x)$, $C(x^2)$, and $C(x^3)$ be given.

(a) Suppose that $C(x) = 0$. Now, either $C(x^k) = 0$ for all k or there is one k for which $C(x^k) \neq 0$. From (38a) it now follows that, if $C(x^\alpha) = 0$, $C(x^\beta) = 0$, and $C(x^K) \neq 0$, then $C(x^{\alpha+\beta}) = 0$.

Proof:

$$C[(x^\alpha + x^K)x^K]$$
$$= \frac{1}{2}\left(\frac{C(x^{2\alpha}) + C(x^{2K}) + 2C(x^{K+\alpha})}{C(x^\alpha) + C(x^K)}\, C(x^K) + \frac{C(x^{2K})}{C(x^K)}\, C(x^K)\right),$$
$$0 = \tfrac{1}{2}C(x^{2\alpha}),$$
$$C[(x^\alpha + x^\beta + x^K)x^K]$$
$$= \tfrac{1}{2}[C(x^{2\alpha}) + C(x^{2\beta}) + C(x^{2K}) + 2C(x^{\alpha+\beta}) + 2C(x^{\alpha+K}) + 2C(x^{\beta+K}) + C(x^{2K})]$$
$$\Rightarrow C(x^{\alpha+\beta}) = 0.$$

Now, using this result, one concludes that, if $C(x) = 0$, it is also zero for all the subalgebra generated by x.

(b) Assume that $C(x) \neq 0$.

(i) If $C(x^2) = 0$ and $C(x^3) = 0$, it follows from the result above that $C(x^p) = 0$ for all $p > 1$.

(ii) If $C(x^2) = 0$ and $C(x^3) \neq 0$, it follows that, for all the even powers of x, $C(x^{2K}) = 0$ and, for all the odd powers, $C(x^{2K+1})$ is uniquely determined from (38b):

$$\frac{C(x^3)}{C(x)} = \frac{C(x^5)}{C(x^3)} = \frac{C(x^7)}{C(x^5)} = \cdots.$$

(iii) If $C(x^2) \neq 0$ and $C(x^3) = 0$, we see first that $C(x^{3p}) = 0$. From (38a)

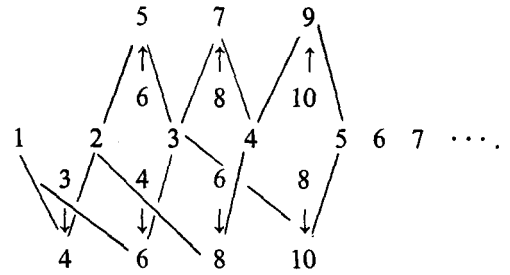
$$C(x^3) = 0 = \frac{1}{2}\left(\frac{C(x^4)}{C(x^2)}\, C(x) + \frac{C(x^2)}{C(x)}\, C(x^2)\right).$$

This implies that $C(x^4) \neq 0$. From (38b) we then write

$$\frac{C(x^4)}{C(x)} = \frac{C(x^5)}{C(x^2)} = \frac{C(x^7)}{C(x^4)} = \frac{C(x^8)}{C(x^5)} = \frac{C(x^{10})}{C(x^7)} = \cdots,$$

implying that $C(x^K) \neq 0$ for $K \neq 3p$ and that its values are uniquely determined from the knowledge of $C(x)$ and $C(x^2)$.

(iv) Let $C(x^2) \neq 0$ and $C(x^3) \neq 0$. From (38a) it is easily seen that, once $C(x)$, $C(x^2)$, and $C(x^3)$ are known and different from zero, the values of $C(x^K)$ for $K > 3$ may be computed according to the following pattern:



The diagram means: Putting in (38a) $\phi_1 = x^1$ and $\phi_2 = x^2$, with $C(x^3)$ being known, one determines $C(x^4)$; putting $\phi_1 = x^1$ and $\phi_2 = x^3$, with $C(x^4)$ known, one determines $C(x^6)$; with $\phi_1 = x^2$ and $\phi_2 = x^3$, $C(x^6)$ being known, $C(x^5)$ is determined, etc.

If in this process no $C(x^K) = 0$ is found, then the process may be carried on and any $C(x^K)$ uniquely determined from the knowledge of $C(x)$, $C(x^2)$, and $C(x^3)$. Suppose now that, for a given $p \geq 4$, $C(x^p) = 0$ and that $C(x^q) \neq 0$ for $q < p$.

From the result proved in (a), $C(x^p) = 0$ implies $C(x^{Kp}) = 0$ for any positive integer K. From (38b) one writes

$$\frac{C(x^{p+1})}{C(x)} = \frac{C(x^{p+2})}{C(x^2)} = \cdots = \frac{C(x^{2p-1})}{C(x^{p-1})}.$$

So, if one of the terms $C(x^{p+1}) \cdots C(x^{2p-1})$ is known, all the others are:

$$0 = C(x^{\frac{1}{2}p+1}x^{\frac{1}{2}p-1})$$
$$= \frac{1}{2}\left(\frac{C(x^{p+2})}{C(x^{\frac{1}{2}p+1})}\, C(x^{\frac{1}{2}p-1}) + \frac{C(x^{p-2})}{C(x^{\frac{1}{2}p-1})}\, C(x^{\frac{1}{2}p+1})\right),$$

p even.

Since $p \geq 4$, $\frac{1}{2}p + 1 < p$ and the second term on the right-hand side is different from zero, thus implying that $C(x^{p+2})$ determined by the above equations is also different from zero:

$$0 = C(x^{\frac{1}{2}(p+1)}x^{\frac{1}{2}(p-1)})$$
$$= \frac{1}{2}\left(\frac{C(x^{p+1})}{C(x^{\frac{1}{2}(p+1)})}\, C(x^{\frac{1}{2}(p-1)}) + \frac{C(x^{p-1})}{C(x^{\frac{1}{2}(p-1)})}\, C(x^{\frac{1}{2}(p+1)})\right),$$

p odd.

As before, $p \geq 5$ implies $\frac{1}{2}(p+1) < p$ and $C(x^{\frac{1}{2}(p+1)}) \neq 0$. Then $C(x^{p+1})$ is determined and different from zero.

In both cases all terms from $C(x^{p+1})$ up to $C(x^{2p-1})$ are determined and different from zero. From (38b)

we may then write

$$\frac{C(x^{p+1})}{C(x)} = \cdots = \frac{C(x^{2p-1})}{C(x^{p-1})} = \frac{C(x^{2p+1})}{C(x^{p+1})} = \cdots = \frac{C(x^{3p-1})}{C(x^{2p-1})} = \frac{C(x^{3p+1})}{C(x^{2p+1})},$$

and the conclusion is that all the $C(x^\alpha)$ are determined and all different from zero except when $\alpha = Kp$. This completes the proof of the result stated above. ∎

The explicit construction of the general solution of Eqs. (38) may now be carried out assuming arbitrary values for $C(x)$, $C(x^2)$, and $C(x^3)$ and iterating the equations. For the sake of brevity, we just present the final result and then verify that it satisfies (38a) and (38b) and the results proved above.

For an algebra of polynomials on one variable $\{\phi(x)\}$, the general solution of the Eq. (26) is

$$C(\phi(x)) = (C/2\eta)[\phi(\epsilon + \eta) - \phi(\epsilon - \eta)], \quad (39)$$

$$Q(\phi(x)) = \tfrac{1}{2}[\phi(\epsilon + \eta) + \phi(\epsilon - \eta)], \quad (40)$$

where C and ϵ are arbitrary real numbers and η is arbitrary and either real or purely imaginary.

It is a trivial matter to verify that (39) and (40) satisfy Eq. (26) or its equivalent form (38). They also satisfy the boundary conditions. We now verify that the three independent parameters C, ϵ, and η allow arbitrary specification of $C(x)$, $C(x^2)$, and $C(x^3)$ in accordance with the previous result, thus proving that (39) and (40) provide the most general solution. From (39)

$$C(x) = C, \quad C(x^2) = 2C\epsilon, \quad C(x^3) = C[3\epsilon^2 + \eta^2].$$

Thus if C and ϵ are real and arbitrary and η is either real or purely imaginary, one may arbitrarily specify $C(x)$, $C(x^2)$, and $C(x^3)$. One also sees that, if $C(x) = 0$, all the $C(x^k)$ are zero and, if $C(x) \neq 0$ and $C(x^2) = C(x^3) = 0$, then all the others are zero, in accordance with what was proved in (a) and (b).

So far we have used Eq. (26) only. Substitution of (39) and (40) in (27) shows that this one is verified too, if

$$\mu = \eta^2/C^2.$$

Consider now the 3-dimensional case,

$$\{\phi\} = \{x_1^p, x_2^q, x_3^r\}.$$

Solving for each one of the three subalgebras $\{\phi_{x_i}\}$, one gets

$$C(\phi_i(x_i)) = (C_i/2\eta_i)[\phi_i(\epsilon_i + \eta_i) - \phi_i(\epsilon_i - \eta_i)],$$

$$Q(\phi_i(x_i)) = \tfrac{1}{2}[\phi_i(\epsilon_i + \eta_i) + \phi_i(\epsilon_i - \eta_i)].$$

Using these solutions with different i's in Eq. (26), one concludes that

$$C_1/\eta_1 = C_2/\eta_2 = C_3/\eta_3.$$

The general solution for the 3-dimensional case is then written

$$C(\phi) = C\left[2\left(\sum_i \eta_i^2\right)^{\frac{1}{2}}\right]^{-1}[\phi(\boldsymbol{\epsilon} + \boldsymbol{\eta}) - \phi(\boldsymbol{\epsilon} - \boldsymbol{\eta})], \quad (41)$$

$$Q(\phi) = \tfrac{1}{2}[\phi(\boldsymbol{\epsilon} + \boldsymbol{\eta}) + \phi(\boldsymbol{\epsilon} - \boldsymbol{\eta})], \quad (42)$$

where C is an arbitrary real number, $\boldsymbol{\epsilon}$ an arbitrary real vector, and $\boldsymbol{\eta}$ an arbitrary vector, whose components are either all real or all purely imaginary.

In this case one obtains from (27)

$$\mu = \left(\sum_i \eta_i^2\right) C^{-2}.$$

This solution was derived for a polynomial algebra. From its form one sees clearly, however, that it still is a solution for any other function algebra with pointwise multiplication, provided that the functions are defined at the evaluation points.

For a general function algebra, $\boldsymbol{\eta}$ will not be allowed to take imaginary values unless all functions of the algebra have analytic extension and the evaluation points are in the common domain of holomorphy; for then $\phi(\boldsymbol{\epsilon} + i\boldsymbol{\alpha}) \pm \phi(\boldsymbol{\epsilon} - i\boldsymbol{\alpha})$ is respectively real or purely imaginary, and $Q(\phi)$ and $C(\phi)$ are again real functionals.

Combining now (24) and (25) with (41) and (42) and using a standard notation for the point evaluation functionals, we write the final result for the reduced matrix elements,

$$A^{I+1}(\phi) = C^I(\phi) = \epsilon_I i\left[\left(\frac{\gamma^2}{(I+1)^2} - \sum_i \eta_i^2\right)\frac{[(I+1)^2 - I_0^2]}{(2I+3)(2I+1)}\right]^{\frac{1}{2}} \times \left[2\left(\sum_i \eta_i^2\right)^{\frac{1}{2}}\right]^{-1}[\delta_{\epsilon+\eta}(\phi) - \delta_{\epsilon-\eta}(\phi)], \quad (43)$$

$$B^I(\phi) = \frac{\gamma I_0}{I(I+1)}\left[2\left(\sum_i \eta_i^2\right)^{\frac{1}{2}}\right]^{-1}[\delta_{\epsilon+\eta}(\phi) - \delta_{\epsilon-\eta}(\phi)] - \tfrac{1}{2}[\delta_{\epsilon+\eta}(\phi) + \delta_{\epsilon-\eta}(\phi)]. \quad (44)$$

The constant C present in (41) was absorbed in γ. Each singleton representation is thus characterized by $(I_0, \gamma, \boldsymbol{\epsilon}, \boldsymbol{\eta})$. From point (vii) in the main result of

Sec. 3, one has

$$\frac{\gamma^2}{(I+1)^2} - \sum_i \eta_i^2 \geq 0 \quad \text{for all } I\text{'s in the space;}$$

this implies that, for η real and different from zero, the representation space is finite dimensional, with

$$I_{\max} = |\gamma| \left[+ \left(\sum_i \eta_i^2 \right)^{\frac{1}{2}} \right]^{-1} - 1$$

and

$$|\gamma| \left[+ \left(\sum_i \eta_i^2 \right)^{\frac{1}{2}} \right]^{-1} = n,$$

where n and I_0 are simultaneously integers or half-integers. For $\eta = 0$ or η purely imaginary, the representation space is infinite dimensional.

For $|\eta| = 0$, the functional $C(\phi)$ should be written

$$C(\phi) = C \nabla \phi \cdot \hat{\eta}. \tag{45}$$

Restricting ourselves to the case of more general utility, that is, the one with η real, one sees that $|\eta|$ plays an essential role determining the structure of the representations. They belong to two different classes:

(I) $|\eta| = 0$: infinite dimensional and γ is not quantized;

(II) $|\eta| \neq 0$: finite dimensional and γ is quantized.

5. CONSTRUCTION OF PHYSICAL SPACES

The representation spaces of the commutator (3) derived in Secs. 3 and 4 are denoted $\mathcal{H}(I_0, \gamma, \epsilon, \eta)$. The vectors in this space will be characterized by $|I_0 \gamma \epsilon \eta I I_3\rangle$. With the exception of I_0, I, and I_3, the physical meaning of the other quantum numbers is as yet unknown. It will be shown in this section that, using the transformation properties of the matrix elements of the charge densities, one is able to identify the nature of ϵ and η and, by forming direct integral spaces, to construct states of well-defined momentum and rotational properties.

The transformation properties to be considered are

$$e^{-i\mathbf{P}\cdot\mathbf{x}} V_0(0) e^{i\mathbf{P}\cdot\mathbf{x}} = V_0(\mathbf{x}), \tag{46a}$$

$$U(R) V_0(\mathbf{x}) U^{-1}(R) = V_0(R\mathbf{x}). \tag{46b}$$

$\mathbf{P}$ is the spatial momentum and R is a proper rotation.

One cannot derive any constraints from the invariance under time translations because the commutators of the charge densities are equal-time commutators; also, one does not consider pure Lorentz transformations because the right-hand side of (46b) would then contain contributions from the space components $V_i(R\mathbf{x})$, whose representations are unknown. One starts by rewriting the matrix elements of the charge densities as distributions on the space variable $\mathbf{x}$:

$$A^{I+1}(\mathbf{x}) = C^I(\mathbf{x}) = \epsilon_I i \left[\left(\frac{\gamma^2}{(I+1)^2} - \sum_i \eta_i^2 \right) \frac{(I+1)^2 - I_0^2}{(2I+3)(2I+1)} \right] \times \frac{1}{2|\eta|} [\delta(\mathbf{x} - \epsilon - \eta) - \delta(\mathbf{x} - \epsilon + \eta)], \tag{47a}$$

$$B^I(x) = \frac{\gamma I_0}{I(I+1)} \frac{1}{2|\eta|} \times [\delta(\mathbf{x} - \epsilon - \eta) - \delta(\mathbf{x} - \epsilon + \eta)] - \tfrac{1}{2}[\delta(\mathbf{x} - \epsilon - \eta) + \delta(\mathbf{x} - \epsilon + \eta)]. \tag{47b}$$

From (47) one sees that (46a) will be verified if the state vectors $|I_0 \gamma \epsilon \eta I I_3\rangle$ transform under translations as follows:

$$e^{-i\mathbf{P}\cdot\mathbf{x}} |I_0 \gamma \epsilon \eta I I_3\rangle = |I_0 \gamma \epsilon + \mathbf{x} \eta I I_3\rangle. \tag{48}$$

They are thus shown to represent localized states, ϵ being the average position.

Forming a direct integral space (in ϵ), one may define momentum states

$$\mathcal{H}(I_0, \gamma, \eta) = \int d\epsilon \mathcal{H}(I_0, \gamma, \epsilon, \eta), \tag{49}$$

$$|I_0 \gamma \mathbf{K} \eta I I_3\rangle = \int e^{i\mathbf{K}\cdot\epsilon} |I_0 \gamma \epsilon \eta I I_3\rangle \, d\epsilon. \tag{50}$$

The direct integral space $\mathcal{H}(I_0, \gamma, \eta)$ is no longer irreducible under the algebra of the charge densities $\{V_i(x)\}$, but is irreducible under the enlarged algebra $\{V^i(x), \mathbf{P}\}$. Choosing for the generalized states $|I_0 \gamma \epsilon \eta I I_3\rangle$ the normalization

$$\langle I_0 \gamma \epsilon' \eta I' I_3' \,|\, I_0 \gamma \epsilon \eta I I_3 \rangle = \delta(\epsilon' - \epsilon) \delta_{I'I} \delta_{I_3' I_3}, \tag{51}$$

one has, for the momentum states (50),

$$\langle I_0 \gamma \mathbf{K}' \eta I' I_3' \,|\, I_0 \gamma \mathbf{K} \eta I I_3 \rangle = (2\pi)^3 \delta(\mathbf{K}' - \mathbf{K}) \delta_{I'I} \delta_{I_3' I_3}.$$

From (47) and (50) one computes the reduced matrix elements between momentum states in the space $\mathcal{H}(I_0, \gamma, \eta)$:

$$\langle \mathbf{K}' \| C^I(\mathbf{x}) \| \mathbf{K} \rangle = \epsilon_I \left[\left(\frac{\gamma^2}{(I+1)^2} - \sum_i \eta_i^2 \right) \frac{(1+I)^2 - I_0^2}{(2I+3)(2I+1)} \right]^{\frac{1}{2}} \times \frac{\sin(\mathbf{K} - \mathbf{K}') \cdot \eta}{|\eta|} e^{i(\mathbf{K} - \mathbf{K}')\cdot\mathbf{x}}, \tag{52a}$$

$$\langle \mathbf{K}' \| B^I(\mathbf{x}) \| \mathbf{K} \rangle = \left(-\frac{i\gamma I_0}{I(I+1)} \frac{\sin(\mathbf{K} - \mathbf{K}') \cdot \eta}{|\eta|} - \cos(\mathbf{K} - \mathbf{K}') \cdot \eta \right) e^{i(\mathbf{K} - \mathbf{K}')\cdot\mathbf{x}}. \tag{52b}$$

For the Fourier transform of the charge density,

$$V_0(\mathbf{q}) = \int V_0(\mathbf{x}) e^{i\mathbf{q}\cdot\mathbf{x}} d^3x, \tag{53}$$

$$\langle \mathbf{K}' \| C^I(\mathbf{q}) \| \mathbf{K} \rangle = \epsilon_I \left[\left(\frac{\gamma^2}{(I+1)^2} - \sum_i \eta_i^2 \right) \frac{(I+1)^2 - I_0^2}{(2I+3)(2I+1)} \right]^{\frac{1}{2}} (2\pi)^3 \times \frac{\sin(\mathbf{K}-\mathbf{K}')\cdot\boldsymbol{\eta}}{|\boldsymbol{\eta}|} \delta^3(\mathbf{K}+\mathbf{q}-\mathbf{K}'), \tag{54a}$$

$$\langle \mathbf{K}' \| B^I(\mathbf{q}) \| \mathbf{K} \rangle = \left(-\frac{i\gamma I_0}{I(I+1)} \frac{\sin(\mathbf{K}-\mathbf{K}')\cdot\boldsymbol{\eta}}{|\boldsymbol{\eta}|} - \cos(\mathbf{K}-\mathbf{K}')\cdot\boldsymbol{\eta} \right) (2\pi)^3 \delta^3(\mathbf{K}+\mathbf{q}-\mathbf{K}'), \tag{54b}$$

going now to the rotation invariance, one sees from Eqs. (52) that $\langle \mathbf{K}' \| C^I(R\mathbf{x}) \| \mathbf{K} \rangle$ and $\langle \mathbf{K}' \| B^I(R\mathbf{x}) \| \mathbf{K} \rangle$ would contain the factors

$$(\mathbf{K}-\mathbf{K}')\cdot R\mathbf{x} = R^{-1}(\mathbf{K}-\mathbf{K}')\cdot\mathbf{x},$$

and

$$(\mathbf{K}-\mathbf{K}')\cdot\boldsymbol{\eta} = R^{-1}(\mathbf{K}-\mathbf{K}')\cdot R^{-1}\boldsymbol{\eta}.$$

This shows that (46b) will be satisfied if under rotations the states transform as follows:

$$U(R)\,|I_0\gamma\mathbf{K}\boldsymbol{\eta}II_3\rangle = e^{il(R)}\,|I_0\gamma R\mathbf{K}R\boldsymbol{\eta}II_3\rangle, \tag{55}$$

where $l(R)$ is an arbitrary phase factor that may be a function of the rotation and of the quantum numbers of the state as well. The transformation law (55) implies that one should extend the space $\mathcal{H}(I_0, \gamma, \boldsymbol{\eta})$ to a space $\mathcal{H}(I_0, \gamma, |\boldsymbol{\eta}|)$, allowing for all directions of $\boldsymbol{\eta}$ but maintaining its modulus fixed (thus maintaining the spectrum of γ fixed—see Sec. 4):

$$\mathcal{H}(I_0, \gamma, |\boldsymbol{\eta}|) = \int d\Omega_\eta \mathcal{H}(I_0, \boldsymbol{\gamma}, \boldsymbol{\eta}). \tag{56}$$

In $\mathcal{H}(I_0, \gamma, |\boldsymbol{\eta}|)$ we construct now states with well-defined rotational properties.

Consider first the case of integral *spin* states. Making $l(R) = 0$, one reduces (55) to

$$U(R)\,|I_0\gamma\mathbf{K}\boldsymbol{\eta}II_3\rangle = |I_0\gamma R\mathbf{K}R\boldsymbol{\eta}II_3\rangle. \tag{55'}$$

We define the states

$$|(I_0\gamma\,|\boldsymbol{\eta}|)\mathbf{K}SMII_3\rangle = \int d\Omega_\eta Y_{SM}(\theta_\eta, \phi_\eta)\,|I_0\gamma\mathbf{K}\boldsymbol{\eta}II_3\rangle, \tag{57}$$

where $Y_{SM}(\theta_\eta, \phi_\eta)$ are the spherical harmonics. Under rotations they transform:

$$\begin{aligned} &U(R)\,|(I_0\gamma\,|\boldsymbol{\eta}|\,\mathbf{K}SMII_3\rangle \\ &= U(R)\int_{-1}^{1} d(\cos\theta)\int_0^{2\pi} d\phi Y_{SM}(\theta,\phi)\,|I_0\gamma\mathbf{K}\,|\boldsymbol{\eta}|\,\theta,\phi II_3\rangle \\ &= \int_{-1}^{1} d(\cos\theta)\int_0^{2\pi} d\phi Y_{SM}(\theta,\phi)\,|I_0\gamma R\mathbf{K}\,|\boldsymbol{\eta}|\,R(\theta,\phi)II_3\rangle \\ &= \int_{-1}^{1} d(\cos\theta)\int_0^{2\pi} d\phi Y_{SM}(R^{-1}(\theta,\phi))\,|I_0\gamma R\mathbf{K}\,|\boldsymbol{\eta}|\,\theta,\phi II_3\rangle \\ &= \sum_{M'} D(R)^{(s)}_{M'M}\,|(I_0\gamma\,|\boldsymbol{\eta}|)R\mathbf{K}SM'II_3\rangle, \end{aligned}$$

where $D(R)^{(s)}$ is the rotation matrix for spin (s).

If the normalization (51) is extended to

$$\langle I_0\gamma\boldsymbol{\epsilon}'\,|\boldsymbol{\eta}|\,\theta'\phi'I'I_3' \,|\, I_0\gamma\boldsymbol{\epsilon}\,|\boldsymbol{\eta}|\,\theta\phi II_3\rangle = \delta(\boldsymbol{\epsilon}'-\boldsymbol{\epsilon})\delta(\cos\theta'-\cos\theta)\delta(\phi'-\phi)\delta_{I'I}\delta_{I_3'}\delta_{I_3}, \tag{51'}$$

then for the states (57) we get

$$\langle (I_0\gamma\,|\boldsymbol{\eta}|)\mathbf{K}'S'M'I'I_3' \,|\, (I_0\gamma\,|\boldsymbol{\eta}|)\mathbf{K}SMII_3\rangle = (2\pi)^3\delta(\mathbf{K}'-\mathbf{K})\delta_{s's}\delta_{M'M}\delta_{I'I}\delta_{I_3'I_3}. \tag{58}$$

With $l(R) = 0$ we have thus constructed a space $\mathcal{H}_1(I_0, \gamma, |\boldsymbol{\eta}|)$, irreducible under the algebra $\{V^i(x), \mathbf{P}, M^{ij}\}$ (where the M^{ij} are the generators of rotations), that contains all integral *spins* but is bounded in isospin when $|\boldsymbol{\eta}| \neq 0$.

For states of half-integral *spin* the construction is analogous. One has to find functions $F_{SM}(\theta, \phi)$ such that the state

$$|(I_0\gamma\,|\boldsymbol{\eta}|)\mathbf{K}SMII_3\rangle = \int d\Omega_\eta F_{SM}(\theta_\eta, \phi_\eta)\,|I_0\gamma\mathbf{K}\boldsymbol{\eta}II_3\rangle_{\frac{1}{2}} \tag{59}$$

has under rotations the transformation property

$$U(R)\,|(I_0\gamma\,|\boldsymbol{\eta}|)\mathbf{K}SMII_3\rangle = \sum_{M'} D(R)^{(s)}_{M'M}\,|(I_0\gamma\,|\boldsymbol{\eta}|)R\mathbf{K}SM'II_3\rangle, \tag{60}$$

where $D(R)^{(s)}$ is now a rotation matrix for half-integral spin. By analogy with the well-known relation

$$Y_{SM}(\theta,\phi) = \left(\frac{2S+1}{4\pi}\right)^{\frac{1}{2}} D^{(s)}_{M0}(\phi,\theta)^* = \left(\frac{2S+1}{4\pi}\right)^{\frac{1}{2}} D^{(s)}_{0M}(0,-\theta,-\phi),$$

we conjecture $F_{SM}(\theta, \phi)$ to be proportional to $D^{(s)}_{\frac{1}{2}M}(0, -\theta, -\phi)$. From the group property

$$D^{(s)}_{\frac{1}{2}M}(R_1R_2) = \sum_{M'} D^{(s)}_{\frac{1}{2}M'}(R_1)D^{(s)}_{M'M}(R_2),$$

one derives the transformation properties of $D^{(s)}_{\frac{1}{2}M}(0, -\theta, -\phi)$:

$$\sum_{M'} D^{(s)}_{\frac{1}{2}M'}(0, -\theta, -\phi) D^{(s)}_{M'M}(\alpha, \beta, \gamma) = D^{(s)}_{\frac{1}{2}M}(\alpha'(R), -R^{-1}\theta, -R^{-1}\phi) = e^{-\frac{1}{2}i\alpha'(R)} D^{(s)}_{\frac{1}{2}M}(0, -R^{-1}\theta, R^{-1}\phi), \quad (61)$$

where R is the rotation defined by the Euler angles (α, β, γ) and

$$\alpha'(R) = \alpha'(\theta, \phi, R) = \operatorname{arccot}\left(\frac{\sin\theta\cot\beta}{\sin(\phi-\alpha)} - \cos\theta\cot(\phi-\alpha)\right).$$

The presence of the phase factor $\alpha'(R)$ shows that, in contrast to $D^{(s)}_{0M}(0, -\theta, -\phi)$, the set $D^{(s)}_{\frac{1}{2}M}(0, -\theta, -\phi)$ is not closed under rotations. One observes, however, that, since the factor $\exp[-\frac{1}{2}i\alpha'(R)]$ is only a function of R, θ, and ϕ and not a function of S or M, one can make the states (59) have the right transformation properties (60), if one uses the freedom of choice of the phase in (55):

$$U(R)\,|I_0\gamma\mathbf{K}\,|\boldsymbol{\eta}|\,\theta,\phi II_3\rangle_{\frac{1}{2}} = e^{-\frac{1}{2}i\alpha'(R\theta, R\phi, R)}\,|I_0\gamma R\mathbf{K}\,|\boldsymbol{\eta}|\,R(\theta,\phi)II_3\rangle_{\frac{1}{2}}. \quad (55'')$$

With (55″) and (61) it is a trivial matter to verify that the states (59) obey the transformation relation (60) with

$$F_{SM}(\theta, \phi) = [(2S+1)/4\pi]^{\frac{1}{2}} D^{(s)}_{\frac{1}{2}M}(0, -\theta, -\phi).$$

The factor $[(2S+1)/4\pi]^{\frac{1}{2}}$ was added to endow the states with the same normalization as in (58).

With $l(R)$ as defined in (55″), a space $\mathcal{H}_2(I_0, \gamma, |\boldsymbol{\eta}|)$ is thus obtained that contains all half-integral *spins*, is irreducible under $\{V^i(\mathbf{x}), \mathbf{P}, M^{ij}\}$, is bounded in isospin for $|\boldsymbol{\eta}| \neq 0$, and is unconnected to $\mathcal{H}_1(I_0, \gamma, |\boldsymbol{\eta}|)$.

The reduced matrix elements of the charge densities in the spaces $\mathcal{H}(I_0, \gamma, |\boldsymbol{\eta}|)$ are now

$$\langle S'M'\mathbf{K}'\|\, C^I(\mathbf{q})\,\|SM\mathbf{K}\rangle = \epsilon_I\left[\left(\frac{\gamma^2}{(I+1)^2} - |\boldsymbol{\eta}|^2\right)\frac{(I+1)^2 - I_0^2}{(2I+3)(2I+1)}\right]^{\frac{1}{2}}\frac{(2\pi)^3}{|\boldsymbol{\eta}|} \times \Gamma(S'M', SM)\delta^3(\mathbf{K}+\mathbf{q}-\mathbf{K}'), \quad (62a)$$

$$\langle S'M'\mathbf{K}'\|\, B^I(\mathbf{q})\,\|SM\mathbf{K}\rangle = \left(-\frac{i\gamma I_0}{|\boldsymbol{\eta}|\, I(I+1)}\Gamma(S'M', SM) - \sum(S'M', SM)\right) \times (2\pi)^3\delta^3(\mathbf{K}+\mathbf{q}-\mathbf{K}'), \quad (62b)$$

with

$$\sum(S'M', SM) = \int_{-1}^{1} d(\cos\theta)\int_0^{2\pi} d\phi K^*_{S'M'}(\theta,\phi) \times \cos(|\mathbf{K}-\mathbf{K}'|\,|\boldsymbol{\eta}|\cos\theta) K_{SM}(\theta,\phi), \quad (63a)$$

$$\Gamma(S'M', SM) = \frac{K_z - K'_z}{|\mathbf{K}-\mathbf{K}'|}\int d\Omega K^*_{S'M'}(\theta,\phi) \times \sin(|\mathbf{K}-\mathbf{K}'|\,|\boldsymbol{\eta}|\cos\theta) K_{SM}(\theta,\phi). \quad (63b)$$

In Eqs. (63) $K_{SM}(\theta, \phi) = Y_{SM}(\theta, \phi)$ for integral *spin*, $K_{SM}(\theta, \phi) = F_{SM}(\theta, \phi)$ for half-integral *spin*, and $\mathbf{K} - \mathbf{K}'$ was taken to lie along the z direction.

We have thus shown the feasibility of the approach proposed in Sec. 1. Matrix elements of the charge densities between states of well-defined momentum and rotational properties were obtained. The states are not necessarily 1-particle states.

Besides the commutator (3) that was our starting point and translational and rotational invariances, our only additional assumption is the singleton character (in isospin) of the representations. The representations belong to two rather different classes, the one with $|\boldsymbol{\eta}| = 0$ and those with $|\boldsymbol{\eta}| \neq 0$. We have already pointed out, in Sec. 3, the distinct behavior of these classes concerning isospin dimensionality and spectrum of γ.

Here too one notes from (63) that, while for $|\boldsymbol{\eta}| = 0$ the matrix elements display a trivial momentum transfer dependence and would give rise to constant form factors like the ones that field-theoretical point particles have, for $|\boldsymbol{\eta}| \neq 0$ the momentum transfer dependence is nontrivial and the representations may be suited to describe particles with structure.

Notice that these nontrivial states appear not as a result of nonlocality or finite-length assumptions, but merely as another representation possibility of an algebra that is strictly local, in much the same way as the Heisenberg algebra possesses representations of quite distinct nature for $\hbar = 0$ and $\hbar \neq 0$. This suggests then that a systematic study of the representations of the infinite-parameter algebras of local operators may be the natural way to generalize the usual field theoretical structures. The appearance of the two kinds of structures is a result of the particular form of the functional equations (26) and (27) that in turn are a consequence of the commutation relations (3).

A natural question to ask is whether all Lie algebras when written in unintegrated form lead to functional equations whose solutions display a nontrivial space dependence. That this does not seem to be so follows from a study of an unintegrated version of the Euclidean algebra.[7] The conclusion is then that states with nontrivial structures are obtained as a result of the algebraic coupling of internal symmetry (integrated Lie algebra) and the space dependence (algebra of the ϕ's) and that this coupling does not seem to occur for all types of Lie algebras.

In this paper we want mainly to show the potentialities of infinite-parameter Lie algebras studied according to the approach proposed in Sec. 1. In Sec. 5 only the model-independent implications of the representations derived in Secs. 3 and 4 were considered. In this way we are still left with the parameters γ and $|\eta|$ physically unidentified. We leave model construction, the derivation of the invariant form factors that is referred to below, and comparison with experimental data and the results of canonical field theory for a future paper. As a last remark we notice that the fact that Eqs. (62) are not manifestly covariant should not come as a surprise. The choice of equal-time commutation relation, which was our starting point, already implies a particular choice of frame. Also the fact that we are dealing with the time components of the currents only prevents us from extracting model-independent information from the invariance under pure Lorentz transformations.

Had we found manifestly covariant expressions for the matrix elements written in (62), the invariance under pure Lorentz transformations would have been guaranteed. This then suggests that the so far unspecified frame be chosen in such a way as to make (62) manifestly covariant. The infinite momentum frame was found to satisfy this condition for spin 0 and $\frac{1}{2}$, and we were able to derive invariant form factors for these cases. The uniqueness of this solution and whether it applies in general remains to be checked.

When this last operation is included in the present approach, the sequence of its steps becomes the reverse of the one in the Dashen–Gell-Mann approach. Whereas in the DG approach the choice of the infinite momentum frame is the initial step, here the choice of a particular frame is the last one. The motivations, however, are rather different. Whereas in the DG approach the infinite momentum frame is chosen to simplify the momentum dependence of the matrix elements and to ensure that the representation space contains only 1-particle states (with this last implication dependent on field theoretical considerations), here a particular frame is picked up to ensure manifest covariance of the final results.

It is our opinion that the approach here proposed, being free of the *ab initio* kinematical constraints, is very suitable for the exploration of the structural richness of the infinite-parameter Lie algebra of local operators, namely for the study of the above discussed algebraic coupling of internal symmetry and space–time dependence, if this turns out to be physically significant. One might also get more definite indications with respect to the physical suitability of the infinite momentum limit.

6. ADDITIONAL REMARKS AND RESULTS

Since this paper was first written (June 1969), further results have been obtained by the authors and by other researchers that helped to shed light on the constructional approach to the representation of current algebra proposed in Sec. 1 and on the structure of the irreducible representations derived in Secs. 3 and 4. The purpose of this section is to make a very brief summary of those developments.

The technique used in this paper for the construction of the irreducible representations of the algebra $\{V^i(\phi)\}$ was an inducing technique on subspaces of the subalgebra $\{V^i(1)\}$. Instead of diagonalizing the integrated subalgebra $\{V^i(1)\}$, one could also think of diagonalizing the generalized Cartan subalgebra [in this case $\{V^3(\phi)\}$].[8] Using this alternative approach and standard techniques of rigged Hilbert space, we were able to prove, under very mild restrictions, that, given any algebra $\{F^i(\phi)\}$ satisfying the commutation relations

$$[F^i(\phi_1), F^j(\phi_2)] = if^{ijk}F^k(\phi_1\phi_2)$$

such that $\{F^i(1)\}$ is semisimple, then in any representation of such an algebra the operators that correspond to $F^i(\phi)$ satisfy the following factorization formula:

$$F^i(\phi) = \int d\nu(x)\phi(x)F^i(x).$$

Although this formula and the defining relation (2) look remarkably alike, the power of the result lies in the essential arbitrariness of the measure ν on the set $\{x\}$ of points in 3-space. The factorization formula suggests that the representations of the current algebra belong to as many classes as the number of possible choices of the measure ν. A natural division in two large classes is obtained immediately: The first one contains the case where ν is a finite discrete measure, the second the case where ν is infinite discrete or continuous. The second class will lead, in general, to representations in nonseparable Hilbert spaces.

A complete classification of the representations of the first class was first worked out in rigorous terms by Joseph.[10] One obtains essentially two subclasses: one corresponding to representations of tensor product algebras $\otimes_n \{F^i(1)\}$, the other corresponding to representations of algebras obtained by contraction of the tensor products. The representations of the tensor product subclass had already been obtained by Roffman.[9]

In the light of the factorization formula, one also

sees that the catalog of representations of Chang, Dashen, and O'Raifeartaigh,[11] because of the nonrigorous nature of its derivation, is not complete,[12] and it leads, in fact, to a classification of representations of the first class. Their results, although less detailed, are essentially those of Joseph's paper.

We are now ready to see where the irreducible representations derived in Secs. 3 and 4 fit in the general scheme outlined above. It turns out that the representations, with $|\eta| \neq 0$ and real, belong to the direct product subclass with an $SU2 \times SU2$ structure and those with $|\eta| = 0$ belong to the contracted subclass with an $E(3)$ structure. This connection was first pointed out by Joseph.[13] The solutions of the functional equations of Sec. 4 that correspond to $|\eta|$ imaginary do not appear in the above classifications because they do not lead to continuous representations if $\{\phi\}$ is a subalgebra of the algebra of continuous functions containing the identity and separating points, with the usual topologies implied.

The above considerations and identifications apply to irreducible representations as those of Secs. 3 and 4. In Sec. 5, to obtain states that are eigenstates of momentum and have well-defined rotational properties, we had to form direct integral spaces that are no longer irreducible under the action of the charge-density operators. It is clear, in fact, that, using current algebra representations of the first class, one will always have to form reducible representations via direct integrals to obtain physical states. It is an open question whether physical states may be formed from irreducible second class representations.

Another point that was studied recently pertains to the hypothetical existence of covariant results in this approach. It turns out that the sequence of operations proposed to form representations of the algebra of current-densities—namely,

(i) study of the irreducible representations of the algebra,

(ii) construction of states of arbitrary momentum and well-defined rotational properties via direct integrals,

(iii) choice of a frame allowing covariance—

is, in fact, possible if the last operation consists in taking the infinite momentum limit. This results from a peculiar relation between sequences of rotations and the operators of the $E(2)$ little group of the infinite momentum frame. We have thus found a partial answer to the open question of Sec. 5 that, as far as we can determine at the present time, seems to display the proposed construction as a probably useful step towards a solution of the problem of representation of the local charge densities. Quasicovariant representations with nontrivial form factors and internal symmetry spectrum are obtained. They may be free from mass spectrum diseases, although this last point needs further checking.

Detailed derivations pertaining to the results discussed in this section will be published in a forthcoming paper.

Note Added in Proof: Because the word *spin* used to qualify the states constructed in Eqs. (57) and (59) may be a source of misunderstanding, it should be emphasized that these states cannot be identified at finite momenta with elements of a Wigner canonical basis of the Poincaré group. Had they been identifiable with physical states, Eqs. (62) would have been manifestly covariant, as pointed out in Sec. 5. We might, however, hope that the $K_y \to \infty$ limit of our construction coincides with the infinite-momentum limit of the charge-densities' matrix-elements between physical states. Thus, it is only at infinite momentum that our S can be related to the eigenvalues of $W_\sigma W^\sigma$, the square of the Pauli–Lubanski operator. The authors are grateful to Professor S. Coleman for raising this point.

* On leave of absence from Laboratório de Física e Engenharia Nucleares, Sacavém, Portugal.

† Supported in part by NSF Grant GU-1598.

‡ Supported in part by the Air Force Office of Scientific Research, Office of Aerospace Research, U.S. Air Force, Under AFOSR Grant EOOAR-68-0010, through the European Office of Aerospace Research.

[1] M. Gell-Mann, Phys. Rev. **125**, 1067 (1962) and Phys. **1**, 63 (1964), hinted at this possibility.

[2] R. F. Dashen and D. H. S. Sharp, Phys. Rev. **165**, 1857 (1968).

[3] R. Dashen and M. Gell-Mann, Phys. Rev. Letters **17**, 340 (1966); the most positive result from this approach to date is described by M. Gell-Mann, D. Horn, and J. Weyers, in *Proceedings of the International Conference on Elementary Particles (Heidelberg, 1967)*, H. Filthuth, Ed. (Interscience, New York, 1969), p. 479; see also H. Leutwyler, Phys. Rev. Letters **20**, 561 (1968), and K. Bardakci and M. B. Halpern, Phys. Rev. **176**, 1686 (1968). The difficulties encountered in this approach have been discussed in many recent articles. See, for example, H. Bebié, F. Ghielmetti, V. Gorgé, and H. Leutwyler, Phys. Rev. **177**, 2133 (1969).

[4] H. Sugawara, Phys. Rev. **170**, 1659 (1968). The connections with canonical field theory have been explained by K. Bardakci, Y. Frishman, and M. B. Halpern, Phys. Rev. **170**, 1353 (1968). This involves a particular limit of a Yang–Mills gauge. The physical content of that limit is somewhat obscure and does not provide a picture of the expected spectrum and interactions.

[5] R. P. Feynman, M. Gell-Mann, and G. Zweig, Phys. Rev. Letters **13**, 678 (1964).

[6] B. Sakita, paper presented at the 14th International Conference on High-Energy Physics, Vienna, 1968.

[7] R. V. Mendes, thesis, University of Texas, 1969.

[8] The diagonalization of the Cartan subalgebra was the technique used by Roffman (Ref. 9) to derive finite-dimensional representations. The method used to derive the factorization formula referred to in Sec. 6 is, in fact, a continuous spectrum version of Roffman's technique.

[9] E. H. Roffman, J. Math. Phys. **8**, 1954 (1967).

[10] A. Joseph (submitted to Commun. Math. Phys.).

[11] S. J. Chang, R. F. Dashen, and L. O'Raifeartaigh, Phys. Rev. **182**, 1805 (1969).

[12] That the catalogue was probably not complete had, in fact, been conjectured by its authors.

[13] A. Joseph, private communication.

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 11, NUMBER 12 DECEMBER 1970

Spinor Treatment of Stationary Space-Times

Z. PERJÉS
Central Research Institute for Physics, Budapest 114, Hungary

(Received 12 June 1970)

A generalized $SU(2)$ spinor calculus is established on the "background space" V_3 of the stationary space-time. The method of spin coefficients is developed in three dimensions. The stationary field equations can be put to a form which in V_3 is analogous to the Newman-Penrose equations. A V_3 filling family of curves is determined by the gravitational field and is called the eigenray congruence. Stationary space-times may be characterized by the geometric properties of eigenrays. The relation of this classification to the algebraic ones is discussed. The method of solving the equations obtainable for various classes is illustrated on the case of nonshearing geodetic eigenrays. Assuming asymptotic flatness, we obtain the Kerr metric.

1. INTRODUCTION

The spinor calculus has recently been widely recognized as an important mathematical tool in the study of general relativity. Probably the most significant approach to melting the generalized notion of spinors into the theory was developed in the paper of Newman and Penrose.[1] The considerable advance in understanding the structure of the gravitational radiational field,[2] discovery of new exact solutions of the gravitational equations,[3] and realization of so far unknown conserved quantities[4] are the most important results based on NP.

Up to now, however, the striking successes of spinor methods did not extend to the theory of gravitational fields admitting Killing motions. Killing vectors could not be fitted suitably in the spinor formalism, and the only attempt to unify spinor and Killing vector techniques[5] apparently did not settle the problem.

In this paper we want to outline a more useful spinor approach to space-times admitting Killing motions. In what follows, for the sake of simplicity, we shall deal with stationary gravitational fields only, noting that the formalism equally well applies to spaces with arbitrary Killing fields. Our method is based on realizing that the stationary space-times permit us to introduce a new "relativity theory" in three dimensions where all methods (like spinor technique) of the theory can once again be put into action.[6]

After a brief survey of the stationary gravitational equations (Sec. 2), we shall develop the spinor algebra and analysis (Secs. 3 and 4, respectively) in stationary space-times.

In curved space, a generalized version of $SL(2, C)$ spinor calculus has long ago been introduced in the literature.[7] The $SL(2, C)$ group of which the spinor representations are dealt with here is the tetrad transformation group. For a detailed treatment of this interpretation we refer the reader to Ref. 8. The timelike Killing vector field distinguishes a direction in every point of the stationary space-time. Consequently, the $SL(2, C) \to SU(2)$ contraction of the tetrad rotation group occurs. Usually this contraction is executed by making spinor components of lower primed index equal to those of upper unprimed index.[9] This procedure will now be generalized in a covariant manner which makes it easy to go over to other subgroups of $SL(2, C)$.

In Sec. 5 we shall develop the method of spin coefficients in the "background" space V_3 and find the relations between the spinor quantities defined in the 4-dimensional space-time and in V_3. The stationary Einstein equations manifest themselves in the form of scalar differential equations which are the 3-dimensional analogs of the Newman-Penrose equations (NP), as will be shown in Sec. 6. Again the equations have strict geometric meaning, which fact will be exploited in establishing an invariant classification of the fields in terms of the "propagation" properties of a certain congruence of curves in the background space. These curves, called the "eigenrays," are uniquely defined in V_3 by the gravitational field itself.

Up to now the Kerr metric appeared an "incident" solution of the stationary gravitational field equations.[10] In Sec. 7 we shall show how this solution can be obtained systematically from the governing equations of stationary gravitational fields. It will be seen that the eigenrays of the Kerr metric are shear free. The results for stationary metrics with shearing eigenrays will be published elsewhere.

2. STATIONARY FIELD EQUATIONS

We now briefly recapitulate the theory of stationary space-times.[11-13]

The stationary gravitational fields are characterized by the existence of a timelike Killing vector field a_μ

satisfying[14]

$$a_{\mu|\nu} + a_{\nu|\mu} = 0. \tag{1}$$

The coordinate system can be chosen such that $x^0 = t$ is the trajectory of motion. Then we have

$$a^\mu = \delta^\mu_0 \tag{2}$$

and $\tilde{g}_{\mu\nu}$ independent of t.

We write the line element of the 4-dimensional space–time V_4 in the form

$$d\tilde{s}^2 = -f^{-1}\,ds^2 + f(dt + \omega_i\,dx^i)^2. \tag{3}$$

Equation (3) may be regarded as the definition of the "background" V_3 with the spacelike line element $ds^2 = g_{ik}\,dx^i\,dx^k$.

The t-independent form (3) of the line element is preserved by the following transformations:

$$x^{i'} = x^{i'}(x^j), \tag{4a}$$

$$t' = t + t^0(x^j). \tag{4b}$$

The (3 + 1) decomposition of the 4-dimensional Ricci tensor yields the relations

$$(\partial_i - G_i + \bar{G}_i)G^i = -f^{-2}\tilde{R}_{00}, \tag{5}$$

$$G_{i;j} - G_{j;i} + G_i\bar{G}_j - \bar{G}_iG_j = -if^{-2}\epsilon_{ijk}\tilde{R}^k_0(g)^{\frac{1}{2}}, \tag{6}$$

$$R_{ij} + G_i\bar{G}_j + \bar{G}_iG_j = f^{-2}(g_{ik}g_{jl}\tilde{R}^{kl} - g_{ij}\tilde{R}_{00}). \tag{7}$$

Here the complex 3-dimensional vector G_i is defined by

$$G_i \overset{\text{def}}{=} \mathcal{E}_i/2f, \tag{8}$$

where, following the notation of Ernst,[11]

$$\mathcal{E}_i \overset{\text{def}}{=} f_{,i} + i\varphi_i, \tag{9}$$

$$\varphi_i \overset{\text{def}}{=} \epsilon_{ijk}\omega^{j;k}(g)^{\frac{1}{2}}f^2. \tag{10}$$

Now the Einstein equations

$$\tilde{R}_{\mu\nu} - \tfrac{1}{2}\tilde{g}_{\mu\nu}\tilde{R} = -kT_{\mu\nu} \tag{11}$$

are to be imposed upon the Ricci tensor $\tilde{R}_{\mu\nu}$ of V_4 in identities (5), (6), and (7). In the absence of matter, the right-hand sides vanish.

Formally, in Eqs. (5)–(7) and (11) a 3-dimensional relativity theory in V_3 in the presence of an additional complex "material" vector field is comprised,[6] (7) providing the Einstein equations and (5) and (6) the "matter equations" in V_3.

3. SPINOR ALGEBRA

Spinors are connected with world tensors by the quantities $\sigma_{\mu AB'}$ satisfying[7]

$$\sigma_{\mu AC'}\sigma_{\nu B}{}^{C'} + \sigma_{\nu AC'}\sigma_{\mu B}{}^{C'} = \tilde{g}_{\mu\nu}\epsilon_{AB} \tag{12}$$

in each point of V_4. Now the coordinate system is chosen according to Eq. (2) such that $\sigma_{\mu AB'}$ need not depend on t. Actually, we shall always take $\sigma_{\mu AB'}$ independent of t. For a fixed μ, $[\sigma_{\mu AB'}]$ is a 2×2 Hermitian matrix:

$$\bar{\sigma}_{\mu AB'} = \sigma_{\mu BA'}. \tag{13}$$

Spinor indices are raised and lowered by means of the real antisymmetric metric spinor ϵ_{AB} according to the formulas

$$\xi^A = \epsilon^{AB}\xi_B, \tag{14}$$

$$\xi_A = \xi^B\epsilon_{BA}. \tag{15}$$

ϵ_{AB} can be chosen such that[7]

$$[\epsilon_{AB}] = [\epsilon_{A'B'}] = \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix}. \tag{16}$$

In stationary space–time the Killing vector a^μ contracted with the connecting quantity $\sigma_{\mu AB'}$ yields a link between primed and unprimed spinor components. With the choice (2) of the coordinate system,

$$\sigma_{\mu AB'}a^\mu = \sigma_{0AB'}. \tag{17}$$

The (3 + 1) decomposition of the defining equation (12) of σ's yields

$$\sigma_{0AC'}\sigma_{0B}{}^{C'} = \tfrac{1}{2}f\epsilon_{AB} \tag{18}$$

and, introducing the quantities σ^i_{AB} by

$$\sigma^i_{AB} \overset{\text{def}}{=} (-2^{\frac{1}{2}}/f)\sigma^i_{AC'}\sigma_{0B}{}^{C'}, \tag{19}$$

$$\sigma^i_{AB} - \sigma^i_{BA} = 0. \tag{20}$$

From the spacelike components of Eq. (12) we get

$$\sigma^A_{iC}\sigma^C_{jB} + \sigma^A_{jC}\sigma^C_{iB} = g_{ij}\delta^A_B. \tag{21}$$

By use of the identity

$$\epsilon_{A[B}\epsilon_{CD]} = 0, \tag{22}$$

it is an easy exercise to derive the useful formula

$$\sigma_{iAB}\sigma^i_{CD} = -\tfrac{1}{2}(\epsilon_{AC}\epsilon_{BD} + \epsilon_{AD}\epsilon_{BC}). \tag{23}$$

The quantities $[\sigma_{iAB}]$ satisfy the commutation relations of the generators of an $SU(2)$ group:

$$\sigma_{iAC}\sigma_{jB}{}^{C} - \sigma_{jAC}\sigma_{iB}{}^{C} = 2^{\frac{1}{2}}i\epsilon_{ijk}\sigma^k_{AB}(g)^{\frac{1}{2}}. \tag{24}$$

This equation can be proved to hold at any point of the space–time by introducing locally Minkowskian coordinates.

The invariant $\sigma_{0AB'}$ relates an arbitrary spinor ξ^A with

$$\xi^{\dagger A} \overset{\text{def}}{=} (2/f)^{\frac{1}{2}}\sigma_0{}^{AB'}\xi_{B'}. \tag{25}$$

$\xi^{\dagger A}$ will be called the adjoint of ξ^A. In stationary space–time a special representation of $[(2/f)^{\frac{1}{2}}\sigma_0{}^{AB'}]$ is the unit matrix. On restricting the spin transformation group to $SU(2)$ which preserves the form of unit matrix only, one would get to the usual $SU(2)$ spinor calculus and could drop the trivial $(2/f)^{\frac{1}{2}}\sigma_0{}^{AB'}$ factors. However, we now will not use this restriction in order to make the method applicable for arbitrary Killing motions.

The adjunction operation, applied twice to a first rank spinor ξ^A, affects as a factor (-1):

$$\xi^{\dagger\dagger}{}_A = -\xi_A. \tag{26}$$

The complex conjugate of a scalar product of two spinors:

$$\bar{\xi}_{A'}\bar{\eta}^{A'} = \xi^\dagger_A\eta^{\dagger A}. \tag{27}$$

Especially,

$$\bar{\xi}^\dagger_{A'}\bar{\eta}^{A'} = \eta^\dagger_A\xi^A. \tag{28}$$

We define the norm of a spinor by

$$\|\xi\|^2 \overset{\text{def}}{=} \xi^{\dagger A}\xi_A. \tag{29}$$

With the special representation of $(2/f)^{\frac{1}{2}}\sigma_0{}^{AB'}$ we have

$$\|\xi\|^2 = |\xi_0|^2 + |\xi_1|^2, \tag{30}$$

so that the norm is real and nonnegative.

Higher-rank spinors under adjunction operation may behave in a definite manner. For example, in our formalism

$$\sigma^\dagger_{AB} = -\sigma_{AB} \tag{31}$$

corresponds to the usual notion of hermiticity [note the minus sign in Eq. (31)]. Actually, by use of Eq. (24) it can be proven that σ^i_{AB} satisfies Eq. (31).

4. SPINOR ANALYSIS

The covariant derivative of a first-rank spinor ξ_A taken in V_4 is defined by[8]

$$\xi_{A|\mu} \overset{\text{def}}{=} \xi_{A,\mu} - \tilde{\Gamma}_\mu{}^B{}_A\xi_B. \tag{32}$$

$\tilde{\Gamma}_\mu{}^B{}_A$ is called the spinor affine connection of V_4. Stipulating that the 4-covariant derivatives of $\sigma_{\mu AB'}$ and ϵ_{AB} should vanish, we get the explicit expression for the quantities $\tilde{\Gamma}_\mu{}^B{}_A$:

$$\tilde{\Gamma}_\mu{}^B{}_A = \tfrac{1}{2}\sigma_\alpha{}^{BF'}(\sigma^\beta{}_{AF'}\tilde{\Gamma}_{\mu\beta}{}^\alpha + \sigma^\alpha_{AF',\mu}). \tag{33}$$

In V_3, the covariant derivative of a first-rank spinor is defined similarly:

$$\xi_{A;i} \overset{\text{def}}{=} \xi_{A,i} - \Gamma_i{}^B{}_A\xi_B. \tag{34}$$

On requiring

$$\sigma_{iAB;j} = \epsilon_{AB;j} = 0, \tag{35}$$

$\Gamma_i{}^B{}_A$ takes the form

$$\Gamma_i{}^B{}_A = -\tfrac{1}{2}\sigma^{BC}_j(\sigma^j_{AC,i} + \Gamma_{ik}{}^j\sigma^k_{AC}). \tag{36}$$

We want to have $\tilde{\Gamma}_\mu{}^B{}_A$ expressed in terms of $\Gamma_i{}^B{}_A$. The results of the rather lengthy calculations are

$$\tilde{\Gamma}_0{}^B{}_A = -\frac{1}{2\sqrt{2}}\sigma^{jB}{}_A\mathfrak{E}_j, \tag{37}$$

$$\tilde{\Gamma}_i{}^B{}_A = \Gamma_i{}^B{}_A - \frac{1}{2\sqrt{2}}\omega_i\sigma^{jB}_A\mathfrak{E}_j - \frac{i}{2\sqrt{2}}\epsilon_{ijk}\frac{\mathfrak{E}^j}{f}\sigma^{kB}_A(g)^{\frac{1}{2}}. \tag{38}$$

Here we note that one has to take care of the commutation of covariant differentiation with adjunction of spinor because $[(2/f)^{\frac{1}{2}}\sigma_0{}^{AB'}]_{;i}$ does not vanish necessarily.

5. SPINOR BASIS

In the spin space, a basic spinor dyad ζ_{aA},

$$\zeta_{0A} \equiv o_A, \quad \zeta_{1A} \equiv \iota_A, \tag{39}$$

is introduced, with the aid of which any spinor can be expressed as the set of its dyad components (NP), e.g., the spinor $\xi_{ABC'}$ has the dyad components

$$\xi_{abc'} = \xi_{ABC'}\zeta^A_a\zeta^B_b\bar{\zeta}^{C'}_{c'}, \tag{40}$$

the algebraic properties being unaltered in the dyadic form. The normalization is chosen to have

$$o_A\iota^A = 1 \tag{41}$$

for the dyad.

The spinor base fixes a basic vector tetrad ($\tilde{l}^\mu$, $\tilde{n}^\mu$, $\tilde{m}^\mu$, $\bar{\tilde{m}}^\mu$) in V_4 by

$$\begin{aligned}\tilde{l}^\mu &= o_A\sigma^{\mu AB'}\bar{o}_{B'},\\ \tilde{n}^\mu &= \iota_A\sigma^{\mu AB'}\bar{\iota}_{B'},\\ \tilde{m}^\mu &= o_A\sigma^{\mu AB'}\bar{\iota}_{B'}.\end{aligned} \tag{42}$$

In V_3, we adopt the vector base z^i_m; $\boldsymbol{m} = 0, +, -$,

$$\begin{aligned}z^i_0 &= l^i,\\ z^i_+ &= m^i,\\ z^i_- &= \bar{m}^i,\end{aligned} \tag{43}$$

with l^i real and with the orthonormality properties

$$[z_{mi}z_n{}^i] \equiv [g_{mn}] = \begin{bmatrix}1 & 0 & 0\\ 0 & 0 & 1\\ 0 & 1 & 0\end{bmatrix}, \quad \boldsymbol{m}, \boldsymbol{n} = 0, +, -. \tag{44}$$

This basis can be traced back to a novel spinor dyad $\eta_{aA} = (\eta_A, \eta^\dagger{}_A)$ with the norm

$$\eta_A\eta^{\dagger A} = 1 \tag{45}$$

via the relations

$$\begin{aligned}l^i &= -2^{\frac{1}{2}}\eta_A\sigma^{iAB}\eta^\dagger_B,\\ m^i &= \eta_A\sigma^{iAB}\eta_B.\end{aligned} \tag{46}$$

The η dyad is essentially a specialized $SL(2, C)$ spinor base and in stationary space-times (with timelike Killing vector field) it may be called the $SU(2)$ base.

The transformation rule between the bases is of the form

$$\zeta_{aA} = \alpha_a{}^b \eta_{bA} \tag{47}$$

with $\det[\alpha_a{}^b] = 1$. In the following, we shall stick to the convenient particular choice

$$\begin{aligned} o_A &= (2/f)^{\frac{1}{2}} \eta_A, \\ \iota_A &= (f/2)^{\frac{1}{2}} \eta_A^\dagger \end{aligned} \tag{48}$$

of the ζ dyad.

Now the (3 + 1) decomposition of the basic tetrad in V_4 can be done. The results are

$$\begin{aligned} \tilde{l}^\mu &= [l^i, f^{-1} - (l_j \omega^j)], \\ \tilde{m}^\mu &= (f)[m^i, -(m_j \omega^j)], \\ \tilde{n}^\mu &= -\tfrac{1}{2} f \tilde{l}^\mu + a^\mu. \end{aligned} \tag{49}$$

The adjunction rules for η-dyad components of spinors follow in a straightforward manner from the above definitions:

$$\begin{aligned} (\bar{\xi}_0) &= (\xi^\dagger)_1, \\ (\bar{\xi}_1) &= -(\xi^\dagger)_0. \end{aligned} \tag{50}$$

As it was found in NP, the Einstein equations can be put down as a set of geometric conditions for the spin coefficients $\tilde{\Gamma}_{abcd'}$. The definition of spin coefficients is taken as follows:

$$\tilde{\Gamma}_{abcd'} \overset{\text{def}}{=} \zeta_{aA|\mu} \zeta_b^A \sigma_{cd'}^\mu, \tag{51}$$

$\tilde{\Gamma}_{abcd'}$ being symmetric in the first pair of indices:

$$\tilde{\Gamma}_{abcd'} = \tilde{\Gamma}_{bacd'}. \tag{52}$$

That means that, in the general $SL(2, C)$ spinor calculus, one has 12 independent complex spin coefficients. In NP, each of the quantities $\tilde{\Gamma}_{abcd'}$ is denoted by a single Greek letter:

$\tilde{\Gamma}_{abcd'} =$

$cd' \backslash ab$	00	01 or 10	11
00′	$\tilde{\kappa}$	$\tilde{\epsilon}$	$\tilde{\pi}$
10′	$\tilde{\rho}$	$\tilde{\alpha}$	$\tilde{\lambda}$
01′	$\tilde{\sigma}$	$\tilde{\beta}$	$\tilde{\mu}$
11′	$\tilde{\tau}$	$\tilde{\gamma}$	$\tilde{\nu}$

. (53)

In our curious "$SU(2)$ formalism," a similar definition can be given for the spin coefficients:

$$\Gamma_{abcd} \overset{\text{def}}{=} \eta_{aA;i} \eta_b^A \sigma_{cd}^i, \tag{54}$$

Γ_{abcd} being now symmetric both in the first and second pair of indices:

$$\Gamma_{abcd} = \Gamma_{bacd} = \Gamma_{abdc}, \tag{55}$$

such that one has now four complex and one pure imaginary (Γ_{0101}) spin coefficient. With the individual nomenclature, we have

$\Gamma_{abcd} =$

$cd \backslash ab$	00	01 10	11
00	$2^{-\frac{1}{2}}\sigma$	$-\frac{1}{2}\bar{\tau}$	$2^{-\frac{1}{2}}\bar{\rho}$
01 10	$-\frac{1}{2}\kappa$	$-\frac{1}{2}(2^{-\frac{1}{2}})\epsilon$	$-\frac{1}{2}\bar{\kappa}$
11	$-2^{-\frac{1}{2}}\rho$	$-\frac{1}{2}\tau$	$-2^{-\frac{1}{2}}\bar{\sigma}$

. (56)

Here the numerical factors are introduced in order to make the final relations as simple as possible.

We now need the notion of complex Ricci rotation coefficients which are the analogs of spin coefficients in vector terms:

$$\gamma_{mnp} \overset{\text{def}}{=} z_{mi;j} z_n^i z_p^j. \tag{57}$$

Using the well-known correspondence between spin coefficients and Ricci rotation coefficients (NP), we can express each of the Ricci coefficients according to the following table:

$\gamma_{mnp} \overset{\text{def}}{=}$

$p \backslash mn$	−0	+0	+−
0	$\bar{\kappa}$	κ	ϵ
+	$\bar{\rho}$	σ	$-\bar{\tau}$
−	$\bar{\sigma}$	ρ	τ

. (58)

For the scalar differential operators $\partial_m \equiv z_m^i \partial_i$, we use the notation

$$\partial_0 = D, \quad \partial_+ = \delta, \quad \partial_- = \bar{\delta}. \tag{59}$$

The correspondence between $SL(2, C)$ spin coefficients and ours can be established by use of the decomposition (37) and (38) of the 4-dimensional

affine connections. We merely report the results:

$$\begin{aligned}
&\tilde{\kappa} = f^{-\frac{1}{2}}(\kappa - 2G_+), \quad \tilde{\epsilon} = (\tfrac{1}{4})(2\epsilon + G_0 - \bar{G}_0),\\
&\tilde{\pi} = (f^{\frac{1}{2}}/2)\bar{\kappa}, \quad \tilde{\rho} = \rho + G_0, \quad \tilde{\sigma} = \sigma,\\
&\tilde{\tau} = -(f^{\frac{1}{2}}/2)\kappa, \quad \tilde{\alpha} = (f^{\frac{1}{2}}/4)(2\tau + G_- - \bar{G}_-),\\
&\tilde{\beta} = -(f^{\frac{1}{2}}/4)(2\bar{\tau} + 3G_+ + \bar{G}_+),\\
&\tilde{\gamma} = -(f/8)(2\epsilon - 3G_0 - \bar{G}_0),\\
&\tilde{\lambda} = (f/2)\bar{\sigma}, \quad \tilde{\mu} = (f/2)(\bar{\rho} + G_0),\\
&\tilde{\nu} = (f^{\frac{3}{2}}/4)(-\bar{\kappa} + 2G_-).
\end{aligned} \tag{60}$$

These relations will be of great value when studying the 3-dimensional ray optics and Petrov types of stationary gravitational fields, in the next section.

6. THE FIELD EQUATIONS IN NEW GUISE

The identities (5), (6), and (7) are rewritten in terms of "triad components", i.e., with all quantities projected onto the vector base z^i_m:

$$G^m{}_{;m} + \gamma_{mn}{}^n G^m - (G_m - \bar{G}_m)G^m = -\lambda, \tag{61}$$

$$G_{m;n} - G_{n;m} - \gamma_m{}^p{}_n G_p + \gamma_n{}^p{}_m G_p - \bar{G}_m G_n + \bar{G}_n G_m = -i\epsilon_{mnp}\chi^p(g)^{\frac{1}{2}}, \tag{62}$$

$$R_{mn} + G_m\bar{G}_n + \bar{G}_m G_n = 2\Phi_{mn}, \tag{63}$$

where the shorthand notation

$$f^{-2}\tilde{R}_{00} \overset{\text{def}}{=} \lambda, \tag{64}$$

$$f^{-2}\tilde{R}^p_0 \overset{\text{def}}{=} \chi^p, \tag{65}$$

$$f^{-2}(\tilde{R}^{ij} - g^{ij}\tilde{R}_{00})z_{im}z_{jn} \overset{\text{def}}{=} 2\Phi_{mn} \tag{66}$$

is being initiated for the decomposed Ricci tensor of V_4. Together with the above relations, the Ricci identities in V_3, acting on the base vectors z^i_m, are

$$R_{mnpq} = \gamma_{mnp;q} - \gamma_{mnq;p} + \gamma^r{}_{mq}\gamma_{rnp} - \gamma^r{}_{mp}\gamma_{rnq} + \gamma_{mnr}(\gamma^r{}_{pq} - \gamma^r{}_{qp}); \tag{67}$$

the decomposition of the 3-dimensional curvature tensor into irreducible parts,

$$R_{mnpq} = -g_{mp}R_{nq} + g_{mq}R_{np} - g_{nq}R_{mp} + g_{np}R_{mq} - \tfrac{1}{2}R(g_{mq}g_{np} - g_{mp}g_{nq}), \tag{68}$$

and the commutator of operations ∂_m acting on a scalar function φ (NP),

$$\varphi_{;m;n} - \varphi_{;n;m} = (\gamma_m{}^r{}_n - \gamma_n{}^r{}_m)\varphi_{;r}, \tag{69}$$

are written out in full detail. From (61) and (62) we have

$$DG_0 + \bar{\delta}G_+ + \delta G_- - (\rho + \bar{\rho})G_0 + (\kappa - \bar{\tau})G_- + (\bar{\kappa} - \tau)G_+ - (G_0 - \bar{G}_0)G_0 - (G_+ - \bar{G}_+)G_- - (G_- - \bar{G}_-)G_+ = -\lambda, \tag{70a}$$

$$\bar{\delta}G_0 - DG_- + \rho G_- + \bar{\sigma}G_+ + \bar{\kappa}G_0 - \epsilon G_- - \bar{G}_0G_- + \bar{G}_-G_0 = -\chi_+, \tag{70b}$$

$$\delta G_0 - DG_+ + \sigma G_- + \bar{\rho}G_+ + \kappa G_0 + \epsilon G_+ - \bar{G}_0G_+ + \bar{G}_+G_0 = \chi_-, \tag{70c}$$

$$\bar{\delta}G_+ - \delta G_- - \rho G_0 - \tau G_+ + \bar{\rho}G_0 + \bar{\tau}G_- - \bar{G}_+G_- + \bar{G}_-G_+ = \chi_0. \tag{70d}$$

The Ricci identities (67) when combined with (68) and (63) yield

$$D\sigma - \delta\kappa - \epsilon\sigma - \bar{\tau}\kappa - \kappa^2 + \sigma(-\epsilon - \bar{\rho}) - \rho\sigma + 2\Phi_{++} - 2G_+\bar{G}_+ = 0, \tag{71a}$$

$$D\rho - \bar{\delta}\kappa + \tau\kappa - \kappa\bar{\kappa} - \sigma\bar{\sigma} - \rho^2 + \Phi_{00} - G_0\bar{G}_0 = 0, \tag{71b}$$

$$D\tau - \bar{\delta}\epsilon + \kappa\bar{\sigma} - \rho\bar{\kappa} + \tau\epsilon - \epsilon\bar{\kappa} + \bar{\tau}\bar{\sigma} - \tau\rho + 2\Phi_{0-} - G_0\bar{G}_- - \bar{G}_0G_- = 0, \tag{71c}$$

$$\bar{\delta}\sigma - \delta\rho - 2\tau\sigma - \kappa(\rho - \bar{\rho}) - 2\Phi_{0+} + \bar{G}_0G_+ + G_0\bar{G}_+ = 0, \tag{71d}$$

$$\delta\tau + \bar{\delta}\bar{\tau} + \sigma\bar{\sigma} - \rho\bar{\rho} - 2\tau\bar{\tau} + \epsilon(\rho - \bar{\rho}) + \Phi_{00} - 2\Phi_{+-} + G_0\bar{G}_0 - G_+\bar{G}_- - \bar{G}_+G_- = 0. \tag{71e}$$

It is a remarkable fact that, while in the 4-dimensional theory the Ricci identities mean 18 independent conditions on the rotation coefficients, we now have five relations only. For φ real, the commutators (69) give two equations:

$$(D\delta - \delta D)\varphi = [(\bar{\rho} + \epsilon)\delta + \sigma\bar{\delta} + \kappa D]\varphi, \tag{72a}$$

$$(\delta\bar{\delta} - \bar{\delta}\delta)\varphi = [\bar{\tau}\bar{\delta} - \tau\delta + (\bar{\rho} - \rho)D]\varphi. \tag{72b}$$

Equations (70), (71), and (72) can equally well be attained in the spinor dyad formalism. To the Newman–Penrose equations the spinor calculus offers the simpler way. Here, on the contrary, the use of vector terms is somewhat more advantageous.

In order to get to the new form of the stationary gravitational equations, of course, one has to impose the Einstein conditions (11) upon the quantities Φ_{mn}, χ_m, and λ in (71), (72), and (73). The geometric meaning of this approach to the stationary space–time problem will be pursued in the remainder of this section.

Most of the rotation coefficients in V_3 have their 4-dimensional analogs, which fact is stressed by singling out the corresponding rotation coefficients by the same Greek letter. Thus, κ, ρ, and σ express in V_3 the same properties of the congruence to which l^i is a tangent vector as $\tilde{\kappa}$, $\tilde{\rho}$, and $\tilde{\sigma}$, respectively, in V_4 for the congruence with $\tilde{l}^\mu$ tangent. In particular, $-\kappa = m^i l_{i;j}l^j$ is the first curvature of the congruence projected onto m^i; κ vanishes if and only if the congruence

is geodetic in the background V_3. Then

$$\mathrm{Re}\,\rho = -\tfrac{1}{2}l^i{}_{;i} \tag{73}$$

gives the divergence,

$$\mathrm{Im}\,\rho = \tfrac{1}{2}[(l_{i,j} - l_{j,i})l^{i;j}]^{\frac{1}{2}} \tag{74}$$

is the rotation, and

$$|\sigma| = 2^{-\frac{1}{2}}[l_{(i;j)}l^{i;j} - \tfrac{1}{2}(l^i{}_{;i})^2]^{\frac{1}{2}} \tag{75}$$

is the shear of the congruence. The imaginary quantity $\epsilon = m_{i;j}\bar{m}^i l^j$ shows how the $(\mathbf{m}, \bar{\mathbf{m}})$ pair rotates when moving along the congruence. As such, it is not characteristic of the geometry of the congruence itself, and in this respect it resembles $\bar{\epsilon}$. The only rotation coefficient not having a 4-dimensional analog is $\tau = m_{i;j}\bar{m}^i\bar{m}^j$.

In the following considerations it will prove useful to introduce the notion of eigenrays of the stationary gravitational fields. To this purpose, we momentarily take $\tilde{l}^\mu$ tangent to a geodetic congruence in V_4, and, making use of the fact that the length of $\tilde{l}^\mu$ is constant (zero), we put the geodetic equation for the congruence in the form

$$(\tilde{l}_{\mu,\nu} - \tilde{l}_{\nu,\mu})\tilde{l}^\nu = 0. \tag{76}$$

From the (3 + 1) decomposition (49) of $\tilde{l}^\mu$, it is seen that, for $\mu = 0$, (76) is identically satisfied. We are left with the 3-covariant equation

$$fl_{i;j}l^j + f_{;i} - (f_{;j}l^j)l_i + \epsilon_{ijk}\varphi^j l^k(g)^{\frac{1}{2}} = 0. \tag{77}$$

Now, if the purely algebraic requirement for l^i,

$$f_{;i} - (f_{;j}l^j)l_i + \epsilon_{ijk}\varphi^j l^k(g)^{\frac{1}{2}} = 0, \tag{78}$$

is fulfilled, then l^i is tangent to a geodesic in V_3. (78) gives two independent conditions on l^i because its l^i projection is identically satisfied. Therefore, given the vectors $f_{;i}$ and φ_i, (78) uniquely determines the direction of l^i. If the space–time is static ($\varphi_i = 0$), (78) means that l^i is in V_3 the unit-normal vector of the equipotential surfaces $f = \text{const}$. For a general stationary field, the geometric content of (78) is more complicated and can be visualized as shown on Fig. 1. The curves defined in V_3 by Eq. (78) will be called the *eigenrays of the gravitational field.*

On projecting Eq. (78) onto the base vector m^i, it takes the simple form

$$G_+ = 0. \tag{79}$$

In many cases, a convenient choice of the base vector l^i can be made by putting it tangent to the eigenray

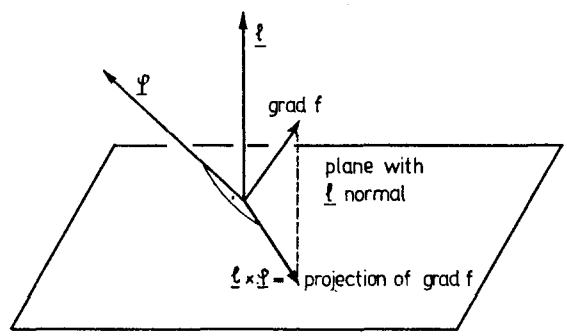


FIG. 1. The "eigendirection" of the gravitational field.

congruence. The only remaining "triad transformation" is then of the form

$$\mathbf{m}' = \mathbf{m}e^{iC}, \tag{80}$$

where C is an arbitrary real function.

An invariant classification of stationary gravitational fields can be achieved leaning on the "optical" properties of the eigenray congruence. In Ref. 6 it has been suggested that stationary axially symmetric fields should be classified by the algebraic structure of the trace-free part $P_i^j = R_i^j - \tfrac{1}{3}\delta_i^j R$ of the 3-dimensional Ricci tensor. It was found that the asymptotically flat fields are either general type (G) or, if two eigenvalues of P_i^j coincide, degenerate type (D). An important example of type D spaces is the Schwarzschild solution, while in this scheme the Kerr metric is of type G.

This invariant classification can be retained unaltered when dropping the restriction to axial symmetry (P_i^j is then defined in V_3). The scheme is now refined by considering the propagation properties of the eigenrays. In particular, the space–times with geodetic eigenrays are characterized by $G_+ = \kappa = 0$ and, if in addition $\mathrm{Re}\,\rho$, $\mathrm{Im}\,\rho$, or σ vanishes, then the eigenrays in V_3 are divergenceless, nonrotating or shear-free, respectively [cf. Eqs. (73), (74), and (75)]. In Sec. 7 we shall point out that the Kerr metric possesses nonshearing geodetic eigenrays and just this property yields the reasonable assumption under which the Kerr solution can be derived from the stationary field equations.

Now, for the rest of this section, we turn to the interrelation between the Petrov types and our classes.

If l^i is fixed by (78), then in general $\tilde{l}^\mu$ is not a geodesic tangent vector in V_4; but if it is, then the eigenrays are geodesics of V_3. Restricting ourselves to vacuum space–times ($\tilde{R}_{\mu\nu} = 0$; for the rest of this section this condition will be assumed to hold), from the Goldberg–Sachs theorem (NP) we find that a

stationary space–time with shear-free geodetic eigenray congruence is algebraically special. This fact reappears when we express the dyad components of the 4-dimensional curvature spinor in terms of 3-covariant quantities[15]:

$$\Psi'_0 \equiv -\tilde{R}_{\alpha\beta\gamma\delta} l^\alpha \tilde{m}^\beta l^\gamma \tilde{m}^\delta$$
$$= 2[\delta G_+ - \sigma G_0 + \bar{\tau} G_+ + (2G_+ + \bar{G}_+)G_+], \quad (81a)$$

$$\Psi'_1 \equiv -\tilde{R}_{\alpha\beta\gamma\delta} l^\alpha \tilde{n}^\beta l^\gamma \tilde{m}^\delta$$
$$= -(f)^{\frac{1}{2}}[DG_+ - \kappa G_0 - \epsilon G_+ + (2G_0 + \bar{G}_0)G_+], \quad (81b)$$

$$\Psi'_2 \equiv -\tfrac{1}{2}\tilde{R}_{\alpha\beta\gamma\delta}(l^\alpha \tilde{n}^\beta l^\gamma \tilde{n}^\delta - l^\alpha \tilde{n}^\beta \tilde{m}^\gamma \bar{\tilde{m}}^\delta)$$
$$= \tfrac{1}{2} f[DG_0 + \bar{\kappa} G_+ + \kappa G_- + (G_0 + \bar{G}_0)G_0 - 2G_+ G_-], \quad (81c)$$

$$\Psi'_3 \equiv \tilde{R}_{\alpha\beta\gamma\delta} l^\alpha \tilde{n}^\beta \tilde{n}^\gamma \bar{\tilde{m}}^\delta$$
$$= \tfrac{1}{2} f^{\frac{3}{2}}[DG_- - \bar{\kappa} G_0 + \epsilon G_- + (2G_0 + \bar{G}_0)G_-], \quad (81d)$$

$$\Psi'_4 \equiv -\tilde{R}_{\alpha\beta\gamma\delta} \tilde{n}^\alpha \bar{\tilde{m}}^\beta \tilde{n}^\gamma \bar{\tilde{m}}^\delta$$
$$= \tfrac{1}{2} f^2[\delta G_- - \bar{\sigma} G_0 + \tau G_- + (2G_- + \bar{G}_-)G_-]. \quad (81e)$$

If a geodetic eigenray congruence exists in V_3 and l^i is chosen to be its tangent vector ($G_+ = \kappa = 0$), then Ψ'_1 vanishes. If, in addition, the eigenray congruence is shear-free ($\sigma = 0$), then $\Psi'_0 = \Psi'_1 = 0$ and the space is algebraically special; l^μ is one of the propagation vectors (NP).

7. GEODETIC EIGENRAYS. THE KERR METRIC

In this section we shall establish the field equations for vacuum stationary space–times with geodetic eigenrays, and, thereafter, the way of solving them will be illustrated on the particular class of non-shearing eigenrays.

We have

$$\Phi_{mn} = \chi_m = \lambda = G_+ = \kappa = 0. \quad (82)$$

The coordinate x^1 will be chosen the affine parameter r of the eigencongruence. The x^a, $a = 2, 3$, label the eigenrays. With this choice, we may write, for the vector base,

$$l^i = \delta^i_1,$$
$$m^i = \omega\delta^i_1 + \xi^a \delta^i_a, \quad a = 2, 3. \quad (83)$$

The scalar differential operators have the form

$$D = \frac{\partial}{\partial r},$$
$$\delta = \omega \frac{\partial}{\partial r} + \xi^a \frac{\partial}{\partial x^a}. \quad (84)$$

(The summation convention is understood to hold for the index a.) The coordinate freedom (4a) thus has been reduced to

$$r' = r + r^0(x^a) \quad \text{(shifting the origin)}, \quad (85a)$$
$$x^{a'} = x^{a'}(x^b) \quad \text{(relabeling eigenrays)}. \quad (85b)$$

By use of the "triad freedom" (80), $\epsilon = m_{i;j}\bar{m}^i l^j$ can be made zero and still remains

$$\mathbf{m}' = \mathbf{m} e^{iC^0} \quad (86)$$

with C^0 real and independent of r.[16]

The connection between the metric tensor and spin coefficients is achieved by applying the commutators (72) to each of the invariants r and x^a:

$$D\omega = \bar{\rho}\omega + \sigma\bar{\omega}, \quad (87a)$$
$$D\xi^a = \bar{\rho}\xi^a + \sigma\bar{\xi}^a, \quad a = 2, 3, \quad (87b)$$
$$\delta\bar{\omega} - \bar{\delta}\omega = \bar{\tau}\bar{\omega} - \tau\omega + \bar{\rho} - \rho, \quad (87c)$$
$$\delta\bar{\xi}^a - \bar{\delta}\xi^a = \bar{\tau}\bar{\xi}^a - \tau\xi^a, \quad a = 2, 3. \quad (87d)$$

From Eqs. (70) and (71) we get

$$D\sigma = (\rho + \bar{\rho})\sigma, \quad (88a)$$
$$D\rho = \rho^2 + \sigma\bar{\sigma} + G_0\bar{G}_0, \quad (88b)$$
$$D\tau = \rho\tau - \bar{\sigma}\bar{\tau} + \bar{G}_0 G_-, \quad (88c)$$
$$DG_0 = 2\rho G_0 + (G_0 - \bar{G}_0)G_0, \quad (88d)$$
$$\delta\rho - \bar{\delta}\sigma = -2\sigma\tau + \bar{G}_+ G_0, \quad (88e)$$
$$\delta\tau + \bar{\delta}\bar{\tau} = \rho\bar{\rho} - \sigma\bar{\sigma} + 2\tau\bar{\tau} - G_0\bar{G}_0 + G_-\bar{G}_+. \quad (88f)$$

The identities (70b)–(70d) yield no new information because they are consequences of the system (87).

Equations (87) and (88) form a complete system for determining the field quantities in a vacuum stationary space–time with geodetic eigencongruence. As it was stated in the preceding section, the invariant Ψ'_1 now vanishes. For Ψ'_0 we have

$$\Psi'_0 = -2\sigma G_0. \quad (89)$$

Thus, we must not expect new nonshearing metrics ($\sigma = 0$) from Eqs. (87) and (88) because this class is trivially contained in the soluble case $\Phi_{mn} = \Psi'_0 = \Psi'_1 = \tilde{\kappa} = \tilde{\sigma} = 0$ of the Newman–Penrose equations. The spaces with nonvanishing σ arise, however, from equations which previously have not been considered. The study of this class deserves, therefore, more attention and is put off to a forthcoming paper. In the following, we will be content to solve the system (87) and (88) for

$$\sigma = 0, \quad (90)$$

assuming asymptotic flatness. The equations like the Newman–Penrose ones offer a well-determined sequence in which they can be solved in turn. First, the

"radial" dependence of the quantities is to be obtained from the equations containing the operator D. Then the r dependence of the quantities is substituted into the remaining equations and the coefficients of the linearly independent r functions are separately made equal to zero. The conditions so obtained yield the "angle" (i.e., x^a) dependence of the radial integration "constants".

Equation (88a) is identically satisfied by the condition (90), and initially we have to solve the coupled system

$$D\rho = \rho^2 + G_0\bar{G}_0, \tag{91a}$$

$$DG_0 = 2\rho G_0 + (G_0 - \bar{G}_0)G_0. \tag{91b}$$

This system can be solved by the matrix method which has been described in Ref. 3. We define the (2×2) matrix N by

$$\mathsf{N} \stackrel{\text{def}}{=} \begin{bmatrix} \rho & |G_0| \\ |G_0| & \bar{\rho} \end{bmatrix}. \tag{92}$$

N satisfies the equation

$$D\mathsf{N} = \mathsf{N}^2. \tag{93}$$

The solutions of this equation can be classified according to whether or not $\det[\mathsf{N}] \equiv \rho\bar{\rho} - G_0\bar{G}_0 = 0$.

For the nonshearing class and for this class only, the Newman–Penrose equations reveal an alternative way of solving the system (88b, d): Namely, from it the equations

$$D(\rho + G_0) = (\rho + G_0)^2, \tag{94a}$$

$$[D - 3(\rho + G_0)]DD\mathcal{E} = 0 \tag{94b}$$

are derived. The solution of (94) is easily found by successive integration. However, the results are to be substituted back to (91b) because (94b) has been obtained by differentiating this equation.

We adopt the weak condition of asymptotic flatness

$$\lim_{r\to\infty} \mathcal{E} = 1. \tag{95}$$

We can write the r dependence of the quantities ρ, G_0, and $\mathcal{E}$, after taking into account the above asymptotic prescription, as

$$\rho = -(r - m + i\alpha)/R^2, \tag{96}$$

$$G_0 = \frac{m}{R^2}\frac{r + i\alpha}{r - i\alpha}, \tag{97}$$

$$\mathcal{E} = 1 - 2m/(r - i\alpha), \tag{98}$$

where α and m are real integration "constants" not depending on r,

$$R^2 = r^2 - 2mr + \alpha^2, \tag{99}$$

and use has been made of the coordinate freedom (85a).

It is now obvious that the split of the solutions of Eq. (93) was unessential, and the occurrence of it must be regarded as a drawback of the first integration method in the nonshearing case.

The remaining radial equations are solved by using the standard integration methods, and yield

$$\frac{\omega}{\omega^0} = \frac{\xi^a}{\xi^{a0}} = \frac{1}{R}e^{i\Omega}, \tag{100}$$

$$\tau = \left[\frac{m^2\bar{\omega}^0}{\beta^2}\left(\frac{r - m}{R^2} + \frac{1}{2i\beta}\ln\frac{r - m - i\beta}{r - m + i\beta}\right) + \tau^0\right]\frac{1}{R}e^{-i\Omega},$$

$$\beta = (\alpha^2 - m^2)^{\frac{1}{2}}, \tag{101}$$

where the phase factor is of the form

$$e^{i\Omega} = \left(\frac{r - m - i\beta}{r - m + i\beta}\right)^{\alpha/2\beta}. \tag{102}$$

By the coordinate transformation (85b), $\xi^{20} = P$ and $\xi^{30} = iP$ can be effected and, by the rotation (86), P can be made real. Thus, the base vector $\mathbf{m}$ is completely fixed and the only coordinate freedom in V_3 is

$$z' = z'(z), \tag{103}$$

where $z = x^2 + ix^3$ and z' is an analytic function of z.[3]

From the nonradial equations we extract the following relations:

$$\delta\mathcal{E} = 0 \to m = \text{const} \quad \text{and} \quad \omega^0 = iP\nabla\alpha, \tag{104a}$$

$$(87\text{d}) \to \tau^0 = \bar{\nabla}P, \tag{104b}$$

$$(87\text{c}) \to \operatorname{Im}(P\bar{\nabla}\omega^0 - \tau^0\omega^0 + i\alpha) = 0, \tag{104c}$$

$$(88\text{f}) \to \operatorname{Re}(P\bar{\nabla}\omega^0 - \tau^0\omega^0) = 0, \tag{104d}$$

and

$$P(\nabla\tau^0 + \bar{\nabla}\bar{\tau}^0) = 1 + 2\tau^0\bar{\tau}^0. \tag{104e}$$

Here $\nabla \equiv 2(\partial/\partial\bar{z})$. Finally, Eq. (88e) is identically satisfied.

Using (104b) and (104e), we get the following equation for P:

$$P^2\nabla\bar{\nabla}\ln P^2 = 1. \tag{105}$$

From (104c) and (104d)[17]:

$$\nabla\bar{\nabla}\alpha = -P^{-2}\alpha. \tag{106}$$

In an appropriate coordinate system,[18] making use of the transformation (103), we can write the solution of Eqs. (105) and (106)

$$P = \frac{1 + z\bar{z}}{2\sqrt{2}}, \quad \alpha = a\frac{1 - z\bar{z}}{1 + z\bar{z}} \tag{107}$$

with $a = \text{const}$; ω^0 and τ^0 are obtained from (104a) and (104b):

$$\omega^0 = -\frac{\sqrt{2}\, ia}{1 + z\bar{z}}, \quad \tau^0 = \frac{\bar{z}}{\sqrt{2}}. \tag{108}$$

Transforming to the real coordinates θ ϕ by

$$z = e^{i\phi} \operatorname{ctg}(\tfrac{1}{2}\theta), \tag{109}$$

we find that $\partial/\partial\phi$ is a Killing vector, and thus we have axial symmetry. By solving Eq. (10) for ω^i and eliminating the arbitrary gradient term in ω^i by (85a), the Kerr metric[19] in the well-known form is reconstructed.

8. CONCLUDING REMARKS

There are several points of the present approach to Killing motions the clarification of which deserves more detailed investigations. First, one could mention the problem of physical interpretation of such abstractions like the eigencongruence or the basic "triad" in the background space. For a^μ timelike, it is clear that a well-established correspondence exists between properly moving observers or laboratory frames and the base triad; the existence of geodetic eigenrays means that the space–time admits a special family of "pale" (in the sense that $T_{\mu\nu} \approx 0$) light rays being excelled by the property that, when perceived in the background V_3, they propagate geodetically.

Second, to avoid unnecessary complications in the presentation, we had to adopt several simplifying assumptions, thus gradually having tightened the class of space–times considered. We do not see any direct method but solving the field equations (as shown in Sec. 7) to decide whether or not so-far unknown metrics can be obtained by dropping some of the restrictions.[20] To summarize the latter, we have assumed that the Killing field was timelike, that there was absence of matter (from Sec. 6 on), and that the space–time possessed a shear-free geodetic eigencongruence and was asymptotically flat (Sec. 7).

The investigation of all classes with geodetic eigenrays needs tedious calculations, which nevertheless do not differ in principle from the illustrative example given in Sec. 7. In the near future we plan to publish the results of these calculations.

It is not obvious whether gravitational fields with "weaker" than Killing symmetries can be involved in our considerations. Even for quite general space–times, the (3 + 1) split of the field equations combined with $SU(2)$ spinor approach might prove of use. The well-known conjecture of Israel[21] on regular event horizons is an example of problems to the research of which our approach might contribute as a useful aid.

ACKNOWLEDGMENTS

I am indebted to J. Kóta for his help in clarifying the rôle of angular coordinate freedom in the calculations of Sec. 7. Further, I would like to thank Dr. G. Györgyi and Dr. A. Sebestyén for several invaluable comments on the original manuscript.

[1] E. T. Newman and R. Penrose, J. Math. Phys. **3**, 566 (1962). This paper, on which we shall heavily rely in the following, will always be referred to as NP.

[2] E. T. Newman and W. J. Unti, J. Math. Phys. **3**, 891 (1962).

[3] E. T. Newman and L. A. Tamburino, J. Math. Phys. **3**, 902 (1962).

[4] E. T. Newman and R. Penrose, Proc. Roy. Soc. (London) **305A**, 175 (1968).

[5] C. D. Collinson and D. C. French, J. Math. Phys. **8**, 701 (1967).

[6] Z. Perjés, Commun. Math. Phys. **12**, 275 (1969).

[7] W. L. Bade and H. Jehle, Rev. Mod. Phys. **25**, 714 (1953).

[8] A. Trautman, *Lectures on General Relativity*, Brandeis Summer Institute in Theoretical Physics I (Prentice-Hall, Englewood Cliffs, N.J., 1964).

[9] J. A. Smorodinksii, Usp. Fiz. Nauk **84**, 3 (1964) [Sov. Phys. Usp. **7**, 637 (1965)].

[10] See, e.g., H. Levy, Nuovo Cimento **56B**, 253 (1958).

[11] F. J. Ernst, Preprint PORS 11T-7, 1968.

[12] A. Lichnerowicz, *Thèories relativistes de la gravitation* (Masson, Paris, 1964).

[13] L. P. Eisenhart, *Riemannian Geometry* (Princeton U.P., Princeton, N.J., 1950).

[14] Four-dimensional quantities bear Greek indices ranging over 0, 1, 2, and 3. Indices in 3-dimensional space are chosen from the Latin letters $i, j, k, \cdots$ and take the values 1, 2, 3 while $\boldsymbol{m}, \boldsymbol{n}, \boldsymbol{p}$ are projection indices with the range 0, +, and −. Spinor and dyad components are labeled by $A, B, C, \cdots$ and a, b, c, respectively, both kind of indexes taking the values 0 and 1. Where necessary, quantities defined in the 4-dimensional space–time V_4 are distinguished by wavy line $\sim$. Covariant differentiation in V_4 is denoted by a stroke, in three dimensions by the symbol ∂ or simply by semicolon; a comma stands for partial differentiation.

[15] A method for obtaining Eqs. (81) is to insert the decomposition (60) of the spin coefficients into the identities (4.2b), (4.2c), (4.2f), (4.2i), and (4.2j) of NP.

[16] Following the custom of NP, we use the zero superscript to denote quantities not depending on r.

[17] It is a surprising fact that Eq. (106) is obtained in quantum mechanics when the angular momentum problem is considered with the eigenvalue $l = 1$.

[18] I. Robinson and A. Trautman, Proc. Roy. Soc. (London) **265A**, 463 (1961).

[19] R. Kerr, Phys. Rev. Letters **11**, 237 (1963).

[20] However, as was shown in Sec. 6, the fields with shear-free geodetic eigenrays are conveniently treated by the Newman–Penrose equations too, and thus we must not hope for essentially new results for this class.

[21] According to Israel's guess, all asymptotically flat vacuum fields with a regular event horizon are algebraically special [see W. Israel, Commun. Math. Phys. **8**, 245 (1968)]. This situation would call forth important consequences for the gravitational collapse.

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 11, NUMBER 12 DECEMBER 1970

Inverse Problem for a Cylindrical Plasma*

ANTONIO DEGASPERIS
Physics Department, Indiana University, Bloomington, Indiana 47401

(Received 7 April 1970)

We study the Maxwell equation for the electric potential inside a neutral nonuniform cylindrical plasma in an external oscillating electric field, from the point of view of the inverse problem. The logarithmic derivatives of the radial Fourier components of the electric potential at the edge of the cylinder are considered as experimental data. We obtain an explicit and exact representation of the electron density in terms of the high frequency behavior of the experimental data. The general analyticity properties in the complex-frequency plane are also discussed.

1. INTRODUCTION

This paper is devoted to the study of a cylindrically symmetric nonuniform plasma from the point of view of the inverse problem. In its general philosophy, the problem is to obtain information about the plasma, namely its local electron density, from certain suitable experimental data. This approach to the relationship between experiment and theoretical model has already been developed in other fields of physics, e.g., the diffusion process of sonic and classical electromagnetic waves and the quantum scattering of elementary particles by a potential. The starting point of this investigation turns out to be, in fact, the mathematical analogy between the stationary radial Schrödinger equation of the theory of scattering by a spherically symmetric potential and the Maxwell equation for the electric potential existing inside a very long (infinite) cylinder of plasma. We shall see, however, that this apparent analogy does not mean a straightforward application of the mathematical techniques used in scattering theory. Our interest is confined to the very simple model which is characterized by the following assumptions: (a) the plasma is cold and (b) neutral, (c) the positions of the ions are fixed, (d) the deviations from the equilibrium position are small so that the continuity equation together with the equation of motion can be linearized, and (e) the plasma magnetic field can be neglected.

The mathematical aspects of this model have been extensively discussed by Barston[1] for a free plasma, where the interest has been focused on the singularities of the radial equation satisfied by the electric potential. These singular points are the zeros of the dielectric constant, and they have been shown to be related to resonant absorption processes when a cylindrical collisionless plasma is driven by an external oscillating electrostatic potential.[2]

We consider now the use of this model in plasma daignostic. Although many techniques have been developed to measure the electron density profile, it seems to us to be worthwhile to investigate the possibility of determining the density profile from completely external measurements, that is to say, from measurements which do not perturb the plasma itself at all.

This point of view has been recently adopted in the case of a plasma slab in the cold approximation[3] where the inverse problem has been solved by taking as experimental data the *ac* resistance of the slab as a function of the frequency.

In our case, due to the infinite length of the cylinder, the problem is defined in a plane perpendicular to the axis of the cylinder where we use polar coordinates, namely the radius r as the distance from the center of the cylinder and the polar angle θ. The experimental information we need to construct the electron density $n(r)$ as a function defined in the interval $0 \leq r \leq R$, where R is the radius of the cylinder, is obtained by placing the plasma in a uniform (z-independent) oscillating external electric field and detecting the angular dependence of the electric potential and of its radial derivative at the edge of the cylinder.[4]

As we show in Appendix A, we are interested in that part of the electric potential that oscillates with the frequency ω of the external electric field. This part can be written

$$\Phi(r, \theta, t) = \exp(i\omega t) \sum_{k=-\infty}^{+\infty} \bar{\Phi}_k \Psi_k^{(1)}(\omega; r) \exp(ik\theta), \quad (1.1)$$

where the constants $\bar{\Phi}_k$ depend on the boundary conditions and $\Psi_k^{(1)}(\omega; r)$ satisfies the radial equation (A18). For future convenience, we consider the following function:

$$y_k(\omega; r) = r^{\frac{1}{2}} \Psi_k^{(1)}(\omega; r), \quad (1.2)$$

which satisfies the equation

$$\frac{d^2}{dr^2} y_k(\omega; r) + V(\omega; r) \frac{d}{dr} y_k(\omega; r) - \left(\frac{k^2 - \frac{1}{4}}{r^2} + \frac{1}{2r} V(\omega; r) \right) y_k(\omega; r) = 0, \quad (1.3)$$

where

$$V(\omega; r) = \frac{1}{\Omega^2(r) - \omega(\omega - i\nu_c)} \frac{d}{dr} \Omega^2(r). \quad (1.4)$$

Equation (1.3) can be derived from Eq. (A18). Any solution of Eq. (1.3) either diverges at the origin as $r^{\frac{1}{2}-|k|}$ or vanishes there as $r^{\frac{1}{2}+|k|}$, $k = \pm 1, \pm 2, \cdots$[5]; however, the boundedness of the electric potential requires the vanishing at the origin of the physical solution $y_k(\omega; r)$. In our notation the solution $\varphi_k(\omega; r) = \varphi_{-k}(\omega; r)$ is understood to be determined by the boundary condition

$$\lim_{r \to 0} r^{-(\frac{1}{2}+|k|)} \varphi_k(\omega; r) = 1, \quad k = 0, \pm 1, \pm 2, \cdots. \quad (1.5)$$

For any physical value of ω, we now define the infinite set of functions

$$\sigma_k(\omega) = \left(\int_0^{2\pi} d\theta \exp(ik\theta) \frac{\partial}{\partial r} \Phi(R, \theta, t) \right) \times \left(\int_0^{2\pi} d\theta \exp(ik\theta) \Phi(R, \theta, t) \right)^{-1} = \frac{d}{dr} \ln \varphi_k(\omega; r)|_{r=R} - \frac{1}{2R} \quad (1.6)$$

as the experimental data to be used to construct the electron density $n(r)$. We note that this definition does not depend on the normalization of the regular solution of Eq. (1.3) so that the choice of the solution $\varphi_k(\omega; r)$ is only a matter of convenience.

In analogy with the quantum scattering theory, starting from the knowledge of the function $\sigma_k(\omega)$ of two variables, the discrete variable[6] $k = 1, 2, 3, \cdots$, and the real positive continuous variable ω, we can consider two different approaches to the inverse problem, namely one that uses the frequency dependence of the function $\sigma_k(\omega)$ at one fixed value of k and the other which uses the sequence of numbers $\{\sigma_k(\omega)\}_{k=1}^{\infty}$ at one fixed value of the frequency ω. The aim of this investigation is to find simple and exact relationships between the electron density $n(r)$ and the function $\sigma_k(\omega)$ rather than to discuss the uniqueness and existence of the solution of both the inverse problems. In Sec. 2 we discuss the analytical structure of the function $\sigma_k(\omega)$ in the complex ω plane, and in Sec. 3 we develop an asymptotic expansion method which allows one to reconstruct the electron density function from the high frequency behavior of the function $\sigma_k(\omega)$, for all possible value of k.

In Sec. 3 we also briefly consider the possibility of applying these results to phenomenology. For completeness we give a short derivation of the radial equation in the Appendix.

2. THE COMPLEX FREQUENCY PLANE

In this section we give some general analyticity properties of the solution $\varphi_k(\omega; r)$ and of the function $\sigma_k(\omega)$, considered as functions of the complex variable ω. First we introduce the following simplifying notation:

$$r = Rt, \quad \Omega_0 = \Omega(0), \quad \Omega^2(Rt) = \Omega_0^2 F(t), \quad \omega(\omega - i\nu_c)\Omega_0^{-2} = z. \quad (2.1)$$

After simple substitutions in Eqs. (1.3) and (1.5), it turns out that the function

$$\Phi_k(z; t) = R^{-\frac{1}{2}-k} \varphi_k(\omega; r), \quad k = 1, 2, 3, \cdots, \quad (2.2)$$

is the solution of the equation

$$\Phi_k''(z; t) + \frac{F'(t)}{F(t) - z} \Phi_k'(z; t) - \left(\frac{k^2 - \frac{1}{4}}{t^2} + \frac{1}{2t} \frac{F'(t)}{F(t) - z} \right) \Phi_k(z; t) = 0, \quad ' \equiv \frac{d}{dt}, \quad (2.3)$$

satisfying the z-independent boundary condition

$$\lim_{t \to 0} t^{-(\frac{1}{2}+k)} \Phi_k(z; t) = 1. \quad (2.4)$$

The function $F(t)$, defined for $t \in I$, where I is the closed interval $[0, 1]$, is what we want to find in some way from the knowledge of the function

$$\eta_k(z) = \frac{d}{dt} \ln \Phi_k(z; t)|_{t=1} - \tfrac{1}{2}, \quad k = 1, 2, \cdots; \quad (2.5)$$

physical arguments can be used to restrict the class of the allowable functions $F(t)$ to those satisfying the conditions

$$F(1) = 0, \quad F'(t) \leq 0, \quad (2.6)$$

together with those conditions already included in the definition (2.1),

$$F(0) = 1, \quad F(t) \geq 0 \quad \text{for} \quad t \in I. \quad (2.7)$$

The function $F(t)$ is then a nonnegative, nonincreasing function such that, for $t \in I$, $F(t) \in I$. This means that for any $z \notin I$, Eq. (2.3) has only one singular but regular point in the interval I, namely $t = 0$. From there we start to integrate the equation with the regular solution boundary condition, obtaining a well-defined function $\Phi_k(z; t)$ in the interval I for every value of $z \notin I$. Furthermore, a well-known theorem states that, if a differential equation depends on a parameter through a function which is holomorphic in a domain of the complex plane of that parameter,

then a solution of that equation, whose boundary conditions are independent of that parameter, is a holomorphic function of that parameter in the same domain. In our case, this theorem says that $\Phi_k(z;t)$ is a holomorphic function of z in the whole complex plane excluding the interval I.

In order to understand what kind of singularities the function $\Phi_k(z;t)$ has for $z \in I$, we note that to each $z \in I$ there corresponds only one point[7] t_z such that $F(t_z) = z$; therefore, t_z is a singular but regular point for Eq. (2.3), and from the general theory[8] we know that in a neighborhood of t_z two linearly independent solutions exist, one of which is analytic in $t = t_z$ while the other has a logarithmic branch point at $t = t_z$.[9] Without going into the details of a proof, it is intuitive that this branch point at $t = t_z$, in the t dependence, generates for fixed t a logarithmic branch point in the z plane at $z = F(t_z)$. Since the function $\Phi_k(z;t)$ is obtained by starting the integration of Eq. (2.3) from $t = 0$, it follows from the above considerations that the function $\Phi_k(z;t)$ has a cut in the z plane determined by the condition $0 \leq t_z \leq t$, that is, for $F(t) \leq z \leq 1$.

We are now in the position to discuss the analyticity properties of the function $\eta_k(z)$, defined by Eq. (2.5), in the complex z plane. In fact, the analyticity properties of the solution $\Phi_k(z;t)$ imply that the function $\eta_k(z)$ has a cut in the complex z plane for $z \in I$ (the discontinuity being singular at $z = 0$ as αz^{-1}). Excluding this cut, the function $\eta_k(z)$ is meromorphic, its poles being due to the zeros z_n of the function $\Phi_k(z;1)$. Since, for each value of $z \notin I$, $\Phi_k'(z;1)$ is analytic and since there is no value of $z \notin I$ which can be a zero of both the function $\Phi_k(z;1)$ and of its derivative $\Phi_k'(z;1)$ [where $\Phi_k(z;t)$ is a solution of a second-order differential equation which is regular at $t = 1$], there is a one-to-one correspondence between the poles of the function $\eta_k(z)$ and the zeros of the function $\Phi_k(z;1)$. Furthermore, if $\Phi_k(z;t)$ is the regular solution of Eq. (2.3), then the function $\Phi_k^*(z^*;t)$ satisfies the same equation with the same boundary condition (2.4), so that we have, because of the uniqueness of this solution,

$$\Phi_k(z^*;t) = \Phi_k^*(z;t), \quad z \notin I. \tag{2.8}$$

This last equation implies that the poles z_n of the function $\eta_k(z)$ occur as pairs of complex conjugate values and, more generally, that $\eta_k(z)$ is a real function

$$\eta_k^*(z) = \eta_k(z^*), \quad z \notin I. \tag{2.9}$$

We now prove that, in fact, the function $\eta_k(z)$ has no poles, that is to say, that it is holomorphic for $z \notin I$. This result is obtained by showing that the function $\Phi_k(z;1)$ cannot have zeros when $z \notin I$. To establish this result, we show that the converse leads to a contradiction. Assume that a value $\bar{z} \notin I$ exists such that

$$\Phi_k(\bar{z};1) = 0. \tag{2.10}$$

This equation, together with definitions (2.1), (2.2), and (1.2), implies that

$$\Psi_k^{(1)}(\bar{\omega};R) = 0, \tag{2.11}$$

where $\bar{\omega}$ is such that

$$f(\bar{\omega};r) \neq 0, \quad \text{for} \quad 0 < r \leq R, \tag{2.12}$$

and the function $f(\omega;r)$ is defined in the Appendix [see formula (A19)]. We now use the differential equation (A18), satisfied by $\Psi_k^{(1)}(\omega;r)$, to obtain the equality

$$\frac{d}{dr}\left(f(\omega;r)\Psi_k^{(1)*}(\omega;r)\frac{d}{dr}\Psi_k^{(1)}(\omega;r)\right) = f(\omega;r)\left(\left|\frac{d}{dr}\Psi_k^{(1)}(\omega;r)\right|^2 + \frac{k^2}{r^2}|\Psi_k^{(1)}(\omega;r)|^2\right). \tag{2.13}$$

Integrating the two terms of this equality from $r = 0$ to $r = R$ and using the condition (2.11), we obtain

$$\int_0^R f(\bar{\omega};r)\left(\left|\frac{d}{dr}\Psi_k^{(1)}(\bar{\omega};r)\right|^2 + \frac{k^2}{r^2}|\Psi_k^{(1)}(\bar{\omega};r)|^2\right)dr = 0, \tag{2.14}$$

which, together with Eq. (2.12), requires $\Psi_k^{(1)}(\bar{\omega};r) \equiv 0$ for $0 < r \leq R$. This shows that the existence of the zero [(2.10)] implies

$$\Phi_k(\bar{z};t) \equiv 0 \quad \text{for} \quad t \in I, \tag{2.15}$$

contradicting the condition (2.4) which defines the solution $\Phi_k(\bar{z};t)$. By excluding the cut $z \in I$, the function $\eta_k(z)$ is then holomorphic; in particular, the point $z = \infty$ is a regular point. In the next section we will use this property of the function $\eta_k(z)$ to study its asymptotic behavior for large z and its relationships with the function $F(t)$.

3. THE HIGH-FREQUENCY EXPANSION

In Sec. 2 we showed that the function $\eta_k(z)$ is regular when z goes to infinity, which means that the asymptotic expansion

$$\eta_k(z) = \sum_{m=0}^{\infty} \eta_k^{(m)} z^{-m} \tag{3.1}$$

holds for large $|z|$ or, more precisely, for $|z| > 1$, if we consider the smallest circle containing the interval I. However, it is worthwhile to give a proof of this asymptotic behavior in which the only requirements

on the function $F(t)$ are that it, together with its first and second derivatives, be finite for $t \in I$.

Using standard methods, we consider the integral equation satisfied by our regular solution $\Phi_k(z;t)$. We find that the function

$$\Phi_k(z;t) = t^{k+\frac{1}{2}}\left(\frac{F(0)-z}{F(t)-z}\right)^{\frac{1}{2}}\chi_k(z;t) \tag{3.2}$$

is the solution of Eq. (2.3) satisfying the condition (2.4) if the function $\chi_k(z;t)$ is the solution of the following integral equation:

$$\chi_k(z;t) = 1 + \frac{1}{8k}\int_0^t\left[1-\left(\frac{x}{t}\right)^{2k}\right]\left(\frac{2[xF'(x)]'}{F(x)-z} - \frac{x[F'(x)]^2}{[F(x)-z]^2}\right)\chi_k(z;x)\,dx, \quad k = 1, 2, \cdots. \tag{3.3}$$

This last equation can be solved by iteration, provided that $|z|$ is sufficiently large. This means that the iterative series converges and defines the solution $\chi_k(z;t)$ for all $t \in I$ and $|z|$ greater than a certain finite value. This shows that the point $z = \infty$ is a regular point for the function $\chi_k(z;t)$ and therefore for the function $\Phi_k(z;t)$, as implied by the formula (3.2). Furthermore, the coefficient of the power z^{-n} of the asymptotic expansion of the function $\Phi_k(z;t)$ is obtained after n iterations, as is evident from the integral equation (3.3). This last feature is important because it leads to the possibility of getting explicit and exact relationships between the high frequency behavior of the experimental data (1.6) and the "plasma frequency" (A9).

In order to get these results, it is convenient to introduce the function

$$\eta_k(z;t) = t\frac{d}{dt}[\ln\Phi_k(z;t)] - \tfrac{1}{2} = t\frac{\Phi_k'(z;t)}{\Phi_k(z;t)} - \frac{1}{2}. \tag{3.4}$$

Equations (2.3) and (2.4) imply that the function $\eta_k(z;t)$ is the solution of the following first-order nonlinear differential equation:

$$\eta_k'(z;t) + \frac{1}{t}[\eta_k(z;t)]^2 + \frac{F'(t)}{F(t)-z}\eta_k(z;t) - \frac{k^2}{t} = 0, \tag{3.5}$$

which satisfies the boundary condition

$$\eta_k(z;0) = k, \quad k = 1, 2, \cdots. \tag{3.6}$$

On the other hand, from the definitions (2.5) and (3.4) we have

$$\eta_k(z) = \eta_k(z;1) \tag{3.7}$$

so that the experimental data can be obtained by integrating directly Eq. (3.5), if the "plasma frequency" (A9) is known.

Let us consider now the asymptotic expansion of the function $\eta_k(z;t)$. It is easy to show, for example, from the zeroth approximation $\chi_k(z;t) = 1$ of Eq. (3.3) in the formulas (3.2) and (2.5), that

$$\lim_{|z|\to\infty}\eta_k(z;t) = k, \quad k = 1, 2, \cdots, \tag{3.8}$$

holds independent of t and the function $F(t)$. By substituting the following asymptotic expansion,

$$\eta_k(z;t) = k + \sum_{m=1}^{\infty}\eta_k^{(m)}(t)z^{-m}, \tag{3.9}$$

into Eq. (3.5), we get the infinite set of coupled equations

$$[\eta_k^{(1)}(t)]' + \frac{2k}{t}\eta_k^{(1)}(t) - kF'(t) = 0, \tag{3.10}$$

$$[\eta_k^{(m)}(t)]' + \frac{2k}{t}\eta_k^{(m)}(t) + \frac{1}{t}\sum_{l=1}^{m-1}\eta_k^{(m-l)}(t)\eta_k^{(l)}(t) - F'(t)\sum_{l=1}^{m-1}[F(t)]^{l-1}\eta_k^{(m-l)}(t) - kF'(t)[F(t)]^{m-1} = 0, \quad m > 1. \tag{3.11}$$

Using the boundary conditions implied by Eqs. (3.6) and (3.9),

$$\eta_k^{(m)}(0) = 0, \quad m = 1, 2, \cdots, \tag{3.12}$$

we obtain, by integrating Eqs. (3.10) and (3.11), the following recursion relations:

$$\eta_k^{(1)}(t) = kt^{-2k}\int_0^t F'(x)x^{2k}\,dx = kF(t) - 2k^2t^{-2k}\int_0^t F(x)x^{2k-1}\,dx, \tag{3.13}$$

$$\eta_k^{(m)}(t) = t^{-2k}\int_0^t\left(kF'(x)[F(x)]^{m-1} + F'(x)\sum_{l=1}^{m-1}[F(x)]^{l-1}\eta_k^{(m-l)}(x) - \frac{1}{x}\sum_{l=1}^{m-1}\eta_k^{(m-l)}(x)\eta_k^{(l)}(x)\right)x^{2k}\,dx, \quad m > 1, \tag{3.14}$$

which could be solved recursively for any m.

In order to obtain explicit relationships between the high frequency behavior of the experimental data $\eta_k(z)$ and the unknown function $F(x)$, we have just to set $t = 1$ in Eqs. (3.13) and (3.14), remembering that $F(1) = 0$ [see (2.6)]. We then obtain an infinite set of integral relations between the function $F(x)$

and the doubly indexed quantities

$$\eta_k^{(m)} \equiv \eta_k^{(m)}(1), \quad k = 1, 2, \cdots, \quad m = 1, 2, \cdots, \tag{3.15}$$

that have to be considered as given by the experiment. As an example, we give the expressions, for $m = 1$ and $m = 2$,

$$\eta_k^{(1)} = -2k^2 \int_0^1 F(x) x^{2k-1}\, dx, \quad k = 1, 2, \cdots, \infty, \tag{3.16}$$

$$\eta_k^{(2)} = -k^2 \int_0^1 F^2(x) x^{2k-1}\, dx + 2k^3 \left(\int_0^1 F(x) x^{2k-1}\, dx \right)^2, \quad k = 1, 2, \cdots, \infty. \tag{3.17}$$

Let us consider now the mathematical implications of such expressions and their possible application to phenomenology. First, we note that, for the purpose of reconstructing the function $F(t)$ for $t \in I$, the knowledge of the quantities (3.15), for all the possible values of k and m, is redundant. This redundancy reflects, in a certain sense, the existence of the two different versions of the inverse problem mentioned in Sec. 1. Namely, the fixed-k and all-frequencies version corresponds to the use of the asymptotic coefficients (3.15) at one fixed value of k and for $m = 1, 2, \cdots$, which determine the Taylor expansion of the function $\eta_k(z)$ in a neighborhood of the point $z = \infty$. On the other hand, the knowledge of one asymptotic coefficient at a fixed value of m for all values of k can be seen as corresponding to the fixed-frequency and all-k version of the inverse problem.

To prove this redundancy in the coefficients $\eta_k^{(m)}$, we limit ourselves to showing that the function $F(t)$ related to the coefficients $\eta_k^{(1)}$, $k = 1, 2, \cdots, \infty$, through Eqs. (3.16), is unique. In fact, the existence of two different functions, $F_1(t)$ and $F_2(t)$, corresponding to the same coefficients $\eta_k^{(1)}$, would imply, for the difference $\Delta F(t) = F_2(t) - F_1(t)$, the vanishing of all these integrals:

$$\int_0^1 \Delta F(t) t^{2k-1}\, dt = \tfrac{1}{2} \int_0^1 \Delta F(\sqrt{x}) x^{k-1}\, dx = 0, \quad k = 1, 2, \cdots, \infty. \tag{3.18}$$

If $\{P_n(x)\}$ is the set of orthonormal polynomials in the interval I, n being the order of the polynomial

$$\int_0^1 P_n(x) P_m(x)\, dx = \delta_{mn}, \quad n, m = 0, 1, 2, \cdots, \infty, \tag{3.19}$$

then the expressions (3.18) imply the vanishing of all the Fourier coefficients:

$$\int_0^1 P_n(x) \Delta F(\sqrt{x})\, dx = 0, \quad n = 0, 1, 2, \cdots, \infty. \tag{3.20}$$

It remains now to use the theorem which states that a square-integrable function in the interval I is identically zero if all its Fourier coefficients corresponding to the polynomials $P_n(x)$ vanish. Then our assertion follows from the fact that our function $\Delta F(\sqrt{x})$ is obviously square integrable. We now go back to the physical quantities defined in Sec. 1 and let v_c go to zero for simplicity. Comparing the definitions (2.2), (2.1), and (2.5) with Eq. (1.6), we get the relation

$$\sigma_k(\omega) = R^{-1} \eta_k(\omega^2/\Omega_0^2), \quad k = 1, 2, \cdots, \tag{3.21}$$

which, together with the asymptotic expansion

$$\sigma_k(\omega) \xrightarrow[\omega \to \infty]{} \frac{k}{R} + \sigma_k^{(1)} \frac{1}{\omega^2} + \sigma_k^{(2)} \frac{1}{\omega^4} + O\left(\frac{1}{\omega^6}\right), \tag{3.22}$$

implies

$$\sigma_k^{(1)} = \frac{\Omega_0^2}{R} \eta_k^{(1)}, \quad \sigma_k^{(2)} = \frac{\Omega_0^4}{R} \eta_k^{(2)}, \quad k = 1, 2, \cdots, \infty. \tag{3.23}$$

We now use the results (3.16) and (3.17), together with the definitions (2.1) and (A9), to get all the odd moments of the electron density function and of its square, in terms of the asymptotic coefficients $\sigma_k^{(1)}$ and $\sigma_k^{(2)}$:

$$\int_0^1 n(Rx) x^{2k-1}\, dx = \frac{-\alpha}{2k^2} \sigma_k^{(1)}, \quad k = 1, 2, \cdots, \infty, \tag{3.24}$$

$$\int_0^1 [n(Rx)]^2 x^{2k-1}\, dx = \frac{\alpha^2}{2k^3} (\sigma_k^{(1)})^2 - \frac{\alpha^2}{Rk^2} \sigma_k^{(2)}, \quad k = 1, 2, \cdots, \infty, \tag{3.25}$$

where

$$\alpha = mR/4\pi e^2. \tag{3.26}$$

From our previous considerations, we know that the sequence of numbers $\sigma_k^{(1)}$ determines completely the electron density $n(r)$. However, in order to give explicitly the function $n(r)$ in terms of the coefficients $\sigma_k^{(1)}$, we have to make some further assumption on the function $n(r)$. In fact, since the odd moments of a function $f(x)$ can be considered as all the moments of the function $f(\sqrt{x})$, one could use the formal completeness of the polynomials $P_n(x)$ to write down an expansion in terms of these polynomials. The coefficients of this expansion would be obtained easily from the known moments. However, even if the function $f(x)$ is analytic in a domain containing the interval

(0, 1), the expansion in polynomials of the function $f(\sqrt{x})$ fails to converge generally for $x = 0$ and $x = 1$, because of the presence of the square root of x. Then it follows that, with the assumption on the density

$$n(Rx) = f(x^2), \quad x \in I, \tag{3.27}$$

where $f(t)$ is a function whose singularities have a nonvanishing distance from the interval I, it is possible to have a convergent expansion of the density function. To obtain this expansion, we introduce the explicit expression[10]

$$P_n(x) = \sum_{l=0}^{n} (2n+1)^{\frac{1}{2}}(-)^{n+l}\binom{l+n}{n}\binom{n}{l}x^l \tag{3.28}$$

so that the Fourier coefficients of the function (3.27) are

$$\int_0^1 P_n(x)f(x)\,dx = -\alpha(2n+1)^{\frac{1}{2}}(-)^n \sum_{l=0}^{n} \frac{(-)^l}{(l+1)^2}\binom{n+l}{n}\binom{n}{l}\sigma_{l+1}^{(1)}, \tag{3.29}$$

as implied by Eqs. (3.27), (3.24), and (2.38). Finally, we obtain the following representation of the density function:

$$n(r) = \alpha \sum_{n=0}^{\infty}(-)^{n+1}(2n+1)^{\frac{1}{2}} \times \sum_{l=0}^{n} \frac{(-)^l}{(l+1)^2}\binom{n+l}{n}\binom{n}{l}\sigma_{l+1}^{(1)}P_n\left(\frac{r^2}{R^2}\right). \tag{3.30}$$

From our previous considerations on the different versions of the inverse problem, it follows that the expression (3.30) is an explicit solution of the fixed-frequency and all-k version of the problem at the very special value of the frequency $\omega = \infty$. A different expression can evidentally be obtained by starting with the moments of the square of the density given by Eq. (3.25). This other representation of the function $n(r)$ could be used to find the relationships between the asymptotic coefficients $\sigma_k^{(1)}$ and $\sigma_k^{(2)}$, due to the expressions (3.24) and (3.25). For example, one of these relationships can be obtained by setting $k = 1$ in Eq. (3.25) and using the Parseval equality

$$\sum_{n=1}^{\infty}(2n+1)\left|\sum_{l=0}^{n}\frac{(-)^l}{(l+1)^2}\binom{n+l}{n}\binom{n}{l}\sigma_{l+1}^{(1)}\right|^2 = -\frac{2}{R}\sigma_1^{(2)}. \tag{3.31}$$

We note that the main virtue of these results is that they explicitly display the relationships between the electron density and the high frequency behavior of the experimental data $\sigma_k(\omega)$. However, the inverse problem for fixed and finite frequency and all values of k is still an open problem. The discussion of the other version of the inverse problem, namely for fixed k, requires a more complete knowledge of the analyticity properties of the function $\eta_k(z)$ in the complex z plane than that sketched in Sec. 2. However, from the results of Sec. 2, we can say that the function $\eta_k(z)$ is known once its discontinuity on the real interval [0, 1] is known.

Starting from the experimental knowledge of the function $\eta_k(z)$ on the real positive axis, one can obtain its discontinuity, so that the main problem is that of obtaining the function $F(t)$ from the singularities of the function $\eta_k(z)$ in the z plane. We note that this problem is completely different from the inverse problem (all energies and fixed angular momentum) in potential scattering theory.

We make now a few remarks on the use of our results from the point of view of an experimental check and application. Although the applicability of these results requires the knowledge of the very high frequency behavior of the function $\sigma_k(\omega)$, we want to emphasize the fact that the function $\sigma_k(\omega)$ has actually a very honest behavior for large ω, so that an extrapolation procedure could be used to get the asymptotic coefficients. This remark is important since a limitation on the high-frequency expansion technique results from the quasistatic assumption (e), made in Sec. 1. According to that assumption, the frequency ω of the electric potential should be such that

$$\omega < 2\pi c/R. \tag{3.32}$$

This inequality implies that, in order to use a frequency ω that is in the holomorphic circle $\omega^2/\Omega_0^2 > 1$ of the $\omega = \infty$ point, the plasma frequency itself must satisfy the inequality

$$\max \Omega(r) = \Omega_0 < 2\pi c/R. \tag{3.33}$$

In ordinary experiments[2] c/R is of the order of magnitude of 10^{10} sec^{-1}, while a dilute plasma has a density of about 10^{12} particles per cubic meter, which means a plasma frequency of the order of magnitude of 10^7 rad $\times$ sec^{-1}, so that the inequality (3.33) holds. For such a plasma[11] a finite range of frequency ω exists between $\Omega(0)$ and $2\pi c/R$ such that the mathematical model under discussion is still valid and the expansion (3.1) is convergent. Therefore, an experimental knowledge of the functions $\sigma_k(\omega)$ in that region of frequency leads, by extrapolating to the vanishing value of the variable $\tau = \omega^{-2}$, to the determination of the derivatives of $\sigma_k(\omega)$ with respect to τ at the origin. The first derivative of the functions $\sigma_k(\omega)$ with respect to τ for vanishing τ is just what we

need to write down the representation (3.30) of the density profile.

Finally, we note that the very simple general conditions can be read, from Eqs. (3.24) and (3.25),

$$\sigma_k^{(1)} < 0, \quad \sigma_k^{(2)} < (R/2k)(\sigma_k^{(1)})^2, \quad k = 1, 2, \cdots, \infty. \tag{3.34}$$

Furthermore, the expressions (3.24) and (3.25) can be used to adjust the parameters entering in the density function. With this aim in mind, we give now a useful formula obtained in the case that the electron density is expressed as a superposition of Gaussian functions:

$$n(r) = \sum_{l=1}^{N} A_l \exp(-\mu_l r^2), \tag{3.35}$$

where the $2N$ real parameters A_l and μ_l must only satisfy the relation

$$\sum_{l=1}^{N} A_l \exp(-\mu_l R^2) = 0. \tag{3.36}$$

Using Eq. (3.24), we obtain an expression which gives the asymptotic coefficients $\sigma_k^{(1)}$ in terms of the parameters describing the density function (3.33):

$$\begin{aligned}\sigma_k^{(1)} = \frac{k^2}{\alpha R^2}\Bigg(&\sum_{l=1}^{N} \frac{A_l}{\mu_l} \exp(-\mu_l R^2) \\ &- (k+1)! \sum_{l=1}^{N} R^{2(1-k)} \mu_l^{-k} A_l \\ &+ (k-1)! \sum_{n=0}^{k-2} [(k-n-2)!]^{-1} \\ &\times \sum_{l=1}^{N} R^{-2(n-1)} \mu_l^{-n-2} A_l \exp(-R^2\mu_l)\Bigg).\end{aligned} \tag{3.37}$$

A similar expression can be obtained for $\sigma_k^{(2)}$ and generally for the asymptotic coefficients $\sigma_k^{(m)}$.

We close this section with two final remarks. The first one is about the limit of collisionless plasma, $\nu_c \to 0$, which we took in stating Eq. (3.21). We note that, because of the definitions (2.1), (3.1), and (3.23), the representation (3.30) of the density profile is independent of the limit $\nu_c \to 0$. In other words, the presence of a nonvanishing collision frequency ν_c affects the coefficients of the powers ω^{-n}, in the asymptotic expansion of the function $\sigma_k(\omega)$, only for $n \geq 3$. The second point is about the presence of a uniform axial magnetic field. In this case the radial equation for our function $\Psi_k^{(1)}(\omega; r)$ (see Appendix) is Eq. (6) of Ref. 2, and it is easy to check, by applying our asymptotic expansion method to that equation, that the representation (3.30) remains unchanged since again the strength of the magnetic field appears in the asymptotic coefficients only from the power ω^{-3} on.

ACKNOWLEDGMENTS

The author wishes to thank Professor R. G. Newton for bringing this problem to his attention and Professor A. N. Kaufman for many clarifying discussions about the physical aspects of the problem.

APPENDIX

In this Appendix we derive[12] the equation for the electric potential using the assumptions stated in Sec. 1.

We start from the continuity equation

$$\nabla \cdot [\rho_e(\mathbf{r}, t)\mathbf{V}(\mathbf{r}, t)] + \frac{\partial}{\partial t}\rho_e(\mathbf{r}, t) = 0, \tag{A1}$$

the Poisson's equation

$$\nabla^2\Phi(\mathbf{r}, t) = 4\pi e[\rho_e(\mathbf{r}, t) - n(r)], \tag{A2}$$

and the equation of motion

$$\left((\mathbf{V}(\mathbf{r}, t) \cdot \nabla)\mathbf{V}(\mathbf{r}, t) + \frac{\partial}{\partial t}\mathbf{V}(\mathbf{r}, t)\right) = \frac{e}{m}\nabla \cdot \Phi(\mathbf{r}, t), \tag{A3}$$

where the electrons are treated as a charged fluid whose velocity field and density are respectively $\mathbf{V}(\mathbf{r}, t)$ and $\rho_e(\mathbf{r}, t)$. The ion density $n(r)$ is assumed to be a time-independent function of the radius r, such that $n(r) = 0$ for $r \geq R$. The electric field is assumed to be irrotational [see assumption (e) in Sec. 1] and is given by $-\nabla\Phi(\mathbf{r}, t)$. Finally, m and e are respectively the mass and the modulus of the charge of the electron.

Introducing the collision frequency constant ν_c in the equation of motion, we read the linearized equations as

$$\nabla \cdot [n(r)\mathbf{V}(\mathbf{r}, t)] + \frac{\partial}{\partial t}\rho(\mathbf{r}, t) = 0, \tag{A4}$$

$$\nabla^2\Phi(\mathbf{r}, t) = 4\pi e\rho(\mathbf{r}, t), \tag{A5}$$

$$\left(\nu_c\mathbf{V}(\mathbf{r}, t) + \frac{\partial}{\partial t}\mathbf{V}(\mathbf{r}, t)\right) = \frac{e}{m}\nabla \cdot \Phi(\mathbf{r}, t), \tag{A6}$$

where we have defined

$$\rho(\mathbf{r}, t) = \rho_e(\mathbf{r}, t) - n(r). \tag{A7}$$

A partial differential equation for the electric potential is then obtained by differentiating the Poisson's equation (A5) twice with respect to t and eliminating the density and the velocity field via Eqs. (A4) and (A6). The resulting equation is

$$\nabla \cdot \left(\nabla \cdot \frac{\partial^2}{\partial t^2} + \Omega^2(r)\nabla + \nu_c\nabla \cdot \frac{\partial}{\partial t}\right)\Phi(\mathbf{r}, t) = 0, \tag{A8}$$

where we have introduced the plasma frequency function

$$\Omega^2(r) = 4\pi(e^2/m)n(r). \tag{A9}$$

Assuming that the initial conditions, as well as the boundary conditions, are z independent, we see that the θ independence of the function (A9) and the linearity of Eq. (A8) imply the following radial equation:

$$\frac{\partial}{\partial r}\left[r\left(\frac{\partial^3}{\partial r\partial t^2}+\Omega^2(r)\frac{\partial}{\partial r}+\nu_c\frac{\partial^2}{\partial r\partial t}\right)\right]\Phi_k(r,t) - \frac{k^2}{r}\left(\frac{\partial^2}{\partial t^2}+\Omega^2(r)+\nu_c\frac{\partial}{\partial t}\right)\Phi_k(r,t)=0, \quad \text{(A10)}$$

where the completeness of the functions $\exp(ik\theta)$, $k=0, \pm1, \pm2, \cdots$, has been used to expand the solution of Eq. (A8):

$$\Phi(\mathbf{r},t)=\sum_{k=-\infty}^{+\infty}\Phi_k(r,t)\exp(ik\theta). \quad \text{(A11)}$$

We now give an example of external electrical driver to discuss the boundary conditions. The source of the electric field is a cylindrical metal pipe, whose radius is $L>R$, coaxial to the plasma, which is cut axially into two halves of angular widths equal to 2α and to $2(\pi-\alpha)$ with $0\leq\alpha<\frac{1}{2}\pi$. In the following we will neglect the gap between the two halves as well as the corresponding fringing fields. The oscillating electric potential on the metal surface is fixed to be

$$\bar{\Phi}(\theta,t)=\exp(i\omega t)\Phi_+, \quad \text{for} \quad \theta\in(\alpha,\pi-\alpha),$$
$$\bar{\Phi}(\theta,t)=\exp(i\omega t)\Phi_-, \quad \text{for} \quad \theta\notin(\alpha,\pi-\alpha), \quad \text{(A12)}$$

where Φ_+ and Φ_- are constant and ω is real.

The boundary conditions for the solution of Eq. (A8) are

$$\Phi(L,\theta,t)=\bar{\Phi}(\theta,t), \quad \text{(A13)}$$

which implies the following condition for the solution of the radial equation (A10),

$$\Phi_k(L,t)=\bar{\Phi}_k\exp(i\omega t), \quad \text{(A14)}$$

where the constants $\bar{\Phi}_k$ are defined by the expansion of the function (A12):

$$\bar{\Phi}(\theta,t)=\exp(i\omega t)\sum_{k=-\infty}^{+\infty}\bar{\Phi}_k\exp(ik\theta). \quad \text{(A15)}$$

They are found to be

$$\bar{\Phi}_0=\tfrac{1}{2}[\Phi_++\Phi_-]-(\alpha/\pi)\Delta, \quad \Delta=\Phi_+-\Phi_-, \quad \text{(A16a)}$$
$$\bar{\Phi}_k=-(\Delta/\pi k)\sin k\alpha \quad \text{if } k \text{ is even},$$
$$\bar{\Phi}_k=(\Delta/i\pi k)\cos k\alpha \quad \text{if } k \text{ is odd}. \quad \text{(A16b)}$$

To find the general solution of Eq. (A10) satisfying the initial conditions and the boundary conditions (A14), we first note that the function

$$\Phi_k^{(1)}(r,t)=\bar{\Phi}_k\exp(i\omega t)\Psi_k^{(1)}(\omega;r) \quad \text{(A17)}$$

is a solution of Eq. (A10), satisfying condition (A14), if $\Psi_k^{(1)}(\omega;r)$ is the solution of the following equation:

$$\frac{d}{dr}\left(f(\omega,r)\frac{d}{dr}\Psi_k^{(1)}(\omega;r)\right)-\frac{k^2}{r^2}f(\omega,r)\Psi_k^{(1)}(\omega;r)=0, \quad \text{(A18)}$$

$$f(\omega,r)=r[\Omega^2(r)-\omega^2+i\omega\nu_c], \quad \text{(A19)}$$

satisfying the boundary condition

$$\Psi_k^{(1)}(\omega;L)=1, \quad \text{(A20)}$$

together with the physical condition[13] at the origin

$$\Psi_k^{(1)}(\omega;0)=0, \quad \text{for} \quad k\neq0, \quad \Psi_0^{(1)}(\omega;0)=1. \quad \text{(A21)}$$

Since Eq. (A18) is regular for $0<r\leq L$, the solution $\Psi_k^{(1)}(\omega;r)$ is well defined and unique.

It remains to show that solutions $\Phi_k^{(0)}(r,t)$ of Eq. (A10) exist such that

$$\Phi_k^{(0)}(L,t)=0. \quad \text{(A22)}$$

We can construct the general solution of our problem by simply adding these solutions to the function (A17). This spectrum problem can be solved following the standard techniques used in Ref. 1, and it is easily proved that the function

$$\Phi_k^{(0)}(r,t) = \exp(-\tfrac{1}{2}\nu_c t)\int_{-\omega_0}^{\omega_0}C_k(\mu)\Psi_k^{(0)}(r,\mu)\exp(i\mu t)\,d\mu, \quad \text{(A23)}$$

where $C_k(\mu)$ is an arbitrary function and

$$\omega_0=[\Omega^2(0)-\tfrac{1}{4}\nu_c^2]^{\frac{1}{2}}, \quad \text{(A24)}$$

is a general solution of Eq. (A10) satisfying (A22). This follows from the fact that the functions $\Psi_k^{(0)}(r,\mu)$ satisfy the equation

$$\frac{d}{dr}\left(g(\mu,r)\frac{d}{dr}\Psi_k^{(0)}(r,\mu)\right)-\frac{k^2}{r^2}g(\mu,r)\Psi_k^{(0)}(r,\mu)=0, \quad \text{(A25)}$$

$$g(\mu,r)=r[\Omega^2(r)-\mu^2-\tfrac{1}{4}\nu_c^2], \quad \text{(A26)}$$

with the boundary conditions

$$\lim_{r\to0}r^{-|k|}\Psi_k^{(0)}(r,\mu)=1, \quad \Psi_k^{(0)}(L,\mu)=0, \quad \text{(A27)}$$

and are uniquely defined. By choosing now the function $C_k(\mu)$ in order to satisfy the initial condition, we write the solution of our problem as

$$\Phi_k(r,t)=\exp(-\tfrac{1}{2}\nu_c t)\int_{-\omega_0}^{\omega_0}C_k(\mu)\Psi_k^{(0)}(r,\mu)\exp(i\mu t)\,d\mu + \bar{\Phi}_k\exp(i\omega t)\Psi_k^{(1)}(\omega;r). \quad \text{(A28)}$$

In this paper we are interested in the steady part of the solution (A28), that is to say, in the function $\Psi_k^{(1)}(\omega; r)$.

We end this appendix by noting that the expression (A28) explicitly shows that the steady part is absent for those values of k which correspond to the case $\overline{\Phi}_k = 0$.

* This work has been partially supported by the U.S. Army Research Office, Durham, North Carolina.

[1] E. M. Barston, Ann. Phys. (N.Y.) **29**, 282 (1964).

[2] D. E. Baldwin and D. W. Ignat, Phys. Fluids **12**, 697 (1969).

[3] D. E. Baldwin and A. N. Kaufman, Phys. Fluids **12**, 1526 (1969).

[4] The choice of the edge of the plasma cylinder is simply due to mathematical simplicity and does not lead to any restriction on the experimental procedure, since our results can be extended to the case in which the measurements take place far from the plasma by matching the solutions of the Poisson's equation and taking into account the dielectric constant of the quartz container or of any other medium.

[5] For $\omega(\omega - i\nu_c) = \Omega^2(0)$, $V(\omega; r)$ has a simple pole at the origin so that the regular and irregular solutions behave there respectively as r^α and r^β where $\alpha = \frac{1}{2}\{[(n+1)^2 + 4k^2]^{\frac{1}{2}} - n\}$ and $\beta = -\frac{1}{2}\{[(n+1)^2 + 4k^2]^{\frac{1}{2}} + n\}$ and n is a nonnegative integer given by the conditions

$$\frac{d^m}{dr^m}\Omega(r)|_{r=0} = 0, \quad \text{for} \quad m = 1, 2, \cdots, n,$$

and

$$\frac{d^{n+1}}{dr^{n+1}}\Omega(r)|_{r=0} \neq 0.$$

This special case has no effect on the considerations which follow.

[6] We note that, for $k = 0$, $\varphi_0(\omega; r) = \sqrt{r}$ so that $\sigma_0(\omega) = 0$.

[7] For the sake of simplicity, we do not consider functions $F(t)$ which are constant in some subinterval of the interval I, since this case does not differ from the one that we discuss, from the point of view of the analyticity properties in the z plane.

[8] See, for example, E. T. Whittaker and C. N. Watson, *A Course of Modern Analysis* (Cambridge U.P., Cambridge, 1965), Chap. X.

[9] We note that the indicial equation corresponding to the regular point $t = t_z$ is $x^2 + nx = 0$ if

$$\frac{d^m}{dt^m}F(t)|_{t=t_z} = 0, \quad \text{for} \quad m = 1, 2, \cdots, n,$$

and

$$\frac{d^{n+1}}{dt^{n+1}}F(t)|_{t=t_z} \neq 0.$$

[10] These polynomials are obtained from the well-known Legendre polynomials by mapping the interval $[-1, +1]$ into the interval $[0, 1]$.

[11] Our plasma must satisfy also the zero-temperature assumption. This means that the Debye radius ρ_D has to be much smaller than the radius R of the cylinder. In this case, since we are considering large values of ω, the cold approximation $R/\rho_D > \Omega_0/\omega$ is well satisfied.

[12] We are following the approach given in Ref. 1.

[13] We note that Eqs. (A18) and (A20) imply $\Psi_0^{(1)}(\omega; r) = 1$.

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 11, NUMBER 12 DECEMBER 1970

Newman–Penrose Constants and Their Invariant Transformations*

EDWARD N. GLASS† AND JOSHUA N. GOLDBERG
Syracuse University, Syracuse, New York

(Received 19 June 1970)

The origin and significance of the Newman–Penrose (N–P) constants of the motion are examined from the point of view that constants of the motion generate invariant transformations. Here the calculation makes use of a generalization of Green's theorem to a situation applicable to the coupled Einstein–Maxwell fields in general relativity. One finds strictly electromagnetic constants generated by an incoming electromagnetic shock wave, with dipole symmetry, at future null infinity. The gravitational constants contain an admixture from the electromagnetic field. They are generated by an incident quadrupole gravitational shock wave at future null infinity ($\mathfrak{I}^+$). Both the electromagnetic dipole field and the gravitational quadrupole field behave like linearized fields at $\mathfrak{I}^+$. All higher-multipole fields do not uncouple from the nonlinear corrections induced by the self-coupling of the gravitational field and its coupling with the electromagnetic field. It is shown that the gravitational constants are related to the rate of shear of the null rays near $\mathfrak{I}^+$ a result which suggests a connection with sources. Finally, the supertranslation invariance of the result is discussed.

1. INTRODUCTION

Radiation of gravitational and electromagnetic energy can take place in an asymptotically flat space–time.[1–3] It was therefore believed that, in the combined Einstein–Maxwell theory, only the total charge is absolutely conserved. However, in 1965 Newman and Penrose[4] announced the discovery of ten constants of the motion in the nonlinear theory of gravitation for an asymptotically flat space–time. With Exton,[5] they later found that in Einstein–Maxwell space there are six Maxwell constants and ten gravitational constants. These quantities are absolutely conserved; they retain their values in the presence of outgoing gravitational and electromagnetic radiation. From the method of their construction, neither the origin of these constants nor their possible physical significance was clear. Therefore, a program to investigate these questions was undertaken.[6–8]

In this paper we are interested in the steady part of the solution (A28), that is to say, in the function $\Psi_k^{(1)}(\omega; r)$.

We end this appendix by noting that the expression (A28) explicitly shows that the steady part is absent for those values of k which correspond to the case $\overline{\Phi}_k = 0$.

* This work has been partially supported by the U.S. Army Research Office, Durham, North Carolina.

[1] E. M. Barston, Ann. Phys. (N.Y.) **29**, 282 (1964).

[2] D. E. Baldwin and D. W. Ignat, Phys. Fluids **12**, 697 (1969).

[3] D. E. Baldwin and A. N. Kaufman, Phys. Fluids **12**, 1526 (1969).

[4] The choice of the edge of the plasma cylinder is simply due to mathematical simplicity and does not lead to any restriction on the experimental procedure, since our results can be extended to the case in which the measurements take place far from the plasma by matching the solutions of the Poisson's equation and taking into account the dielectric constant of the quartz container or of any other medium.

[5] For $\omega(\omega - i\nu_c) = \Omega^2(0)$, $V(\omega; r)$ has a simple pole at the origin so that the regular and irregular solutions behave there respectively as r^α and r^β where $\alpha = \frac{1}{2}\{[(n+1)^2 + 4k^2]^{\frac{1}{2}} - n\}$ and $\beta = -\frac{1}{2}\{[(n+1)^2 + 4k^2]^{\frac{1}{2}} + n\}$ and n is a nonnegative integer given by the conditions

$$\frac{d^m}{dr^m}\Omega(r)|_{r=0} = 0, \quad \text{for} \quad m = 1, 2, \cdots, n,$$

and

$$\frac{d^{n+1}}{dr^{n+1}}\Omega(r)|_{r=0} \neq 0.$$

This special case has no effect on the considerations which follow.

[6] We note that, for $k = 0$, $\varphi_0(\omega; r) = \sqrt{r}$ so that $\sigma_0(\omega) = 0$.

[7] For the sake of simplicity, we do not consider functions $F(t)$ which are constant in some subinterval of the interval I, since this case does not differ from the one that we discuss, from the point of view of the analyticity properties in the z plane.

[8] See, for example, E. T. Whittaker and C. N. Watson, *A Course of Modern Analysis* (Cambridge U.P., Cambridge, 1965), Chap. X.

[9] We note that the indicial equation corresponding to the regular point $t = t_z$ is $x^2 + nx = 0$ if

$$\frac{d^m}{dt^m}F(t)|_{t=t_z} = 0, \quad \text{for} \quad m = 1, 2, \cdots, n,$$

and

$$\frac{d^{n+1}}{dt^{n+1}}F(t)|_{t=t_z} \neq 0.$$

[10] These polynomials are obtained from the well-known Legendre polynomials by mapping the interval $[-1, +1]$ into the interval $[0, 1]$.

[11] Our plasma must satisfy also the zero-temperature assumption. This means that the Debye radius ρ_D has to be much smaller than the radius R of the cylinder. In this case, since we are considering large values of ω, the cold approximation $R/\rho_D > \Omega_0/\omega$ is well satisfied.

[12] We are following the approach given in Ref. 1.

[13] We note that Eqs. (A18) and (A20) imply $\Psi_0^{(1)}(\omega; r) = 1$.

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 11, NUMBER 12 DECEMBER 1970

Newman–Penrose Constants and Their Invariant Transformations*

EDWARD N. GLASS† AND JOSHUA N. GOLDBERG
Syracuse University, Syracuse, New York

(Received 19 June 1970)

The origin and significance of the Newman–Penrose (N–P) constants of the motion are examined from the point of view that constants of the motion generate invariant transformations. Here the calculation makes use of a generalization of Green's theorem to a situation applicable to the coupled Einstein–Maxwell fields in general relativity. One finds strictly electromagnetic constants generated by an incoming electromagnetic shock wave, with dipole symmetry, at future null infinity. The gravitational constants contain an admixture from the electromagnetic field. They are generated by an incident quadrupole gravitational shock wave at future null infinity ($\mathfrak{I}^+$). Both the electromagnetic dipole field and the gravitational quadrupole field behave like linearized fields at $\mathfrak{I}^+$. All higher-multipole fields do not uncouple from the nonlinear corrections induced by the self-coupling of the gravitational field and its coupling with the electromagnetic field. It is shown that the gravitational constants are related to the rate of shear of the null rays near $\mathfrak{I}^+$ a result which suggests a connection with sources. Finally, the supertranslation invariance of the result is discussed.

1. INTRODUCTION

Radiation of gravitational and electromagnetic energy can take place in an asymptotically flat space–time.[1–3] It was therefore believed that, in the combined Einstein–Maxwell theory, only the total charge is absolutely conserved. However, in 1965 Newman and Penrose[4] announced the discovery of ten constants of the motion in the nonlinear theory of gravitation for an asymptotically flat space–time. With Exton,[5] they later found that in Einstein–Maxwell space there are six Maxwell constants and ten gravitational constants. These quantities are absolutely conserved; they retain their values in the presence of outgoing gravitational and electromagnetic radiation. From the method of their construction, neither the origin of these constants nor their possible physical significance was clear. Therefore, a program to investigate these questions was undertaken.[6–8]

Newman and Penrose[9] have shown that these constants (N–P constants) inhibit an initially stationary gravitational configuration from emitting gravitational radiation for a period and then returning to an arbitrary stationary configuration once again. This result can be understood by considering the axially symmetric stationary solutions. There are only two gravitational constants in this case, and they have the structure

$$G = MQ - D^2,$$

where M is the mass and Q and D are the quadrupole and dipole moments, respectively. After a radiative phase all of these quantities change, but G remains constant.

Since constants of the motion generate invariant transformations, the method of study undertaken has been to look for the transformations generated by the N–P constants. An initial examination revealed that they do not generate any type of coordinate transformation and hence are unlike energy and momentum. For the electromagnetic field in Minkowski space, Goldberg found that the constants generate a zero change in the field variables at all finite points of space. This result was to be expected as the constants themselves are given by 2-dimensional surface integrals at null infinity.

Further study of linear field equations in Minkowski space[7] showed that these constants are related to the linear superposition of solutions. An invariant mapping of solutions onto solutions is obtained by adding a specific solution of the homogeneous field equations to all solutions. The particular additional solution generated by the N–P constants is an incoming shock wave in the limit that the wavefront moves out to future null infinity. Since the shock wave may have an arbitrary shape, there exist an infinite number of N–P constants for linear fields.

The important question is how these results carry over to the nonlinear theory of gravitation. There are two different approaches:

(a) D. C. Robinson has used the conformal technique developed by Penrose which allowed him to calculate "at infinity." Robinson extended the results of Goldberg to curved space and to mappings of solutions of the field equations themselves.[8] Using generalized Hertz potentials (superpotentials) for zero mass fields, Robinson proved that the Maxwell N–P constants generate an incoming dipole pulse with support at future null infinity and the gravitational constants generate a quadrupole pulse with similar support and that these pulses are solutions of the field equations "at infinity."

This approach does not develop as much physical understanding as one would like, since the calculations are all performed in the unphysical conformal space. Furthermore, in a curved space–time superpotentials exist everywhere only for the electromagnetic field; for other fields superpotentials may exist at infinity in asymptotically flat space–times. Finally, the gauge adopted by Robinson for the generalized Hertz potentials is not conformally invariant, and so it is extremely difficult to map his calculation back to the physical space–time in order to develop a physical interpretation.

This critique leads to the second approach:

(b) The present work obtains the invariant transformations generated by the N–P constants by proceeding in the physical space–time. By identifying Noether's equation with a generalization of Green's identity, the invariant transformations are constructed without using the superpotentials. Thus we can use the null gauge in electromagnetism and the Bondi–Sachs coordinate conditions in gravitation. These conditions are conformally invariant so that, if we chose, the entire calculation could be taken over to the conformal space, where future null infinity is a well-defined 3-surface.

The result is that, in the absence of an electromagnetic field at null infinity, the constants are obtained directly from the shear of the outgoing null rays, thus providing a direct link between the multipole structure of sources, which gives rise to part of the shear, and the N–P constants. A connection is thus also obtained with the result of Exton, who has proven that the gravitational constants vanish for algebraically special fields.[10]

In Sec. 2 a metric and associated null tetrad are introduced. The Einstein and Maxwell field equations are to be treated within the Newman–Penrose formalism and the necessary spin coefficient restrictions are stated. Section 3 extends the Goldberg–Newman generalization of Green's identity to any number of interacting fields whose equations of motion (linear or nonlinear) are derivable from an action principle. The relation between constants of the motion and invariant transformations is discussed, and a connection is drawn between Noether's theorem and Green's identity. The method of constructing the N–P constants from Green's identity is then presented. In Sec. 4 the N–P constants are constructed for the Maxwell and Einstein fields in an asymptotically flat Einstein–Maxwell space. The pure gravitational constants are related to the geometrical behavior of null rays and in turn to the multipole structure of the

sources. Supertranslation invariance is proven in Sec. 5, and the nonexistence of higher-order constants for curved space is discussed in Sec. 6.

2. FORMALISM AND CONVENTIONS

The Bondi–Sachs metric[1,2] defines the asymptotically flat space-time which is the arena for our investigation of the N–P constants:

$$ds^2 = \frac{Ve^{2b}}{r}\,du^2 + 2e^{2b}\,du\,dr - r^2h_{ij}(dx^i - U^i\,du)(dx^j - U^j\,du), \quad (2.1)$$

where a null coordinate system has been constructed in the usual manner[1,2,11] with outgoing null hypersurfaces labeled by $x^0 = u = \text{const}$. The rays of these hypersurfaces (lines with tangent vector $u_{,\nu}$) are null geodesics which are parametrized by the luminosity distance $x^1 = r$. Coordinates $x^2 = \theta$ and $x^3 = \phi$ are constant along each ray. The luminosity distance is defined by

$$r^4 \sin^2\theta = g_{22}g_{33} - (g_{23})^2, \quad (2.2)$$

and h_{ij} is chosen[12] to satisfy (2.2):

$$h_{ij} = \begin{pmatrix} e^{2\nu}\cosh(2q) & \sinh(2q)\sin\theta \\ \sinh(2q)\sin\theta & e^{-2\nu}\cosh(2q)\sin^2\theta \end{pmatrix},$$

with the result that

$$\sqrt{-g} = e^{2b}r^2\sin\theta. \quad (2.3)$$

Working in the Newman–Penrose formalism,[13] we construct a tetrad from the Bondi–Sachs metric which satisfies the completeness relation

$$g_{\mu\nu} = 2l_{(\mu}n_{\nu)} - 2m_{(\mu}\bar{m}_{\nu)}$$

and orthogonality properties

$$l_\mu n^\mu = -m_\mu\bar{m}^\mu = 1,$$

with all other contractions vanishing. From (2.1) the tetrad vectors are chosen to be

$$l_\mu = \delta^0_\mu, \qquad l^\mu = e^{-2b}\delta^\mu_1, \quad (2.4)$$

$$n_\mu = e^{2b}\delta^1_\mu + \frac{Ve^{2b}}{2r}\delta^0_\mu, \quad n^\mu = \delta^\mu_0 - \frac{V}{2r}\delta^\mu_1 + U^j\delta^\mu_j, \quad (2.5)$$

$$m_\mu = \xi_i\delta^i_\mu - U^i\xi_i\delta^0_\mu, \quad m^\mu = \xi^i\delta^\mu_i, \quad (2.6)$$

where

$$\xi_i = \frac{-r}{2^{\frac{1}{2}}(1+i)} \times [(e^{-q} + ie^{q})e^{\nu}\delta^2_i + i\sin\theta(e^{q} + ie^{-q})e^{-\nu}\delta^3_i],$$

$$\xi^i = \frac{r^{-1}}{2^{\frac{1}{2}}(1+i)} \times \left((e^{q} + ie^{-q})e^{-\nu}\delta^i_2 + \frac{i}{\sin\theta}(e^{-q} + ie^{q})e^{\nu}\delta^i_3\right)$$

satisfy

$$\xi^i\xi_i = 0, \quad \xi_i\bar{\xi}^i = -1,$$

and

$$\xi_i\bar{\xi}_j + \bar{\xi}_i\xi_j = r^2h_{ij}.$$

This tetrad is different from those of Newman and Penrose[11] and Sachs.[2] It is chosen so that m^μ and $\bar{m}^\mu$ are surface forming (they lie in 2-surfaces of constant u and r), and resembles a tetrad used by Hawking[14] except that he uses an affine parameter and we use a luminosity distance.

The spin coefficients (see Appendix B) are restricted by the relations

$$\kappa = \epsilon = 0, \quad \tau = \bar{\alpha} + \beta,$$
$$\mu = \bar{\mu}, \quad \rho = \bar{\rho}, \quad (2.7)$$

where the propagation of m^μ along l^μ is chosen to be[15] $m^\mu{}_{;\nu}l^\nu = \bar{\pi}l^\mu$.

The technique of Newman and Unti[16] is used to integrate the Newman–Penrose field equations, and the asymptotic solutions together with the definitions of the field variables are collected in Appendix B. Intrinsic tetrad derivatives are defined as

$$D := l^\mu\nabla_\mu, \quad \Delta := n^\mu\nabla_\mu, \quad \delta := m^\mu\nabla_\mu.$$

3. GREEN'S IDENTITY AND INVARIANT TRANSFORMATIONS

Goldberg and Newman[17] have constructed a generalization of Green's identity which is applicable to nonlinear differential equations derivable from a variational principle. The generalization says that if a system of field equations is derivable from a variational principle whose Lagrangian density $L(y_A, y_{A,\rho})$ is homogeneous of degree n_1 in the field variables y_A and of degree n_2 in their first derivatives $y_{A,\rho}$, then the following identity holds:

$$y_AM^A(y, z) - (n_1 + n_2 - 1)L^Az_A \equiv t^\rho{}_{,\rho}, \quad (3.1)$$

where $z_A := \delta y_A$ is[18] an arbitrary variation of the field variables and $M^A(y, z)$ is the corresponding first variation of the field equations[19]

$$L^A := \partial^AL - (\partial^{A\rho}L)_{,\rho}. \quad (3.2)$$

Actually, the proof given in Ref. 17 treats only the case $n_2 = 2$. Furthermore, the Langranglan density may be composed additively of parts with different homogeneity properties as in the case of the Einstein–Maxwell field. The generalization of the theorem to include these properties is easily carried out via the methods of Streudel.[20,21] In the following proof we shall restrict the generality to the case of two interacting fields, which is equivalent to the Einstein–Maxwell field. The method of proof can clearly be extended to any number of additive parts.

Consider a Lagrangian density which consists of two parts with field variables g_a and y_A:

$$L(g_a, y_A, g_{a,\mu}, y_{A,\mu}) = L_0(g_a, g_{a,\mu}) + L_1(y_A, y_{A,\mu}, g_a, g_{a,\mu}),$$

where $a = 1 \cdots n$, $A = 1 \cdots N$, and $\mu = 0, 1, 2, 3$ over the coordinate labels. We assume that L_0 is homogeneous of degree n_1 in g_a and n_2 in $g_{a,\mu}$; L_1 is homogeneous of degree n_3 in y_A, n_4 in $y_{A,\mu}$, n_5 in g_a, and n_6 in $g_{a,\mu}$. Using Euler's theorem for homogeneous functions, the definition for the field equations given in (3.2), and $L^a = L_0^a + L_1^a$, $L^A = L_1^A$, we get the identities

$$g_a L^a + (g_a \partial^{a\mu} L)_{,\mu} \equiv (n_1 + n_2)L_0 + (n_5 + n_6)L_1, \tag{3.3a}$$

$$y_A L^A + (y_A \partial^{A\mu} L)_{,\mu} \equiv (n_3 + n_4)L_1. \tag{3.3b}$$

Now form the variation of Eqs. (3.3) and combine the resulting expressions to give ($m := n_1 + n_2 - n_5 - n_6$)

$$(1 - n_1 - n_2)\tilde{\delta} g_a L^a + g_a \tilde{\delta} L^a - \frac{m(n_3 + n_4 - 1) + (n_3 + n_4)(n_5 + n_6)}{(n_3 + n_4)} \tilde{\delta} y_A L_1^A + \frac{m}{n_3 + n_4} y_A \tilde{\delta} L_1^A \equiv t^\rho{}_{,\rho}, \tag{3.4}$$

$$t^\rho := -\tilde{\delta}(g_a \partial^{a\rho} L) + (n_1 + n_2)(\tilde{\delta} g_a \partial^{a\rho} L + \tilde{\delta} y_A \partial^{A\rho} L_1) - \frac{m}{n_3 + n_4} \tilde{\delta}(y_A \partial^{A\rho} L_1). \tag{3.5}$$

Equation (3.4) is the identity we have been working toward.

For the specific case of the Einstein–Maxwell field,

$$L_0 = \tfrac{1}{8}(2\hat{g}^{\rho\sigma}\hat{g}_{\lambda\alpha}\hat{g}_{\beta\tau} - \hat{g}^{\rho\sigma}\hat{g}_{\alpha\beta}\hat{g}_{\lambda\tau} - 4\delta^\sigma_\beta \delta^\rho_\lambda \hat{g}_{\alpha\tau})\hat{g}^{\alpha\beta}{}_{,\rho}\hat{g}^{\lambda\tau}{}_{,\sigma}, \tag{3.6}$$

where

$$\hat{g}^{\alpha\beta} := (-g)^{\frac{1}{2}} g^{\alpha\beta}, \quad \hat{g}^{\alpha\beta}\hat{g}_{\beta\gamma} = \delta^\alpha_\gamma,$$
$$L_1 = -\tfrac{1}{4}(-g)^{\frac{1}{2}}(g^{\mu\rho}g^{\nu\sigma} - g^{\nu\rho}g^{\mu\sigma})A_{\mu,\nu}A_{\rho,\sigma}. \tag{3.7}$$

For the field variables $\hat{g}^{\mu\nu}$ and A_μ, the homogeneity conditions are $n_1 = -1$, $n_2 = 2$, $n_3 = 0$, $n_4 = 2$, $n_5 = n_6 = 0$, and $m = 1$. Equation (3.4) now becomes

$$-\hat{g}^{\mu\nu}\tilde{\delta}(R_{\mu\nu} + T_{\mu\nu}) + \tfrac{1}{2}\{\tilde{\delta}A_\mu(-g)^{\frac{1}{2}}F^{\mu\nu}{}_{;\nu} - A_\mu \tilde{\delta}[(-g)^{\frac{1}{2}}F^{\mu\nu}{}_{;\nu}]\} \equiv t^\rho{}_{,\rho}. \tag{3.8}$$

Introducing the notation $z_{\mu\nu} := \tilde{\delta} g_{\mu\nu}$, we can write t^ρ as

$$t^\rho = (-g)^{\frac{1}{2}}(g^{\alpha\sigma}g^{\beta\rho} - g^{\alpha\beta}g^{\sigma\rho})z_{\alpha\beta;\sigma} + \tfrac{1}{2}\{\tilde{\delta}A_\mu(-g)^{\frac{1}{2}}F^{\mu\rho} - A_\mu\tilde{\delta}[(-g)^{\frac{1}{2}}F^{\mu\rho}]\}. \tag{3.9}$$

To obtain (3.9), we have used

$$\tilde{\delta}R_{\mu\nu} = (\tilde{\delta}\Gamma^\alpha_{\mu\alpha})_{;\nu} - (\tilde{\delta}\Gamma^\rho_{\mu\nu})_{;\rho}$$

and

$$\tilde{\delta}\Gamma^\nu_{\alpha\beta} = \tfrac{1}{2}g^{\mu\nu}(z_{\mu\alpha;\beta} + z_{\mu\beta;\alpha} - z_{\alpha\beta;\mu}).$$

For convenience, in the following discussion relating Noether's theorem to the generalized Green identity, we shall use y_A as a generic symbol including g_a and will understand that Eq. (3.1) would have to be replaced by (3.4) for the more general situation in which we are actually interested.

Now, if a set of field equations

$$L^A = 0 \tag{3.10}$$

is derivable from a variational principle with Lagrangian density $L(y_A, y_{A,\mu})$, then, for an invariant transformation $y_A \to y_A + \tilde{\delta}y_A$, Noether's theorem says that the following identity holds[22,23]:

$$\tilde{\delta}y_A L^A \equiv t^\rho{}_{,\rho}. \tag{3.11}$$

That is, when the field equations (3.10) are satisfied, we have the conservation law

$$t^\rho{}_{,\rho} = 0. \tag{3.12}$$

The converse of this statement is also true: If a relationship of the form (3.11) holds, then $\tilde{\delta}y_A$ is an invariant transformation.

Equation (3.1), the generalization of Green's identity obtained above, arrives at the Noether equation whenever $\tilde{\delta}y_A =: z_A$ is a solution of the first variation of the field equations, i.e., whenever $\tilde{\delta}L^A =: M^A = 0$. Therefore such z_A define infinitesimal invariant transformations. We shall show that the N–P constants generate a particular class of such transformations.

The generator of a given invariant transformation is that constant of the motion defined by the corresponding conservation law (3.12). To obtain the generator, one integrates the conservation law over an appropriate 4-dimensional region R_4 and defines the generator as an integral over a segment of the boundary of the region. For our purposes it is convenient to take R_4 (see Fig. 1) as that region bounded by two outgoing null surfaces $\mathcal{N}_1$ and $\mathcal{N}_2$ and two surfaces Σ_1 and Σ_2 which may be timelike or null. Σ_2 is to be taken eventually to future null infinity. It is convenient to take Σ_2 to a surface $v = u + 2r = \text{const}$, which in fact becomes null in the limit of future null infinity. The surface Σ_1 in principle need only bound R_4 away from any sources. We take it to be an $r = \text{const}$ surface.

From Stokes' theorem, an integral over the conservation law (3.12) gives an integral over the

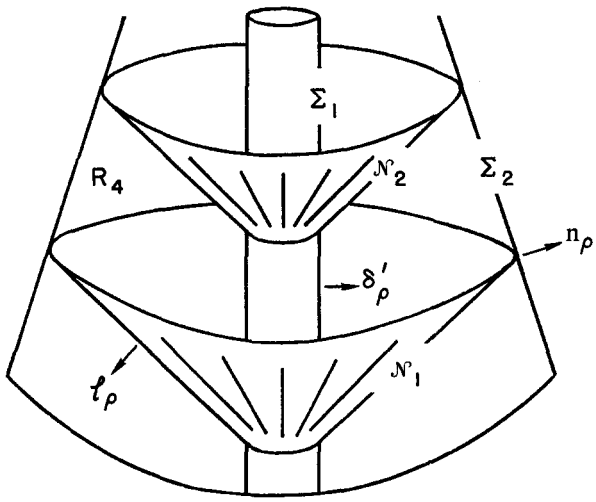


FIG. 1. Region of integration.

boundary of R_4, which we write

$$\int_{\mathcal{N}_2} t^\rho l_\rho \, d\tau_{(3)} - \int_{\mathcal{N}_1} t^\rho l_\rho \, d\tau_{(3)} = -\int_{\Sigma_2} t^\rho \, dS_\rho + \int_{\Sigma_1} t^\rho \, dS_\rho. \quad (3.13)$$

If the integrals over Σ_1 and Σ_2 can be shown to vanish, then

$$G[z] := \int_{\mathcal{N}} t^\rho l_\rho \, d\tau_{(3)} \quad (3.14)$$

is a constant of the motion.

In flat space a solution of the wave equation can be found which is sharply confined between two null surfaces. One can choose the surfaces to be incoming null cones between Σ_1 and Σ_2, and, indeed, we can choose the solution to have a delta-function shape (see Fig. 2). Clearly, by choosing $z_A \not\equiv 0$ only on an incoming cone so defined, we make the right-hand side of (3.13) vanish and thus define constants of the motion. If the support for the incoming shock wave is taken in the limit of future null infinity, the constants so defined, when they exist, are the N–P constants. Of course, to carry out this limit, the Σ_2 boundary must itself be taken out to future null infinity.

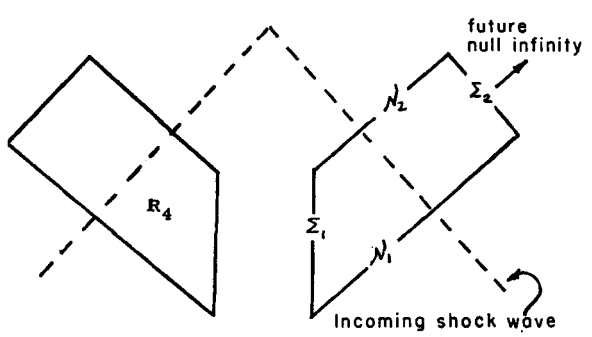


FIG. 2. Support of the incoming wave.

Therefore, for the remainder of the paper, the region of integration will be defined with Σ_2 at future null infinity.

In the following sections, we show that to a limited extent something similar can be done in asymptotically flat space–time. In particular, since a discontinuity in the electromagnetic or gravitational field can appear across a null surface, one can always choose z_A to be zero on the inside of the surface of discontinuity. In this case, the contribution from the Σ_1 *flux integral* will always vanish. The problem then is to investigate the falloff of the field on the Σ_2 side of the surface of discontinuity. If the falloff is sufficiently fast, the integral on the left-hand side of (3.13) may have a finite limit when the shock front moves out to future null infinity, while the Σ_2 integral goes to zero. We shall see that this is indeed the case for a dipole shock wave for the Maxwell field and for a quadrupole shock wave for the gravitational field.

4. N–P CONSTANTS IN EINSTEIN-MAXWELL SPACE

One can generate transformations by initiating the variation of the degrees of freedom of the electromagnetic and gravitational fields separately. When this is done, certain combinations of the field quantities will remain constant. That the electromagnetic constants are different from the gravitational ones seems natural when we recall that the degrees of freedom of the two fields are independent. This is made particularly clear by remembering that the lowest order of electromagnetic radiation is dipole in character, whereas for the gravitational case it is quadrupole. Therefore, one might expect that constants associated with an electromagnetic dipole field would not couple strongly with the gravitational field, whereas the gravitational quadrupole field might couple strongly with the electromagnetic field. As a result, the N–P Maxwell constants would not be altered by the presence of the gravitational field, while the gravitational constants would be modified by the presence of an electromagnetic field. We shall see that this is indeed the case.

The Maxwell and Einstein fields are determined by the coupled nonlinear field equations

$$F^{(-)\mu\nu}{}_{;\nu} = 0,$$
$$R_{\mu\nu} = -2F^{(-)}{}_{\mu}{}^{\rho} F^{(+)}{}_{\rho\nu}, \quad (4.1)$$

together with asymptotic conditions which restrict the system to asymptotically flat space–time, and where[24]

$$F^{(\mp)\mu\nu} := \tfrac{1}{2}(F^{\mu\nu} \pm i\eta^{\mu\nu\rho\sigma} F_{\rho\sigma}).$$

The asymptotic solutions of Eqs. (4.1) are collected in Appendix B. Here we shall also need the physical

components of the vector potential A^μ:

$$A^\mu =: A_0 l^\mu + A_1 n^\mu + A_2 m^\mu + A_3 \bar{m}^\mu, \quad (4.2)$$

where A_0 and A_1 are real and $A_2 = \bar{A}_3$. The most convenient gauge in which to proceed is a null or "Bondi" type gauge:

$$A^\mu l_\mu = A_1 = 0 \quad (4.3)$$

(a conformally invariant choice). From (B24), (B25), and (B26), we obtain the following asymptotic forms[25]:

$$\begin{aligned} A_0 &= A_0^0 - \frac{1}{r}(\phi_1^0 + \bar{\phi}_1^0) + O_2, \\ A_3 &= \frac{A_3^0}{r} + \frac{1}{r^2}(\phi_0^0 - \sigma^0 \bar{A}_3^0) \\ &\quad + \frac{1}{2r^3}(\phi_0^1 - \sigma^0 \bar{\phi}_0^0 + \sigma^0 \bar{\sigma}^0 A_3^0) + O_4, \\ A_2 &= \bar{A}_3, \end{aligned} \quad (4.4)$$

where

$$A_0^0 = -(\bar{\eth}^{-1}\phi_2^0 + \eth^{-1}\bar{\phi}_2^0)$$

and

$$\bar{\eth} A_3^0 = \phi_1^0.$$

From Eq. (3.9) of the previous section, we have the conserved vector density

$$\begin{aligned} t^\rho = {} & 2(-g)^{\frac{1}{2}} g^{\alpha[\sigma} g^{\beta]\rho} z_{\alpha\beta;\sigma} - \tfrac{1}{2} A_\mu F^{\mu\rho} \delta(-g)^{\frac{1}{2}} \\ & + \tfrac{1}{2}(-g)^{\frac{1}{2}}(F^{\mu\rho}\delta A_\mu - A_\mu \delta F^{\mu\rho}). \end{aligned} \quad (4.5)$$

$z_{\mu\nu}$ satisfies the equation $\delta(R_{\mu\nu} + T_{\mu\nu}) = 0$, and δA_μ and $\delta F_{\mu\nu}$ satisfy $\delta(F^{\mu\nu}{}_{;\nu}) = 0$.

For our needs the solutions of the varied equations fall into two cases. To explain what they mean, we shall briefly consider the flat space N-P constants for the electromagnetic field. In flat space, there exists an infinite number of constants given by

$$F_m^l = \oint {}_1\bar{Y}_{l,m} \phi_0^l \, d\Omega, \quad l = 1, 2, 3, \cdots, \quad -l \le m \le l,$$

where ϕ_0^l is defined by the asymptotic expansion

$$\phi_0 = \frac{\phi_0^0}{r^3} + \frac{\phi_0^1}{r^4} + \frac{\phi_0^2}{r^5} + \cdots.$$

Robinson has shown that these constants generate incoming multipole pulses which are exact solutions of Maxwell's equations in Minkowski space. Newman and Penrose found the first member of this infinite set, $l = 1$, to be constant also in curved space. Therefore, we are led to seek a solution $\delta F_{\mu\nu}$ which has the leading behavior of an incoming dipole wave. The behavior of $z_{\mu\nu}$ is then driven by this requirement.

To construct the constants related to the gravitational field, we similarly must look for $z_{\mu\nu}$ whose leading behavior is an incoming quadrupole pulse. $\delta F_{\mu\nu}$ is then driven by this requirement. Below, we treat these two possibilities separately.

A. Maxwell N-P Constants

The first variation of the Maxwell equations (see Appendix B for the definitions of all tetrad components and spin coefficients which appear in these equations) are solved by an iterative procedure. First, one starts with $\delta\phi_0$ as an incoming dipole shock wave (see Appendix C):

$$\delta\phi_0 = {}_1Y_{1,m}\partial_r^2[B^m(v)/r],$$

where

$$v = u + 2r.$$

The "radial" equations allow all other pertinent quantities to be determined. The "nonradial" equation permits corrections to the ansatz to be calculated. One can then start again with a corrected $\delta\phi_0$. Fortunately, we do not have to go beyond the first correction, as we are interested only in asymptotic solutions in an asymptotically flat space-time.

The propagation equation for $\delta\phi_0$ shows the correction terms fall off in powers of $1/r$; thus[26]

$$\delta\phi_0 = {}_1Y_{1,m}\partial_r^2[B^m(v)/r] + O_5. \quad (4.6)$$

Here $\partial_r^2(B/r)$ is taken to be a term of asymptotic order $1/r^3$. This is to be understood by taking the convolution of $\partial_r^2(B/r)$ with a test function (a function of r alone) and considering $B(v)$ to be $O(1)$. In this way $\partial_r^n B(v) = O_n$ for all n required. In particular, we shall take $B(v)$ to be an incoming shock wave with a delta-function radial profile,

$$B^m(v) = a^m R^4 \delta(u + 2r - 2R), \quad (4.7)$$

where R is a radial parameter which defines the location of the shock wave and the a^m are complex numerical coefficients. (It should be noted here that $v = u + 2r$ is not an incoming null surface in curved space, and discontinuities in the field occur only across true characteristic surfaces. However, in asymptotically flat space, as one approaches future null infinity, $u + 2r$ does become an incoming null surface with the curvature terms falling off in powers of $1/r$. This behavior is sufficient for our purposes.)

With $\delta\phi_0$ above, the varied field components are found to be

$$\begin{aligned} \delta\phi_1 &= {}_0Y_{1,m}\partial_r\left(\frac{B^m}{r^2}\right) + {}_1Y_{1,m}(\eth\bar{\sigma}^0)\frac{1}{r^2}\partial_r\left(\frac{B^m}{r}\right) + O_5, \\ \delta\phi_2 &= {}_{-1}Y_{1,m}\frac{B^m}{r^3} - \frac{\partial\bar{\sigma}^0}{\partial u}\,{}_1Y_{1,m}\frac{1}{r}\partial_r\left(\frac{B^m}{r}\right) \\ &\quad - {}_0Y_{1,m}(\eth\bar{\sigma}^0)\frac{B^m}{r^4} + {}_1Y_{1,m}(\bar{\eth}\eth\bar{\sigma}^0)\frac{1}{r^2}\partial_r\left(\frac{B^m}{r}\right) \\ &\quad - {}_1Y_{1,m}\bar{\sigma}^0\frac{1}{2r}\partial_r\left(\frac{B^m}{r^2}\right) + O_5 \end{aligned} \quad (4.8)$$

and

$$\tilde{\delta}A_0 = ({}_0Y_{1,m} + {}_0\bar{Y}_{1,m})\frac{B^m}{r} - ({}_1Y_{1,m}\bar{\eth}\bar{\sigma}^0 + {}_1\bar{Y}_{1,m}\bar{\eth}\sigma^0)\frac{B^m}{r^3} + O_4, \quad (4.9)$$

$$\tilde{\delta}A_3 = -{}_1Y_{1,m}r\partial_r\left(\frac{B^m}{r^2}\right) - \sigma^0\,{}_1\bar{Y}_{1,m}\frac{B^m}{r^3} + O_4,$$

$$\tilde{\delta}A_2 = \tilde{\delta}\bar{A}_3.$$

As anticipated in the introductory paragraphs of this section, we find that the gravitational effects first appear in order O_5 in the spin coefficients. This order is too far down in $1/r$ to make a contribution to the conserved quantity defined by Eq. (3.14) with the transformation characterized by (4.6). Therefore, we need only be concerned with the wholly electromagnetic portion of t^ρ which can be written

$$t^\rho_{(e)} = \tfrac{1}{2}(-g)^{\frac{1}{2}}(F^{\mu\rho}\tilde{\delta}A_\mu - A_\mu\tilde{\delta}F^{\mu\rho}). \quad (4.10)$$

Now inserting Eqs. (4.6)–(4.9) into (4.10), one easily shows that the flux integrals on the right-hand side of (3.13) vanish in the limit of $R \to +\infty$. The constants F_m are given by

$$F_m = \lim_{R\to\infty}\int_{\mathcal{N}}(A_3\tilde{\delta}\bar{\phi}_0 - \phi_0\tilde{\delta}A_2)e^{2b}r^2\sin\theta\,dr\,d\theta\,d\phi,$$

$$F_m = \oint {}_1Y_{1,m}\phi_0^1\sin\theta\,d\theta\,d\phi. \quad (4.11)$$

B. Gravitational N–P Constants

To obtain the gravitational constants, we start with the ansatz taken from Appendix C:

$$\tilde{\delta}\psi_0 = {}_2Y_{2,m}\partial_r^4[B^m(v)/r], \quad (4.12)$$

with $B^m(v) = a^m R^6\delta(u + 2r - 2R)$. By the iterative process described in connection with the search for the Maxwell constants, we develop an asymptotic infinitesimal solution. In principle, we require this solution for both the Maxwell field and for the gravitational field. There is, however, a trick by which we can avoid evaluating the varied (infinitesimal) Maxwell field.

In the conserved vector density t^ρ, there is a purely gravitational part and a mixed part which contains the metric tensor as well as the electromagnetic field:

$$t^\rho_{(g)} = (-g)^{\frac{1}{2}}2g^{\alpha[\sigma}g^{\beta]\rho}z_{\alpha\beta;\sigma},$$

$$t^\rho_{(g,e)} = \tfrac{1}{2}[(-g)^{\frac{1}{2}}F^{\mu\rho}\tilde{\delta}A_\mu - A_\mu\tilde{\delta}((-g)^{\frac{1}{2}}F^{\mu\rho})]. \quad (4.13)$$

One can show easily that

$$t^\rho_{(g,e),\rho} = -\hat{g}^{\mu\nu}\tilde{\delta}T_{\mu\nu}, \quad (4.14)$$

when the Maxwell equations and their first variation are presumed satisfied. From the fact that the trace of the Maxwell stress–energy tensor vanishes, the right-hand side of (4.14) can be wholly expressed in terms of the varied metric and the original (unvaried) metric and Maxwell fields:

$$t^\rho_{(g,e),\rho} = T_{\mu\nu}\tilde{\delta}\hat{g}^{\mu\nu} = -(-g)^{\frac{1}{2}}z_{\mu\nu}T^{\mu\nu}, \quad (4.15)$$

where $z_{\mu\nu}$ is defined following Eq. (3.8). Therefore, the quantity to be integrated over the 4-dimensional volume illustrated in Fig. 1 is

$$t^\rho_{(g),\rho} - (-g)^{\frac{1}{2}}z_{\mu\nu}T^{\mu\nu} = 0. \quad (4.16)$$

We shall have to show that the 4-dimensional integral over $z_{\mu\nu}T^{\mu\nu}$ becomes just a 2-dimensional surface integral in the limit $R \to \infty$, with the solution determined by (4.12).

$z_{\mu\nu}$ is expressed in terms of the null tetrad given in Sec. 2:

$$z_{\mu\nu} = (e^{2b}/r)(\tilde{\delta}V)l_\mu l_\nu + 4(\tilde{\delta}b)l_{(\mu}n_{\nu)} + 2\bar{\xi}_i(\tilde{\delta}U^i)l_{(\mu}m_{\nu)} + 2\xi_i(\tilde{\delta}U^i)l_{(\mu}\bar{m}_{\nu)} - 2\xi_i(\tilde{\delta}\xi^i)\bar{m}_\mu\bar{m}_\nu - 2\bar{\xi}_i(\tilde{\delta}\bar{\xi}^i)m_\mu m_\nu. \quad (4.17)$$

The six functions $\tilde{\delta}b$, $\tilde{\delta}\xi^i$, $\tilde{\delta}U^i$, and $\tilde{\delta}V$ are determined from the first variation of the Newman–Penrose equations (4.2), (4.4), and (4.5) of Ref. 11. These equations are solved by the iterative procedure.

With $\tilde{\delta}\psi_0$ given in (4.12), the infinitesimal metric components are found to be

$$\tilde{\delta}b = \tfrac{1}{2}(\sigma^0\,{}_2\bar{Y}_{2,m} + \bar{\sigma}^0\,{}_2Y_{2,m})\partial_r\left[\frac{1}{r}\partial_r\left(\frac{B^m}{r}\right)\right] + O_5, \quad (4.18a)$$

$$\tilde{\delta}\xi^i = (\xi^{i0}\,{}_2Y_{2,m})r\partial_r\left[\frac{1}{r^2}\partial_r\left(\frac{B^m}{r}\right)\right] + O_5, \quad (4.18b)$$

$$\tilde{\delta}U^i = -2(\xi^{i0}\bar{\eth}\,{}_2\bar{Y}_{2,m} + \bar{\xi}^{i0}\bar{\eth}\,{}_2Y_{2,m})\partial_r\left(\frac{B^m}{r^3}\right) + O_5, \quad (4.18c)$$

$$\tilde{\delta}V = 2(\eth^2\,{}_2\bar{Y}_{2,m} + \bar{\eth}^2\,{}_2Y_{2,m})\left(\frac{B^m}{r^2}\right) + O_3. \quad (4.18d)$$

It turns out that only the leading terms of the infinitesimal metric components are necessary for the calculation if one chooses r to be the luminosity distance. When the same calculation is performed with r an affine parameter, additional asymptotic orders must be computed for each component of $z_{\mu\nu} := \tilde{\delta}g_{\mu\nu}$. The advantage of choosing r the luminosity distance is that a large number of terms are gathered together in the factor e^{2b} which otherwise appear in the metric of the 2-space $u = \text{const}$, $r = \text{const}$. This simplifies the appearance of the equations, as well as their solution.

Now we are in a position to examine $z_{\mu\nu}T^{\mu\nu}$; $T^{\mu\nu}$ is taken from (B37), and $z_{\mu\nu}$ is given in (4.17) and (4.18).

Therefore, after some manipulation and use of the field equations, we obtain (X_s are expressions of spin weight s whose exact form is not needed)

$$(-g)^{\frac{1}{2}}z_{\mu\nu}T^{\mu\nu} = -2\sin\theta\left(\partial_r X_0 + \frac{B^m(v)}{r^6}(\eth X_{-1} + \bar{\eth}X_1 + 4\,{}_2\bar{Y}_{2,m}\dot{S}) + \text{c.c.} + O_7\right), \quad (4.19)$$

where[27]

$$S = F\bar{\eth}^{-1}(\phi_1^0 - \bar{E}) - \phi_1^0\bar{\eth}^{-1}(\phi_0^1 - F) \quad (4.20)$$

and

$$F := \sum_{m=-1}^{1} F_m\,{}_1Y_{1,m}, \quad \bar{E} := \bar{e}\,{}_0Y_{0,0}.$$

In the course of the integration over R_4, the two terms involving the differential operators $\eth$ and $\bar{\eth}$ vanish by Eq. (A9). X_0 is linear in $B(v)$ [see Eq. (4.12)] and hence has no support on the boundaries Σ_1 and Σ_2 of Fig. 2.

Therefore,

$$\int_{R_4}(-g)^{\frac{1}{2}}z_{\mu\nu}T^{\mu\nu}d^4x = -8\int_{\mathcal{N}_2-\mathcal{N}_1}({}_2\bar{Y}_{2,m}S + \text{c.c.} + O_7)\sin\theta\,dr\,d\theta\,d\phi.$$

Clearly the delta function in S will give us just a 2-dimensional surface integral in the limit $R \to \infty$.

Going back to (4.16) and treating the divergence term as in Eq. (3.13), we find

$$\int_{\mathcal{N}_2-\mathcal{N}_1} t^\rho l_\rho\,dr\,d\theta\,d\phi = 2\int_{\mathcal{N}_2-\mathcal{N}_1}[2r\eth b + r^2\sigma\bar{\xi}_i\eth\bar{\xi}^i + 4\,{}_2\bar{Y}_{2,m}S + \text{c.c.} + O_7]\sin\theta\,dr\,d\theta\,d\phi.$$

Again the residual flux integrals over Σ_2 vanish in the limit, and we obtain the constants of the motion

$$G_m := \oint {}_2\bar{Y}_{2,m}(\psi_0^1 + 4S)\sin\theta\,d\theta\,d\phi. \quad (4.21)$$

In the absence of the electromagnetic field, $S = 0$ and we have simply

$$G_m = \oint {}_2\bar{Y}_{2,m}\psi_0^1\sin\theta\,d\theta\,d\phi. \quad (4.22)$$

From Eq. (B5) we find that ψ_0^1 can be expressed in terms of the rate of shear σ as

$$\psi_0^1 = -r^2\partial_r r^3\partial_r r^2\sigma \quad (4.23)$$

and therefore

$$G_m = -\lim_{r\to\infty}\oint {}_2\bar{Y}_{2,m}r^2\partial_r r^3\partial_r r^2\sigma\sin\theta\,d\theta\,d\phi. \quad (4.24)$$

This way of writing the constants is very suggestive. First of all, it shows clearly that the N–P constants are independent of the outgoing news which is defined by $\dot{\sigma}^0$. Furthermore, since the shear is determined by the distribution of matter, one imagines that the constants should be intimately related to the multipole structure in the infinite past. This interpretation is further supported by examining the static Weyl–Levi-Civita solution (translated into outgoing null coordinates by Bondi *et al.*[1]). This solution has one real N–P constant

$$G_0 = MQ - D^2, \quad (4.25)$$

where M, D, and Q are the real mass, dipole, and quadrupole moments. The shear of this exact solution (with $q = U^3 = 0$)[12] is

$$\sigma = -e^{-2b}\partial_r y = \frac{\sigma^0(\theta)}{r^2} + \frac{3Q\sin^2\theta}{2r^4} + \frac{5\sin^2\theta}{r^5} \times [(T - M^2D)\cos\theta + (MQ - D^2)] + O_6,$$

where T is the octupole moment. A BMS "frame" exists[28] in which σ^0 can be transformed to zero by a supertranslation [i.e., $\sigma^0(\theta) = \eth^2\alpha(\theta)$]. With σ^0 transformed away, it is clear that the shear arises solely from the structure of the sources.

One would have hoped that even when $S \neq 0$, the N–P constants G_m could be related to the geometrical behavior of null rays in the combined gravitational and electromagnetic fields. Unfortunately, the Maxwell field modifies the rate of shear only in a higher order than is selected in Eq. (4.23). We have not yet been able to identify the N–P constants of the combined fields with the geometry alone.

5. SUPERTRANSLATION INVARIANCE

The symmetry group of the asymptotically flat space–time under consideration is the Bondi–Metzner–Sachs (BMS) group (pseudogroup). This set of transformations preserves the Bondi–Sachs metric and the asymptotic boundary conditions. The transformations are the conformal transformation of the (θ, ϕ)-sphere into itself with conformal factor K and the transformation of one system of null hypersurfaces, $u = \text{const}$, into another by

$$u' = K(\theta, \phi)[u + \alpha(\theta, \phi)],$$

where α is an arbitrary real function on the sphere. The BMS transformations with $(\theta', \phi') = (\theta, \phi)$ are the supertranslations $u' = u + \alpha$. (Sachs[29] has shown that the supertranslations are an invariant Abelian subgroup of the BMS group with a factor group isomorphic to the orthochronous homogeneous Lorentz group $L^{\uparrow}_{+}$.)

Penrose, working in conformal space, proved that the N–P constants are supertranslation invariant.[9] Our calculation provides a simple proof in the physical space–time. The N–P constants are obtained from an expression of the form

$$\int_{R_4} t^{\rho}{}_{,\rho} d^4x = \int_{\mathcal{N}_2-\mathcal{N}_1} t^{\rho} l_{\rho}\, dr\, d\theta\, d\phi + \text{flux} = 0,$$

where in the limit of future null infinity the flux vanishes. Suppose $\mathcal{N}_2$ does not belong to the same coordinate system as $\mathcal{N}_1$ (i.e., $\mathcal{N}_2$ is a $u' = \text{const}$ surface, $\mathcal{N}_1$ a $u = \text{const}$ surface). Their coordinates will be connected by a BMS transformation. If $\mathcal{N}_2$ is a $u' = u - \alpha(\theta, \phi) = \text{const}$ surface, the normal to $\mathcal{N}_2$, $l'_{\rho} = u'_{,\rho}$, will be a null vector up to terms of order $1/r^2$ (i.e., $l'_{\rho} l'^{\rho} = -2r^{-2}|\eth\alpha|^2 + O_3$). The proof is completed by noting that $t^{\rho} l'_{\rho}$ is nonzero only at future null infinity where $l'_{\rho} \to l_{\rho}$.

Under a conformal transformation of the (θ, ϕ) sphere, l_{ρ} is unchanged, but the angular part of the incoming quadrupole perturbation transforms as a $D(2, 0)$ representation of $L^{\uparrow}_{+}$. Thus, in the limit of future null infinity, the N–P constants transform as a $D(2, 0)$ representation of $L^{\uparrow}_{+}$ and are unchanged by supertranslations.

6. HIGHER-ORDER CONSTANTS IN CURVED SPACE

Since there are an infinite number of N–P constants in flat space with only the leading members of this infinite set discovered to be constant in curved space, it is natural to look for higher-order constants or determine why they do not exist. The Green's theorem technique is used to search for the generators of higher-order incoming perturbations. In particular, we will examine the Maxwell field for generators of an incoming quadrupole perturbation.

The Maxwell equations for the incoming perturbation will be solved by the iterative method of Sec. 4. Again, the results of the flat-space theory guide us in choosing a trial solution. In flat space the N–P constants $F^2_m = \oint {}_1\bar{Y}_{2,m}\phi_0^2\, d\Omega$ generate an incoming quadrupole pulse. Thus, the trial solution will be an exact flat–space incoming quadrupole wave plus correction term. The flat-space solution is obtained from Appendix C:

$$\begin{aligned} \eth A_0 &= ({}_0Y_{2,m} + {}_0\bar{Y}_{2,m}) r^2 \partial_r (B^m/r^4), \\ \eth A_3 &= -(\tfrac{1}{3})^{\frac{1}{2}}\, {}_1Y_{2,m} r^2 \partial_r^2 (B^m/r^3), \\ \eth A_2 &= \eth \bar{A}_3, \end{aligned} \tag{6.1}$$

$$\begin{aligned} \eth\phi_0 &= (\tfrac{1}{3})^{\frac{1}{2}}\, {}_1Y_{2,m} r \partial_r^3 (B^m/r^2), \\ \eth\phi_1 &= {}_0Y_{2,m} r \partial_r^2 (B^m/r^3), \\ \eth\phi_2 &= (3)^{\frac{1}{2}}\, {}_{-1}Y_{2,m} r \partial_r (B^m/r^4). \end{aligned} \tag{6.2}$$

To see how far to carry the process of iteration, we first examine the integrand of the null surface integral. From (4.10),

$$(-g)^{\frac{1}{2}} t^{\rho} l_{\rho} = e^{2b} r^2 \sin\theta (A_3 \eth\bar{\phi}_0 - \phi_0 \eth A_2) + \text{c.c.} \tag{6.3}$$

It turns out that terms of $\eth A_2$ with curvature corrections through O_5 and terms of $\eth\phi_0$ through O_7 contribute. These are calculated from the field equations

$$(\Delta + \mu - 2\gamma)\eth\phi_0 = (\delta - 2\tau)\eth\phi_1 + \sigma\eth\phi_2$$

and

$$\eth\phi_0 = -(D - \rho)\eth A_3 + \sigma\eth A_2$$

by substituting (6.1) and (6.2) and using the iterative technique to obtain the corrections.

Equation (6.3) becomes

$$\begin{aligned} (-g)^{\frac{1}{2}} t^{\rho} l_{\rho} = \partial_r [X_0] &+ \frac{B^m(v)}{r^6} [(2\phi_0^2 - \sigma^0\bar{\sigma}^0\phi_0^0)\, {}_1\bar{Y}_{2,m} + \cdots] \\ &+ O_7 + \text{c.c.}, \end{aligned} \tag{6.4}$$

where $\cdots$ abbreviates several more terms involving curvature quantities. The result in (6.4) fixes the weight of the parameter R to be R^6. Thus $B(v)$ is chosen as

$$B^m(v) := a^m R^6 \delta(u + 2r - 2R). \tag{6.5}$$

Now it is possible to determine if the terms in the flux fall off rapidly enough for it to vanish. From (4.10) the flux integrand is (recall that n_{ρ} is the normal to Σ_2 at future null infinity)

$$\begin{aligned} (-g)^{\frac{1}{2}} t^{\rho} n_{\rho} = e^{2b} r^2 \sin\theta (&\phi_2 \eth A_3 + \phi_1 \eth A_0 \\ &- A_3 \eth\phi_2 - A_0 \eth\phi_1) + \text{c.c.} \end{aligned} \tag{6.6}$$

Using one of Maxwell's equations

$$(D - \rho)\eth\phi_2 = (\eth + 2\pi)\eth\phi_1 - \lambda\eth\phi_0$$

and one of the O_6 correction terms of $\eth\phi_0$,

$$\sigma^0\bar{\sigma}^0\, {}_1Y_{2,m} r^{-1} \partial_r^4 (B^m/r),$$

we can examine part of the tail of $\eth\phi_2$. Its r dependence is

$$\frac{1}{r}\int_{\infty}^{r} \frac{1}{x}\, \partial_x^4 \left(\frac{B^m}{x}\right) dx,$$

with $B(v)$ given in (6.5). Performing the integration, one obtains $\theta(r - R)/r$, where θ is the unit step function. Examining the third term of the flux integrand

$$e^{2b} r^2 \sin\theta A_3 \eth\phi_2,$$

we find

$$e^{2b} r^2 \sin\theta \left(\frac{A_3^0}{r} + \cdots\right) \times \left(\cdots + {}_1Y_{2,m} \frac{\partial\bar{\sigma}^0}{\partial u} \sigma^0\bar{\sigma}^0 \frac{\theta(r - R)}{r} + \cdots\right).$$

It is clear that this tail term (among others) leaves a contribution in the flux integral. Hence the result of using Green's identity here is that, at future null infinity,

$$\oint [(2\phi_0^2 - \sigma^0\bar{\sigma}^0\phi_0^0)\,{}_1\bar{Y}_{2,m} + \cdots] \sin\theta\, d\theta\, d\phi$$

$$= \text{nonzero flux.}$$

This same analysis holds true in the gravitational case as well. Thus, it is clear that the nonexistence of higher-order N–P constants is due to the dispersion of the space–time curvature. For the higher-order incoming pulses, the curvature produces wavetails which fall off too slowly for further constants to exist. Physically, this result can be understood as follows: The lowest incoming multipole perturbations, both for electrodynamics and gravitation, are exact solutions of the field equations at future null infinity. This is no longer true for the higher multipole solutions. For, in that case, the correction terms (wavetails) interfere with the linearized multipole fields even at future null infinity.

7. SUMMARY

The aim of this research has been to investigate the N–P constants and their associated symmetries. The initial work of Goldberg and Robinson has been supported and extended.

Of the infinite set of flat space N–P constants, only the first members are preserved as constants in curved space. The invariant transformations which they generate have been found to be incoming multipole waves (dipole for the electromagnetic constants, quadrupole for the gravitational constants) with support restricted to future null infinity. The nonexistence of higher-order constants has been discussed, and the invariant transformations they would have generated have been examined. Higher multipole wave perturbations have been obtained from the infinitesimal form of the field equations by an iteration process. The incoming higher multipole pulses are found to have correction terms (added to the exact flat-space solutions) which are wavetails that fall off too slowly for their contribution in the flux to vanish at future null infinity. Thus the space–time curvature is seen to allow only the lowest multipole incoming perturbations as invariant transformations generated by constants.

The Goldberg–Newman generalization of Green's identity has been used to obtain the Maxwell and Einstein–Maxwell N–P constants, and the invariant transformations they generate have been identified by comparing the generalized Green's identity and Noether's equation. The pure gravitational constants have been obtained from the shear of the outgoing null geodesics. This shows a close connection between the constants and physical sources, since part of the shear arises from the multipole structure of sources. Indeed, the presence of mass is always signaled by the shear of the surrounding null hypersurfaces. Even a Schwarzschild mass, unless viewed from a preferred frame, is to be observed via a shearing null congruence. It is when the preferred shear-free congruences exist that the N–P constants vanish. Let us stress that we find it unsatisfactory that a simple geometrical explanation has been obtained only for the pure gravitational field and not for the combined Einstein–Maxwell fields.

A null gauge has been used in the electromagnetic calculations, and Bondi–Sachs coordinate conditions adopted for gravitation. These algebraic, rather than differential, gauges are conformally invariant and allow the entire presentation to be taken over intact to Penrose's unphysical conformal space. Only in the conformal space is the limit of future null infinity a well-defined geometric object, and the arguments in this paper can be made rigorous by taking them over to the conformal space.

The supertranslation invariance of the constants has been obtained by examining the calculation over null surfaces belonging to different coordinate systems. The BMS transformation between the different surfaces leads directly to the supertranslation invariance of the constants.

No additional understanding of the selection rules imposed by the constants on radiative processes has been gained in these studies, and this topic along with an investigation of the effect of modifying the space–time boundary conditions requires further investigation for deeper insight into the N–P constants.

APPENDIX A: THE DIFFERENTIAL OPERATOR $\eth$ AND SPIN-s SPHERICAL HARMONICS

Here we will list the properties of $\eth$ and the spin-s spherical harmonics ${}_sY_{l,m}$, which were introduced by Newman and Penrose.[28,30–32]

Let the real and imaginary parts of the complex null vector field m^μ be spacelike vectors in the tangent plane of the sphere. Under a rotation of these vectors through the angle ψ, the complex vector transforms as

$$m'^\mu = e^{i\psi} m^\mu. \tag{A1}$$

A function defined on the sphere is said to have spin weight s if, under the above rotation,

$$\eta \to \eta' = e^{is\psi}\eta. \tag{A2}$$

The operator $\eth(\bar{\eth})$ raises (lowers) by one the spin weight of a function η with definite spin weight s. In general, s can be integral or half-integral. For s an integer, the ${}_sY_{l,m}$ can be related to the ordinary spherical harmonics by[33]

$$\begin{aligned} {}_sY_{l,m} &:= [(l-s)!/(l+s)!]^{\frac{1}{2}}(2^{\frac{1}{2}}\eth)^s Y_{l,m}, \quad 0 \le s \le l, \\ &:= (-1)^s[(l+s)!/(l-s)!]^{\frac{1}{2}}(2^{\frac{1}{2}}\bar{\eth})^{-s}Y_{l,m}, \\ &\qquad -l \le s \le 0. \end{aligned} \tag{A3}$$

For each value of s, the ${}_sY_{l,m}$ form a complete orthonormal set of functions on the sphere:

$$\oint {}_s\bar{Y}_{l',m'}\,{}_sY_{l,m}\,d\Omega = \delta_{ll'}\delta_{mm'}, \tag{A4}$$

where $d\Omega$ is the spherical area element. The following properties are easily verified:

$${}_s\bar{Y}_{l,m} = (-1)^{m+s}\,{}_{-s}Y_{l,-m}, \tag{A5}$$

$$\begin{aligned} 2^{\frac{1}{2}}\eth\,{}_sY_{l,m} &= [(l-s)(l+s+1)]^{\frac{1}{2}}\,{}_{s+1}Y_{l,m}, \\ 2^{\frac{1}{2}}\bar{\eth}\,{}_sY_{l,m} &= -[(l+s)(l-s+1)]^{\frac{1}{2}}\,{}_{s-1}Y_{l,m}. \end{aligned} \tag{A6}$$

Hence

$$(2\eth\bar{\eth} + 2\bar{\eth}\eth - s^2)\,{}_sY_{l,m} = -l(l+1)\,{}_sY_{l,m}. \tag{A7}$$

The commutation properties of $\eth$ and $\bar{\eth}$ are given by

$$(\bar{\eth}\eth - \eth\bar{\eth})\eta = s\eta. \tag{A8}$$

As a consequence of (A6), $\eth$ annihilates ${}_sY_{s,m}$ while $\bar{\eth}$ annihilates ${}_s\bar{Y}_{s,m}$. One further important property is that if α and β have spin weights one and minus one, respectively, then

$$\oint \bar{\eth}\alpha\,d\Omega = \oint \eth\beta\,d\Omega = 0. \tag{A9}$$

If the quantity τ is defined as a solution of the differential equation

$$\bar{\eth}\tau = \eta, \tag{A10}$$

where η has spin weight s, then τ exists and has spin weight $s+1$ if (a) η has spin weight $s < 0$ or if (b) η has spin weight $s \ge 0$ and $\oint {}_s\bar{Y}_{s,m}\eta\,d\Omega = 0$ (i.e., η does not possess its lowest possible l).

When τ exists, it can be written[5] as

$$\tau = \bar{\eth}^{-1}\eta. \tag{A11}$$

Finally, note that if the usual spherical angles (θ, ϕ) are introduced on the sphere, then $\eth$ acting on a spin-s function has the representation

$$\eth\eta := -\frac{1}{\sqrt{2}}(\sin\theta)^s\left(\frac{\partial}{\partial\theta} + \frac{i}{\sin\theta}\frac{\partial}{\partial\phi}\right)[(\sin\theta)^{-s}\eta], \tag{A12a}$$

and similarly

$$\bar{\eth}\eta := -\frac{1}{\sqrt{2}}(\sin\theta)^{-s}\left(\frac{\partial}{\partial\theta} - \frac{i}{\sin\theta}\frac{\partial}{\partial\phi}\right)[(\sin\theta)^s\eta]. \tag{A12b}$$

With

$$\alpha^0 = -\frac{1}{2\sqrt{2}}\cot\theta \quad \text{and} \quad \xi^{i0} = \frac{1}{\sqrt{2}}\left(1, \frac{i}{\sin\theta}\right),$$

the leading term of the differential operator $\delta + 2s\alpha$ is $-\eth$ when acting on a quantity of spin weight s. That is,

$$(\delta^0 + 2s\alpha^0)\eta = -\eth\eta, \tag{A13}$$

and similarly

$$(\bar{\delta}^0 - 2s\alpha^0)\eta = -\bar{\eth}\eta, \tag{A14}$$

where η has spin weight s.

APPENDIX B: ASYMPTOTIC SOLUTIONS OF THE FIELD EQUATIONS

Asymptotic solutions of the Einstein–Maxwell field equations have been obtained subject to the conditions that space–time be asymptotically flat and that the Maxwell fields vanish asymptotically. In Ref. 11 it is demonstrated that, for the vacuum field equations, a very weak condition which insures the asymptotic flatness of space–time is $\psi_0 = O(1/r^5)$ (along with certain smoothness conditions). Similarly one can show that the condition for the flat-space Maxwell field to vanish asymptotically is $\phi_0 = O(1/r^3)$. For the combined Einstein–Maxwell field, Kozarzewski[34] has shown that the combined assumptions imply the asymptotic flatness of space–time and the asymptotic vanishing of the Maxwell field. The asymptotic order of the spin coefficients is the same for both the vacuum field and the Einstein–Maxwell field.

To obtain the asymptotic solutions using our choice of tetrad, we followed the techniques of Newman and Unti.[16] First, one integrates the "radial" equations to obtain the asymptotic r dependence of the solution on a given $u = \text{const}$ hypersurface. This solution will contain "constants" of integration (actually functions of θ, ϕ on a given null hypersurface). The "nonradial" field equations will determine the propagation of the solution off the given hypersurface and relate the "constants" of integration to the initial data. The method is straightforward, and so here we list some definitions and then give the asymptotic solutions.

The twelve complex spin coefficients (linear combinations of the Ricci rotation coefficients) are

defined below:

$$\begin{aligned}
\kappa &:= m^\mu Dl_\mu, \quad \epsilon := \tfrac{1}{2}(n^\mu Dl_\mu - \bar{m}^\mu Dm_\mu),\\
\pi &:= -\bar{m}^\mu Dn_\mu,\\
\rho &:= m^\mu \delta l_\mu, \quad \alpha := \tfrac{1}{2}(n^\mu \delta l_\mu - \bar{m}^\mu \bar{\delta} m_\mu),\\
\lambda &:= -\bar{m}^\mu \bar{\delta} n_\mu,\\
\sigma &:= m^\mu \delta l_\mu, \quad \beta := \tfrac{1}{2}(n^\mu \delta l_\mu - \bar{m}^\mu \delta m_\mu),\\
\mu &:= -\bar{m}^\mu \delta n_\mu,\\
\tau &:= m^\mu \Delta l_\mu, \quad \gamma := \tfrac{1}{2}(n^\mu \Delta l_\mu - \bar{m}^\mu \Delta m_\mu),\\
\nu &:= -\bar{m}^\mu \Delta n_\mu.
\end{aligned} \tag{B1}$$

The Weyl tensor components are the five complex scalars

$$\begin{aligned}
\psi_0 &:= -C_{\mu\nu\rho\sigma} l^\mu m^\nu l^\rho m^\sigma,\\
\psi_1 &:= -C_{\mu\nu\rho\sigma} l^\mu n^\nu l^\rho m^\sigma,\\
\psi_2 &:= -C_{\mu\nu\rho\sigma} \bar{m}^\mu n^\nu l^\rho m^\sigma,\\
\psi_3 &:= -C_{\mu\nu\rho\sigma} \bar{m}^\mu n^\nu l^\rho n^\sigma,\\
\psi_4 &:= -C_{\mu\nu\rho\sigma} \bar{m}^\mu n^\nu \bar{m}^\rho n^\sigma.
\end{aligned} \tag{B2}$$

The Ricci tensor components are the four real and three complex scalars

$$\begin{aligned}
\Phi_{00} = \bar{\Phi}_{00} &:= -\tfrac{1}{2} R_{\mu\nu} l^\mu l^\nu,\\
\Phi_{11} = \bar{\Phi}_{11} &:= -\tfrac{1}{4} R_{\mu\nu} (l^\mu n^\nu + m^\mu \bar{m}^\nu),\\
\Phi_{22} = \bar{\Phi}_{22} &:= -\tfrac{1}{2} R_{\mu\nu} n^\mu n^\nu,\\
\Lambda = \bar{\Lambda} &:= \tfrac{1}{24} R,\\
\Phi_{01} = \bar{\Phi}_{10} &:= -\tfrac{1}{2} R_{\mu\nu} l^\mu m^\nu,\\
\Phi_{02} = \bar{\Phi}_{20} &:= -\tfrac{1}{2} R_{\mu\nu} m^\mu m^\nu,\\
\Phi_{12} = \bar{\Phi}_{21} &:= -\tfrac{1}{2} R_{\mu\nu} n^\mu m^\nu.
\end{aligned} \tag{B3}$$

The Maxwell field components are the three complex scalars

$$\begin{aligned}
\phi_0 &:= F_{\mu\nu} l^\mu m^\nu,\\
\phi_1 &:= \tfrac{1}{2} F_{\mu\nu} (l^\mu n^\nu - m^\mu \bar{m}^\nu),\\
\phi_2 &:= F_{\mu\nu} \bar{m}^\mu n^\nu.
\end{aligned} \tag{B4}$$

We present below the asymptotic solutions of the N-P field equations. Abbreviating $O(1/r^n)$ by O_n, we see that the spin coefficients are

$$\sigma = \sigma^0 r^{-2} - \tfrac{1}{2}\psi_0^0 r^{-4} - \tfrac{1}{3}\psi_0^1 r^{-5} + O_6, \tag{B5}$$

$$\rho = -r^{-1} - \tfrac{1}{2}(\sigma^0\bar{\sigma}^0) r^{-3} + O_5, \tag{B6}$$

$$\tau = -(\bar{\delta}\sigma^0) r^{-2} + [\tfrac{1}{2}\delta(\sigma^0\bar{\sigma}^0) - \psi_1^0 + 2\sigma^0 \delta\bar{\sigma}^0] r^{-3} + O_4, \tag{B7}$$

$$\pi = -(\delta\bar{\sigma}^0) r^{-2} + [\bar{\delta}(\sigma^0\bar{\sigma}^0) - \bar{\psi}_1^0 + 2\bar{\sigma}^0 \bar{\delta}\sigma^0] r^{-3} + O_4, \tag{B8}$$

$$\alpha = \alpha^0 r^{-1} + (\bar{\alpha}^0\bar{\sigma}^0 - \delta\bar{\sigma}^0) r^{-2} + \tfrac{1}{2}[\alpha^0\sigma^0\bar{\sigma}^0 + 2\bar{\sigma}^0\bar{\delta}\sigma^0 + \bar{\delta}(\sigma^0\bar{\sigma}^0) - \bar{\psi}_1^0] r^{-3} + O_4, \tag{B9}$$

where

$$\alpha^0 = -\frac{1}{2\sqrt{2}} \cot\theta,$$

$$\beta = -\bar{\alpha}^0 r^{-1} - \alpha^0\sigma^0 r^{-2} - \tfrac{1}{2}(\bar{\alpha}^0\sigma^0\bar{\sigma}^0 - 2\sigma^0\delta\bar{\sigma}^0 + \psi_1^0) r^{-3} + O_4, \tag{B10}$$

$$\mu = -\tfrac{1}{2} r^{-1} - \frac{1}{2}\left(\sigma^0 \frac{\partial\bar{\sigma}^0}{\partial u} + \bar{\sigma}^0 \frac{\partial\sigma^0}{\partial u} + \psi_2^0 + \bar{\psi}_2^0 + \bar{\delta}^2\sigma^0 + \delta^2\bar{\sigma}^0\right) r^{-2} + O_3, \tag{B11}$$

$$\lambda = \frac{\partial\bar{\sigma}^0}{\partial u} r^{-1} + \tfrac{1}{2}(\bar{\sigma}^0 - 2\bar{\delta}\delta\bar{\sigma}^0) r^{-2} + O_3, \tag{B12}$$

$$\gamma = -(\tfrac{1}{2}\psi_2^0 + \bar{\alpha}^0\delta\bar{\sigma}^0 - \alpha^0\bar{\delta}\sigma^0) r^{-2} + O_3, \tag{B13}$$

$$\nu = -\frac{1}{2}\left[\bar{\delta}\left(\sigma^0 \frac{\partial\bar{\sigma}^0}{\partial u} + \bar{\sigma}^0 \frac{\partial\sigma^0}{\partial u} + \psi_2^0 + \bar{\psi}_2^0\right)\right] r^{-2} + O_3. \tag{B14}$$

The metric components are the following:

$$\xi^i = \xi^{i0} r^{-1} - \sigma^0 \bar{\xi}^{i0} r^{-2} + \tfrac{1}{2}\sigma^0\bar{\sigma}^0 \xi^{i0} r^{-3} + O_4, \tag{B15}$$

where

$$\xi^{i0} = \frac{1}{\sqrt{2}}\left(1, \frac{i}{\sin\theta}\right),$$

$$V = r + \left(\sigma^0 \frac{\partial\bar{\sigma}^0}{\partial u} + \bar{\sigma}^0 \frac{\partial\sigma^0}{\partial u} + \psi_2^0 + \bar{\psi}_2^0\right) + O_1, \tag{B16}$$

$$U^i = (\bar{\xi}^{i0}\bar{\delta}\sigma^0 + \xi^{i0}\delta\bar{\sigma}^0) r^{-2} + \tfrac{1}{6}[\bar{\xi}^{i0}(4\psi_1^0 - 3\delta(\sigma^0\bar{\sigma}^0) - 12\sigma^0\delta\bar{\sigma}^0) + \text{c.c.}] r^{-3} + O_4, \tag{B17}$$

$$b = -\tfrac{1}{4}\sigma^0\bar{\sigma}^0 r^{-2} + \tfrac{1}{16}[2(\sigma^0\bar{\sigma}^0)^2 + \sigma^0\bar{\psi}_0^0 + \bar{\sigma}^0\psi_0^0 - 2\phi_0^0\bar{\phi}_0^0] r^{-4} + O_5. \tag{B18}$$

The Weyl tensor components are

$$\psi_0 = \psi_0^0 r^{-5} + \psi_0^1 r^{-6} + O_7, \tag{B19}$$

$$\psi_1 = \psi_1^0 r^{-4} + (\bar{\delta}\psi_0^0 + 3\phi_0^0\bar{\phi}_1^0) r^{-5} + O_6, \tag{B20}$$

$$\psi_2 = \psi_2^0 r^{-3} + (\bar{\delta}\psi_1^0 + 2\phi_1^0\bar{\phi}_1^0) r^{-4} + O_5, \tag{B21}$$

$$\psi_3 = \psi_3^0 r^{-2} + (\bar{\delta}\psi_2^0 + \phi_2^0\bar{\phi}_1^0) r^{-3} + O_4, \tag{B22}$$

$$\psi_4 = \psi_4^0 r^{-1} + (\bar{\delta}\psi_3^0) r^{-2} + O_3. \tag{B23}$$

The Maxwell field components are

$$\phi_0 = \phi_0^0 r^{-3} + \phi_0^1 r^{-4} + O_5, \tag{B24}$$

$$\phi_1 = \phi_1^0 r^{-2} + (\bar{\delta}\phi_0^0) r^{-3} + O_4, \tag{B25}$$

$$\phi_2 = \phi_2^0 r^{-1} + (\bar{\delta}\phi_1^0) r^{-2} + O_3. \tag{B26}$$

Finally, we have the further equations involving u derivatives:

$$\psi_2^0 - \bar{\psi}_2^0 = \bar{\eth}^2\sigma^0 - \eth^2\bar{\sigma}^0 + \bar{\sigma}^0\frac{\partial\sigma^0}{\partial u} - \sigma^0\frac{\partial\bar{\sigma}^0}{\partial u}, \quad \text{(B27)}$$

$$\psi_3^0 = \eth\frac{\partial\bar{\sigma}^0}{\partial u}, \quad \text{(B28)}$$

$$\psi_4^0 = -\frac{\partial^2\bar{\sigma}^0}{\partial u^2}, \quad \text{(B29)}$$

$$\partial_u\psi_0^0 + \eth\psi_1^0 - 3\sigma^0\psi_2^0 - 3\phi_0^0\bar{\phi}_2^0 = 0, \quad \text{(B30)}$$

$$\partial_u\psi_1^0 + \eth\psi_2^0 - 2\sigma^0\psi_3^0 - 2\phi_1^0\bar{\phi}_2^0 = 0, \quad \text{(B31)}$$

$$\partial_u\psi_0^1 + \bar{\eth}(\eth\psi_0^0 - 4\sigma^0\psi_1^0) + 4(\bar{\phi}_1^0\eth\phi_0^0 - 2\sigma^0\phi_1^0\bar{\phi}_1^0 - \phi_0^1\bar{\phi}_2^0) = 0, \quad \text{(B32)}$$

$$\partial_u\psi_2^0 + \eth\psi_3^0 - \sigma^0\psi_4^0 - \phi_2^0\bar{\phi}_2^0 = 0, \quad \text{(B33)}$$

$$\partial_u\phi_0^0 + \eth\phi_1^0 - \sigma^0\phi_2^0 = 0, \quad \text{(B34)}$$

$$\partial_u\phi_0^1 + \bar{\eth}(\eth\phi_0^0 - 2\sigma^0\phi_1^0) = 0, \quad \text{(B35)}$$

$$\partial_u\phi_1^0 + \eth\phi_2^0 = 0. \quad \text{(B36)}$$

For purposes of reference, we include the tetrad form of the Maxwell stress–energy tensor:

$$\begin{aligned}\tfrac{1}{2}T^{\mu\nu} = {} & \phi_2\bar{\phi}_2 l^\mu l^\nu + \phi_0\bar{\phi}_0 n^\mu n^\nu \\ & + \bar{\phi}_0\phi_2 m^\mu m^\nu + \phi_0\bar{\phi}_2\bar{m}^\mu\bar{m}^\nu \\ & - 2\bar{\phi}_1\phi_2 l^{(\mu}m^{\nu)} - 2\phi_1\bar{\phi}_2 l^{(\mu}\bar{m}^{\nu)} \\ & - 2\bar{\phi}_0\phi_1 n^{(\mu}m^{\nu)} - 2\phi_0\bar{\phi}_1 n^{(\mu}\bar{m}^{\nu)} \\ & + 2\phi_1\bar{\phi}_1(l^{(\mu}n^{\nu)} + m^{(\mu}\bar{m}^{\nu)}). \end{aligned} \quad \text{(B37)}$$

APPENDIX C: EXACT MULTIPOLE SOLUTIONS IN FLAT SPACE

The exact flat-space multipole solutions are well known both for the Maxwell field and the linearized Weyl field. For the convenience of the reader we list the advanced solutions in terms of the coordinates and notation of this paper. The incoming solutions are given in terms of an arbitrary function $B = B(v)$ whose argument is constant on the past null cones $v = u + 2r$. Specifying the function is equivalent to giving the incoming news.

The Maxwell components are

$$\phi_0 = \left(\frac{2}{l(l+1)}\right)^{\frac{1}{2}} {}_1Y_{l,m}r^{l-1}\partial_r^{l+1}\left(\frac{B^{lm}}{r^l}\right), \quad \text{(C1)}$$

$$\phi_1 = {}_0Y_{l,m}r^{l-1}\partial_r^l\left(\frac{B^{lm}}{r^{l+1}}\right), \quad \text{(C2)}$$

$$\phi_2 = \left(\frac{l(l+1)}{2}\right)^{\frac{1}{2}} {}_{-1}Y_{l,m}r^{l-1}\partial_r^{l-1}\left(\frac{B^{lm}}{r^{l+2}}\right), \quad \text{(C3)}$$

where $l \geq 1$ and $-l \leq m \leq +l$.

The advanced Weyl tensor components are

$$\psi_0 = K_{+2\,2}Y_{l,m}r^{l-2}\partial_r^{l+2}\left(\frac{B^{lm}}{r^{l-1}}\right), \quad \text{(C4)}$$

$$\psi_1 = K_{+1\,1}Y_{l,m}r^{l-2}\partial_r^{l+1}\left(\frac{B^{lm}}{r^l}\right), \quad \text{(C5)}$$

$$\psi_2 = {}_0Y_{l,m}r^{l-2}\partial_r^l\left(\frac{B^{lm}}{r^{l+1}}\right), \quad \text{(C6)}$$

$$\psi_3 = K_{-1\,-1}Y_{l,m}r^{l-2}\partial_r^{l-1}\left(\frac{B^{lm}}{r^{l+2}}\right), \quad \text{(C7)}$$

$$\psi_4 = K_{-2\,-2}Y_{l,m}r^{l-2}\partial_r^{l-2}\left(\frac{B^{lm}}{r^{l+3}}\right), \quad \text{(C8)}$$

where

$$K_p := [2^p(l-p)!/(l+p)!]^{\frac{1}{2}}$$

and

$$l \geq 2, \quad -l \leq m \leq +l.$$

* Supported in part by the National Science Foundation; based partly on the first author's doctoral dissertation at Syracuse University.

† National Academy of Sciences Postdoctoral Resident Research Associate. Present address: Aerospace Research Laboratories, Wright–Patterson Air Force Base, Ohio 45433.

[1] H. Bondi, M. van der Burg, and A. Metzner, Proc. Roy. Soc. (London) **A269**, 21 (1962).
[2] R. K. Sachs, Proc. Roy. Soc. (London) **A270**, 103 (1962).
[3] E. T. Newman and T. W. J. Unti, J. Math. Phys. **6**, 1806 (1965).
[4] E. T. Newman and R. Penrose, Phys. Rev. Letters **15**, 231 (1965).
[5] A. Exton, E. T. Newman, and R. Penrose, J. Math. Phys. **10**, 1566 (1969).
[6] J. N. Goldberg, J. Math. Phys. **8**, 2161 (1967).
[7] J. N. Goldberg, J. Math. Phys. **9**, 674 (1968).
[8] D. C. Robinson, J. Math. Phys. **10**, 1745 (1969).
[9] E. T. Newman and R. Penrose, Proc. Roy. Soc. (London) **A305**, 175 (1968).
[10] A. Exton, private communication.
[11] E. T. Newman and R. Penrose, J. Math. Phys. **3**, 566 (1962); **4**, 998 (1963).
[12] The correspondence between the notation here and that of Sachs (Ref. 2) is $2\gamma = \gamma + \delta$, $2q = \gamma - \delta$, $b = \beta$.
[13] We assume the Newman–Penrose spin coefficient formalism is well known, and we follow the sign conventions and notation of Ref. 11, with the exception that Greek indices run through 0, 1, 2, 3 and Latin indices through 2, 3.
[14] S. W. Hawking, J. Math. Phys. **9**, 598 (1968).
[15] ∂_α and $,_\alpha$ denote the ordinary derivative, ∇_α and $;_\alpha$ the covariant derivative.
[16] E. T. Newman and T. Unti, J. Math. Phys. **3**, 891 (1962).
[17] J. N. Goldberg and E. T. Newman, J. Math. Phys. **10**, 369 (1969).
[18] We use $\tilde{\delta}$ since the more usual $\bar{\delta}$ is reserved here for an intrinsic tetrad derivative.
[19] Here we are using the notation

$$\partial^A L := \frac{\partial L}{\partial y_A}, \quad \partial^{A\rho}L := \frac{\partial L}{\partial y_{A,\rho}}.$$

This notation will also be used with the variables g_a which are introduced in the following paragraph; thus,

$$\partial^a L = \frac{\partial L}{\partial g_a}, \quad \partial^{a\rho}L = \frac{\partial L}{\partial g_{a,\rho}}.$$

[20] H. Streudel, Nuovo Cimento **39**, 395 (1965).
[21] H. Streudel, Z. Naturforsch. **21a**, 1826 (1965).
[22] A. Trautman, "Conservation Laws," in *Gravitation*, L. Witten, Ed. (Wiley, New York, 1962).

[23] A. Trautman, Commun. Math. Phys. **6**, 248 (1967).
[24] $\eta_{\mu\nu\rho\sigma}$ is the completely antisymmetric tensor and $\eta_{0123} = (-g)^{\frac{1}{2}}$.
[25] O_n abbreviates $O(1/r^n)$ for $n \geq 1$.
[26] Summation over m is understood.
[27] It is important to recognize that the Maxwell equations allow one to write

$$\partial_u S = \bar{\delta}[\phi_2^{I0}\bar{\delta}^{-1}(\phi_0^1 - F)] + \phi_1^{I0}\delta\phi_0^0 - 2\sigma^0\phi_1^0\phi_1^{I0} - \phi_0^1\phi_2^{I0}$$

(see Appendix A for the definition of $\bar{\delta}^{-1}$).
[28] E. T. Newman and R. Penrose, J. Math. Phys. **7**, 863 (1966).
[29] R. Sachs, Phys. Rev. **128**, 2851 (1962).
[30] J. N. Goldberg, A. J. Macfarlane, E. T. Newman, F. Rohrlich, and E. C. G. Sudarshan, J. Math. Phys. **8**, 2155 (1967).
[31] M. Carmeli, J. Math. Phys. **10**, 569 (1969).
[32] A. Staruszkiewicz, J. Math. Phys. **8**, 2221 (1967).
[33] The operator δ defined here is related to the usual Newman-Penrose operator by $\delta = 2^{-\frac{1}{2}}\delta_{NP}$.
[34] B. Kozarzewski, Acta Phys. Polon. **27**, 775 (1965).

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 11, NUMBER 12 DECEMBER 1970

Perturbation Theory for Damped Nonlinear Oscillations

KENNETH S. MENDELSON
Physics Department, Marquette University, Milwaukee, Wisconsin 53233

(Received 4 June 1970)

A perturbation theory has been worked out for the decay of autonomous, nonlinear oscillations in the case where there is large linear damping. The solution reduces to a solution obtained by Kryloff and Bogoliuboff for small damping and to the perturbation solution for periodic oscillations for vanishing damping. The solution is applied to the decay of oscillations in Duffing's equation. In this case it shows good agreement with a solution obtained by numerical integration.

I. INTRODUCTION

In most treatments of nonlinear oscillations by perturbation methods, only periodic oscillations are treated; transients are not considered. Kryloff and Bogoliuboff[1] have used a perturbation method to discuss transients in the equation

$$\ddot{x} + \nu^2 x = -\epsilon f(x, \dot{x}), \tag{1}$$

where ϵ is a small parameter. In this equation the damping terms are small.

It is of interest to extend the perturbation method to the case where there is a large linear damping in addition to small nonlinear terms. In the present paper the method of Kryloff and Bogoliuboff is extended to the equation

$$\ddot{x} + 2\gamma\dot{x} + \nu^2 x = -\epsilon f(x, \dot{x}), \tag{2}$$

where again ϵ is a small parameter. It is assumed that $\gamma < \nu$ so that the system is underdamped in the linear approximation.

The perturbation method is developed in Sec. II. In Sec. III it is applied to the decay of oscillations described by Duffing's equation.

II. THE PERTURBATION METHOD

For $\epsilon = 0$, Eq. (2) has the solution

$$x = a_0 e^{-\gamma t}\cos(\omega_0 t + \psi_0), \tag{3}$$

where

$$\omega_0 = (\nu^2 - \gamma^2)^{\frac{1}{2}} \tag{4}$$

and a_0 and ψ_0 are constants. We seek a solution of Eq. (2) that reduces to Eqs. (3) and (4) in the limit $\epsilon \to 0$.

Following Kryloff and Bogoliuboff, we look for a solution

$$x = x(a, \psi), \tag{5}$$

where x is periodic in ψ and where

$$\frac{da}{dt} = \xi(a), \quad \frac{d\psi}{dt} = \omega(a). \tag{6}$$

Substituting Eqs. (5) and (6) in Eq. (2) yields

$$\omega^2\frac{\partial^2 x}{\partial\psi^2} + 2\omega\xi\frac{\partial^2 x}{\partial a\partial\psi} + \xi^2\frac{\partial^2 x}{\partial a^2} + \left(\xi\frac{d\omega}{da} + 2\gamma\omega\right)\frac{\partial x}{\partial\psi} + \left(\xi\frac{d\xi}{da} + 2\gamma\xi\right)\frac{\partial x}{\partial a} + \nu^2 x = -\epsilon f\left(x, \omega\frac{\partial x}{\partial\psi} + \xi\frac{\partial x}{\partial a}\right). \tag{7}$$

To obtain a perturbation solution of Eq. (7), x, ξ, and ω are expanded in powers of ϵ; thus

$$\begin{aligned} x &= x_0 + \epsilon x_1 + \cdots, \\ \xi &= -\gamma a + \epsilon\xi_1 + \cdots, \\ \omega &= \omega_0 + \epsilon\omega_1 + \cdots, \end{aligned} \tag{8}$$

where ω_0 is given by Eq. (4). The leading terms in the expansions for ξ and ω are chosen to yield the solution (3) and (4) in the limit $\epsilon \to 0$.

[23] A. Trautman, Commun. Math. Phys. **6**, 248 (1967).
[24] $\eta_{\mu\nu\rho\sigma}$ is the completely antisymmetric tensor and $\eta_{0123} = (-g)^{\frac{1}{2}}$.
[25] O_n abbreviates $O(1/r^n)$ for $n \geq 1$.
[26] Summation over m is understood.
[27] It is important to recognize that the Maxwell equations allow one to write

$$\partial_u S = \bar{\delta}[\phi_2^{I0}\bar{\delta}^{-1}(\phi_0^1 - F)] + \phi_1^{I0}\delta\phi_0^0 - 2\sigma^0\phi_1^0\phi_1^{I0} - \phi_0^1\phi_2^{I0}$$

(see Appendix A for the definition of $\bar{\delta}^{-1}$).
[28] E. T. Newman and R. Penrose, J. Math. Phys. **7**, 863 (1966).
[29] R. Sachs, Phys. Rev. **128**, 2851 (1962).
[30] J. N. Goldberg, A. J. Macfarlane, E. T. Newman, F. Rohrlich, and E. C. G. Sudarshan, J. Math. Phys. **8**, 2155 (1967).
[31] M. Carmeli, J. Math. Phys. **10**, 569 (1969).
[32] A. Staruszkiewicz, J. Math. Phys. **8**, 2221 (1967).
[33] The operator δ defined here is related to the usual Newman-Penrose operator by $\delta = 2^{-\frac{1}{2}}\delta_{NP}$.
[34] B. Kozarzewski, Acta Phys. Polon. **27**, 775 (1965).

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 11, NUMBER 12 DECEMBER 1970

Perturbation Theory for Damped Nonlinear Oscillations

KENNETH S. MENDELSON
Physics Department, Marquette University, Milwaukee, Wisconsin 53233

(Received 4 June 1970)

A perturbation theory has been worked out for the decay of autonomous, nonlinear oscillations in the case where there is large linear damping. The solution reduces to a solution obtained by Kryloff and Bogoliuboff for small damping and to the perturbation solution for periodic oscillations for vanishing damping. The solution is applied to the decay of oscillations in Duffing's equation. In this case it shows good agreement with a solution obtained by numerical integration.

I. INTRODUCTION

In most treatments of nonlinear oscillations by perturbation methods, only periodic oscillations are treated; transients are not considered. Kryloff and Bogoliuboff[1] have used a perturbation method to discuss transients in the equation

$$\ddot{x} + \nu^2 x = -\epsilon f(x, \dot{x}), \tag{1}$$

where ϵ is a small parameter. In this equation the damping terms are small.

It is of interest to extend the perturbation method to the case where there is a large linear damping in addition to small nonlinear terms. In the present paper the method of Kryloff and Bogoliuboff is extended to the equation

$$\ddot{x} + 2\gamma\dot{x} + \nu^2 x = -\epsilon f(x, \dot{x}), \tag{2}$$

where again ϵ is a small parameter. It is assumed that $\gamma < \nu$ so that the system is underdamped in the linear approximation.

The perturbation method is developed in Sec. II. In Sec. III it is applied to the decay of oscillations described by Duffing's equation.

II. THE PERTURBATION METHOD

For $\epsilon = 0$, Eq. (2) has the solution

$$x = a_0 e^{-\gamma t}\cos(\omega_0 t + \psi_0), \tag{3}$$

where

$$\omega_0 = (\nu^2 - \gamma^2)^{\frac{1}{2}} \tag{4}$$

and a_0 and ψ_0 are constants. We seek a solution of Eq. (2) that reduces to Eqs. (3) and (4) in the limit $\epsilon \to 0$.

Following Kryloff and Bogoliuboff, we look for a solution

$$x = x(a, \psi), \tag{5}$$

where x is periodic in ψ and where

$$\frac{da}{dt} = \xi(a), \quad \frac{d\psi}{dt} = \omega(a). \tag{6}$$

Substituting Eqs. (5) and (6) in Eq. (2) yields

$$\omega^2\frac{\partial^2 x}{\partial\psi^2} + 2\omega\xi\frac{\partial^2 x}{\partial a\partial\psi} + \xi^2\frac{\partial^2 x}{\partial a^2} + \left(\xi\frac{d\omega}{da} + 2\gamma\omega\right)\frac{\partial x}{\partial\psi} + \left(\xi\frac{d\xi}{da} + 2\gamma\xi\right)\frac{\partial x}{\partial a} + \nu^2 x = -\epsilon f\left(x, \omega\frac{\partial x}{\partial\psi} + \xi\frac{\partial x}{\partial a}\right). \tag{7}$$

To obtain a perturbation solution of Eq. (7), x, ξ, and ω are expanded in powers of ϵ; thus

$$\begin{aligned} x &= x_0 + \epsilon x_1 + \cdots, \\ \xi &= -\gamma a + \epsilon\xi_1 + \cdots, \\ \omega &= \omega_0 + \epsilon\omega_1 + \cdots, \end{aligned} \tag{8}$$

where ω_0 is given by Eq. (4). The leading terms in the expansions for ξ and ω are chosen to yield the solution (3) and (4) in the limit $\epsilon \to 0$.

Substitution of Eq. (8) into Eq. (7) yields a system of equations for the terms in Eq. (8). The equation for the zero-order terms is

$$\omega_0^2\frac{\partial^2 x_0}{\partial\psi^2} - 2\gamma\omega_0 a\frac{\partial^2 x_0}{\partial a\partial\psi} + \gamma^2 a^2\frac{\partial^2 x_0}{\partial a^2} + 2\gamma\omega_0\frac{\partial x_0}{\partial\psi} - \gamma^2 a\frac{\partial x_0}{\partial a} + (\omega_0^2+\gamma^2)x_0 = 0, \quad (9)$$

which has the solution

$$x_0 = a\cos\psi. \quad (10)$$

The first-order equation is

$$\omega_0^2\frac{\partial^2 x_1}{\partial\psi^2} - 2\gamma\omega_0 a\frac{\partial^2 x_1}{\partial a\partial\psi} + \gamma^2 a^2\frac{\partial^2 x_1}{\partial a^2} + 2\gamma\omega_0\frac{\partial x_1}{\partial\psi} - \gamma^2 a\frac{\partial x_1}{\partial a} + (\omega_0^2+\gamma^2)x_1 = 2\omega_0\omega_1 a\cos\psi + 2\omega_0\xi_1\sin\psi - \gamma a\frac{d\omega_1}{da}a\sin\psi + \left(\gamma a\frac{d\xi_1}{da} - \gamma\xi_1\right)\cos\psi - f(a\cos\psi, -\omega_0 a\sin\psi - \gamma a\cos\psi). \quad (11)$$

Since the function f is periodic in ψ, it can be expanded in a Fourier series:

$$f = \tfrac{1}{2}F_0(a) + \sum_{n=0}^{\infty}[F_n(a)\cos n\psi + G_n(a)\sin n\psi]. \quad (12)$$

The coefficients in Eq. (12) can in turn be expanded in Maclaurin's series in a;

$$F_n(a) = \sum_{k=0}^{\infty}F_k^{(n)}a^k, \quad G_n(a) = \sum_{k=0}^{\infty}G_k^{(n)}a^k. \quad (13)$$

In the limit $\gamma\to 0$, the terms in $\cos\psi$ and $\sin\psi$ on the right side of Eq. (11) must vanish identically to prevent the appearance of secular terms in the solution.[1] We require that these terms vanish for all γ. This leads to the equations

$$2\omega_0\omega_1 a + \gamma a\frac{d\xi_1}{da} - \gamma\xi_1 = F_1(a),$$
$$2\omega_0\xi_1 - \gamma a^2\frac{d\omega_1}{da} = G_1(a). \quad (14)$$

Expanding ξ_1 and ω_1 in power series in a,

$$\xi_1 = \sum_{k=0}^{\infty}A_k^{(1)}a^k, \quad \omega_1 = \sum_{k=0}^{\infty}B_k^{(1)}a^{k-1}, \quad (15)$$

leads to the equations

$$(k-1)\gamma A_k^{(1)} + 2\omega_0 B_k^{(1)} = F_k^{(1)},$$
$$2\omega_0 A_k^{(1)} - (k-1)\gamma B_k^{(1)} = G_k^{(1)},$$

from which we obtain

$$A_k^{(1)} = \frac{(k-1)\gamma F_k^{(1)} + 2\omega_0 G_k^{(1)}}{4\omega_0^2 + (k-1)^2\gamma^2},$$
$$B_k^{(1)} = \frac{2\omega_0 F_k^{(1)} - (k-1)\gamma G_k^{(1)}}{4\omega_0^2 + (k-1)^2\gamma^2}. \quad (16)$$

To determine x_1, it is written as a Fourier series in ψ,

$$x_1(a,\psi) = \tfrac{1}{2}y_0(a) + \sum_{n=2}^{\infty}[y_n(a)\cos n\psi + z_n(a)\sin n\psi], \quad (17)$$

and the coefficients are expanded in power series in a,

$$y_n(a) = \sum_{k=0}^{\infty}A_k^{(n)}a^k, \quad z_n(a) = \sum_{k=0}^{\infty}B_k^{(n)}a^k. \quad (18)$$

Substituting Eqs. (17) and (18) into Eq. (11) and using Eqs. (12) and (13) then gives

$$[(1-n^2)\omega_0^2 + (k-1)^2\gamma^2]A_k^{(n)} - 2n(k-1)\gamma\omega_0 B_k^{(n)} = -F_k^{(n)},$$
$$2n(k-1)\gamma\omega_0 A_k^{(n)} + [(1-n^2)\omega_0^2 + (k-1)^2\gamma^2]B_k^{(n)} = -G_k^{(n)},$$

which have the solutions

$$A_k^{(n)} = -\frac{[(1-n^2)\omega_0^2+(k-1)^2\gamma^2]F_k^{(n)}+2n(k-1)\gamma\omega_0 G_k^{(n)}}{[(1-n^2)\omega_0^2+(k-1)^2\gamma^2]^2+4n^2(k-1)^2\gamma^2\omega_0^2}$$
$$B_k^{(n)} = \frac{2n(k-1)\gamma\omega_0 F_k^{(n)}-[(1-n^2)\omega_0^2+(k-1)^2\gamma^2]G_k^{(n)}}{[(1-n^2)\omega_0^2+(k-1)^2\gamma^2]^2+4n^2(k-1)^2\gamma^2\omega_0^2}. \quad (19)$$

This completes the determination of the first-order corrections to the solution. The procedure can be carried to higher orders in the same way.

III. EXAMPLE: DUFFING'S EQUATION

As an example of the above procedure, we consider the decay of oscillations in Duffing's equation,

$$\ddot{x} + 2\gamma\dot{x} + \nu^2 x + \epsilon x^3 = 0. \quad (20)$$

In this equation

$$f(x_0,\dot{x}_0) = x_0^3 = a^3\cos^3\psi = a^3(\tfrac{3}{4}\cos\psi + \tfrac{1}{4}\cos 3\psi) \quad (21)$$

or

$$F_3^{(1)} = \tfrac{3}{4}, \quad F_3^{(3)} = \tfrac{1}{4}, \quad (22)$$

with all other coefficients vanishing. The coefficients in the solution are obtained from Eqs. (16) and (19).

The nonvanishing coefficients are

$$A_3^{(1)} = \frac{3\gamma}{8\nu^2}, \quad B_3^{(1)} = \frac{3\omega_0}{8\nu^2},$$
$$A_3^{(3)} = \frac{2\omega_0^2 - \gamma^2}{16\nu^2(4\omega_0^2 + \gamma^2)}, \tag{23}$$
$$B_3^{(3)} = \frac{3\gamma\omega_0}{16\nu^2(4\omega_0^2 + \gamma^2)}.$$

Combining the results obtained above yields

$$\frac{da}{dt} = -\gamma a + \frac{3\gamma\epsilon}{8\nu^2} a^3 \tag{24}$$

and

$$\frac{d\psi}{dt} = \omega_0 + \frac{3\omega_0\epsilon}{8\nu^2} a^2. \tag{25}$$

Equation (24) has the solution

$$a = a_0 e^{-\gamma t} \Big/ \left(1 + \frac{3\epsilon a_0^2}{8\nu^2}(e^{-2\gamma t} - 1)\right)^{\frac{1}{2}}. \tag{26}$$

Substituting Eq. (26) into Eq. (25) and integrating gives

$$\psi = \psi_0 + \omega_0 t - \frac{\epsilon\omega_0}{2\gamma} \ln\left(1 + \frac{3\epsilon a_0^2}{8\nu^2}(e^{-2\gamma t} - 1)\right). \tag{27}$$

Finally, to first order in ϵ, the solution to Eq. (20) is

$$x = a\cos\psi + \epsilon A_3^{(3)} a^3 \cos 3\psi + \epsilon B_3^{(3)} a^3 \sin 3\psi. \tag{28}$$

As a check on the solution of Eq. (20) obtained above, a second solution was obtained by numerical integration using a fourth-order Runge-Kutta formula.[2] The results are plotted in Fig. 1 for $\gamma = 1/\sqrt{2}$, $\nu = 1$, and $\epsilon = 1$. The two curves agree very closely even in the region where the function is changing rapidly.

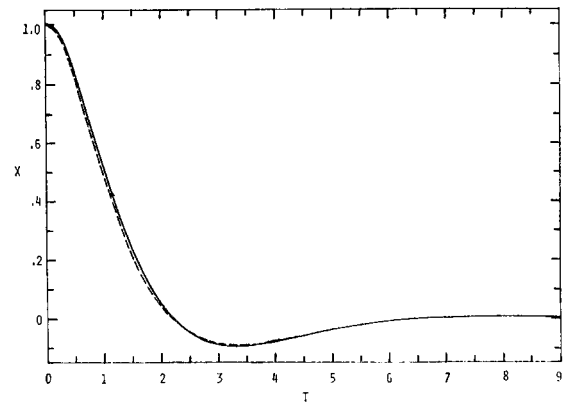


FIG. 1. Solutions of Eq. (20) obtained by perturbation theory (solid line) and numerical integration (dashed line) for $\gamma = 1/\sqrt{2}$, $\nu = 1$, and $\epsilon = 1$.

ACKNOWLEDGMENTS

The author wishes to express his appreciation to Dr. T. Bronikowski, Dr. P. Bender, and Dr. C. Braunschweiger of the Marquette University Mathematics Department for many discussions concerning this work.

[1] N. Kryloff and N. Bogoliuboff, *Introduction to Nonlinear Mechanics* (Princeton U.P., Princeton, N.J., 1947), Sec. 32, p. 51.

[2] M. L. James, G. M. Smith, and J. C. Wolford, *Applied Numerical Methods for Digital Computation with FORTRAN* (International Textbook, Scranton, Pa., 1967), p. 339.

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 11, NUMBER 12 DECEMBER 1970

Half-Range Expansion Theorems in Studies of Polarized Light*

E. E. BURNISTON AND C. E. SIEWERT
Departments of Mathematics and Nuclear Engineering, North Carolina State University
Raleigh, North Carolina 27607

(Received 19 June 1970)

The established normal modes of the vector equation of transfer describing the transport of polarized light are used to construct solutions to typical half-space problems. The half-range completeness theorem required by this method is discussed in the context of systems of singular integral equations. Although the Riemann–Hilbert problem encountered here is defined in terms of continuous rather than Hölder-continuous functions, the existence of a canonical solution is established, and the developed properties of this canonical matrix are used to complete the proof of the necessary half-range expansion theorem.

I. INTRODUCTION

We consider here the vector equation of transfer

$$\mu \frac{\partial}{\partial \tau} \mathbf{I}(\tau, \mu) + \mathbf{I}(\tau, \mu) = \tfrac{1}{2}\omega \mathsf{Q}(\mu) \int_{-1}^{1} \tilde{\mathsf{Q}}(\mu') \mathbf{I}(\tau, \mu')\, d\mu' \tag{1}$$

applicable to several studies of the scattering of polarized light.[1] Relying principally on Chandrasekhar's formulation of this mathematical model,[1] we denote by $\mathbf{I}(\tau, \mu)$ a vector whose two components $I_l(\tau, \mu)$ and $I_r(\tau, \mu)$ are the angular intensities in the two states of polarization. Further, τ is the optical variable, and μ is the direction cosine (as measured from the positive τ axis) of the propagating radiation.

The scattering process considered here is characterized in Eq. (1) by the single-scatter albedo ω and the square matrix $\mathsf{Q}(\mu)$, with $\tilde{\mathsf{Q}}(\mu)$ denoting the transpose of $\mathsf{Q}(\mu)$. Although much of the analysis presented in this paper is valid for a general Q matrix of polynomials, we are concerned primarily with the form

$$\mathsf{Q}(\mu) = \frac{3(c+2)^{\frac{1}{2}}}{2(c+2)} \begin{bmatrix} c\mu^2 + \frac{2}{3}(1-c) & (2c)^{\frac{1}{2}}(1-\mu^2) \\ \frac{1}{3}(c+2) & 0 \end{bmatrix}; \tag{2}$$

we thus allow the right-hand side of Eq. (1) to contain the two parameters ω and c so that the following special cases can be readily identified:

For $c = 1$ and $\omega = 1$, Eq. (1) with Eq. (2) yields Chandrasekhar's conservative Rayleigh-scattering model[1]

$$\mu \frac{\partial}{\partial \tau} \mathbf{I}(\tau, \mu) + \mathbf{I}(\tau, \mu) = \tfrac{1}{2} \int_{-1}^{1} \mathsf{K}(\mu, \mu') \mathbf{I}(\tau, \mu')\, d\mu', \tag{3}$$

where

$$\mathsf{K}(\mu, \mu') = \tfrac{3}{4} \begin{bmatrix} 2(1-\mu^2)(1-\mu'^2) + \mu^2\mu'^2 & \mu^2 \\ \mu'^2 & 1 \end{bmatrix}. \tag{4}$$

For the case $c \in [0, 1]$ and $\omega = 1$, Eqs. (1) and (2) yield Chandrasekhar's conservative model[1] for a mixture of scattering laws,

$$\mu \frac{\partial}{\partial \tau} \mathbf{I}(\tau, \mu) + \mathbf{I}(\tau, \mu) = \tfrac{1}{2} \int_{-1}^{1} [c\mathsf{K}(\mu, \mu') + (1-c)\mathsf{E}] \mathbf{I}(\tau, \mu')\, d\mu', \tag{5}$$

where

$$\mathsf{E} = \tfrac{1}{2} \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix}. \tag{6}$$

Finally, observing the choices $c = 1$ and $\omega \in [0, 1)$, we note that Eqs. (1) and (2) yield the nonconservative version of Eq. (3), as considered, for example, by Simmons,[2] Mullikin,[3] Abhyankar and Fymat,[4] and Schnatz and Siewert,[5] whereas, if we allow the values $c \in [0, 1]$ and $\omega \in [0, 1)$, we obtain the analogous nonconservative version of Eq. (5).

In order to establish the elementary solutions of Eq. (1), we introduce the proposed form

$$\mathbf{I}_\eta(\tau, \mu) = \mathbf{\Phi}(\eta, \mu) e^{-\tau/\eta} \tag{7}$$

to obtain

$$(\eta - \mu)\mathbf{\Phi}(\eta, \mu) = \tfrac{1}{2}\omega\eta \mathsf{Q}(\mu)\mathbf{M}(\eta), \tag{8}$$

where the normalization vector $\mathbf{M}(\eta)$ is given by

$$\mathbf{M}(\eta) = \int_{-1}^{1} \tilde{\mathsf{Q}}(\mu) \mathbf{\Phi}(\eta, \mu)\, d\mu. \tag{9}$$

In the usual manner,[6] we first consider the discrete spectrum, $\eta \notin [-1, 1]$, and solve Eq. (8) to find

$$\mathbf{\Phi}(\pm\eta_0, \mu) = \tfrac{1}{2}\omega\eta_0 \frac{1}{\eta_0 \mp \mu} \mathsf{Q}(\mu)\mathbf{M}(\pm\eta_0), \tag{10}$$

where $\pm\eta_0$ are the two zeros (in the complex plane cut from -1 to 1 along the real axis) of the dispersion function

$$\Lambda(z) = \det \mathbf{\Lambda}(z), \tag{11}$$

where

$$\mathbf{\Lambda}(z) = \mathsf{I} + z \int_{-1}^{1} \mathbf{\Psi}(\mu) \frac{d\mu}{\mu - z}; \tag{12}$$

here I denotes the unit matrix, and the "characteristic" matrix is

$$\Psi(\mu) = \tfrac{1}{2}\omega\tilde{Q}(\mu)Q(\mu). \tag{13}$$

Clearly since $\Psi(\mu)$ is a symmetic matrix, $\Lambda(z)$ is symmetric; further, we note that $\Lambda(z) = \Lambda(-z)$. Although the normalization vectors can be established from

$$\Lambda(\eta_0)\mathbf{M}(\eta_0) = \mathbf{0}, \tag{14}$$

we do not require any explicit forms here.

Solving Eq. (8) now for $\eta \in (-1, 1)$, we write

$$\Phi(\eta, \mu) = \tfrac{1}{2}\omega\left(\eta\frac{P}{\eta - \mu} + \lambda(\eta)\delta(\eta - \mu)\right)Q(\mu)\mathbf{M}(\eta), \tag{15}$$

where the symbol P denotes that all ensuing integrals are to be evaluated in the Cauchy principal-value sense and $\delta(x)$ is the Dirac δ functional. If we multiply Eq. (15) by $\tilde{Q}(\mu)$ and integrate over μ from -1 to 1, we find

$$[\lambda(\eta) - \lambda(\eta)\Psi(\eta)]\mathbf{M}(\eta) = \mathbf{0}; \tag{16}$$

and hence from

$$\det[\lambda(\eta) - \lambda(\eta)\Psi(\eta)] = 0, \tag{17}$$

where

$$\lambda(\eta) = \mathrm{I} + \eta\int_{-1}^{1}\Psi(\mu)\frac{P}{\mu - \eta}\,d\mu, \tag{18}$$

we obtain, in general, a quadratic equation in $\lambda(\eta)$, which yields two solutions $\lambda_1(\eta)$ and $\lambda_2(\eta)$. There is thus a twofold degeneracy for the continuum, $\eta \in (-1, 1)$; there result then two solutions to Eq. (8),

$$\Phi_\alpha(\eta, \mu) = \tfrac{1}{2}\omega\left(\eta\frac{P}{\eta - \mu} + \lambda_\alpha(\eta)\delta(\eta - \mu)\right)Q(\mu)\mathbf{M}_\alpha(\eta), \quad \eta \in (-1, 1), \quad \alpha = 1 \text{ and } 2. \tag{19}$$

Since the normal modes are now explicitly available, we write our general solution to Eq. (1) in the form

$$\mathbf{I}(\tau, \mu) = A(\eta_0)\Phi(\eta_0, \mu)e^{-\tau/\eta_0} + A(-\eta_0)\Phi(-\eta_0, \mu)e^{\tau/\eta_0} + \int_{-1}^{1}[A_1(\eta)\Phi_1(\eta, \mu) + A_2(\eta)\Phi_2(\eta, \mu)]e^{-\tau/\eta}\,d\eta, \tag{20}$$

where $A(\pm\eta_0)$ and $A_\alpha(\eta)$, $\alpha = 1$ and 2, are the arbitrary expansion coefficients to be determined once the boundary conditions of a given problem are specified. The full-range expansion theorem for the eigenvectors considered here has been established by Schnatz and Siewert[5]; it is, however, the considerably more important half-range expansion theorem we wish to discuss.

II. ANALYSIS

In order to illustrate the need for the half-range expansion theorem, we consider a typical half-space problem: We seek a bounded solution to Eq. (1) for $\tau \in [0, \infty)$ such that the radiation incident at the surface may be specified, i.e.,

$$\mathbf{I}(0, \mu) = \mathbf{I}(\mu), \quad \mu \in (0, 1), \tag{21}$$

where $\mathbf{I}(\mu)$ is given. Clearly, a bounded solution can be readily obtained from Eq. (20) by requiring $A(-\eta_0) \equiv 0$ and $A_1(\eta) = A_2(\eta) \equiv 0$, $\eta < 0$. Thus the desired solution can be written as

$$\mathbf{I}(\tau, \mu) = A(\eta_0)\Phi(\eta_0, \mu)e^{-\tau/\eta_0} + \int_0^1[A_1(\eta)\Phi_1(\eta, \mu) + A_2(\eta)\Phi_2(\eta, \mu)]^{-\tau/\eta}\,d\eta, \tag{22}$$

where the expansion coefficients must be chosen such that Eq. (21) is satisfied. We must solve, therefore, the system of singular integral equations

$$\mathbf{I}(\mu) = A(\eta_0)\Phi(\eta_0, \mu) + \int_0^1[A_1(\eta)\Phi_1(\eta, \mu) + A_2(\eta)\Phi_2(\eta, \mu)]\,d\eta, \quad \mu \in (0, 1). \tag{23}$$

A statement to the effect that Eq. (23) admits a solution for an arbitrary Hölder[7] vector $\mathbf{I}(\mu)$ is the required half-range expansion theorem; it is this statement we wish to establish.

For the sake of notational convenience, we now introduce the matrix

$$\Psi(\eta, \mu) = [\Phi_1(\eta, \mu)\ \Phi_2(\eta, \mu)] \tag{24}$$

and a vector $\mathbf{A}(\eta)$, with elements $A_1(\eta)$ and $A_2(\eta)$, in order to write Eq. (23) as

$$\mathbf{I}'(\mu) = \int_0^1\Psi(\eta, \mu)\mathbf{A}(\eta)\,d\eta, \quad \mu \in (0, 1), \tag{25}$$

where temporarily we have taken the discrete term to the left-hand side of the equation and defined

$$\mathbf{I}'(\mu) = \mathbf{I}(\mu) - A(\eta_0)\Phi(\eta_0, \mu). \tag{26}$$

We note that, when Eq. (15) is premultiplied by $\tilde{Q}(\mu)$ and Eq. (16) is used, there results

$$\tilde{Q}(\mu)\Psi(\eta, \mu) = \left(\eta\frac{P}{\eta - \mu}\Psi(\mu) + \delta(\eta - \mu)\lambda(\eta)\right)V(\eta), \tag{27}$$

where $\mathsf{V}(\eta)$ is the normalization matrix:

$$\mathsf{V}(\eta) = \int_{-1}^{1} \tilde{\mathsf{Q}}(\mu)\mathbf{\Psi}(\eta, \mu)\, d\mu. \tag{28}$$

A more convenient form of Eq. (25) can now be established by premultiplying that equation by $\tilde{\mathsf{Q}}(\mu)$ and using Eq. (27):

$$\tilde{\mathsf{Q}}(\mu)\mathbf{I}'(\mu) = \boldsymbol{\lambda}(\mu)\mathbf{B}(\mu) + \mathbf{\Psi}(\mu)\int_0^1 \eta\mathbf{B}(\eta)\frac{P}{\eta - \mu}\, d\eta, \quad \mu \in (0, 1), \tag{29}$$

where we have defined

$$\mathbf{B}(\eta) = \mathsf{V}(\eta)\mathbf{A}(\eta). \tag{30}$$

In the usual manner, Eq. (29) can be converted to an equivalent inhomogeneous Riemann–Hilbert problem.[5–7] To this end, we introduce the sectionally holomorphic function

$$\mathbf{N}(z) = \frac{1}{2\pi i}\int_0^1 \eta\mathbf{B}(\eta)\frac{d\eta}{\eta - z}, \tag{31}$$

with boundary values, from above (+) and below (−) the cut, given by the Plemelj formulas[7]

$$\mathbf{N}^{\pm}(\mu) = \frac{1}{2\pi i}\int_0^1 \eta\mathbf{B}(\eta)\frac{P}{\eta - \mu}\, d\eta \pm \tfrac{1}{2}\mu\mathbf{B}(\mu), \quad \mu \in (0, 1); \tag{32}$$

and thus

$$\pi i[\mathbf{N}^{+}(\mu) + \mathbf{N}^{-}(\mu)] = \int_0^1 \eta\mathbf{B}(\eta)\frac{P}{\eta - \mu}\, d\eta, \quad \mu \in (0, 1), \tag{33a}$$

and

$$\mathbf{N}^{+}(\mu) - \mathbf{N}^{-}(\mu) = \mu\mathbf{B}(\mu), \quad \mu \in (0, 1). \tag{33b}$$

The boundary values of the $\mathbf{\Lambda}$ matrix, as given by Eq. (12), are related by

$$\mathbf{\Lambda}^{+}(\mu) + \mathbf{\Lambda}^{-}(\mu) = 2\boldsymbol{\lambda}(\mu), \qquad \mu \in (-1, 1), \tag{34a}$$

and

$$\mathbf{\Lambda}^{+}(\mu) - \mathbf{\Lambda}^{-}(\mu) = 2\pi i\mu\mathbf{\Psi}(\mu), \quad \mu \in (-1, 1); \tag{34b}$$

these relations can be used with Eqs. (33) to write Eq. (29) in the form

$$\mu\tilde{\mathsf{Q}}(\mu)\mathbf{I}'(\mu) = \mathbf{\Lambda}^{+}(\mu)\mathbf{N}^{+}(\mu) - \mathbf{\Lambda}^{-}(\mu)\mathbf{N}^{-}(\mu), \quad \mu \in (0, 1). \tag{35}$$

The general solution to the inhomogeneous Eq. (35) may be written as[7]

$$\mathbf{N}(z) = \mathsf{X}^{-1}(z)\left(\frac{1}{2\pi i}\int_0^1 \mathbf{\Gamma}(\mu)\mathbf{I}'(\mu)\frac{d\mu}{\mu - z} + \mathbf{P}(z)\right), \tag{36}$$

where $\mathbf{P}(z)$ is a vector with polynomial elements,

$$\mathbf{\Gamma}(\mu) = \mu\mathsf{X}^{+}(\mu)[\mathbf{\Lambda}^{+}(\mu)]^{-1}\tilde{\mathsf{Q}}(\mu), \tag{37}$$

and $\mathsf{X}(z)$ is the canonical solution to the homogeneous problem

$$\mathsf{X}^{+}(\mu) = \mathsf{X}^{-}(\mu)[\mathbf{\Lambda}^{-}(\mu)]^{-1}\mathbf{\Lambda}^{+}(\mu), \quad \mu \in (0, 1). \tag{38}$$

Clearly, then, to complete the desired proof,we must argue that a matrix $\mathsf{X}(z)$, analytic in the complex plane cut from 0 to 1 along the real axis, exists and has properties such that $\mathbf{N}(z)$ as given by Eq. (36) can be made consistent with the original definition introduced by Eq. (31).

In order to be consistent with the notational convention established by Muskhelishvili[7] and Vekua,[8] we define $\mathbf{\Phi}(z)$ to be the transpose of $\mathsf{X}(z)$ and thus write the transpose of Eq. (38) as

$$\mathbf{\Phi}^{+}(\mu) = \mathsf{G}(\mu)\mathbf{\Phi}^{-}(\mu), \quad \mu \in (0, 1), \tag{39a}$$

where the symmetry properties of $\mathbf{\Lambda}(z)$ allow us to write

$$\mathsf{G}(\mu) = \mathbf{\Lambda}^{+}(\mu)[\mathbf{\Lambda}^{-}(\mu)]^{-1}, \quad \mu \in (0, 1). \tag{39b}$$

It is clear that,by adding an arbitrary arc C_1 to the real-line segment [0, 1], we need only deal with a closed Lyapanov contour C. On C_1 we define $\mathsf{G}(\mu) = \mathsf{I}$; thus, since Eq. (12) yields continuous boundary values on (0, 1) while at the end points of the line segment $\lim_{\mu\to 0+}\mathsf{G}(\mu) = \lim_{\mu\to 1-}\mathsf{G}(\mu) = \mathsf{I}$, the matrix $\mathsf{G}(\mu)$ is continuous for all $\mu \in C$. This function, however, fails to be Hölder continuous at $\mu = 1$, as can be seen from the special case $c = 0$. Here $\mathsf{G}(\mu)$ becomes

$$\mathsf{G}_0(\mu) = \begin{bmatrix} g(\mu) & 0 \\ 0 & 1 \end{bmatrix}, \quad \mu \in (0, 1), \quad c = 0, \tag{40}$$

with $g(\mu)$ being equivalent to the one-speed result discussed by Case and Zweifel[9]:

$$g(\mu) = \left[1 + \tfrac{1}{2}\omega\mu \ln\left(\frac{1 - \mu}{1 + \mu}\right) + \tfrac{1}{2}\omega\mu\pi i\right] \times \left[1 + \tfrac{1}{2}\omega\mu \ln\left(\frac{1 - \mu}{1 + \mu}\right) - \tfrac{1}{2}\omega\mu\pi i\right]^{-1}. \tag{41}$$

In order for $g(\mu)$ to be Hölder continuous at $\mu = 1$, we require $|g(\mu) - g(1)|/|\mu - 1|^{\alpha}$ to be bounded for some $\alpha \in [0, 1)$. However,

$$\frac{|g(\mu) - 1|}{|\mu - 1|^{\alpha}} = \frac{\omega\mu\pi}{|\mu - 1|^{\alpha}}\left|1 + \tfrac{1}{2}\omega\mu \ln\left(\frac{1 - \mu}{1 + \mu}\right) - \tfrac{1}{2}\omega\mu\pi i\right| \tag{42}$$

is clearly unbounded for all appropriate α. For the same reason, the matrix $\mathsf{G}(\mu)$ fails to be piecewise Hölder continuous on C. Thus, without modification, it is apparent that neither the theory given by Muskhelishvili[7] nor that of Vekua[8] is sufficient for the solution of

$$\mathbf{\Phi}^{+}(\mu) = \mathsf{G}(\mu)\mathbf{\Phi}^{-}(\mu), \quad \mu \in C, \tag{43}$$

and so we base our reasoning for the existence of a canonical solution on the theory given by Mandžavidze and Hvedelidze.[10] These authors prove that if $\mathsf{G}(\mu)$ is a nonsingular continuous matrix and C a simple closed Lyapanov curve, then there exists a so-called canonical matrix $\mathbf{\Phi}_0(z)$ such that: (i) The matrices $\mathbf{\Phi}_0(z)$ and $[\mathbf{\Phi}_0(z)]^{-1}$ are representable by Cauchy integrals with polynomial principal parts at infinity; (ii) the matrix $\mathbf{\Phi}_0(z)$ has normal form at infinity; (iii) the boundary values on C of $\mathbf{\Phi}_0(z)$ are L_p functions ($p > 1$) and those of $[\mathbf{\Phi}_0(z)]^{-1}$ are L_q functions, p and q being conjugate indices; in addition, these boundary values satisfy

$$\mathsf{G}(\mu) = \mathbf{\Phi}_0^{+}(\mu)[\mathbf{\Phi}_0^{-}(\mu)]^{-1} \tag{44}$$

almost everywhere on C. The procedure reported by Mandžavidze and Hvedelidze is not concerned with the solubility and equivalence of a certain quasi-Fredholm equation as is the theory given by Muskhelishvili (see Eq. 126.5, p. 386 of Ref. 7). Basically their method is to show that the problem

$$\mathbf{\Theta}^{+}(\mu) - \mathbf{\Theta}^{-}(\mu) = \mathsf{G}_1(\mu)\mathbf{\Theta}^{-}(\mu) + \mathsf{E}(\mu), \quad \mu \text{ on } C, \tag{45}$$

where each component of the matrix $\mathsf{G}_1(\mu)$ is sufficiently small and $\mathsf{E}(\mu)$ is a given matrix of L_p functions, can be solved by the following sequence:

$$\mathbf{\Theta}_0^{-}(\mu) = \mathbf{0} \tag{46a}$$

and

$$\mathbf{\Theta}_m(z) = \frac{1}{2\pi i}\int_C \mathsf{G}_1(\mu)\mathbf{\Theta}_{m-1}^{-}(\mu)\frac{d\mu}{\mu - z} + \frac{1}{2\pi i}\int_C \mathsf{E}(\mu)\frac{d\mu}{\mu - z}, \quad m = 1, 2, 3, \cdots. \tag{46b}$$

This sequence has been shown to be Cauchy in the L_p norm[10] and hence convergent to an L_p function which satisfies Eq. (45) almost everywhere.

In order to establish the required properties of this canonical solution, we need to determine the index of $\mathsf{G}(\mu)$, namely

$$\kappa = \frac{1}{2\pi i}[\arg \det \mathsf{G}(\mu)]_C$$

which is easily seen, for the case of $\mathsf{Q}(\mu)$ as given by Eq. (2), to be unity. Thus the partial indices[7,10] satisfy

$$\kappa_1 + \kappa_2 = 1. \tag{47}$$

In actual fact, these partial indices turn out to be zero and unity. A proof of this may be modeled on one given by Kuščer.[11] Let $\mathbf{\Phi}_0(z)$ be the canonical solution to Eq. (43); then it is easily shown[7,10] that any other solution of finite degree at infinity can be expressed as

$$\mathbf{\Phi}(z) = \mathbf{\Phi}_0(z)\mathsf{P}(z), \tag{48}$$

where $\mathsf{P}(z)$ is a matrix of polynomials. Considering now the function

$$\mathbf{\Psi}(z) = \mathbf{\Lambda}(z)[\tilde{\mathbf{\Phi}}_0(-z)]^{-1}, \tag{49}$$

we note that the boundary values of $\mathbf{\Psi}(z)$ on C clearly satisfy Eq. (43) almost everywhere. If we now change z to $-z$ in Eq. (49), we can take boundary values of the resulting equation to obtain

$$\mathbf{\Psi}^{-}(-\mu) = \mathbf{\Lambda}^{+}(\mu)[\tilde{\mathbf{\Phi}}_0^{+}(\mu)]^{-1}, \quad \mu \in (0, 1), \tag{50a}$$

while

$$\mathbf{\Psi}^{+}(-\mu) = \mathbf{\Lambda}^{-}(\mu)[\tilde{\mathbf{\Phi}}_0^{-}(\mu)]^{-1}, \quad \mu \in (0, 1), \tag{50b}$$

since, as noted previously, $\mathbf{\Lambda}(z) = \mathbf{\Lambda}(-z)$. Equation (43) and the fact that $\mathbf{\Lambda}(z) = \tilde{\mathbf{\Lambda}}(z)$ can now be used to show that Eqs. (50) yield $\mathbf{\Psi}^{-}(-\mu) = \mathbf{\Psi}^{+}(-\mu)$, $\mu \in (0, 1)$, almost everywhere. Clearly, then, $\mathbf{\Psi}(z)$ is analytic in the plane cut from 0 to 1 on the real axis and of finite degree at infinity, and thus $\mathbf{\Psi}(z)$ is a solution to Eq. (43). Consequently, it can be expressed in the form of Eq. (48). We now suppose that one of the partial indices is negative, say κ_1 [note from Eq. (47) that only one can be negative]; then the first element of the first column of $\mathbf{\Phi}_0(z)$ has a pole at infinity. This implies, however, from Eq. (49) that the first column of $\mathbf{\Psi}(z)$ vanishes at infinity. But, recalling Eq. (48), we note that this is impossible. Thus the partial indices are zero and unity.

Since the existence of a canonical matrix $\mathbf{\Phi}_0(z)$ has been established, it is a simple matter to complete the proof of half-range completeness. The fact that $\mathbf{\Phi}_0(z)$ must be of normal form at infinity requires that

$$\lim_{z\to\infty} \mathbf{\Phi}_0(z)\begin{bmatrix} z^{\kappa_1} & 0 \\ 0 & z^{\kappa_2} \end{bmatrix} = \mathsf{I}, \tag{51}$$

and, since we have shown that the partial indices must be nonnegative and sum to unity, without loss of generality, we select $\kappa_1 = 0$ and $\kappa_2 = 1$.

We note that the analytic properties of $\mathbf{N}(z)$, as given by Eq. (36), are correct if we make the identification

$$\mathsf{X}(z) = \tilde{\mathbf{\Phi}}_0(z); \tag{52}$$

on the other hand, we observe from Eq. (31) that $z\mathbf{N}(z)$ must be bounded as z tends to infinity. The required behavior of $\mathbf{X}^{-1}(z)$ for large z can be deduced from Eqs. (51) and (52):

$$\mathbf{X}^{-1}(z) \sim z \begin{bmatrix} \frac{1}{z} + \cdots & \frac{a}{z} + \cdots \\ \frac{b}{z^2} + \cdots & 1 + \cdots \end{bmatrix}. \tag{53}$$

Considering now Eq. (36), we find that, in order for $z\mathbf{N}(z)$ to be bounded at infinity, we must take $\mathbf{P}(z) \equiv \mathbf{0}$; the behavior of $\mathbf{X}^{-1}(z)$ for large z, as given by Eq. (53), indicates that $z\mathbf{N}(z)$ will not be bounded at infinity unless we impose on $\mathbf{I}'(\mu)$ the constraint

$$\begin{bmatrix} 0 \\ 1 \end{bmatrix}^{\mathrm{T}} \int_0^1 \mathbf{\Gamma}(\mu)\mathbf{I}'(\mu)\, d\mu = 0, \tag{54}$$

where the superscript T denotes transpose. Recalling Eq. (26), we see that Eq. (54) can be satisfied for all appropriate $\mathbf{I}'(\mu)$ simply by choosing the correct discrete coefficient $A(\eta_0)$:

$$A(\eta_0)\begin{bmatrix} 0 \\ 1 \end{bmatrix}^{\mathrm{T}} \int_0^1 \mathbf{\Gamma}(\mu)\mathbf{\Phi}(\eta_0, \mu)\, d\mu = \begin{bmatrix} 0 \\ 1 \end{bmatrix}^{\mathrm{T}} \int_0^1 \mathbf{\Gamma}(\mu)\mathbf{I}(\mu)\, d\mu. \tag{55}$$

The desired expansion theorem (23) is thus established.

ACKNOWLEDGMENT

The authors wish to express their appreciation to J. F. Heyda of the General Electric Company, Philadelphia, who so kindly translated from Russian the required sections of the paper by Mandžavidze and Hvedelidze.

* This work was supported in part by National Science Foundation Grant GK 11935 and Joint Services Advisory Group Grant AFOSR-69-1779.

[1] S. Chandrasekhar, *Radiative Transfer* (Oxford U.P., London, 1950).
[2] G. M. Simmons, thesis, Stanford University, 1966.
[3] T. W. Mullikin, *Multiple Scattering of Partially Polarized Light, SIAM-AMS Proceedings* (American Mathematical Society, Providence, R.I., 1969), Vol. I.
[4] K. D. Abhyankar and A. L. Fymat, Astron. Astrophys. **4**, 101 (1970).
[5] T. W. Schnatz and C. E. Siewert, J. Math. Phys. **11**, 2733 (1970).
[6] K. M. Case, Ann. Phys. (N.Y.) **9**, 1 (1960).
[7] N. Muskhelishvili, *Singular Integral Equations* (Noordhoff, Groningen, The Netherlands, 1953).
[8] N. P. Vekua, *Systems of Singular Integral Equations* (Noordhoff, Groningen, The Netherlands, 1967).
[9] K. M. Case and P. F. Zweifel, *Linear Transport Theory* (Addison-Wesley, Reading, Mass., 1967).
[10] G. F. Mandžavidze and B. V. Hvedelidze, Dokl. Akad. Nauk SSSR **123**,791 (1958).
[11] I. Kuščer, in *Developments in Transport Theory*, E. Inönü and P. F. Zweifel, Eds. (Academic, New York, 1967).

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 11, NUMBER 12 DECEMBER 1970

Diagrammatic Technique for Constructing Matrix Elements

ROBERT GILMORE
Physics Department and Center for Materials Science and Engineering, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139
and
Department of Mathematics and Physics, Alabama Agricultural and Mechanical University Normal, Alabama 35762

(Received 26 May 1970)

A diagrammatic technique is presented for computing the matrix elements of the generators of the unitary, orthogonal, symplectic, and symmetric groups (A_n, D_n, B_n, C_n, and S_n) within any of their unitary irreducible representations. Examples are worked out.

1. INTRODUCTION

The simple classical Lie groups, with algebras A_n, D_n, B_n, and C_n, have a great many properties in common. So many, in fact, that it is useful to search for similarities among them rather than differences between them.[1] The symmetric group S_n is also closely related to the classical groups, particularly $U(n)$, and may be treated analogously.[1]

In this paper we exploit the similarities among these five series of groups in order to find a convenient vehicle for describing their unitary irreducible representations (UIR's). We present a diagrammatic technique for constructing the matrix elements of the generators for each of these groups within any UIR.

2. THE COMMUTATION RELATIONS

The commutation relations among the generators of the classical and symmetric groups are

$$U(n): A_{n-1}[U^i{}_j, U^r{}_s] = U^i{}_s\delta^r{}_j - U^r{}_j\delta^i{}_s, \qquad ijrs = 1, 2, \cdots, n,$$

on the other hand, we observe from Eq. (31) that $z\mathbf{N}(z)$ must be bounded as z tends to infinity. The required behavior of $\mathbf{X}^{-1}(z)$ for large z can be deduced from Eqs. (51) and (52):

$$\mathbf{X}^{-1}(z) \sim z \begin{bmatrix} \frac{1}{z} + \cdots & \frac{a}{z} + \cdots \\ \frac{b}{z^2} + \cdots & 1 + \cdots \end{bmatrix}. \tag{53}$$

Considering now Eq. (36), we find that, in order for $z\mathbf{N}(z)$ to be bounded at infinity, we must take $\mathbf{P}(z) \equiv \mathbf{0}$; the behavior of $\mathbf{X}^{-1}(z)$ for large z, as given by Eq. (53), indicates that $z\mathbf{N}(z)$ will not be bounded at infinity unless we impose on $\mathbf{I}'(\mu)$ the constraint

$$\begin{bmatrix} 0 \\ 1 \end{bmatrix}^{\mathrm{T}} \int_0^1 \mathbf{\Gamma}(\mu)\mathbf{I}'(\mu)\, d\mu = 0, \tag{54}$$

where the superscript T denotes transpose. Recalling Eq. (26), we see that Eq. (54) can be satisfied for all appropriate $\mathbf{I}'(\mu)$ simply by choosing the correct discrete coefficient $A(\eta_0)$:

$$A(\eta_0)\begin{bmatrix} 0 \\ 1 \end{bmatrix}^{\mathrm{T}} \int_0^1 \mathbf{\Gamma}(\mu)\mathbf{\Phi}(\eta_0, \mu)\, d\mu = \begin{bmatrix} 0 \\ 1 \end{bmatrix}^{\mathrm{T}} \int_0^1 \mathbf{\Gamma}(\mu)\mathbf{I}(\mu)\, d\mu. \tag{55}$$

The desired expansion theorem (23) is thus established.

ACKNOWLEDGMENT

The authors wish to express their appreciation to J. F. Heyda of the General Electric Company, Philadelphia, who so kindly translated from Russian the required sections of the paper by Mandžavidze and Hvedelidze.

* This work was supported in part by National Science Foundation Grant GK 11935 and Joint Services Advisory Group Grant AFOSR-69-1779.

[1] S. Chandrasekhar, *Radiative Transfer* (Oxford U.P., London, 1950).
[2] G. M. Simmons, thesis, Stanford University, 1966.
[3] T. W. Mullikin, *Multiple Scattering of Partially Polarized Light, SIAM-AMS Proceedings* (American Mathematical Society, Providence, R.I., 1969), Vol. I.
[4] K. D. Abhyankar and A. L. Fymat, Astron. Astrophys. **4**, 101 (1970).
[5] T. W. Schnatz and C. E. Siewert, J. Math. Phys. **11**, 2733 (1970).
[6] K. M. Case, Ann. Phys. (N.Y.) **9**, 1 (1960).
[7] N. Muskhelishvili, *Singular Integral Equations* (Noordhoff, Groningen, The Netherlands, 1953).
[8] N. P. Vekua, *Systems of Singular Integral Equations* (Noordhoff, Groningen, The Netherlands, 1967).
[9] K. M. Case and P. F. Zweifel, *Linear Transport Theory* (Addison-Wesley, Reading, Mass., 1967).
[10] G. F. Mandžavidze and B. V. Hvedelidze, Dokl. Akad. Nauk SSSR **123**,791 (1958).
[11] I. Kuščer, in *Developments in Transport Theory*, E. Inönü and P. F. Zweifel, Eds. (Academic, New York, 1967).

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 11, NUMBER 12 DECEMBER 1970

Diagrammatic Technique for Constructing Matrix Elements

ROBERT GILMORE
Physics Department and Center for Materials Science and Engineering, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139
and
Department of Mathematics and Physics, Alabama Agricultural and Mechanical University Normal, Alabama 35762

(Received 26 May 1970)

A diagrammatic technique is presented for computing the matrix elements of the generators of the unitary, orthogonal, symplectic, and symmetric groups (A_n, D_n, B_n, C_n, and S_n) within any of their unitary irreducible representations. Examples are worked out.

1. INTRODUCTION

The simple classical Lie groups, with algebras A_n, D_n, B_n, and C_n, have a great many properties in common. So many, in fact, that it is useful to search for similarities among them rather than differences between them.[1] The symmetric group S_n is also closely related to the classical groups, particularly $U(n)$, and may be treated analogously.[1]

In this paper we exploit the similarities among these five series of groups in order to find a convenient vehicle for describing their unitary irreducible representations (UIR's). We present a diagrammatic technique for constructing the matrix elements of the generators for each of these groups within any UIR.

2. THE COMMUTATION RELATIONS

The commutation relations among the generators of the classical and symmetric groups are

$$U(n): A_{n-1}[U^i{}_j, U^r{}_s] = U^i{}_s\delta^r{}_j - U^r{}_j\delta^i{}_s, \qquad ijrs = 1, 2, \cdots, n,$$

$$
\begin{aligned}
&SO(2n): D_n[O_{ij}, O_{rs}] \\
&\quad = O_{is}\delta_{jr} + O_{jr}\delta_{is} - O_{ir}\delta_{js} - O_{js}\delta_{ir}, \\
&\qquad\qquad ijrs = 1, 2, \cdots, 2n, \\
&SO(2n+1): B_n[O_{ij}, O_{rs}] \\
&\quad = O_{is}\delta_{jr} + O_{jr}\delta_{is} - O_{ir}\delta_{js} - O_{js}\delta_{ir}, \\
&\qquad\qquad ijrs = 1, 2, \cdots, 2n+1, \qquad (2.1) \\
&Sp(2n): C_n[Z^i{}_j, Z^r{}_s] \\
&\quad = \operatorname{sgn}(jr)\{Z^i{}_s\delta^{-j}{}_{-r} + Z^{-j}{}_{-r}\delta^i{}_s + Z^i{}_{-r}\delta^{-j}{}_s \\
&\qquad + Z^{-j}{}_s\delta^r{}_{-i}\}, \quad ijrs = \pm 1, \pm 2, \cdots, \pm n, \\
&S_n: S_{ij}S_{rs}S_{ij}{}^{-1} \\
&\quad = S_{rs}\{1 - [\delta_{ir} + \delta_{is} + \delta_{jr} + \delta_{js}]\} \\
&\qquad + S_{js}\delta_{ir} + S_{jr}\delta_{is} + S_{is}\delta_{jr} + S_{ir}\delta_{js} \\
&\qquad + (\delta_{ir} + \delta_{is})(S_{ii} - S_{ij}) + (\delta_{jr} + \delta_{js})(S_{jj} - S_{ij}).
\end{aligned}
$$

These generators have the additional properties

$$
\begin{aligned}
&U^i{}_j = U^j{}_i{}^\dagger, \\
&O_{ij} = -O^\dagger{}_{ij} = -O_{ji}, \\
&Z^i{}_j = Z^j{}_i{}^\dagger = -\operatorname{sgn}(ij)Z^{-j}{}_{-i}, \\
&S_{ij} = S_{ij}{}^\dagger = S_{ji} = S^{-1}{}_{ij}.
\end{aligned} \qquad (2.2)
$$

These commutation relations (2.1) are used to construct the matrix elements of the generators within UIR's. The similarity among these commutation relations indicates the extensive similarities which exist in the construction of the UIR's.

3. STATE LABELING

If a nested sequence of canonically embedded subgroups

$$
\begin{aligned}
&G(n) \downarrow G(n-1) \downarrow \cdots \downarrow G(i+1) \downarrow G(i) \\
&\qquad \downarrow G(i-1) \downarrow \cdots \downarrow G(2) \downarrow G(1)
\end{aligned} \qquad (3.1)
$$

obeys the following two conditions,

(1) the UIR's of $G(i)$ are contained at most once in the branching of any UIR of $G(i+1)$ under the subgroup restriction $G(i+1) \downarrow G(i)$,

(2) the last nontrivial subgroup is Abelian,

then we reach the following conclusion:

Conclusion: Every basis in any UIR of $G(n)$ is contained in exactly one sequence of UIR's in the descending series of subgroups $G(n) \downarrow G(n-1) \downarrow \cdots \downarrow G(1)$. Such a specification of representation labels provides a mechanism for a complete labeling of bases within any UIR of $G(n)$.

The groups $U(n)$, $SO(n)$, and S_n obey conditions (1) and (2) above. Thus it is possible to label bases within any UIR by specifying a chain of partitions.[2] The partition at *level* i, $i = n, n-1, \cdots, 2, 1$, is a representation label for the canonically nested subgroup $G(i)$. The ith level partition contains i rows for the unitary, groups $[\frac{1}{2}i]$ rows for the orthogonal groups, and exactly i boxes for the symmetric groups.

The symplectic groups obey neither condition (1) nor (2) above. Nevertheless, it is known[2,3,4] that the bases in any UIR of $USp(2n)$ are labeled by a descending chain of partitions, $i = 2n, 2n-1, \cdots, 2, 1$. The partition at the ith level has $[\frac{1}{2}(i+1)]$ rows.

4. BASIC SHIFT OPERATORS

The group generators in general connect bases with different partition sequences. They may be viewed, therefore, as operators which alter the partition structure of a basis.

It is useful to introduce a set of basic shift operators defined by

$$
I_i^{\pm} = \sum_{r=1} I_i^{(r)\pm}. \qquad (4.1)
$$

Here $I_i^{(r)\pm}$ is an operator which has nonzero matrix elements only between a basis $|b\rangle$ and a basis $|b'\rangle = |\alpha_i^{(r)\pm} b\rangle$. Here $\alpha_i^{(r)\pm}$ creates (annihilates) one box in the rth row of the ith level partition. $I_i^{(r)+}$ and $I_i^{(r)-}$ are mutually adjoint. It is also useful to introduce a diagonal operator I_i^3, which is self-adjoint.

The explicit structure of the matrix element involving $I_i^{(r)\pm}$ and I_i^3 depends on i and the particular series of groups involved. These structures are given in Table I. The important point is that *all generators for the unitary, orthogonal, symplectic, and symmetric groups can easily be expressed in terms of the operators* $I_i^{\pm}$ *and* I_i^3. Moreover, the basic shift operators $I_i^{\pm,3}$ are *local* in the sense that *their matrix elements depend on the structure of only the* $(i-1)$*st*, *ith*, *and* $(i+1)$*st level partitions.*

5. AUGMENTED PARTITIONS

Partitions $\bar{\lambda}$ are convenient for the purposes of representation and state labeling, as well as discussing branching rules. Another quantity, the *augmented partition* $\bar{l}$, is more naturally involved in explicit computational procedures:

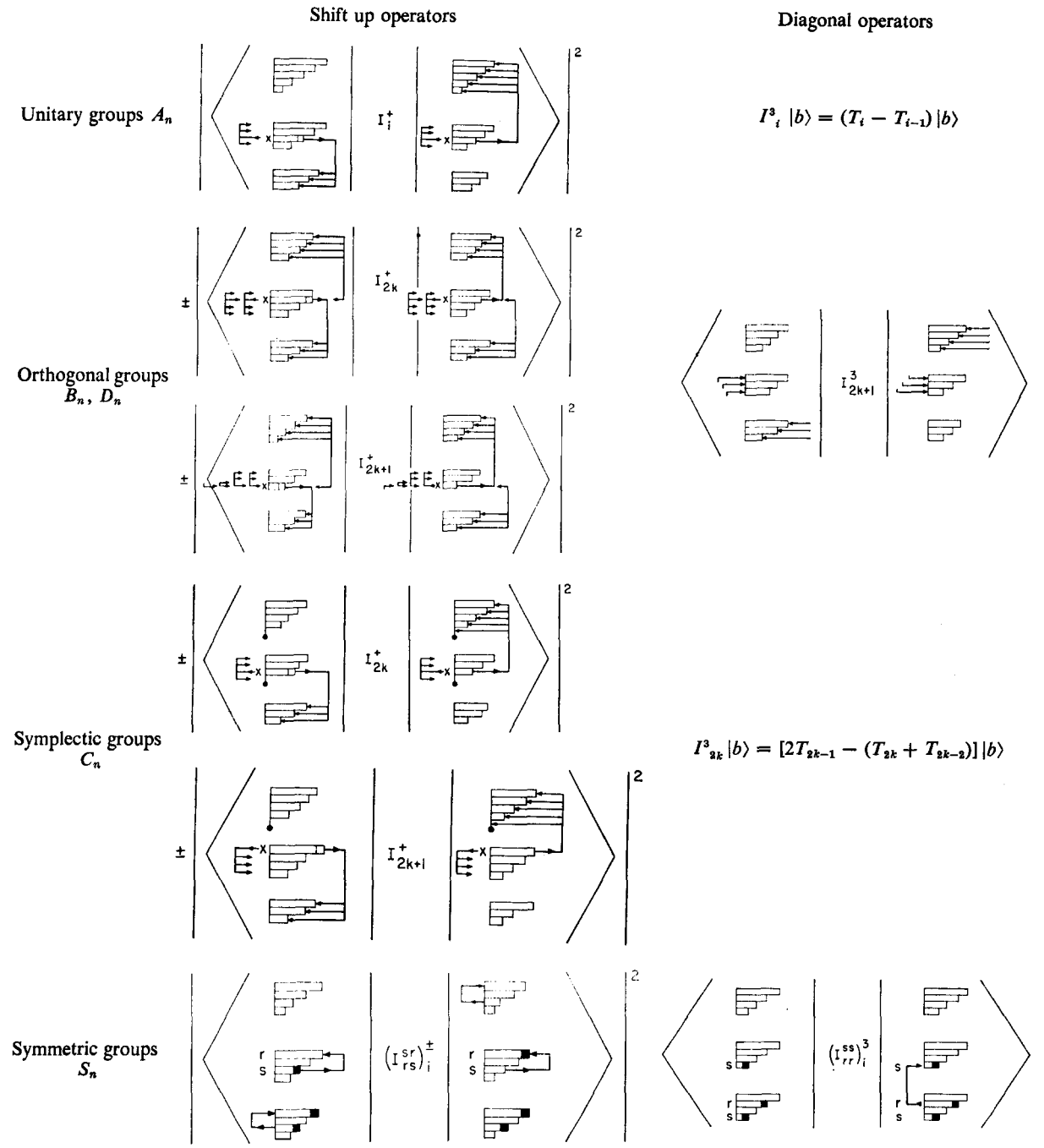
$$
\bar{l} = \bar{\lambda} + \bar{R}, \qquad (5.1)
$$

$$
\bar{R} = \tfrac{1}{2}\sum_{\bar{\alpha}>0} \bar{\alpha}. \qquad (5.2)
$$

Here $\bar{R}$ is, as usual, half the sum of the algebra's positive roots. $\bar{R}$ for the symmetric group is chosen to express the duality between S_{n+1} and $U(n+1)$:

$$
\begin{aligned}
&S_{n+1}: R_i = \tfrac{1}{2}n + 1 - i, \\
&A_n: R_i = \tfrac{1}{2}n + 1 - i, \quad i = 1, 2, \cdots, n+1, \\
&D_n: R_j = \tfrac{1}{2}(2n+0) - j, \\
&B_n: R_j = \tfrac{1}{2}(2n+1) - j, \\
&C_n: R_j = \tfrac{1}{2}(2n+2) - j, \quad j = 1, 2, \cdots, n.
\end{aligned} \qquad (5.3)
$$

TABLE I.



For the symplectic partitions at level $2n - 1$, $R_j = \frac{1}{2}(2n + 1) - j$.

In all subsequent work, when a partition is used to label a UIR or a basis, or to describe branching rules, the partitions $\bar{\lambda}$ are to be understood. But, when partitions are used in an explicitly indicated computational sense, the augmented partitions $\bar{l} = \bar{\lambda} + \bar{R}$ are to be understood.

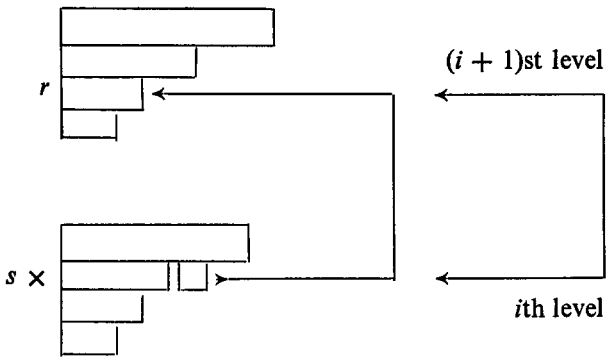
6. THE MATRIX ELEMENT CALCULUS

The matrix elements of the $I_i^{\pm,3}$ basic shift operators are presented in a diagrammatic manner in Table I. A dictionary for converting these diagrams to a more familiar analytic representation is given.

(1) A row s (marked with an x) in the ith partition undergoes a length change by having one more (□)

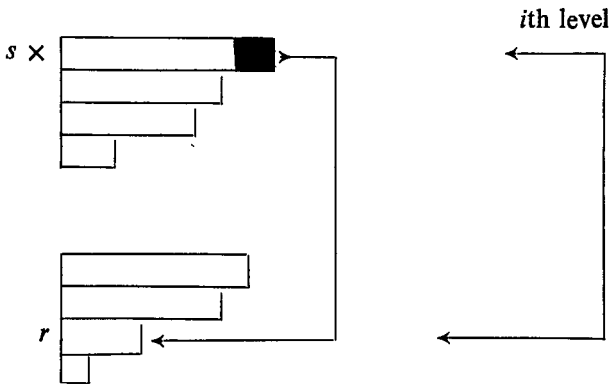
or one less (■) box. It may be connected to row r in:

(i) the $(i+1)$st level partition



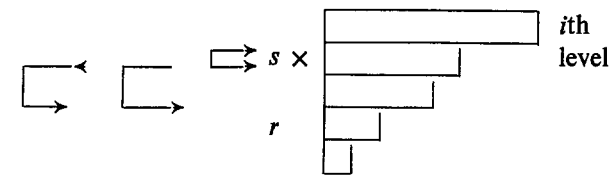
$$[l^r_{i+1} - (l_i{}^s + 1) - \tfrac{1}{2}] \qquad [l^r_{i+1} + (l_i{}^s + 1) - \tfrac{1}{2}];$$

(ii) the $(i-1)$st level partition



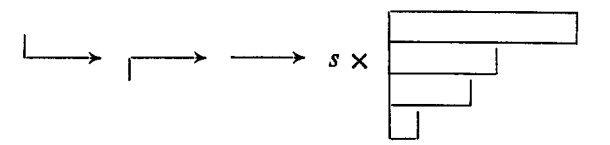
$$[l^r_{i-1} - (l_i{}^s - 1) + \tfrac{1}{2}] \qquad [l^r_{i-1} + (l_i{}^s - 1) + \tfrac{1}{2}];$$

(iii) the ith level partition



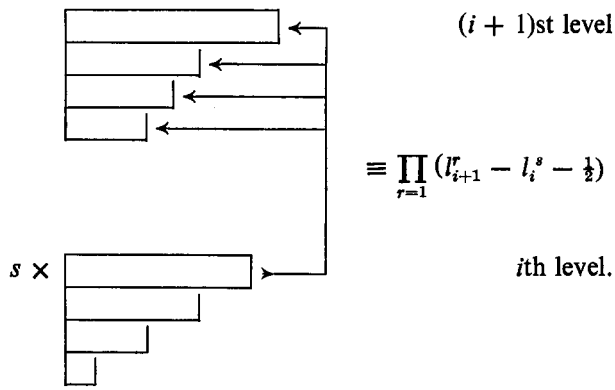
$$(l_i^r - l_i^s)\ (l_i^r + l_i^s)\ (l_i^s + l_i^s = 2l_i^s);$$

(iv) or may not be connected to another row at all



$$(l_i^s - \tfrac{1}{2})\ \ (l_i^s + \tfrac{1}{2}) \qquad l_i^s.$$

(2) Products of factors involving the same row are indicated



$$\equiv \prod_{r=1} (l^r_{i+1} - l_i{}^s - \tfrac{1}{2})$$

(3) All line segments appearing on the right of a partition sequence indicate factors appearing in the numerator; those on the left, in the denominator.

(4) $T_i = \sum_{r=1} l_i{}^r$, where these terms occur, they may equally well be defined by $T_i = \sum \lambda_i{}^r$. This is the total number of boxes in the ith level partition.

7. THE UNITARY GROUPS

The generators $U^i{}_j$, basic shift operators $I_i{}^{\pm,3}$, and Cartan generators $\bar{H}$ and $E_{\bar{\alpha}}$ for the unitary groups[5] are related by

$$\begin{aligned} U^i{}_i &= H_i \qquad = I_i{}^3, \\ U^i{}_{i+1} &= E_{e_i - e_{i+1}} = I_i{}^+, \\ U^{i+1}{}_i &= E_{e_{i+1} - e_i} = I_i{}^-. \end{aligned} \tag{7.1}$$

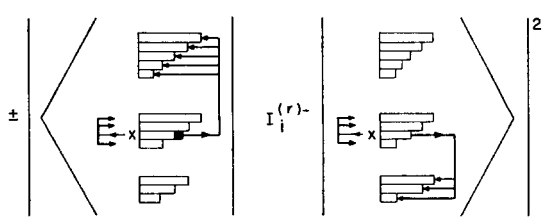
From Table I we can immediately read off the matrix element of $I_i{}^{r+}$; it is the positive square root of the quantity given there:

$$\langle \alpha_i{}^{(r)+} b |\, I_i{}^{(r)+} \,| b \rangle = \left(\pm \frac{\prod_{s=1}^{i-1} [l^s_{i-1} - (l_i^r + 1) + \tfrac{1}{2}]}{\prod_{l \neq r}^{i} [l_i^l - (l_i^r + 1)]} \frac{\prod_{t=1}^{i+1} (l^t_{i+1} - l_i^r - \tfrac{1}{2})}{\prod_{l \neq r}^{i} (l_i^l - l_i^r)} \right)^{\frac{1}{2}}. \tag{7.2}$$

The sign $\pm$ is chosen here and subsequently to make the product within the radical positive.

The matrix elements of $I_i{}^{(r)-}$ are simply related to those of $I_i{}^{(r)+}$ by transposition (Fig. 1). The relations

$$U^i{}_{i+2} = [U^i{}_{i+1},\ U^{i+1}{}_{i+2}] = [I^+{}_i,\ I_{i+1}{}^+] \tag{7.3}$$



$$\equiv \frac{\prod_{s=1}^{i+1} [l^s_{i+1} - (l^r_i - 1) - \tfrac{1}{2}]}{\prod_{l \neq r}^{i} [l^l_i - (l^r_i - 1)]} \otimes \frac{\prod_{t=1}^{i-1} (l^t_{i-1} - l^r_i + \tfrac{1}{2})}{\prod_{l \neq r}^{i} (l^l_i - l^l_i)}$$

FIG. 1. The matrix elements of $I^{(r)-}{}_i$ are related to those of $I^{(r)+}{}_i$ by transposition.

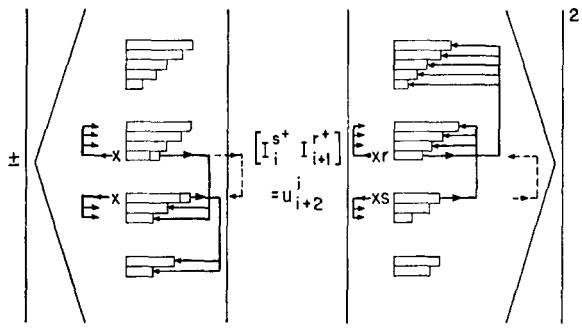


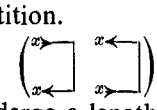
FIG. 2. The matrix elements for the commutator $[I^+_i, I^+_{i+1}]$ differ from the product of the matrix elements of I^+_i and I^+_{i+1} only by (not having) the factors shown as dashed lines. These are just the difference brackets connecting rows undergoing a length change in adjacent partitions. The matrix element itself

$$\langle \alpha^{s^+}{}_i \alpha^{r^+}{}_{i+1} b | \, [I^{(s)^+}{}_i, I^{(r)^+}{}_{i+1}] \, | b \rangle$$

is $\pm$ the square root of the product above, depending on whether $r \leq s$ or $r > s$.

lead easily to matrix elements for U^i_{i+2}, shown in Fig. 2. From the commutation relations (7.3) and Figs. 1 and 2 we easily see the algorithm for constructing the matrix elements of U^i_j by inspection:

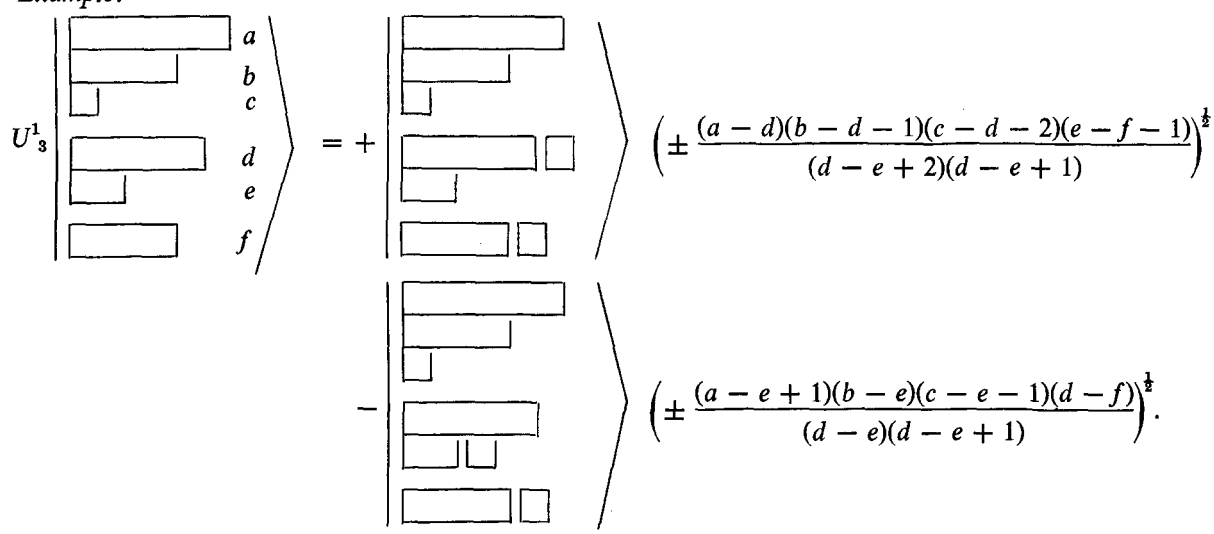
(i) Connect all rows undergoing a length change to rows in the appropriate contiguous partition.

(ii) Remove all difference factors  connecting two rows, both of which undergo a length change.

(iii) The over-all sign of the matrix element $\langle \alpha_i{}^{r_i+} \cdots \alpha_j{}^{r_j+} b | \, U^i_j \, | b \rangle$ is $(-)^n$. Here n is the number of times $r_{k+1} > r_k$, $k = i, i+1, \cdots, j-1$.

(iv) This algorithm holds for the commutators $[I^+_i, I^+_{i\pm1}]$ and $[I^-_i, I^-_{i\pm1}]$ in the orthogonal and symplectic groups as well.

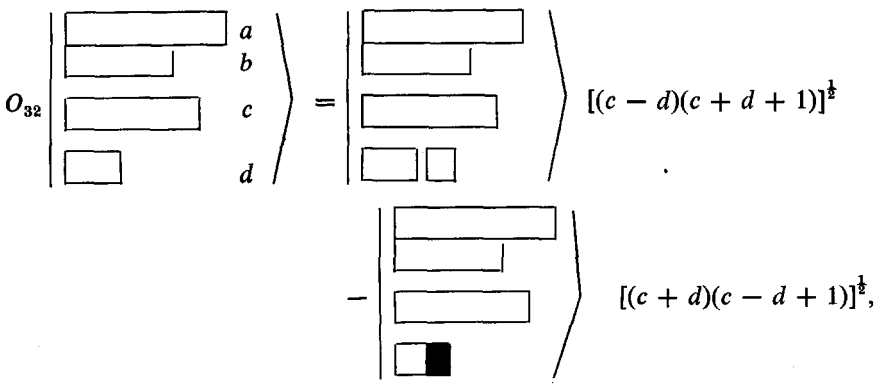
Example:

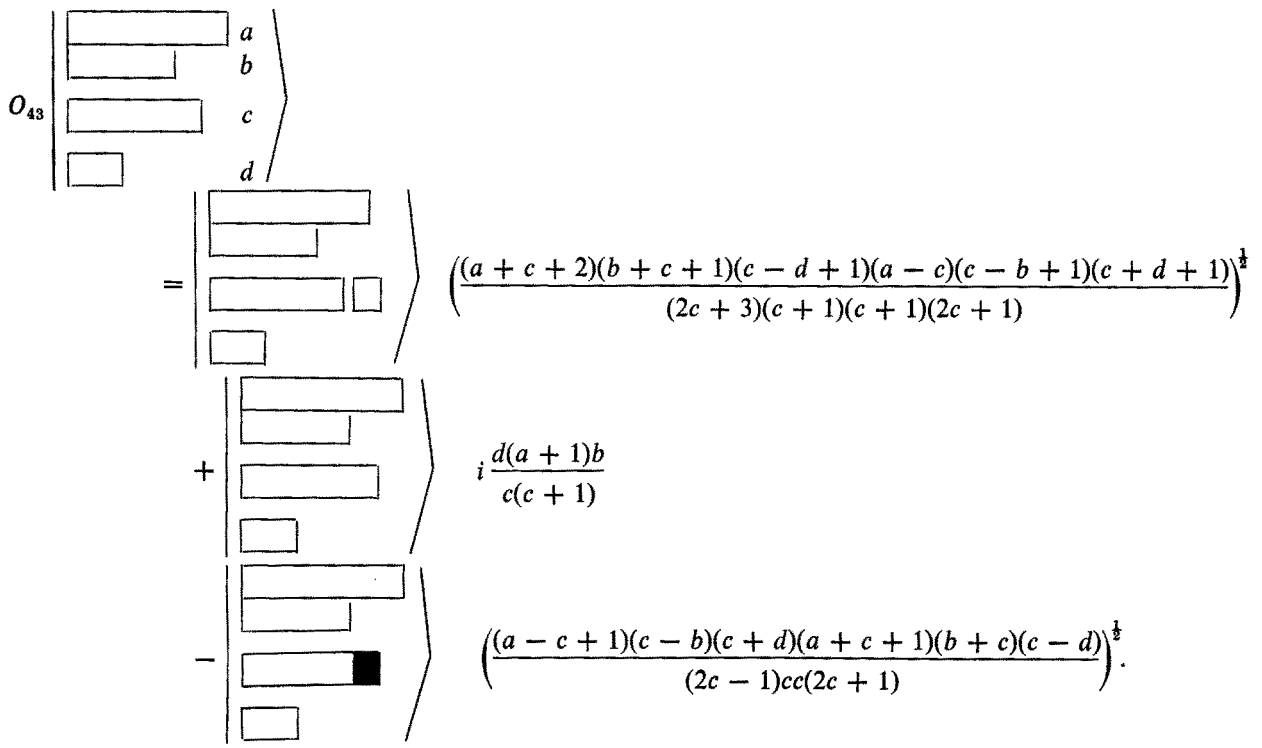


8. THE ORTHOGONAL GROUPS[6]

The relation between the anti-Hermitian infinitesimal generators O_{rs} and the basic shift operators is

$$O_{2k+2,2k+1} = I^+{}_{2k+1} + iI^3{}_{2k+1} - I^-{}_{2k+1}, \quad O_{2k+1,2k} = I^+{}_{2k} - I^-{}_{2k}. \tag{8.1}$$





9. THE SYMPLECTIC GROUP

It is useful to reindex the generators $Z^{i'}{}_{j'}$ of the symplectic group as follows:

$$\begin{aligned} Z^{i'}{}_{j'} \to Z^{i}{}_{j}, \quad i &= 2i', & i' &> 0, \\ &= 2(-i') - 1, & i' &< 0. \end{aligned}$$

The relation between $Z^{i}{}_{j}$; $\bar{H}$, $E_{\bar{\alpha}}$; and $I_i^{\pm\cdot 3}$ is

$$\begin{aligned} Z^{2k}{}_{2k} &= -Z^{2k-1}{}_{2k-1} = H_k = I^{(3)}{}_{2k}, \\ Z^{2k}{}_{2k-2} &= Z^{2k-2^{\dagger}}{}_{2k} = E_{e_k - e_{k-1}} \\ &= [I^{+}{}_{2k-2}, I^{+}{}_{2k-1}] + [I^{-}{}_{2k-2}, I^{-}{}_{2k-3}], \\ Z^{2k}{}_{2k-1} &= Z^{2k-1^{\dagger}}{}_{2k} = E_{2e_k} = I^{+}{}_{2k-1}. \end{aligned} \tag{9.1}$$

For purposes of computing matrix elements of the basic shift operators $I_i^{\pm}$, it is useful to assume the partitions in the $(i-1)$st, ith, and $(i+1)$st level have $[\frac{1}{2}i]$, $[\frac{1}{2}i]+1$, and $[\frac{1}{2}i]+2$ rows, respectively. Some of these rows are necessarily of length 0, as indicated (by $\downarrow$) in Table I. The computation of the commutators in Eq. (9.1) is formally identical to the computation shown in Fig. 2. The results are shown in Figs. 3 and 4.

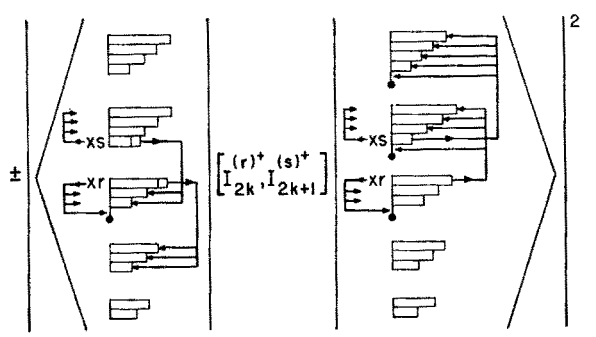


FIG. 3. The value of $\langle \alpha^{r^+}{}_{2k} \alpha^{s^+}{}_{2k+1} b | [I^{r^+}{}_{2k}, I^{s^+}{}_{2k+1}] | b \rangle$ is $\pm$ the square root of the absolute magnitude of the product of factors shown above. The $\pm$ is chosen depending on whether $s \leq r$ or $s > r$.

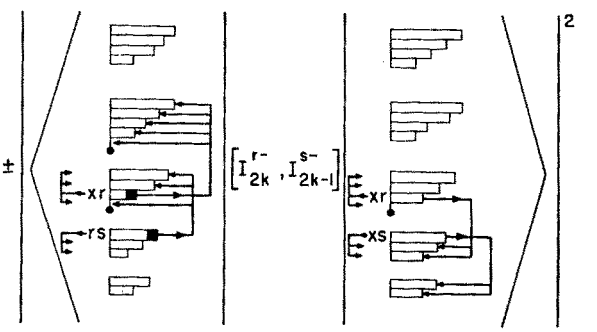
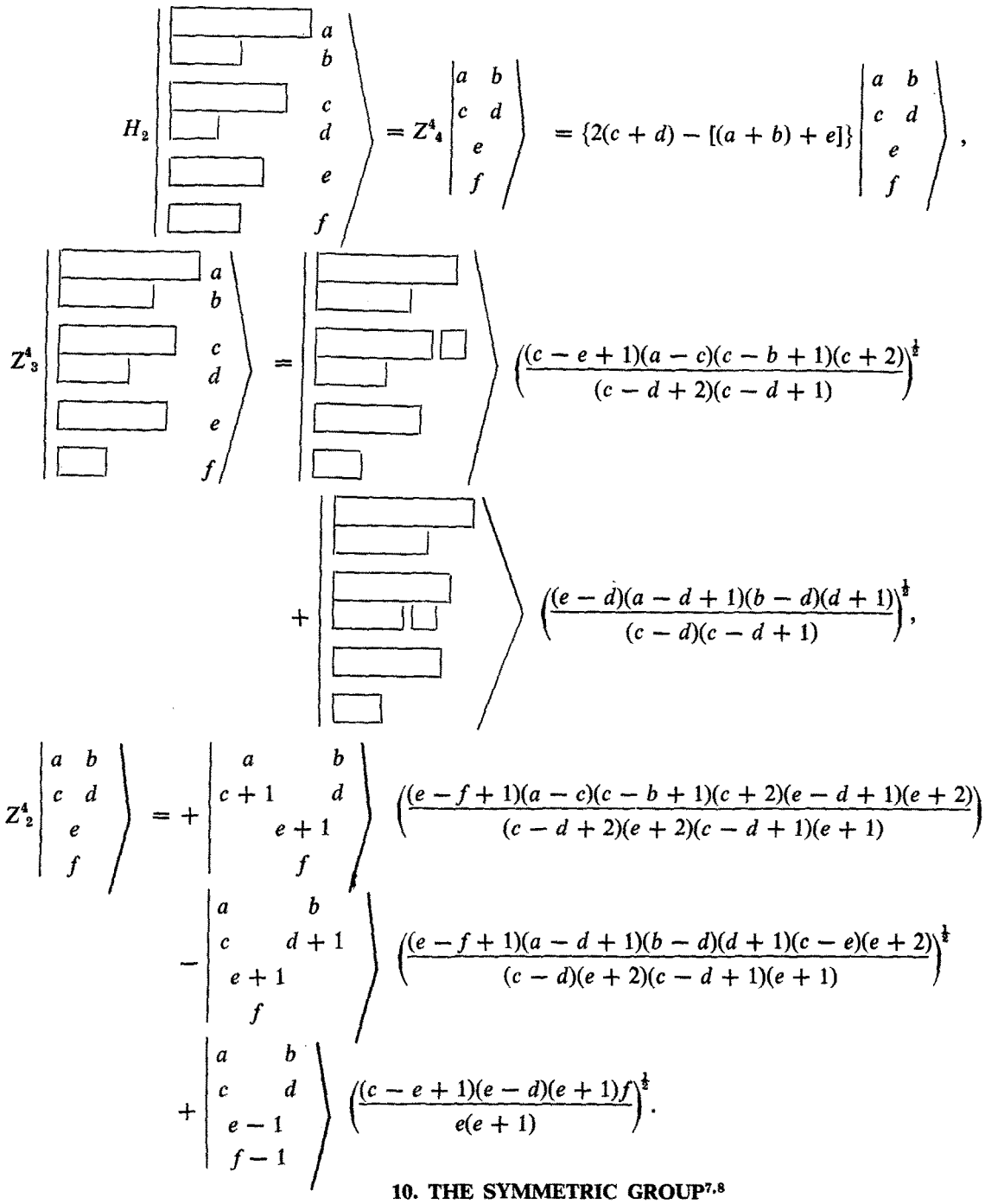


FIG. 4. The value of $\langle \alpha^{r^-}{}_{2k} \alpha^{s^-}{}_{2k-1} b | [I^{r^-}{}_{2k}, I^{s^-}{}_{2k-1}] | b \rangle$ is $\pm$ the square root of the absolute magnitude of the product of factors shown above. The $\pm$ sign is chosen depending on whether $s \geq r$ or $s < r$.

Example: For the representation $\langle a, b\rangle_4$ of $USp(4)$:



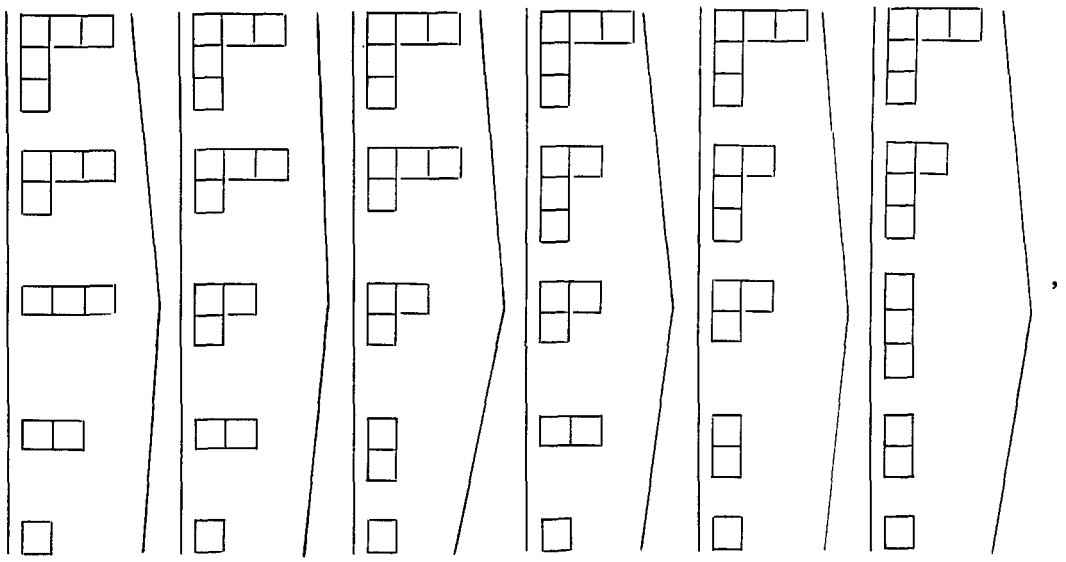
10. THE SYMMETRIC GROUP[7,8]

The exchange operator $P_{i,i+1}$ connects bases in which only the ith level partitions may differ. They can differ at most by having one additional node in the sth row, one less in the rth. The operator $P_{i,i+1}$ consists of a nondiagonal part $(I^{sr}{}_{rs})^{\pm}{}_i$ and a diagonal part $(I^{ss}{}_{rr})^3{}_i$:

$$P_{i,i+1} = \sum_{r\neq s}(I^{sr}{}_{rs})^{\pm}{}_i + \sum(I^{ss}{}_{rr})^3{}_i, \tag{10.1}$$

whose values are indicated in Table I. We observe that the diagonal matrix elements are of the form $\pm(\text{integer})^{-1}$, depending on the order of box removal. In the limiting case of removal from the same row (column), the limiting value $+1$ (-1) is reached.

Example: With respect to the bases



the matrix representative for P_{34} is found immediately by inspection

$$P_{34} = \begin{bmatrix} -\frac{1}{3} & \frac{2}{3}\sqrt{2} & & & & \\ \frac{2}{3}\sqrt{2} & +\frac{1}{3} & & & & \\ & & +1 & & & \\ & & & -1 & & \\ & & & & -\frac{1}{3} & \frac{2}{3}\sqrt{2} \\ & & & & \frac{2}{3}\sqrt{2} & +\frac{1}{3} \end{bmatrix}.$$

11. COMMENTS

The matrix elements of the basic shift operators $I_i^{\pm,3}$ show a great deal of symmetry in their component difference (and sum) factors involving contiguous partitions. Thus, if one factor occurs involving a row undergoing a length change and a row in a contiguous partition, then additional factors involving the remaining rows of the contiguous partition also occur.

These products vanish whenever addition or removal of a box in the ith partition renders it impossible either to obtain the new ith partition from the branching of the $(i+1)$st or to obtain the $(i-1)$st from the branching of the new ith level partition. The products become undefined when the box added to or removed from the ith level partition creates a nonstandard (i.e., unallowed) partition.

Partitions[9] describing representations of D_n may have $l^n{}_{2n} = \lambda^n{}_{2n} < 0$. The considerations of the two preceding paragraphs then lead, in a very natural way, to the existence of sum as well as difference products for the orthogonal groups. The absence of negative length rows for the unitary and symplectic groups corresponds to the presence of only difference factors in their matrix elements.

The local nature of the $I^{\pm,3}{}_i$ shift operators (cf. Sec. 4) indicates that their matrix elements depend *only* on the structure of the $(i-1)$st, ith, and $(i+1)$st level partitions. The matrix element is unchanged if we consider a different representation of $G(n)$ $(n > i+1)$ with the same $i-1$, i, $i+1$ partition structure, or even if we go to a different group within the same chain, provided the i, $i \pm 1$ level structure of the bases is unaltered.

12. CONCLUSION

A simple diagrammatic technique has been presented for the construction of the generators of the unitary, orthogonal, symplectic, and symmetric groups within any of their UIR's. The main results are contained in Table I and Eqs. (7.1), (8.1), (9.1), and (10.1). Examples have been given for each series of groups.

ACKNOWLEDGMENTS

The author wishes to thank Professor F. T. Arecchi and Professor D. M. Kim, at whose urging this work was begun.

[1] H. Weyl, *Classical Groups* (Princeton U.P., Princeton, N.J., 1946); G. Racah, Ergeb. Exakt. Naturw. **37**, 28 (1965).
[2] D. P. Zhelobenko, Russ. Math. Surv. **17**, 1 (1962); A. U. Klimyk, Transl. Am. Math. Soc. (2) **76**, 75 (1968).
[3] G. C. Hegerfeldt, J. Math. Phys. **8**, 731 (1967).
[4] R. Gilmore, J. Math. Phys. **11**, 513 (1970).
[5] I. M. Gel'fand and M. L. Cetlin, Dokl. Akad. Nauk SSSR **71** 825 (1950); G. E. Baird and L. C. Biedenharn, J. Math. Phys. **4**, 1449 (1963). L. C. Biedenharn and J. D. Louck, Commun. Math. Phys. **8**, 89 (1968), have developed a similar pattern calculus for the description of a class of isoscalar factors among the unitary groups.
[6] I. M. Gel'fand and M. L. Cetlin, Dokl. Akad. Nauk SSSR **71**, 1017 (1950); U. Ottoson, Commun. Math. Phys. **8**, 228 (1968).
[7] T. Yamanouchi, Proc. Phys. Math. Soc. (Japan) **19**, 436 (1937).
[8] M. Hamermesh, *Group Theory* (Addison-Wesley, Reading, Mass., 1962). J. R. Gabriel, Proc. Cambridge Phil. Soc. **57**, 330 (1961).
[9] R. Gilmore, J. Math. Phys. **11**, 1853 (1970).

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 11, NUMBER 12 DECEMBER 1970

Group Representations and Geometry

G. DE B. ROBINSON
Department of Mathematics, University of Toronto, Toronto, Ontario

(Received 29 May 1970)

We confine our attention here to simply reducible groups and show how six of the seven points of a finite projective plane $PG(2, 2)$ constitute a "Pasch" diagram representing a $6j$ symbol. The class of all equivalent symbols may thus be represented by the seventh point in the plane. Analyzing the symmetries of such configurations, we derive two theorems, the first of which is the geometrical analog of Regge's result while the second gives the geometrical analog of the multiplication of two $6j$ symbols. In these terms the analogs of Eqs. (I1), (I2), and (I3) of Appendix I of *Irreducible Tensorial Sets* by Fano and Racah are very simply expressed. In particular, the Biedenharn identity (I3) becomes a vector equation (mod 2), and the relation with Desargues' theorem is clarified. The advantage of this geometrical model is that the structure alone survives and all summations and complicated coefficients disappear.

1. INTRODUCTION

Ever since Weyl's *Gruppentheorie und Quanten Mechanik*[1] showed the way, physicists have become increasingly interested in group representations. If we denote two irreducible representations of a group G by j_i and j_k, interest focuses on the reduction of the Kronecker product

$$j_i j_k = \sum c_{ikl} j_l . \tag{1.1}$$

If we assume that the characters of the j's are real and further that the coefficients c_{ikl} are 0 or 1, the group G is said to be *simply reducible*. We shall confine our attention to this case though much of what we say can be generalized.[2] Under these conditions c_{ikl} is invariant under any permutation of the suffixes.

The purpose of the present paper is to amplify and further develop the relationship of this representation theory to projective geometry as suggested by Fano and Racah[3] in their Appendix I. We begin by introducing the familiar axioms[4] and interpret them in terms of the representation theory.

Closely related to this discussion is the analysis of the significant vector diagrams.[5] As a result of (1.1), three vectors concur in every vertex of such a diagram. If we dualize in the plane, we obtain a figure in which three points (and only three) lie on every line, so that the dual figure can be embedded in a finite projective geometry $PG(n, 2)$.[6] In such a context the formulas (I1), (I2), and (I3) of Ref. 3 are of particular interest.

As a result of these considerations two theorems emerge. Theorem 1 yields the analog of Regge's determination[7] of the symmetries of the $6j$ symbol, while Theorem 2 provides the geometrical interpretation of the *multiplication* of such symbols, based on their representation by exponential functions as given at the end of Sec. 4.

2. FOUNDATIONS

Let us begin by denoting $j_i \neq I$ by a *point* in our geometry and assume:

(i) There are at least two distinct points j_i and j_k;

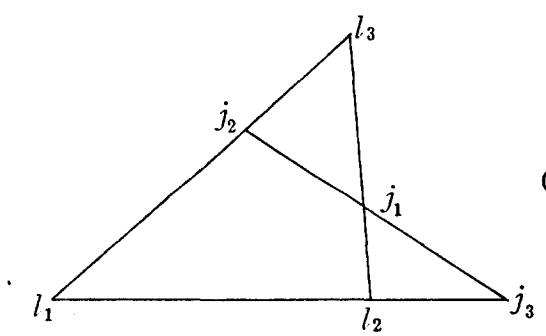
(ii) two points j_i and j_k determine one and only one line $j_i j_k = j_k j_i$;

(iii) if j_i and j_k are distinct, there exists at least one point $j_l \neq I$ such that $c_{ikl} \neq 0$ in (1.1).

It follows from the symmetry of the c_{ikl} that the same line is determined by any two of its points. If we assume that

(iv) there is at least one point *not* on the line $j_i j_k$, then, in order to define a plane and prove the intersection of any two coplanar lines, it is sufficient to assume that:

(v) (Pasch) If l_1, l_2, l_3 are three noncollinear points and j_1 is a point on $l_2 l_3$, j_2 a point on $l_1 l_3$, then there is a point j_3 on $l_1 l_2$ such that j_1, j_2, and j_3 are collinear.



(2.1)

This assumption corresponds in group theoretic terms to the nonvanishing of the $6j$ symbol[8]

$$\begin{Bmatrix} j_i & j_2 & j_3 \\ l_1 & l_2 & l_3 \end{Bmatrix}. \tag{2.2}$$

If we assume further that

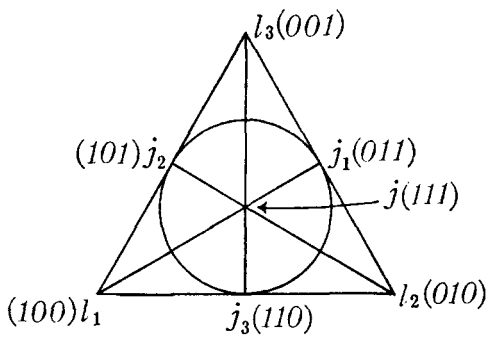
(vi) there is at least one point *not* on the plane $l_1\, l_2\, l_3$, we can *prove* Desargues' theorem in space.[9]

In general, it is necessary to make a further assumption to obtain Pappus' theorem or the fundamental theorem, but, if we assume there are only a finite number of points on a line, then these theorems follow from Desargues' theorem.

We now make this assumption and limit attention to $PG(n, 2)$, in which there are $1 + 2 = 3$ points on every line and

$$1 + 2 + 2^2 + \cdots + 2^n$$

points in $PG(n, 2)$. We can redraw figure (2.1) as follows[10]:



(2.3)

with a coordinate system introduced as indicated. The equation of the line $j_1\, j_2\, j_3$ is

$$x_1 + x_2 + x_3 = 0 \pmod 2,$$

and we shall denote the (mod 2) analog of the ordinary $6j$ symbol (2.2) by

$$\left\| \begin{matrix} j_1 & j_2 & j_3 \\ l_1 & l_2 & l_3 \end{matrix} \right\|. \tag{2.4}$$

It should be observed that the *vectorial* sum (mod 2) of any three collinear points in figure (2.3) is the zero vector or, alternatively, such a sum of two points is the third point on the line. We shall write

$$j_i + j_k = j_l \pmod 2,$$

and this relationship enables us to define *projection* in $PG(n, 2)$.

3. GEOMETRICAL SYMMETRIES

We turn now to a brief survey of the symmetry properties of the $6j$ symbol (2.4) as represented geometrically in figure (2.3). Since a linear transformation or a "projectivity" in $PG(2, 2)$ transforms points into points and lines into lines, it is only necessary to write down all such transformations which leave $j(1\,1\,1)$ in figure (2.3) invariant. To begin with,

$$I \sim \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}, \quad (j_1\, j_2)(l_1\, l_2) \sim \begin{pmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix},$$

$$(j_2\, j_3)(l_2\, l_3) \sim \begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{pmatrix},$$

$$(j_1\, j_3)(l_1\, l_3) \sim \begin{pmatrix} 0 & 0 & 1 \\ 0 & 1 & 0 \\ 1 & 0 & 0 \end{pmatrix},$$

$$(j_1\, j_2\, j_3)(l_1\, l_2\, l_3) \sim \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \end{pmatrix},$$

$$(j_1\, j_3\, j_2)(l_1\, l_3\, l_2) \sim \begin{pmatrix} 0 & 0 & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{pmatrix}$$

yield a permutation representation of S_3 corresponding to permuting the columns of the $6j$ symbol. Further symmetries interchange the $j_i\, l_i$ by pairs:

$$(j_1\, l_1)(j_2\, l_2) \sim \begin{pmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 1 & 1 & 1 \end{pmatrix},$$

$$(j_2\, l_2)(j_3\, l_3) \sim \begin{pmatrix} 1 & 1 & 1 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{pmatrix},$$

$$(j_1\, l_1)(j_3\, l_3) \sim \begin{pmatrix} 0 & 0 & 1 \\ 1 & 1 & 1 \\ 1 & 0 & 0 \end{pmatrix},$$

so that there are 24 symmetries in all.

We have, in fact, a modular representation of S_4 which can be reduced by transforming by

$$\begin{pmatrix} 1 & 0 & 0 \\ 1 & 1 & 0 \\ 1 & 0 & 1 \end{pmatrix}$$

to yield

$$(j_1\, j_2)(l_1\, l_2) \sim \left(\begin{array}{c|cc} 1 & 1 & 0 \\ \hline 0 & 1 & 0 \\ 0 & 1 & 1 \end{array}\right),$$

$$(j_2\, j_3)(l_2\, l_3) \sim \left(\begin{array}{c|cc} 1 & 0 & 0 \\ \hline 0 & 0 & 1 \\ 0 & 1 & 0 \end{array}\right),$$

$$(j_1\, j_3)(l_1\, l_3) \sim \left(\begin{array}{c|cc} 1 & 0 & 1 \\ \hline 0 & 1 & 1 \\ 0 & 0 & 1 \end{array}\right),$$

$$(j_1 l_1)(j_2 l_2) \sim \left(\begin{array}{c|cc} 1 & 1 & 0 \\ \hline 0 & 1 & 0 \\ 0 & 0 & 1 \end{array}\right),$$

$$(j_2 l_2)(j_3 l_3) \sim \left(\begin{array}{c|cc} 1 & 1 & 1 \\ \hline 0 & 1 & 0 \\ 0 & 0 & 1 \end{array}\right),$$

$$(j_1 l_1)(j_3 l_3) \sim \left(\begin{array}{c|cc} 1 & 0 & 1 \\ \hline 0 & 1 & 0 \\ 0 & 0 & 1 \end{array}\right).$$

It should be noted that this representation is reducible, but not completely reducible, with the expected[11] irreducible modular components of S_4.

4. FUNCTIONAL SYMMETRIES

In order to discuss the further functional symmetries of the $6j$ symbol, Regge introduced[12] the following transformation of the ordinary $6j$ symbol:

$$\begin{Bmatrix} j_1 & j_2 & j_3 \\ l_1 & l_2 & l_3 \end{Bmatrix} = \begin{Bmatrix} j_1, & j_2 + x, & j_3 - x \\ l_1, & l_2 - x, & l_3 + x \end{Bmatrix}, \quad (4.1)$$

where $x = \frac{1}{2}(-j_2 + j_3 + l_2 - l_3)$. This allows the consideration of degenerate $6j$ symbols as well. The analog of (4.1) in our geometrical interpretation is the following:

Theorem 1:

$$\begin{Vmatrix} j_1 & j_2 & j_3 \\ l_1 & l_2 & l_3 \end{Vmatrix} = \begin{Vmatrix} j_1 + y + z, & j_2 + x + z, & j_3 + x + y \\ l_1 + y + z, & l_2 + x + z, & l_3 + x + y \end{Vmatrix} \pmod 2, \quad (4.2)$$

for all points x, y, z in $PG(n, 2)$.

As in (4.1), row and column sums (mod 2) are not affected by these changes, and it is of interest to note that every linear transformation of Sec. 3 can be expressed in these terms. In particular,

$$\begin{aligned} (j_1 j_2)(l_1 l_2) &\sim x = y = 0, \quad z = j_3, \\ (j_1 j_3)(l_1 l_3) &\sim x = z = 0, \quad y = j_2, \\ (j_2 j_3)(l_2 l_3) &\sim y = z = 0, \quad x = j_1, \\ (j_1 l_1)(j_2 l_2) &\sim x = y = 0, \quad z = j, \\ (j_1 l_1)(j_3 l_3) &\sim x = z = 0, \quad y = j, \\ (j_2 l_2)(j_3 l_3) &\sim y = z = 0, \quad x = j. \end{aligned}$$

As in the application of (4.1), we may set $x = y = 0$, $z = j_1$, to yield

$$\begin{Vmatrix} j_1 & j_2 & j_3 \\ l_1 & l_2 & l_3 \end{Vmatrix} = \begin{Vmatrix} 0 & j_3 & j_3 \\ j & l_3 & l_3 \end{Vmatrix} = \begin{Vmatrix} j_3 & l_3 & j \\ l_3 & j_3 & 0 \end{Vmatrix} \pmod 2.$$

This is in effect a *projection* of the configuration on the line $j_3 l_3$. We may similarly project on to $j_1 l_1$ and $j_2 l_2$ to yield

$$\begin{Vmatrix} j_1 & j_2 & j_3 \\ l_1 & l_2 & l_3 \end{Vmatrix} = \begin{Vmatrix} j & j_1 & l_1 \\ 0 & l_1 & j_1 \end{Vmatrix} = \begin{Vmatrix} j_2 & j & l_2 \\ l_2 & 0 & j_2 \end{Vmatrix} \pmod 2,$$

for $x = j_2$, $y = z = 0$ and $x = z = 0$, $y = j_3$, respectively. Each of these $6j$ symbols is subject to 12 symmetries as described above.

Finally, we may project the $6j$ symbol on the point j by setting $x = j_1$, $y = j_2$, and $z = j_3$ so that

$$\begin{Vmatrix} j_1 & j_2 & j_3 \\ l_1 & l_2 & l_3 \end{Vmatrix} = \begin{Vmatrix} 0 & 0 & 0 \\ j & j & j \end{Vmatrix} \pmod 2, \quad (4.3)$$

which is subject to four symmetries. It is not difficult to relate these degenerate cases to the application of Regge's transformation in the ordinary theory.

We conclude this section by observing that all symmetries of the $6j$ symbol (2.4) leave the remaining point j in the plane $PG(2, 2)$ invariant, and so *it is natural to designate the class of equivalent* $6j$ *symbols* (2.4) *by this point.* In order to make explicit the vectorial addition of points, we write this relation exponentially,

$$\begin{Vmatrix} j_1 & j_2 & j_3 \\ l_1 & l_2 & l_3 \end{Vmatrix} = e^j \pmod 2, \quad (4.4)$$

though the choice of the base $e \neq 0$ does not seem important. Note that a set of points *not* satisfying axiom (v) could only be interpreted by setting $e = 0$, as we do in (5.2).

5. DESARGUES' THEOREM

In order to see the full significance of the geometry $PG(n, 2)$ which we have described, it is necessary to produce an interpretation for the *multiplication* of $6j$ symbols. This is provided by the following:

Theorem 2:

$$\begin{Vmatrix} j_1 & j_2 & j_3 \\ l_1 & l_2 & l_3 \end{Vmatrix} \begin{Vmatrix} j_1' & j_2' & j_3' \\ l_1' & l_2' & l_3' \end{Vmatrix} = \begin{Vmatrix} j_1 + j_1', & j_2 + j_2', & j_3 + j_3' \\ l_1 + l_1', & l_2 + l_2', & l_3 + l_3' \end{Vmatrix} \pmod 2. \quad (5.1)$$

Proof: Since the top row sum is zero (mod 2) and the lower row sum, as well as each column sum, in (2.4) is j, it is clear that we are only rewriting the relation

$$j + j' = (j + j') \pmod 2$$

in exponential form.

Corresponding[13] to (I1) of Ref. 3,

$$\begin{Vmatrix} j_1 & j_2 & j_3 \\ l_1 & l_2 & l_3 \end{Vmatrix} \begin{Vmatrix} j_1 & j_2 & j_3 \\ l_1 & l_2 & l_3' \end{Vmatrix} = \begin{cases} l^{2j} = 1, & \text{for } l_3' = l_3 \\ 0, & \text{for } l_3' \neq l_3 \end{cases} \pmod 2. \quad (5.2)$$

Combining Eqs. (5.1) and (5.2), we recognize the further degenerate symbol

$$\begin{Vmatrix} j_1 & j_2 & j_3 \\ l_1 & l_2 & l_3 \end{Vmatrix} \begin{Vmatrix} j_1 & j_2 & j_3 \\ l_1 & l_2 & l_3 \end{Vmatrix} = \begin{Vmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \end{Vmatrix} = 1 \pmod 2$$

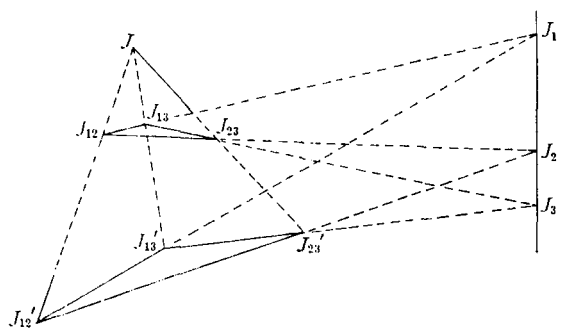
and its equivalents by Theorem 1.

Corresponding to (I2), we have

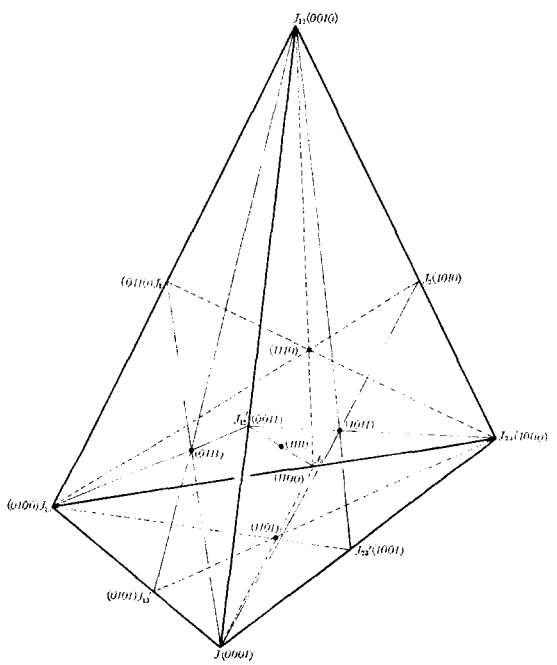
$$\begin{Vmatrix} l_1 & j_3 & l_2 \\ j_2 & j & j_1 \end{Vmatrix} \begin{Vmatrix} j & j_1 & l_1 \\ l_3 & j_2 & l_2 \end{Vmatrix} = \begin{Vmatrix} j_1 & j_2 & j_3 \\ l_1 & l_2 & l_3 \end{Vmatrix} = \begin{Vmatrix} j_1 & l_1 & j \\ l_2 & j_2 & j_3 \end{Vmatrix} \begin{Vmatrix} j_1 & l_1 & j \\ j_2 & l_2 & l_3 \end{Vmatrix} \pmod 2 \quad (5.3)$$

without summing over j.

Referring to our geometrical axioms in Sec. 2, we stated that it was possible to *prove* Desargues' theorem in $PG(3, 2)$. We adopt a slightly different notation using double suffixes to emphasize the collinearity of points.



(5.4)



(5.5)

If we coordinatize figure (5.4) as in the accompanying figure (5.5), the coordinates of the centroids of the plane faces of the tetrahedron incident in J are

$$(0111), \quad (1011), \quad (1101),$$

while that of the face opposite J is (1110). The "plane" containing $J_1, J_2, J_3, J'_{23}, J'_{13}, J'_{12}$ has equation

$$x_1 + x_2 + x_3 + x_4 = 0 \pmod 2$$

with centroid (1111), so that

$$(1111) = (0111) + (1011) + (1101) + (1110) \pmod 2. \quad (5.6)$$

Adding (1110) to both sides, we have

$$(1110) + (1111) = (0111) + (1011) + (1101). \quad (5.7)$$

In terms of $6j$ symbols, (5.7) yields Desargues' theorem

$$\begin{Vmatrix} J_1 & J_2 & J_3 \\ J_{23} & J_{13} & J_{12} \end{Vmatrix} \begin{Vmatrix} J_1 & J_2 & J_3 \\ J'_{23} & J'_{13} & J'_{12} \end{Vmatrix} = \begin{Vmatrix} J_1 & J_{13} & J_{12} \\ J & J'_{12} & J'_{13} \end{Vmatrix} \begin{Vmatrix} J_{23} & J_2 & J_{12} \\ J'_{12} & J & J'_{23} \end{Vmatrix} \begin{Vmatrix} J_{23} & J_{13} & J_3 \\ J'_{13} & J'_{23} & J \end{Vmatrix} \pmod 2. \quad (5.8)$$

Clearly the choice of a coordinate system in figure (5.5) is unimportant, but it should be observed that, while a $6j$ symbol is always represented by a unique point P in its plane, P also represents the class of all $6j$ symbols lying in different planes through P and equivalent under (4.2).

If we apply Theorem 2 to reduce (5.8), we obtain

$$\begin{Vmatrix} 0 & 0 & 0 \\ J & J & J \end{Vmatrix} = \begin{Vmatrix} J_1 & J_2 & J_3 \\ (0111) & (1011) & (1101) \end{Vmatrix} \pmod 2,$$

which is (4.3) in the plane $x_1 + x_2 + x_3 = 0 \pmod 2$ of figure (5.5). Desargues' theorem is degenerate in $PG(2, 2)$, but it may be recognized in (4.3) by setting $j_1 = J_1 = J_{23}$, $j_2 = J_2 = J_{13}$, $j_3 = J_3 = J_{12}$, $l_1 = (0111) = J'_{23}$, $l_2 = (1011) = J'_{13}$, $l_3 = (1101) = J'_{12}$, and $j = J$.

Since the fundamental theorem of projective geometry follows from Desargues' theorem in $PG(n, 2)$ for $n > 2$, we conclude that any relation among $6j$ symbols (2.4) must be provable in $PG(n, 2)$. Furthermore, since any $3mj$ symbol is expressible in terms of $6j$ symbols, any relation between ordinary $3mj$ symbols is derivable from (I1), (I2), and (I3) as stated by Fano and Racah.

6. $9j$ AND $12j$ SYMBOLS

We do not propose to pursue this symbolical analysis further, but it is perhaps worth commenting briefly on the $9j$ and $12j$ symbols.

Without giving details, the reader will find it easy to embed the appropriate diagram for the $9j$ symbol[14] in $PG(3, 2)$. The $12j$ symbol[15] of the "second kind" can also be embedded in $PG(3, 2)$, but it is interesting to observe that the interchange which yields the $12j$ symbol of the "first kind" requires $PG(4, 2)$. In order to see the reason for this, we coordinatize[16] thus:

$$\begin{Vmatrix} j_1 & j_2 & j_3 & j_4 & \\ & l_1 & l_2 & l_3 & l_4 \\ k_1 & k_2 & k_3 & k_4 & \end{Vmatrix}$$

$$= \begin{Vmatrix} (10000) & (01000) & (00100) & (00010) & \\ & (11000) & (01100) & (00110) & (10011) \\ (10001) & (01001) & (00101) & (00011) & \end{Vmatrix}$$

$$= \begin{Vmatrix} j_1 & k_1 & x \\ k_2 & j_2 & l_1 \end{Vmatrix} \begin{Vmatrix} j_2 & k_2 & x \\ k_3 & j_3 & l_2 \end{Vmatrix}$$

$$\times \begin{Vmatrix} j_3 & k_3 & x \\ k_4 & j_4 & l_3 \end{Vmatrix} \begin{Vmatrix} j_4 & k_4 & x \\ j_1 & k_1 & l_4 \end{Vmatrix} \quad (\text{mod } 2),$$

where x has coordinates (00001). If we project this configuration from the point j(11111) on the space of the tetrahedron $j_1 j_2 j_3 x$ by adding j to each point outside this space, we see that the interchange of j_1 and k_1 is impossible in $PG(3, 2)$ since in the projection l_4 projects into l_2.

7. COMMENTS

In conclusion, reference should be made to a paper by Giovannini and Smith[17] which uses some of these ideas, but in a different way, and to papers of Shelepin;[18] who introduces topological ideas and Betti numbers.

After reading the manuscript of this paper in June 1970, Professor Biedenharn suggested to me that, instead of developing the geometrical model and showing its significance for the physical pattern, one could, alternatively, use this pattern explicitly to motivate the model.

We could seek first to eliminate summation by assuming in (1.1) that $j_i j_k = j_l$ *uniquely*. Interpreted additively, this would imply that there should be just three points on every line, so that we would be dealing with a finite projective geometry (mod 2). The relation $(I1)^3$ then implies that the $6j$ symbol is mapped on 0 or 1, while $(I2)^3$ implies the content of Theorem 2. Since $(I3)^3$ is provable in its physical context, it is natural that Desargues' theorem should be provable in $PG(n, 2)$, provided $n > 2$. As we have seen, (5.8) degenerates to (5.9) or (4.3) for $n = 2$. Dualizing, we have demonstrated the validity of the diagrams of Yutsis *et al.*[5]

ACKNOWLEDGMENTS

The author discussed most of these geometrical interpretations with B. R. Judd in Christchurch, New Zealand in 1968 and subsequently with W. T. Sharp in Toronto. To both of them he is indebted for many useful comments.

[1] H. Weyl, *Group Theory in Quantum Mechanics* (Dover, New York, 1931).
[2] J.-R. Derome and W. T. Sharp, J. Math. Phys. **6**, 1584 (1965).
[3] U. Fano and R. Racah, *Irreducible Tensorial Sets* (Academic, New York, 1959).
[4] G. de B. Robinson, *Foundations of Geometry* (U. of Toronto Press, Toronto, 1959), Chap. II.
[5] A. P. Yutsis, I. B. Levinson, and V. V. Vanagas, *Theory of Angular Momentum* (NSF, Jerusalem, 1962).
[6] Ref. 4, p. 106.
[7] T. Regge, Nuovo Cimento **11**, 116 (1959).
[8] Ref. 3, p. 57 and Ref. 5, p. 50.
[9] Ref. 4, p. 14.
[10] Ref. 4, p. 108.
[11] G. de B. Robinson, *Representation Theory of the Symmetric Group* (U. of Toronto Press, Toronto, 1961), Appendix Tables 2–4.
[12] Ref. 5, p. 53, and Ref. 7.
[13] Ref. 3, p. 159.
[14] Ref. 5, p. 59, Eq. (18.2).
[15] Ref. 5, p. 62, Eq. (19.2).
[16] Ref. 5, p. 62, Eq. (19.1).
[17] A. Giovannini and D. A. Smith, *Spectroscopic and Group-Theoretical Methods in Physics* (North-Holland, Amsterdam, 1968).
[18] L. A. Shelepin, Sov. J. Nucl. Phys. **8**, 481 (1969).

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 11, NUMBER 12 DECEMBER 1970

Generalized Operators*

GEORGE SVETLICHNY
Department of Physics and Astronomy, University of Maryland, College Park, Maryland 20740

(Received 29 May 1970)

A formal expression T in creation and annihilation operators (e.g., the Hamiltonian for a field theory model) is generally not a densely defined bona fide Hilbert space operator but is usually a densely defined sesquilinear form; as such it is convenient to consider it as a linear map from a dense domain Φ_- of a Hilbert space Φ_0 to a still larger space Φ_+ of antilinear functionals on Φ_-; that is, $T:\Phi_- \to \Phi_+ \supset \Phi_0$. We give here the basis of a mathematical structure theory of such generalized operators. The idea which we explore is that, associated with T, there is a (not necessarily unique) analytic family R_λ of generalized operators called the *resolvent* of T. Formally, $R_\lambda = (\lambda - T)^{-1}$, an equation to which we give more precise interpretations. The ambiguities in determining R_λ are associated with the arbitrary adjustments that are characteristic of renormalization programs. When appropriate conditions are met, we can construct from R_λ a new Hilbert space Ψ_0 and a bona fide operator T_R (the renormalized T) which is related to T by a formal intertwining equation $T_R\Delta = \Delta T$, where Δ maps Φ_- into a space containing Ψ_0. Given several generalized operators, we outline a procedure by which a subset of these can be renormalized to bona fide operators while the rest are reinterpreted as new generalized operators in the new Hilbert space. These are the rudiments of a multiplicity theory. Numerous examples illustrate the methods; in particular, the $N\theta$ sector of the Lee model with arbitrary cutoff (including none) is treated in detail.

I. INTRODUCTION

Quantum field theory or more generally the quantum theory of physical systems having infinitely many degrees of freedom is known to be beset with many mathematical difficulties. One very common source of such difficulties is the fact that in following the quantum mechanical procedures of representing physical quantities by "operators" one is led by physical necessity to mathematical objects which strictly speaking are not operators on a Hilbert space. For example, the quantum field amplitude $\phi(x)$ at a space–time point x is not an operator; furthermore, if we take any of the common field theory models such as the Yukawa model, the $\lambda\phi^4$ model, or quantum electrodynamics and construct in terms of the bare creation and annihilation operators the formal Hamiltonian, then one does not obtain a bona fide operator on a Hilbert space. Similar difficulties appear with the formal operators of quantum statistical mechanics and solid state theory. The difficulty with the field amplitude $\phi(x)$ has been somewhat overcome by considering the field to be an operator valued generalized function [1-4];however, the difficulties with the other objects have never been fully met. What has usually been done up to now is to approximate these objects in some way by bona fide operators, the so-called cutoff versions of these objects, and then to study the limiting situation as the cutoffs are removed. This approach has disadvantages. In the first place, the cutoff and limiting procedures are quite complex; no systematic methods are known and each problem requires its own tricks. In the second place, the original object often possesses a certain degree of symmetry such as definite transformation properties under various physical groups. The final results are also required to possess definite symmetries but the cutoff procedure violates these symmetries leaving one with the problem of recovering it in the limit. One needs a method which does not violate the symmetries of the problem.

Concerning the mathematical nature of a formal operator T of a physical theory, it must not be thought that it is mathematically ambiguous. What happens is that though T may not be a densely defined operator in a Hilbert space $\mathcal{H}$ in the usual sense that there is a dense domain $\mathfrak{D} \subset \mathcal{H}$ such that T maps $\mathfrak{D}$ into $\mathcal{H}$, it is true nevertheless that in practically all cases there is a dense subspace $\mathfrak{D} \subset \mathcal{H}$ for which the expectation values (f, Tg), where $f, g \in \mathfrak{D}$ are well defined. Now $f \rightsquigarrow (f, Tg)$ is an antilinear functional on $\mathfrak{D}$, and thus it is convenient to consider T as mapping $\mathfrak{D}$ into a space of antilinear functionals on $\mathfrak{D}$. More precisely, we shall consider triplets of spaces $\Phi_- \subset \Phi_0 \subset \Phi_+$, where Φ_- is a dense subspace of the Hilbert space Φ_0 and Φ_+ is a space of antilinear functionals on Φ_-. By a generalized operator T we shall mean a linear map $T:\Phi_- \to \Phi_+$ and most formal operators of physics can in fact be easily defined as generalized operators of this sort. This suggests the mathematical program of constructing a useful structure theory for classes of generalized operators. For the class of generalized operators defined in terms of creation and annihilation operators, the only extensive structure theory known is the combinatorial and diagrammatic approach of perturbation theory. This approach has been very fruitful in bringing to light

many difficulties and in solving a few physical problems. As a structure theory it fails in many respects, mostly in that one is usually at a loss as to how to combine the various diagrams to effect a solution to a given problem. Our approach here will be to give meaning to the inhomogeneous eigenvalue problem for T, which is expressed by the formal equation $(\lambda - T)\phi = \psi$. That this is fruitful can be seen from the difficult but beautiful work of Friedrichs[5] and the more recent important works of Glimm[6-8] and Glimm and Jaffe.[9,10] These works use diagrammatic perturbation techniques essentially but the introduction there of dressing or intertwining operators strongly hints at an underlying spectral structure, since an intertwining operator may be looked upon as one transforming the spectral structure of one operator to that of another, in the above cases that of the full Hamiltonian to that of the free Hamiltonian. The above mentioned works of Friedrichs and Glimm have been the inspirational source for the present work.

Now even for a bona fide operator, the eigenvectors need not lie in Hilbert space. Even keeping this in mind we find that for generalized operators further difficulties appear. Thus, take for example the generalized operator on the space of sequences defined by

$$(Ta)_0 = \sum_{k=1}^{\infty} a_k,$$
$$(Ta)_k = a_0, \quad k \geq 1,$$

where $a = (a_0, a_1, a_2, \cdots)$ is a sequence. Formally the eigenvalue problem reads

$$\sum_{k=1}^{\infty} a_k = \lambda a_0,$$
$$a_0 = \lambda a_k, \quad k \geq 1,$$

and assuming $\lambda \neq 0$ we find again formally

$$a_k = a_0/\lambda, \quad k \geq 1;$$

and therefore,

$$\sum_{k=1}^{\infty} a_0/\lambda = a_0/\lambda \sum_{1}^{\infty} 1 = \lambda a_0.$$

Thus $\lambda = \pm(\sum_1^\infty 1)^{\frac{1}{2}}$ are the eigenvalues and $a = (\pm(\sum_1^\infty 1)^{\frac{1}{2}}, 1, 1, 1, \cdots)$ are the eigenvectors. Now $(\sum_1^\infty 1)^{\frac{1}{2}}$ is a meaningless expression and such expressions occur abundantly in any formal treatment of generalized operators. It is usually said that by an appropriate regularization or renormalization technique these expressions will be replaced by numbers which *a priori*, however, can be arbitrary; thus, any sequence of the form $a = (\alpha, 1, 1, 1, \cdots)$, where α is an arbitrary complex number, should in some sense be considered a generalized eigenvector of T with eigenvalue α. The question then arises whether one can describe the precise relationship that these generalized eigenvectors have to the generalized operator T without going through the formal intermediate calculations involving meaningless expressions. The answer is yes, these vectors can be reached by appropriate limits. One learns in elementary quantum mechanics texts that even though the continuum eigenfunctions ψ_λ of a physical operator A are not square integrable, there is a sequence of square integrable functions ψ_n such that $\psi_n \to \psi_\lambda$ in some sense and $A\psi_n \to \lambda\psi_\lambda$ in some sense. Likewise, an element $\phi \in \Phi_+$ will be called a generalized eigenvector of a generalized operator T if for certain nets $f_\alpha \in \Phi_-$ such that $f_\alpha \to \phi$ in an appropriate sense Tf_α appears more and more as a multiple of ϕ again in an appropriate sense. Even more generally we shall say that ϕ solves the formal resolvent equation $(\lambda - T)\phi = \psi$ if there is a net $f_\alpha \in \Phi_-$ such that $f_\alpha \to \phi$ and $-Tf_\alpha$ differs from ψ by a vector that appears more and more as a multiple of ϕ. The next question is whether one can pick from this class of solutions an appropriate subclass that reflects the mathematical and physical properties of T. We shall formulate various criteria for doing this but unfortunately we shall not have a complete set of criteria; further insight is needed here either from physics or from a different structure approach. Having surmounted the above two obstacles, there still remains in most problems an additional step before we can say we have a reasonable solution to a physical problem. In order to conform to the usual assumptions of quantum mechanics, any generalized operator describing an observable must somehow be transformed into a bona fide operator in a Hilbert space. We shall call this process renormalization and see how a knowledge of the spectral structure sheds light on this problem.

The central idea which will emerge from our investigations is that associated to a generalized operator T there is a (not necessarily unique) analytic family of generalized operators R_λ called the *resolvent* of T. Formally $R_\lambda = (\lambda - T)^{-1}$ but due to the ambiguities present in working with generalized operators there are numerous interpretations of the above expression. We feel that among all such interpretations there are those that are in a sense the best in that they involve the least deviations from purely formal calculations.

We introduce two approaches to the construction of the resolvent. The first may be called the multivalued approach as it involves the use of multivalued linear maps. A multivalued linear map from a vector

space V to a vector space W is nothing more than a linear subspace of $V \times W$. It is of particular interest to consider a generalized operator $T:\Phi_- \to \Phi_+$ as being the linear subspace $\{(f, Tf) \mid f \in \Phi_-\} \subset \Phi_+ \times \Phi_+$. The closure $\bar{T}$ of the above subspace in an appropriate topology plays an important role, allowing us to define the generalized eigenvectors of T and to introduce a family $\mathcal{R}_\lambda(\bar{T})$, $\lambda \in \mathbb{C}$, of multivalued linear maps called the resolvent relations, which in an appropriate sense is given by $\mathcal{R}_\lambda(\bar{T}) = (\lambda - \bar{T})^{-1}$. We now try to pick linear subspaces R_λ, one from each $\mathcal{R}_\lambda$ and each one defining a single-valued transformation such that collectively they come as close as possible to having the properties of a bona fide resolvent of a bona fide self-adjoint operator. This approach works well for simple examples but is often too crude in that the possibilities for R_λ are too numerous and it is not clear how to pick the relevant ones. This difficulty is in part overcome by the second approach, which we call the orthogonal method since it involves orthogonal decompositions of the Hilbert space. Such a decomposition presents the generalized operator in terms of a matrix of linear transformations. By formally expressing the resolvent in terms of these matrix entries, one obtains an expression which under appropriate interpretations also leads to resolvents which, however, now can come from a more restricted class than can be obtained from the multivalued approach.

Having obtained a resolvent, we are able to introduce two methods of renormalization, one of which we discuss in detail in a separate subsection of Sec. III. We do not claim that these mathematical renormalization methods provide us directly with a physical interpretation but they do allow us to pass to bona fide operators and should at least in part embody important physical notions.

We devote one section to the problem of several generalized operators. We show how several generalized operators can be treated so that a certain subset of them are renormalized to bona fide operators while the remaining ones are appropriately modified to become new generalized operators so as to still retain their relationship among themselves and to the renormalized set. This is a very important situation physically since a physical theory is most often described by a set of generalized operators having various relations among themselves.

Throughout this work we often make use of common abuses of language. Thus we sometimes use as a name of a function an expression for its value at an arbitrary point; for example, we shall sometimes speak of the function $f(x)$ rather than of the function f or the function $x \rightsquigarrow f(x)$. Furthermore, we often perpetrate the fiction that generalized functions have the same independent variable for an argument as do bona fide functions; we then use $\tau(x)$ as another name for the generalized function τ and write the value of this generalized function on the test function f by means of a symbolic integral $\tau(f) = \int \tau(x) f(x)\, dx$. Other similar transgressions will be engaged in whenever the exact expressions are too awkward and the context supplies the necessary clarification.

We shall use the black square symbol ■ as an emphatic period which will usually signify the end of an argument that establishes a certain point.

II. THE ELEMENTARY THEORY OF GENERALIZED OPERATORS

A. Rigged Hilbert Spaces and Generalized Operators

We consider pairs (Φ_-, Φ_+) of complex vector spaces in which Φ_+ is some space of antilinear functionals on Φ_-. It is also useful to introduce the space Φ'_-, the space of *all* antilinear functionals on Φ_-. By definition $\Phi_+ \subset \Phi'_-$. The pairing of Φ_- with either Φ_+ or Φ'_- we will denote by $\langle \cdot, \cdot \rangle$; thus $\langle f, \phi \rangle$ is the value of the functional $\phi \in \Phi_+$ or $\phi \in \Phi'_-$ on the vector $f \in \Phi_-$.

A pair such as the above will be called a *rigged Hilbert space* if there is given a linear inclusion $j:\Phi_- \subset \Phi_+$ such that the sesquilinear form $f, g \rightsquigarrow (f, g) = \langle f, jg \rangle$ has a positive-definite quadratic form $f \rightsquigarrow (f, f)$ and, furthermore, such that if Φ_0 is the Hilbert space which is the completion of Φ_- with respect to $(\cdot, \cdot)$, then for any $h \in \Phi_0$ the antilinear functional $f \rightsquigarrow (f, h)$, $f \in \Phi_-$, is in Φ_+. We thus arrive at a triple (Φ_-, Φ_0, Φ_+) of complex vector spaces and inclusions

$$\Phi_- \overset{j_-}{\subset} \Phi_0 \overset{j_+}{\subset} \Phi_+, \quad j_+ j_- = j,$$

such that Φ_- is a dense subspace of the Hilbert space Φ_0 and for $f \in \Phi_-$ and $h \in \Phi_0$ we have $\langle f, j_+ h \rangle = (j_- f, h)$. The elements of Φ_- we shall call the *smooth vectors* and the elements of Φ_+ the *generalized vectors*. These adjectives may at times be dropped.

We shall denote the triple (Φ_-, Φ_+, j) by the single letter Φ and say that Φ is a rigged Hilbert space. We shall also say that the Hilbert space Φ_0 is *rigged by* Φ_- and Φ_+, or that it is *rigged by* Φ. If $\mathcal{H}$ is a Hilbert space then we shall sometimes consider it as the rigged Hilbert space $(\mathcal{H}, \mathcal{H}, \mathrm{id})$.

By an abuse of language we shall henceforth often drop writing the symbols j, j_-, j_+; thus we would write $\langle f, h \rangle = (f, h)$ for $f \in \Phi_-$, $h \in \Phi_0$.

The *weak$^+$ topology* $\mathfrak{W}^+$ on Φ_+ is defined by the pairing of Φ_+ with Φ_-. A basis for the neighborhoods of zero consists of the sets

$$U(f_1, f_2, \cdots, f_n; \epsilon) = \{\phi \in \Phi_+ \mid |\langle f_i, \phi\rangle| < \epsilon; i = 1, \cdots, n\},$$

where $f_i \in \Phi_-$. The weak$^+$ topology can, of course, be restricted to Φ_0 and Φ_- to define the induced topologies $\mathfrak{W}^+ \mid \Phi_0$ and $\mathfrak{W}^+ \mid \Phi_-$.

The space Φ_- is $\mathfrak{W}^+$ dense in Φ_+ since given $\phi \in \Phi_+$ and a $\mathfrak{W}^+$ neighborhood of zero $U(f_1, \cdots, f_n; \epsilon)$ in Φ_+ we can certainly choose an $h \in \Phi_0$ such that $\langle f_i, h\rangle = \langle f_i, \phi\rangle$; but since Φ_- is weakly dense in Φ_0, we can choose a $g \in \Phi_-$ such that

$$|\langle f_i, g\rangle - \langle f_i, h\rangle| < \epsilon$$

and then $g \in \phi + U(f_1, \cdots, f_n; \epsilon)$. We can thus approximate any element $\phi \in \Phi_+$ by a net $f_\alpha \in \Phi_-$ such that $f_\alpha \to \phi$ in the $\mathfrak{W}^+$ topology and where $\alpha \in A$ and A is some directed set. Such nets will play an important role. Note that by passing to subnets we can always choose the directed set A to be the directed set of $\mathfrak{W}^+$ neighborhoods of zero in Φ_+ ordered by inclusion. In subsequent discussions we shall, without any further comment, pass freely from the considerations of arbitrary nets to that of such special nets and vice versa.

We note that by the above argument the $\mathfrak{W}^+$ completion of Φ_+ is Φ'_-. We could,however,choose to have Φ_+ be *sequentially* $\mathfrak{W}^+$ complete since every element of the sequential $\mathfrak{W}^+$ completion Φ_+^ω of Φ_+ can be interpreted as an antilinear functional on Φ_-. It is, in fact, very advantageous for many purposes to have Φ_+ be sequentially $\mathfrak{W}^+$ complete.

We now give some examples of rigged Hilbert spaces.

Example 1: $\Phi_- = \mathfrak{D}(\mathbb{R}^n) \subset \mathfrak{D}'(\mathbb{R}^n) = \Phi_+$, where $\mathfrak{D}$ is the space of C^∞ functions of compact support on $\mathbb{R}^n$ and $\mathfrak{D}'$ is the space of distributions. The inclusion is the natural one and the pairing is $\langle f, \phi\rangle = \phi(\bar{f})$, where $\bar{f}(x) = \overline{f(x)}$, the bar denoting the complex conjugate. Clearly $\Phi_0 = L^2(\mathbb{R}^n)$.

Example 2: $\Phi_- = \mathcal{S}(\mathbb{R}^n) \subset \mathcal{S}'(\mathbb{R}^n) = \Phi_+$, where $\mathcal{S}$ is the Schwartz space of functions on $\mathbb{R}^n$ and $\mathcal{S}'$ is the space of tempered distributions. The inclusion is the natural one, the pairing is $\langle f, \phi\rangle = \phi(\bar{f})$, and $\Phi_0 = L^2(\mathbb{R}^n)$.

Example 3: $\Phi_- = d^n \subset d^{n'} = \Phi_+$, where d^n is the space of all n-fold sequences $\{f_{i_1 i_2 \cdots i_n}\}$, $i_j = 1, 2, \cdots$, which vanish whenever any index is sufficiently large and $d^{n'}$ is the space of *all* n-fold sequences. The inclusion is the natural one and the pairing is $\langle f, \phi\rangle = \sum_{i_1=1}^\infty \cdots \sum_{i_n=1}^\infty \overline{f_{i_1 \cdots i_n}} \phi_{i_1 \cdots i_n}$. When $n = 1$ we shall occasionally let the index i in f_i start with zero rather than one.

Example 4: Let Φ_- be a nuclear space with a continuous nondegenerate sesquilinear form $(\cdot, \cdot)$, let Φ_+ be Φ_-^* the space of *continuous* antilinear functionals on Φ_-, and let the inclusion $j: \Phi_- \subset \Phi_+$ be the one provided by the sesquilinear form; then we have a rigged Hilbert space Φ as defined by Gel'fand and Vilenkin.[11]

Example 5: Let $\Phi_\alpha = (\Phi_{-\alpha}, \Phi_{+\alpha}, j_\alpha)$, $\alpha \in A$, be a family of rigged Hilbert spaces. We now define $\Phi = \oplus \Phi_\alpha$, the *direct sum* of the family $\{\Phi_\alpha\}_{\alpha \in A}$. For Φ_- we take the subspace of the *algebraic* direct sum $\oplus \Phi_{-\alpha}$ whose elements have all but a finite number of components zero. For Φ_+ we take the *algebraic* direct sum $\oplus \Phi_{+\alpha}$. For j we take the algebraic direct sum $\oplus j_\alpha$ restricted to Φ_-. The pairing is taken to be $\langle f, \phi\rangle = \sum_{\alpha \in A} \langle f_\alpha, \phi_\alpha\rangle_\alpha$. One easily shows that Φ is a rigged Hilbert space in which $\Phi_0 = \oplus \Phi_{0\alpha}$, the *Hilbert space direct sum.*

Example 6: This is a very involved example which we write out in detail because the construction is new and it plays an important role.

Let $\Phi_\alpha = (\Phi_{-\alpha}, \Phi_{+\alpha}, j_\alpha)$, $\alpha \in A$, be a family of rigged Hilbert spaces; we say $\{\Phi_\alpha\}_{\alpha \in A}$ is an *inductive family* if it satisfies the following properties:

(IF1) A is a directed set.

(IF2) If $\alpha, \beta \in A$ and $\beta \geq \alpha$, then we have linear maps $k_{-\beta\alpha}: \Phi_{-\alpha} \to \Phi_{-\beta}$ and $k_{+\beta\alpha}: \Phi_{+\alpha} \to \Phi_{+\beta}$ such that the diagram

$$\begin{array}{ccc} \Phi_{-\alpha} & \overset{j_\alpha}{\subset} & \Phi_{+\alpha} \\ {\scriptstyle k_{-\beta\alpha}}\downarrow & & \downarrow{\scriptstyle k_{+\beta\alpha}} \\ \Phi_{-\beta} & \overset{j_\beta}{\subset} & \Phi_{+\beta} \end{array}$$

is commutative; that is, $j_\beta k_{-\beta\alpha} = k_{+\beta\alpha} j_\alpha$.

(IF3) For $\alpha, \beta, \gamma \in A$ and $\gamma \geq \beta \geq \alpha$, we have the transitivity relations

$$k_{-\gamma\beta} k_{-\beta\alpha} = k_{-\gamma\alpha},$$
$$k_{+\gamma\beta} k_{+\beta\alpha} = k_{+\gamma\alpha}.$$

(IF4) For $f \in \Phi_-$, $\phi \in \Phi_{+\alpha}$, and $\beta \geq \alpha$ we have

$$\langle f, \phi\rangle_\alpha = \langle k_{-\beta\alpha} f, k_{+\beta\alpha} \phi\rangle_\beta.$$

Note that this implies that $k_{-\beta\alpha}$ and $k_{+\beta\alpha}$ are both inclusions, and subsequently by (IF3) $k_{-\alpha\alpha}$ and $k_{+\alpha\alpha}$ are both identities.

(IF5) We finally require that whenever $\beta \geq \alpha$ the map $(k_{-\beta\alpha})' : \Phi_{+\beta} \to \Phi'_{-\alpha}$ defined by $\langle f, (k_{-\beta\alpha})'\phi\rangle_\alpha = \langle k_{-\beta\alpha} f, \phi\rangle_\beta$ map $\Phi_{+\beta}$ into $\Phi^\omega_{+\alpha}$, the sequential $\mathfrak{W}^+$ completion of $\Phi_{+\alpha}$; that is:

$$(k_{-\beta\alpha})'\Phi_{+\beta} \subset \Phi^\omega_{+\alpha}.$$

We now define a new rigged Hilbert space Φ called the *inductive limit* $\Phi = {}_\to\lim \Phi_\alpha$ of the inductive family $\{\Phi_\alpha\}_{\alpha\in A}$.

For Φ_- we take ${}_\to\lim \Phi_{-\alpha}$, the inductive limit of the family $\{\Phi_{-\alpha}\}_{\alpha\in A}$ considered as an inductive family of complex vector spaces. Specifically, the underlying set of Φ_- is the disjoint union $\bigvee_{\alpha\in A} \Phi_{-\alpha}$ modulo the equivalence relation $\sim$ by which $f_\alpha \sim g_\beta$ if there is a γ, $\gamma \geq \alpha$, $\gamma \geq \beta$ such that $k_{-\gamma\alpha} f_\alpha = k_{-\gamma\beta} g_\beta$. Let $f = [f_\alpha]$ denote the equivalence class of $f_\alpha \in \bigvee \Phi_{-\alpha}$. Note that, for any $\beta \geq \alpha$, $[f_\alpha] = [k_{-\beta\alpha} f_\alpha]$ by (IF3). If $f = [f_\alpha]$, $g = [g_\beta]$ are in Φ_-, if λ and μ are two complex numbers, and if we pick a γ, $\gamma \geq \alpha$, $\gamma \geq \beta$, then we can define the linear operation $\lambda f + \mu g$ by $\lambda f + \mu g = [\lambda k_{-\gamma\alpha} f_\alpha + \mu k_{-\gamma\beta} g_\beta]$. This is easily checked to be well defined by virtue of (IF3) and thus endows the quotient $\bigvee \Phi_{-\alpha}/\sim$ with a linear structure to define the complex vector space Φ_-.

We can now define the canonical linear inclusions $k_{-\alpha} : \Phi_{-\alpha} \subset \Phi_-$ by setting $k_{-\alpha} f_\alpha = [f_\alpha]$ for $f_\alpha \in \Phi_{-\alpha}$. This is in fact an inclusion by (IF4). We can thus consider each $\Phi_{-\alpha}$ as a subspace of Φ_-.

For $\beta \geq \alpha$ we easily have $k_{-\beta\alpha} k_{-\alpha} = k_{-\beta}$.

For Φ_+ we now take that subspace of Φ'_- each element of which when restricted to $\Phi_{-\alpha}$ via the inclusion $k_{-\alpha}$ is an element of $\Phi^\omega_{+\alpha}$. Explicitly, $\phi \in \Phi'_-$ belongs to Φ_+ if and only if for every $\alpha \in A$ the anti-linear functional $f_\alpha \rightsquigarrow \langle k_{-\alpha} f_\alpha, \phi\rangle$, $f_\alpha \in \Phi_{-\alpha}$, belongs to $\Phi^\omega_{+\alpha}$.

This definition provides us, therefore, with the canonical linear restrictions $r_\alpha : \Phi_+ \to \Phi^\omega_{+\alpha}$.

We now define linear inclusions $k_{+\alpha} : \Phi_{+\alpha} \subset \Phi_+$ as follows: Let $f \in \Phi_-$, $f = [f_\beta]$; let $\phi_\alpha \in \Phi_{+\alpha}$, and pick a γ, $\gamma \geq \alpha$, and $\gamma \geq \beta$. We define $k_{+\alpha}\phi_\alpha$ by $\langle f, k_{+\alpha}\phi_\alpha\rangle = \langle k_{-\gamma\beta} f_\beta, k_{+\gamma\alpha}\phi_\alpha\rangle_\gamma$. This is well defined, for, if $f_{\beta'}$ and γ' are possibly different choices, then there is a δ, $\delta \geq \beta$, and $\delta \geq \beta'$, such that $k_{-\delta\beta} f_\beta = k_{-\delta\beta'} f_{\beta'}$; furthermore, there is a δ', $\delta' \geq \gamma$, $\delta' \geq \gamma'$, and $\delta' \geq \delta$, and we have, using (IF3) and (IF4),

$$\begin{aligned}\langle k_{-\gamma\beta} f_\beta, k_{+\gamma\alpha}\phi_\alpha\rangle_\gamma &= \langle k_{-\delta\gamma} k_{-\gamma\beta} f_\beta, k_{+\delta\gamma} k_{+\gamma\alpha}\phi_\alpha\rangle_\delta \\ &= \langle k_{-\delta\beta} f_\beta, k_{+\delta\alpha}\phi_\alpha\rangle_\delta \\ &= \langle k_{-\delta\beta'} f_{\beta'}, k_{+\delta\alpha}\phi_\alpha\rangle_\delta \\ &= \langle k_{-\gamma'\beta'} f_{\beta'}, k_{+\gamma'\alpha}\phi_\alpha\rangle_{\gamma'}.\end{aligned}$$

Thus, $k_{+\alpha}\phi_\alpha$ is well defined as an element of Φ'_-, and we need now show it to be, in fact, an element of Φ_+. Now for $\gamma \geq \beta$, $\gamma \geq \alpha$, we have, on $k_{-\beta}\Phi_{-\beta}$, $\langle k_{-\beta} f_\beta, k_{+\alpha}\phi_\alpha\rangle = \langle k_{-\gamma\beta} f_\beta, k_{+\gamma\alpha}\phi_\alpha\rangle_\gamma = \langle f_\beta, (k_{-\gamma\beta})' k_{+\gamma\alpha}\phi_\alpha\rangle_\beta$, but $k_{+\gamma\alpha}\phi_\alpha \in \Phi_{+\gamma}$ and so by (IF5) $(k_{-\gamma\beta})' k_{+\gamma\alpha}\phi_\alpha \in \Phi^\omega_{+\beta}$, showing by definition of Φ_+ that $k_{+\alpha}\phi_\alpha \in \Phi_+$. Thus, $k_{+\alpha}\Phi_{+\alpha} \subset \Phi_+$. Suppose now that $k_{+\alpha}\phi_\alpha = 0$; this means, in particular, that $\langle k_{-\alpha} f_\alpha, k_{+\alpha}\phi_\alpha\rangle = 0$ for all $f_\alpha \in \Phi_{-\alpha}$; but then by definition of $k_{+\alpha}$ and by (IF4) we have

$$0 = \langle k_{-\alpha} f_\alpha, k_{+\alpha}\phi_\alpha\rangle = \langle k_{-\alpha\alpha} f_\alpha, k_{+\alpha\alpha}\phi_\alpha\rangle_\alpha = \langle f_\alpha, \phi_\alpha\rangle_\alpha$$

for all $f_\alpha \in \Phi_{-\alpha}$, and this implies $\phi_\alpha = 0$, proving that $k_{+\alpha}$ is an inclusion.

It is clear by construction that for $\beta \geq \alpha$ we have $k_{+\beta\alpha} k_{+\alpha} = k_{+\beta}$.

We can now consider the family of linear inclusions $l_\alpha = k_{+\alpha} j_\alpha : \Phi_{-\alpha} \subset \Phi_+$. Since, for $\beta \geq \alpha$, $l_\beta k_{-\beta\alpha} = l_\alpha$, we see that the family $\{l_\alpha\}$ defines a linear map $j : \Phi_- \to \Phi_+$ such that $jk_{-\alpha} = l_\alpha$; namely, take $j[f_\alpha] = l_\alpha f_\alpha$. The map j is, in fact, an inclusion, for if $jf = 0$, then $f = [f_\alpha]$ and $0 = jf = l_\alpha f_\alpha$ implies that $f_\alpha = 0$ for l_α is an inclusion; subsequently $f = 0$.

We now proceed to show that the triple $\Phi = (\Phi_-, \Phi_+, j)$ is a rigged Hilbert space.

Consider the quadratic form $f \rightsquigarrow \langle f, jf\rangle$. Let $f = [f_\alpha]$; then $\langle f, jf\rangle = \langle f, l_\alpha f_\alpha\rangle = \langle f, k_{+\alpha} j_\alpha f_\alpha\rangle = \langle k_{-\alpha\alpha} f_\alpha, k_{+\alpha\alpha} j_\alpha f_\alpha\rangle_\alpha = \langle f_\alpha, j_\alpha f_\alpha\rangle_\alpha \geq 0$ and this vanishes if and only if $f_\alpha = 0$, that is if and only if $f = 0$. Thus, the quadratic form is positive definite.

Now, let Φ_0 be the Hilbert space which is the completion of Φ_- with respect to the inner product $f, g \rightsquigarrow (f, g) = \langle f, jg\rangle$, and let $h \in \Phi_0$; then h is given by a Cauchy sequence $h_n \in \Phi_-$ and we can set $h_n = [h_{n\alpha(n)}]$ for $h_{n\alpha(n)} \in \Phi_{-\alpha(n)}$. Consider now $k_{-\beta}\Phi_{-\beta}$ and pick $\gamma(n)$, $\gamma(n) \geq \alpha(n)$, and $\gamma(n) \geq \beta$; then

$$\begin{aligned}(k_{-\beta} f_\beta, h) &= \lim_{n\to\infty} \langle k_{-\beta} f_\beta, j h_n\rangle \\ &= \lim_{n\to\infty} \langle k_{-\beta} f_\beta, k_{+\alpha(n)} j_{\alpha(n)} h_{n\alpha(n)}\rangle \\ &= \lim_{n\to\infty} \langle k_{-\gamma(n)\beta} f_\beta, k_{+\gamma(n)\alpha(n)} j_{\alpha(n)} h_{n\alpha(n)}\rangle_{\gamma(n)} \\ &= \lim_{n\to\infty} \langle f_\beta, (k_{-\gamma(n)\beta})' k_{+\gamma(n)\alpha(n)} j_{\alpha(n)} h_{n\alpha(n)}\rangle_\beta.\end{aligned}$$

Now this last limit exists for all $f_\beta \in \Phi_{-\beta}$ since it is equal to $(k_{-\beta} f_\beta, h)$. By (IF5),

$$(k_{-\gamma(n)\beta})' k_{+\gamma(n)\alpha(n)} j_{\alpha(n)} h_{n\alpha(n)} \in \Phi^\omega_{+\beta};$$

since $\Phi^\omega_{+\beta}$ is $\mathfrak{W}^+$ sequentially complete, we see that this last limit in fact defines an element of $\Phi^\omega_{+\beta}$ and so $f \rightsquigarrow (f, h)$ for $f \in \Phi_-$ defines an element of Φ_+, concluding the proof that Φ is a rigged Hilbert space ■

Let $\{\Phi_\alpha\}_{\alpha\in A}$ be an inductive family of rigged Hilbert spaces and let $\Phi = {}_\to\lim \Phi_\alpha$. Let V be a

complex vector space and for each $\alpha \in A$ let there be given a linear map $\Delta_\alpha : V \to \Phi_{+\alpha}$ such that, for $\beta \geq \alpha, k_{+\beta\alpha}\Delta_\alpha = \Delta_\beta$. There is then a unique map $\Delta : V \to \Phi_+$ such that $r_\alpha \Delta = \Delta_\alpha$. This map will also be denoted by $\underrightarrow{\lim}\, \Delta_\alpha$ and will be called the *inductive limit* of the family of maps $\{\Delta_\alpha\}$.

We define Δ as follows: Let $f \in \Phi_-$, $f = [f_\alpha]$, and let $v \in V$; then we set $\langle f, \Delta v\rangle = \langle f_\alpha, \Delta_\alpha v\rangle_\alpha$. The stated properties of Δ follow readily.

Given a rigged Hilbert space Φ by a *generalized operator* T, we shall mean any linear map $T : \Phi_- \to \Phi_+$. If T is a generalized operator, then the set of numbers $\langle g, Tf\rangle$ for $f, g \in \Phi_-$ we shall call the *matrix elements* of T and in particular the numbers $\langle f, Tf\rangle$ we shall call the *expectation values* of T.

A generalized operator defines a sesquilinear form on Φ_-, namely $f, g \rightsquigarrow \langle f, Tg\rangle$. The converse is not necessarily true for, given any sesquilinear form $f, g \rightsquigarrow \{f, g\}$ on Φ_-, we can certainly define a linear map $S : \Phi_- \to \Phi'_-$ by $Sg : f \rightsquigarrow \{f, g\}$; but the range of S need not lie within Φ_+. This problem arises whenever we perform operations on the matrix elements of generalized operators to obtain new sesquilinear forms. Often we want the new object to also be a generalized operator. This question must involve relations, among Φ_-, Φ_+, j and the generalized operators, which we have not investigated. We have not yet made any regularity assumptions on T nor any topological or other restrictions on Φ_-, Φ_+, and j. Such questions eventually have to be faced, but we shall get a better perspective on them only after we further develop the formalism. In this work we shall usually assume that the relevant maps are in fact generalized operators when such is the need. All the circumstances under which this question will arise will be sufficiently natural both mathematically and physically that the assumption will not be unreasonable.

It is now appropriate to point out that there is a certain ambiguity in defining generalized operators, namely, we are often able to pass to other rigged Hilbert spaces without loss of any information about the original situation. In the first place, we can always pass to the sequential $\mathfrak{W}^+$ completion of Φ_+; so assume Φ_+ to be sequentially $\mathfrak{W}^+$ complete. We may now be able to find a subspace Ψ_- of Φ_0 such that $j_-\Phi_- \subset \Psi_-$ and such that $j_-\Phi_-$ is dense in Ψ_- in some topology τ. Furthermore, certain functionals in Φ_+ may be extendable from Φ_- to Ψ_- by virtue of continuity with respect to τ, thus giving rise to a new space Ψ_+. Assume, in addition, that $\Psi_- \subset \Psi_+$ and that this forms a new rigged Hilbert space Ψ. If T is continuous with respect to the topology τ on Φ_- and some topology τ_+ on Φ_+, then it can be extended to a map $\hat{T} : \Psi_- \to \Phi_+$ and if in addition $\hat{T}\Psi_- \subset \Psi_+$, we have a new generalized operator from which by virtue of continuity assumptions the original operator T can be recovered. One can also pass in another direction. Suppose we take a subspace $\Theta_- \subset \Phi_-$ such that Θ_- is dense in Φ_- with respect to some topology τ. The space Φ_+ now acquires a new $\mathfrak{W}^+$ topology induced by Θ_- and with respect to this new topology it may not be sequentially complete; we pass to the sequential completion thus arriving at a new space $\Theta_+ \supset \Phi_+$. The triple $\Theta = (\Theta_-, \Theta_+, j \mid \Theta_-)$ can form a new rigged Hilbert space and if T is continuous with respect to the topology τ on Θ_- and some topology τ_+ on Φ_+, then, passing to the restriction $T \mid \Theta_-$, we again get a new generalized operator from which the original generalized operator can be recovered. One can combine any number of such extensions and restrictions and possibly other similar operations to obtain a whole class of generalized operators. Now for physical and most mathematical applications any one of the generalized operators in this class may be taken as representing the same situation and thus it becomes reasonable to ask which among this class is in some way mathematically natural. We shall not look very deeply into this problem as it is mathematically subtle but in concrete situations does not often present any difficulties. We shall at times, however, give indications of the issues that it raises. All such problems related to the various choices of riggings we shall call *candidacy* problems since they relate to the choice of a mathematical candidate for a given situation.

The following are examples of generalized operators.

Example 7: If $T\Phi_- \subset \Phi_-$, then T is a densely defined operator in Φ_0 with an invariant domain. If the rigging is of the Gel'fand–Vilenkin type and T is essentially self-adjoint or unitary on Φ_-, then there is a detailed spectral theory for T as is given in Gel'fand and Vilenkin.

Example 8: Suppose there is a sesquilinear map $M : \Phi_- \times \Phi_- \to \Phi_-$ antilinear in the first variable. If $\tau \in \Phi'_-$, then we can define a linear map $M_\tau : \Phi_- \to \Phi'_-$ by the formula

$$\langle g, M_\tau f\rangle = \langle M(f, g), \tau\rangle.$$

If $M_\tau \Phi_- \subset \Phi_+$, then M_τ defines a generalized operator. A useful particular case of this in $\mathfrak{D}(\mathbb{R}^n) \subset \mathfrak{D}'(\mathbb{R}^n)$ is the following: Let $M(f, g)(x) = \overline{f(x)}g(x)$ and let $\tau \in \mathfrak{D}'$; then $M_\tau f = f\tau \in \mathfrak{D}'$ and M_τ is a generalized operator which can be called *multiplication by a distribution* and is a generalization of the familiar

bona fide operator which is that of multiplication by a function.

Example 9: In $\mathbb{C} \oplus \mathfrak{D}(\mathbb{R}^n) \subset \mathbb{C} \oplus \mathfrak{D}'(\mathbb{R}^n)$ define T as

$$T(f_0, f_1(\mathbf{k})) = \left(-q_0 \int \frac{f_1(\mathbf{k}) d^n k}{[2\omega(\mathbf{k})]^{\frac{1}{2}}}, \omega(\mathbf{k}) f_1(\mathbf{k}) - q_0 \frac{f_0}{[2\omega(\mathbf{k})]^{\frac{1}{2}}}\right),$$

where $\omega(\mathbf{k}) = (m^2 + \mathbf{k}^2)^{\frac{1}{2}}$ and q_0 and m are real constants. This example is the *bare Hamiltonian* for the $N\Theta$ sector of the Lee model in n dimensions without cutoff.

Example 10: In $\mathbb{C} \oplus \mathfrak{D}(\mathbb{R}) \subset \mathbb{C} \oplus \mathfrak{D}'(\mathbb{R})$ define T as

$$T(f_0, f_1(x)) = (-g f_1(0), -g f_0 \delta(x)),$$

where $\delta(x)$ is the Dirac distribution and g a real constant.

Example 11: In $d \subset d'$ define T by

$$(Tf)_0 = \lambda_0 f_0 + \sum_{k=1}^{\infty} f_k \rho_k,$$
$$(Tf)_k = \lambda_k f_k + f_0 \rho_k \quad \text{for} \quad k \geq 1.$$

Here the λ_n, $n = 0, 1, \cdots$, and the ρ_k, $k = 1, 2, \cdots$, are fixed complex numbers. This example is a discrete version of Example 9 and will be called the discrete Lee model. It will prove to be very instructive since everything can be computed explicitly and it is sufficiently complicated to exhibit most of the phenomena in which we are interested.

Example 12: All Fock spaces are a direct sum of the multiparticle subspaces. One of these, the vacuum, is $\mathbb{C}$ which we rig in the manner $(\mathbb{C}, \mathbb{C}, \mathrm{id})$. All the others are of the form $L^2_P(\mathbb{R}^n)$ for some n, where the subscript P denotes that the vectors of this space have certain specified symmetries under permutations of the independent variables, said symmetries arising from physical statistics. We rig these subspaces by $\mathfrak{D}_P(\mathbb{R}^n) \subset \mathfrak{D}'_P(\mathbb{R}^n)$ and then rig the Fock space by taking the direct sum of the resulting family of rigged multiparticle subspaces.

With Fock space rigged in this way then, as was pointed out in the introduction, most of the Fock space "operators" of quantum field theory, quantum statistical mechanics, and solid state theory are in fact generalized operators. Among these are included the Hamiltonians and Lagrangians for most models, their densities, Wick polynomials in the free fields at a fixed point, and in general any formal normal ordered expression in creation and annihilation operators with distribution kernels.

Example 13: Let T be a generalized operator. We define the map $T^*: \Phi_- \to \Phi'_-$ by $T^* f: g \rightsquigarrow \overline{\langle f, Tg \rangle}$. If $T^* \Phi_- \subset \Phi_+$, then T^* is a generalized operator called the *dual* of T.

If T is a generalized operator and $T^* = T$, then we say T is *symmetric*. In terms of matrix elements this means that $\langle g, Tf \rangle = \overline{\langle f, Tg \rangle}$.

If T is a generalized operator and $\langle f, Tf \rangle \geq 0$ for all $f \in \Phi_-$, then we say T is *positive*. Note that if T is positive, then T must be symmetric.

Example 14: Let $\phi \in \Phi_+$, $\psi \in \Phi'_-$; then we define the generalized operator $|\phi\rangle\langle\psi|$ by $|\phi\rangle\langle\psi| f = \overline{\langle f, \psi \rangle} \phi$.

The set of all generalized operators on a given rigged Hilbert space clearly forms a complex vector space in the usual manner. We shall place the weak topology on this space; thus a basis for the neighborhoods of zero is given by

$$U(f_1, f_2, \cdots, f_n; g_1, g_2, \cdots, g_n; \epsilon) = \{T \mid |\langle f_i, Tg_i \rangle| < \epsilon, i = 1, \cdots, n\},$$

where f_i, $g_i \in \Phi_-$. Furthermore, when speaking of any notion relating to generalized operators, we shall, unless otherwise stated, mean the weak notion; thus, for example, a family $T(\lambda)$, $\lambda \in \mathbb{C}$, of generalized operators is analytic in a domain $\mathfrak{D}$ if every matrix element $\langle f, T(\lambda) g \rangle$ is analytic in $\mathfrak{D}$.

B. Linear Relations

It turns out that it is necessary to study multivalued linear maps, which we call *linear relations*.

Let V and W be two complex vector spaces then we define a linear relation $\mathfrak{Q}$ from V to W to be map from V to the power set of W, $\mathfrak{Q}: V \to \mathfrak{P}(W)$ satisfying the following three conditions:

(LR1) If $\phi, \psi \in V$, then $\mathfrak{Q}(\phi + \psi) \supset \mathfrak{Q}\phi + \mathfrak{Q}\psi$.

(LR2) If $\lambda \neq 0$ is a complex number and $\psi \in V$, then $\mathfrak{Q}(\lambda\psi) = \lambda \mathfrak{Q}\psi$.

(LR3) $\mathfrak{Q}0 \neq \varnothing$, where $\varnothing$ is the empty set.

Consider now the set $V \times W$ and the subset $\Gamma(\mathfrak{Q}) \subset V \times W$ called the *graph* of $\mathfrak{Q}$ and defined by

$$\Gamma(\mathfrak{Q}) = \{(\psi, \psi') \mid \psi' \in \mathfrak{Q}\psi\}.$$

Now conditions (LR1), (LR2), and (LR3) imply that $\Gamma(\mathfrak{Q})$ is a linear subspace of $V \times W$ considered as a vector space in the usual componentwise manner.

To see this, let $(\phi, \phi'), (\psi, \psi') \in \Gamma(\mathfrak{Q})$;then

$$\mathfrak{Q}(\phi + \psi) \supset \mathfrak{Q}\phi + \mathfrak{Q}\psi \ni \phi' + \psi',$$

which shows that $(\phi + \psi, \phi' + \psi') \in \Gamma(\mathfrak{Q})$. Furthermore, let $(\psi, \psi') \in \Gamma(\mathfrak{Q})$ and $\lambda \neq 0$ be a complex number; then $\mathfrak{Q}(\lambda\psi) = \lambda\mathfrak{Q}\psi \ni \lambda\psi'$, which shows that $(\lambda\psi, \lambda\psi') \in \Gamma(\mathfrak{Q})$; lastly, $\mathfrak{Q}0 \neq \varnothing$ and

$$\mathfrak{Q}0 = \mathfrak{Q}(0 - 0) \supset \mathfrak{Q}0 + \mathfrak{Q}(-0) = \mathfrak{Q}0 - \mathfrak{Q}0 \ni 0$$

shows that $(0, 0) \in \Gamma(\mathfrak{Q})$■

Conversely, given any linear subspace $L \subset V \times W$, we construct the map

$$\mathfrak{R}(L): V \to \mathfrak{P}(W) \quad \text{by} \quad \mathfrak{R}(L)\psi = \{\psi' \mid (\psi, \psi') \in L\}.$$

To show that $\mathfrak{R}(L)$ is a linear relation,we have first

$$\begin{aligned}\mathfrak{R}(L)(\phi + \psi) &= \{\eta \mid (\phi + \psi, \eta) \in L\} \supset \{\phi' \mid (\phi, \phi') \in L\} \\ &\quad + \{\psi' \mid (\psi, \psi') \in L\} = \mathfrak{R}(L)\phi + \mathfrak{R}(L)\psi,\end{aligned}$$

so (LR1) is satisfied.Furthermore, assume $\lambda \neq 0$ is a complex number;then

$$\begin{aligned}\mathfrak{R}(L)(\lambda\psi) &= \{\psi' \mid (\lambda\psi, \psi') \in L\} = \{\psi' \mid (\psi, 1/\lambda\psi') \in L\} \\ &= \lambda\{\psi' \mid (\psi, \psi') \in L\} = \lambda\mathfrak{R}(L)\psi,\end{aligned}$$

so (LR2) is satisfied; lastly, $(0, 0) \in L$ shows that $0 \in \mathfrak{R}(L)0$,so $\mathfrak{R}(L)0 \neq \varnothing$ and (LR3) is satisfied.

It is immediate that $\mathfrak{R}(\Gamma(\mathfrak{Q})) = \mathfrak{Q}$ and $\Gamma(\mathfrak{R}(L)) = L$, and thus linear relations are in 1-to-1 correspondence with linear subspaces of $V \times W$. This is very helpful in many considerations. Two facts that follow immediately from this are that $\mathfrak{Q}0$ is a linear subspace of W and that if $\psi' \in \mathfrak{Q}\psi$, then $\mathfrak{Q}\psi = \psi' + \mathfrak{Q}0$.

The set of all linear relations from V to W we shall denote by $M(V, W)$ and in the case $V = W$ by $M(V)$.

If $\mathfrak{Q}$ is a linear relation, we define the *domain* of $\mathfrak{Q}$ to be $\mathfrak{D}(\mathfrak{Q}) = \{\psi \in V \mid \mathfrak{Q}\psi \neq \varnothing\}$. The domain is clearly a linear subspace of V since it is the canonical projection of $\Gamma(\mathfrak{Q})$ onto V.

If $\mathfrak{Q}_1$ and $\mathfrak{Q}_2$ are linear relations and $\Gamma(\mathfrak{Q}_1) \subset \Gamma(\mathfrak{Q}_2)$, then we say that $\mathfrak{Q}_1$ is a *restriction* of $\mathfrak{Q}_2$ and that $\mathfrak{Q}_2$ is an *extension* of $\mathfrak{Q}_1$, and we express this by $\mathfrak{Q}_1 \subset \mathfrak{Q}_2$ or $\mathfrak{Q}_2 \supset \mathfrak{Q}_1$. The relation $\mathfrak{Q}_1 \subset \mathfrak{Q}_2$ is equivalent to $\mathfrak{Q}_1\psi \subset \mathfrak{Q}_2\psi$ for all $\psi \in V$.

A restriction or an extension is called an *operator restriction* or an *operator extension* if it is in fact an operator, that is,if it is single valued on its domain.

If $\mathfrak{Q}_1$ and $\mathfrak{Q}_2$ are linear relations, then we define the linear relation $\mathfrak{Q}_1 + \mathfrak{Q}_2$ by

$$(\mathfrak{Q}_1 + \mathfrak{Q}_2)\psi = \mathfrak{Q}_1\psi + \mathfrak{Q}_2\psi.$$

One easily shows that $\mathfrak{Q}_1 + \mathfrak{Q}_2$ is a linear relation.

If $\mathfrak{Q}$ is a linear relation and λ a complex number, define the linear relation $\lambda\mathfrak{Q}$ by $(\lambda\mathfrak{Q})\psi = \lambda(\mathfrak{Q}\psi)$.

If $\mathfrak{Q} \in M(V, W)$ and $\mathfrak{R} \in M(W, Z)$, we define $\mathfrak{R}\mathfrak{Q} \in M(V, Z)$ by $\mathfrak{R}\mathfrak{Q}\psi = \mathfrak{R}(\mathfrak{Q}\psi) = \bigcup_{\psi' \in \mathfrak{Q}\psi} \mathfrak{R}\psi'$.

Of particular interest for function theory of generalized operators is the case when $V = W$, for then $\mathfrak{R}\mathfrak{Q}$ is again in $M(V)$.

If $\mathfrak{Q} \in M(V, W)$,define $\mathfrak{Q}^{-1} \in M(W, V)$ by

$$\Gamma(\mathfrak{Q}^{-1}) = \{(\psi', \psi) \mid \psi' \in \mathfrak{Q}\psi\},$$

which is clearly a linear subspace of $W \times V$. Equivalently $\mathfrak{Q}^{-1}\omega = \{\psi \mid \omega \in \mathfrak{Q}\psi\}$ for $\omega \in W$.

Let now $\mathfrak{Q}_\alpha \in M(V, W)$, $\alpha \in A$,be a family of linear relations. We define $\bigvee_{\alpha \in A} \mathfrak{Q}_\alpha \in M(V, W)$ by setting $\Gamma(\bigvee \mathfrak{Q}_\alpha) = \bigvee \Gamma(\mathfrak{Q}_\alpha)$, where the last set is the linear span of the family of subspaces $\Gamma(\mathfrak{Q}_\alpha)$.

If $\mathfrak{Q}_1 \subset \mathfrak{Q}$, $\mathfrak{Q}_2 \subset \mathfrak{Q}$, and if $\psi \in \mathfrak{D}(\mathfrak{Q}_1 - \mathfrak{Q}_2)$,we have

$$(\mathfrak{Q}_1 - \mathfrak{Q}_2)\psi = \mathfrak{Q}_1\psi - \mathfrak{Q}_2\psi \subset \mathfrak{Q}\psi - \mathfrak{Q}\psi \subset \mathfrak{Q}0$$

and if, on the other hand, $\psi \notin \mathfrak{D}(\mathfrak{Q}_1 - \mathfrak{Q}_2)$, then $(\mathfrak{Q}_1 - \mathfrak{Q}_2)\psi = \varnothing \subset \mathfrak{Q}0$. This observation leads to a new notion.

Suppose $\mathfrak{K}$ and $\mathfrak{Q}$ are two linear relations such that, for all ψ, $\mathfrak{K}\psi \subset \mathfrak{Q}0$; then we shall say that $\mathfrak{K}$ is a *counter* of $\mathfrak{Q}$. This terminology is introduced since as will be seen later certain so-called counterterms which occur in the formal function theory of physical "operators" arise in this way. As we saw above, any two restrictions of the same linear relation differ by a counter on the intersection of their domains, and conversely, if $\mathfrak{T} \subset \mathfrak{Q}$ and $\mathfrak{K}$ is a counter of $\mathfrak{Q}$, then $\mathfrak{T} + \mathfrak{K} \subset \mathfrak{Q}$.

It is useful now to introduce certain topological considerations. We assume that V and W are both topological vector spaces and we endow $V \times W$ with the product topology.

We say that $\mathfrak{Q} \in M(V, W)$ is *closed* if its graph $\Gamma(\mathfrak{Q})$ is closed. In terms of limits this means that if $\psi_\alpha \in \mathfrak{D}(\mathfrak{Q})$ is a convergent net, $\psi_\alpha \to \psi \in V$, and if there exist $\psi'_\alpha \in \mathfrak{Q}\psi_\alpha$ such that $\psi'_\alpha \to \psi' \in W$, then $\psi' \in \mathfrak{Q}\psi$.

If $\mathfrak{Q} \in M(V, W)$, then it has a canonical closed extension $\bar{\mathfrak{Q}}$ called the *closure* of $\mathfrak{Q}$ and is defined by $\Gamma(\bar{\mathfrak{Q}}) = \overline{\Gamma(\mathfrak{Q})}$. In terms of limits, we have that $\psi' \in \bar{\mathfrak{Q}}\psi$ if and only if there are convergent nets $\psi_\alpha \to \psi$ in V and $\psi'_\alpha \to \psi'$ in W such that $\psi'_\alpha \in \mathfrak{Q}\psi_\alpha$. The closure is in a sense the best attempt to extend a linear relation by topological means.

If $\mathfrak{T}$, $\mathfrak{Q}$ are linear relations, we say that $\mathfrak{T}$ is a *version* of $\mathfrak{Q}$ if $\mathfrak{T} \subset \bar{\mathfrak{Q}}$; this is equivalent to $\Gamma(\mathfrak{T}) \subset \Gamma(\bar{\mathfrak{Q}}) = \overline{\Gamma(\mathfrak{Q})}$ and is thus equivalent to $\overline{\Gamma(\mathfrak{T})} \subset \overline{\Gamma(\mathfrak{Q})}$ which means that $\bar{\mathfrak{T}}$ is also a version of $\mathfrak{Q}$. *Version of*

is a pseudo-order, that is, it is reflexive and transitive. The relation of being a version of each other is an equivalence relation since it is the same as having equal closures. Among closed relations, *version of* is in fact a partial order.

Our next topological construction has to do with limits of linear relations. Let Q_ξ, $\xi \in X$, be a family of linear relations and let N be a set of nets in X all defined on the same directed set A. We define the linear relation $\mathfrak{Q} = \lim_N \mathfrak{Q}_\xi$ by taking $\Gamma(\mathfrak{Q})$ to be the set of all limit points of nets $\gamma_\alpha \in \Gamma(\mathfrak{Q}_{\xi_\alpha})$, where ξ_α is a net in N.

Concerning this limit procedure, we want to make the following remarks. Suppose the net of neighborhoods of zero of $V \times W$ admits a cofinal subnet indexed by A, say $\{U_\alpha\}$; then we claim that $\lim_N \mathfrak{Q}_\xi = \lim_N \bar{\mathfrak{Q}}_\xi$. The inclusion $\lim_N \mathfrak{Q}_\xi \subset \lim_N \bar{\mathfrak{Q}}_\xi$ is obvious and we need prove only the opposite inclusion. Let $\bar{\gamma}_\alpha \in \Gamma(\bar{\mathfrak{Q}}_{\xi_\alpha})$ and $\bar{\gamma}_\alpha \to \bar{\gamma}$; then there are points $\gamma_\alpha \in \Gamma(\mathfrak{Q}_{\xi_\alpha})$ such that $\gamma_\alpha - \bar{\gamma}_\alpha \in U_\alpha$, but since $\{U_\alpha\}$ is cofinal, we see that $\lim \gamma_\alpha = \lim \bar{\gamma}_\alpha = \bar{\gamma}$ so $\bar{\gamma} \in \lim_N \mathfrak{Q}_\xi$. Precisely the same reasoning also shows that $\lim_N \bar{\mathfrak{Q}}_\xi$ is closed and so in the case that the directed set A is "large enough" we have $\lim_N \mathfrak{Q} = \lim_N \bar{\mathfrak{Q}}_\xi = \overline{\lim_N \bar{\mathfrak{Q}}_\xi}$.

Because the map $(\psi, \psi') \rightsquigarrow (\psi', \psi): V \times W \to W \times V$ is continuous, we have that

$$\left(\lim_N \mathfrak{Q}_\xi\right)^{-1} = \lim_N \mathfrak{Q}_\xi^{-1}.$$

Let now $T: \Phi_- \to \Phi_+$ be a generalized operator. Since j is an inclusion of Φ_- into Φ_+ we can view T as an operator from Φ_+ to Φ_+ defined on a domain $\mathfrak{D}(T) = \Phi_-$. Finally, we view T as a linear relation in $M(\Phi_+)$. Since Φ_+ is endowed with the $\mathfrak{W}^+$ topology, it becomes meaningful to take the closure $\bar{T}$ of T. This closure will in general be multivalued, but it is an object of great interest. For rigged Hilbert spaces we are thus interested in $M(\Phi_+)$ and particularly in the closed relations.

Let $\mathfrak{Q} \in M(\Phi_+)$; we call a restriction of $\mathfrak{Q}$ a *g-restriction* if its domain is contained in Φ_- and if it is maximal with respect to this property. If in addition we require the g-restriction to be an operator restriction, again imposing among such restrictions the condition of maximality, then we speak of *g-operator restrictions*. The g-operator restrictions of $\mathfrak{Q}$ are those that come closest to being generalized operators; they will be generalized operators if their domain coincides with Φ_-. We shall also speak of *g-versions* and *g-operator versions*, meaning that we first pass to $\bar{\mathfrak{Q}}$ and then take restrictions. We shall also consider additional properties, such as symmetry, that restrictions and operator restrictions must satisfy. It is the imposition of these additional properties that provides constraints that are more severe than could be obtained from weak topologies alone.

C. Orthogonal Projections

Let Φ be a rigged Hilbert space and let L be a $\mathfrak{W}^+$ closed subspace of Φ_+. We are interested in the case that $L \cap \Phi_-$ has L for its $\mathfrak{W}^+$ closure and furthermore that no element of L vanishes on all of $L \cap \Phi_-$. In this case we define π_L, the *orthogonal projection onto* L, to be a linear relation on Φ_+ given by requiring $\pi_L\psi$ to be that element of L, if it exists, which coincides with ψ on $L \cap \Phi_-$; namely, $\langle f, \pi_L\psi\rangle = \langle f, \psi\rangle$ for $f \in L \cap \Phi_-$. Note that if $\pi_L\psi$ exists, then it is unique and furthermore L is contained in the domain of π_L and π_L restricted to L is the identity. We have $\pi_L^2 = \pi_L$.

If L is finite dimensional, then $L \subset \Phi_-$ and any finite dimensional subspace of Φ_- satisfies our hypotheses above. Let L be generated by $f_1, \cdots, f_n$ all linearly independent; then π_L is defined on all of Φ_+ and is given explicitly by the formula

$$\pi_L\psi = \sum_{i,j=1}^{n} f_i M_{ij} \langle f_j, \psi\rangle,$$

where M_{ij} is the inverse of the matrix $\langle f_i, f_j\rangle$.

Let $\mathfrak{F}$ be any cofinal subset of the directed set of all finite-dimensional subspaces of Φ_-. Now for any $\psi \in \Phi_+$ the net $\{\pi_F\psi\}_{F \in \mathfrak{F}}$ converges to ψ in the $\mathfrak{W}^+$ topology: $\pi_F\psi \to \psi$. In fact if $U(g_1, \cdots, g_n; \epsilon)$ is a $\mathfrak{W}^+$ neighborhood of zero in Φ_+ and if F is such that $g_i \in F$, then $\pi_F\psi \in U(g_1, \cdots, g_n; \epsilon) + \psi$.

Let I be the identity linear relation on Φ_+; that is, $I\psi = \{\psi\}$. If π_L is an orthogonal projection, then $I - \pi_L$ is an idempotent, $(I - \pi_L)^2 = I - \pi_L$, and a map of some interest, though it is not necessarily an orthogonal projection. In certain special cases it may, however, be an orthogonal projection.

Suppose that Φ is a direct sum $\Phi_1 \oplus \Phi_2$ and let $L = \Phi_{+1} \oplus 0$. Then L satisfies our hypotheses and π_L exists and is given by $\pi_L\psi_1 \oplus \psi_2 = \psi_1 \oplus 0$. Furthermore, $0 \oplus \Phi_{+2}$ also satisfies our hypotheses and the orthogonal projection onto it is $I - \pi_L$.

Let T be a generalized operator and π_L an orthogonal projection. Associated with T are also the following maps: $\pi_L T \pi_L$, $\pi_L T(I - \pi_L)$, $(I - \pi_L)T\pi_L$, and $(I - \pi_L)T(I - \pi_L)$. In case Φ is a direct sum $\Phi_1 \oplus \Phi_2$, then, if we set $\pi_1 = \pi_{\Phi_{+1} \oplus 0}$, $\pi_2 = \pi_{0 \oplus \Phi_{+2}}$, we have $\pi_2 = I - \pi_1$. We can now write T in the

form of a matrix

$$T = \begin{pmatrix} T_{11} & T_{12} \\ T_{21} & T_{22} \end{pmatrix},$$

where

$$T_{ij} = \pi_i T \pi_j \,|\, \Phi_{+j}.$$

One gains from the above considerations only if L is somehow specifically related to T and this means that we should single out certain special subspaces. Now in all cases we have the inclusions

(i) $\overline{\pi_L T \pi_L} \subset \pi_L \bar{T} \pi_L$,

(ii) $\pi_L T(I - \pi_L)\Phi_+ \subset \pi_L \Phi_+$.

The special cases hold when these inclusions are strict, that is, an inequality holds. Heuristically speaking, if (i) is a strict inequality, then part of the "singular" nature of T is expressed in either the off-diagonal elements or in the other diagonal element $(I - \pi_L)T(I - \pi_L)$. A strict inequality for (ii) is to be considered as a generalization of T being diagonal.

Suppose now L to be finite dimensional; then $\overline{\pi_L T \pi_L}$ is single valued and defined on all of Φ_+ while $\pi_L \bar{T} \pi_L$, also defined on all of Φ_+, may be multivalued. A measure of this multivaluedness is given by the dimension of $\pi_L \bar{T} 0$. Among all L of a fixed dimension there are those for which the dimension of $\pi_L \bar{T} 0$ is minimum and these are especially singled out.

Likewise, among all the finite-dimensional L of a fixed dimension those are specially singled out for which the dimension of $\pi_L T(I - \pi_L)\Phi_+$ is minimum.

We shall not develop here the theory of special subspaces such as the above to any great extent but we shall use such considerations in various examples.

As an illustration consider the discrete Lee model with an infinite number of ρ_k different from zero. We shall see in the next section that

$$\bar{T} 0 = \{(\alpha, 0, 0, \cdots) \,|\, \alpha \in \mathbb{C}\}$$

and so for $\pi_F \bar{T} 0$ to be zero dimensional for a finite-dimensional F we must have the vector

$$e_0 = (1; 0, 0, \cdots, 0, \cdots)$$

be orthogonal to F. The subspace generated by all such F is $\{f \in d \,|\, f_0 = 0\}$ and this is in fact an orthogonal subspace of Φ_-. Thus the decomposition $\Phi = \mathbb{C} \oplus \Phi_2$, where $\Phi_{2+} = \{\psi \in d' \,|\, \psi_0 = 0\}$ is specifically adapted to T.

Assume now for convenience that all the λ_k are different and none of ρ_k vanish. Let

$$\rho = (0; \rho_1, \rho_2, \cdots, \rho_k, \cdots)$$

and let $e_k = (0; 0, \cdots, 0, 1, 0, \cdots)$, where the 1 is in the kth place. We have

$$T = \sum_{n=0}^{\infty} \lambda_n |e_n\rangle\langle e_n| + |e_0\rangle\langle\rho| + |\rho\rangle\langle e_0|.$$

Let F be a finite-dimensional subspace of d and consider

$$\begin{aligned} \pi_F T(I - \pi_F) = \pi_F \sum_{n=0}^{\infty} \lambda_n |e_n\rangle\langle e_n| (I - \pi_F) \\ + \pi_F |e_0\rangle\langle\rho| (I - \pi_F) \\ + \pi_F |\rho\rangle\langle e_0| (I - \pi_F). \end{aligned}$$

We note that if F is generated by a finite number of the e_k, $k \geq 1$, then $\pi_F T(I - \pi_F)\Phi_+$ is one dimensional and a tedious argument shows that these are the only subspaces for which this is true and that for no subspace is this space zero dimensional. Thus, these particular subspaces are singled out. These subspaces form a cofinal set in the set of all finite-dimensional subspaces of $\{f \in d \,|\, f_0 = 0\}$.

D. Elementary Function Theory of Generalized Operators

Here we develop only the function theory dealing with simple expressions, particularly those involving products sums and inverses. A deeper version of function theory will come only after the development of a structure theory.

The basic idea here is to use the closures of the generalized operators, manipulate with the linear relations so obtained according to the definitions introduced in the last section, and then take a g-operator version of the result as the answer. Which g-operator version is appropriate to take must be decided by other considerations.

The simplest example of this is that of passing from the generalized operator T to any of its g-operator versions; furthermore, if T is symmetric, we may want to restrict ourselves only to the symmetric g-operator versions.

A slightly more complicated example is that of interpreting the product TS of two generalized operators to be any of the g-operator versions of $\bar{T}\bar{S}$. Now in formally defining TS one often runs into certain undefined divergent expressions which are removed by appropriate counter terms. One may then consider the formal product TS as a version of $\bar{T}\bar{S}$ with divergent coefficients and this differs from the finite version by a counter with divergent coefficients. All other finite versions then differ by finite counters, that is, generalized operators with values in $\overline{\bar{T}\bar{S}} 0$.

To illustrate the above ideas, consider the discrete Lee model

$$T: f \rightsquigarrow \left(\lambda_0 f_0 + \sum_1^{\infty} f_k \rho_k; \cdots, \lambda_k f_k + f_0 \rho_k, \cdots\right).$$

Let a net $f_\alpha \to \psi$ in the $\mathfrak{W}^+$ topology; then $(f_\alpha)_n \to \psi_n$, $n \geq 0$, and thus $\lambda_k(f_\alpha)_k + (f_\alpha)_0\rho_k \to \lambda_k\psi_k + \psi_0\rho_k$, $k \geq 1$. What are the possible limits of

$$\lambda_0(f_\alpha)_0 + \sum_1^\infty (f_\alpha)_k\rho_k?$$

There are two cases, depending on whether only a finite number of the ρ_k are different from zero or not. In the first case, the vector $\rho = (0, \rho_1, \rho_2, \cdots) \in d$, and we have

$$\lambda_0(f_\alpha)_0 + \sum_1^\infty (f_\alpha)_k\rho_k \to \lambda_0\psi_0 + \sum_1^\infty \psi_k\rho_k;$$

in the second case, $\rho \notin d$, and by choosing the net f_α appropriately we can make the sum approach any complex value. Thus, for $\rho \in d$, T is continuous and $\bar{T}$ is an operator

$$\bar{T}\psi = \left\{\left(\lambda_0\psi_0 + \sum_1^\infty \psi_k\rho_k; \cdots, \lambda_k\psi_k + \psi_0\rho_k, \cdots\right)\right\},$$

where $\psi \in d'$. For $\rho \notin d$, T is not continuous and

$$\bar{T}\psi = \{(\alpha; \cdots, \lambda_k\psi_k + \psi_0\rho_k, \cdots) \mid \alpha \in \mathbb{C}\},$$

where $\psi \in d'$.

For $\rho \notin d$, a g-operator version of T is thus given by

$$f \rightsquigarrow \left(\alpha_0 f_0 + \sum_1^\infty \alpha_k f_k; \cdots, \lambda_k f_k + f_0\rho_k, \cdots\right)$$

and $f \in d$ and α_0 and α_k are complex numbers. Now the dual of this version is

$$f \rightsquigarrow \left(\bar{\alpha}_0 f_0 + \sum_1^\infty \bar{\rho}_k f_k; \cdots, \bar{\lambda}_k f_k + f_0\bar{\alpha}_k, \cdots\right)$$

and the dual of T is obtained by the change $\lambda_k \rightsquigarrow \overline{\lambda_n}$, $n \geq 0$ and $\rho_k \rightsquigarrow \overline{\rho_k}$. Thus, in order to preserve the duality properties as much as possible, an appropriate choice for α_n, $n \geq 0$, may be $a_k = \rho_k$, $k \geq 1$ and α_0 arbitrary. If in addition λ_n, $n \geq 0$, and ρ_k, $k \geq 1$, are real, then T is symmetric and we should then restrict α_0 to be real. All symmetric g-operator versions of T are thus obtained by a change of λ_0, still remaining real.

Let us now consider products. In the case $\rho \in d$, $\bar{T}\bar{T}\psi$ is seen to be

$$\bar{T}\bar{T}\psi = \left\{\left(\lambda_0^2\psi_0 + \lambda_0\sum_1^\infty \psi_k\rho_k + \sum_1^\infty \lambda_k\psi_k\rho_k + \psi_0\sum_1^\infty \rho_k^2; \cdots, \lambda_k^2\psi_k + \lambda_k\psi_0\rho_k + \lambda_0\psi_0\rho_k + \left(\sum_{l=1}^\infty \psi_l\rho_l\right)\rho_k, \cdots\right)\right\}$$

and since in this case again only finite sums are involved, we have $\bar{\bar{T}}\bar{\bar{T}} = \bar{T}\bar{T}$ and is thus an operator. In the case $\rho \notin d$, we have

$$\bar{T}\bar{T}\psi = \{(\alpha; \cdots, \lambda_k^2\psi_k + \lambda_k\psi_0\rho_k + \beta\rho_k, \cdots) \mid \alpha, \beta \in \mathbb{C}\}$$

and since no infinite sums are involved, we have $\bar{\bar{T}}\bar{\bar{T}} = \bar{T}\bar{T}$. A g-operator version of $\bar{T}\bar{T}$ is given by

$$f \rightsquigarrow \left(\alpha_0 f_0 + \sum_1^\infty \alpha_k f_k; \cdots, \lambda_k^2 f_k + \lambda_k f_0\rho_k + \left(\beta_0 f_0 + \sum_1^\infty \beta_l f_l\right)\rho_k, \cdots\right).$$

The dual of this version is

$$f \rightsquigarrow \left(\bar{\alpha}_0 f_0 + \sum_1^\infty \bar{\lambda}_k\bar{\rho}_k f_k + \bar{\beta}_0\sum_1^\infty \bar{\rho}_k f_k; \cdots, \bar{\lambda}_k^2 f_k + f_0\bar{\alpha}_k + \left(\sum_1^\infty \bar{\rho}_l f_l\right)\bar{\beta}_k, \cdots\right).$$

Thus, again to preserve duality properties, it may be appropriate to take $a_k = \lambda_k\rho_k + \bar{\beta}_0\rho_k$ and $\beta_k = \mu\rho_k$ with α_0, β_0, and μ arbitrary; in particular, if T is symmetric, then in order to preserve symmetry we must take α_0 and μ real.

In formally computing T^2 in the case $\rho \notin d$, we would come across the term $f_0\sum_1^\infty \rho_k^2$ in $(T^2f)_0$. This term is divergent in case $\rho \notin l^2$ and so we introduce the counter $K{:}\Phi_- \to \bar{T}\bar{T}0$, $K{:}f \rightsquigarrow (f_0; 0, 0, \cdots)$ and we now have

$$(T^2)_{\text{formal}} - \left(\sum \rho_k^2\right)K{:}f \rightsquigarrow \left(\lambda_0^2 f_0 + \lambda_0\sum_1^\infty \rho_k f_k + \sum_1^\infty \lambda_k\rho_k f_k; \cdots, \lambda_k^2 f_k + \lambda_k f_0\rho_k + \lambda_0 f_0\rho_k + \left(\sum_1^\infty \rho_l f_l\right)\rho_k, \cdots\right),$$

which is a nondivergent version of $\bar{T}\bar{T}$. This corresponds to the previous versions by the choice $\beta_0 = \lambda_0$, $\alpha_0 = \lambda_0^2$, and $\mu = 1$. We see that the most general g-operator version of $\bar{T}\bar{T}$ contains more adjustments than is absolutely necessary but whether to make these or not must be decided by other considerations. We note here, however, that

$$\overline{\bar{T}\bar{T}}\psi = \left\{\left(\alpha; \cdots, \lambda_k^2 f_k + f_0\lambda_k\rho_k + \lambda_0 f_0\rho_k + \left(\sum_1^\infty \rho_l f_l\right)\rho_k, \cdots\right) \,\middle|\, \alpha \in \mathbb{C}\right\}$$

gives rise to symmetric g-operator versions for which $\beta_0 = \lambda_0$ and $\mu = 1$ and contain only the minimal adjustments necessary.

It should be pointed out that elementary function theory allows considerable latitude of interpretation. This is desirable since formal expressions involving generalized operators must often be taken very loosely. Thus, even the equality $T = S$ may be taken now in the very loose sense that a g-operator version of T equals a g-operator version of S. In particular, $TS - ST = 0$ can be interpreted as meaning that $\overline{\bar{T}\bar{S}} - \overline{\bar{S}\bar{T}}$ has 0 restricted to Φ_- as a version; this is the loosest sense of the expression. For the product of three generalized operators TSR, the loosest interpretation is a g-operator version of $\overline{\bar{S}(\bar{T}\bar{R})} \vee \overline{(\bar{S}\bar{T})\bar{R}}$. The loosest interpretation of $(d/dt)T(t)$, where $T(t)$ is a one-parameter family of generalized operators, is that $(d/dt)T(t)$ is a g-operator version of

$$\lim (1/h)[\bar{T}(t+h) - \bar{T}(t)]$$

as $h \to 0$. Of course, one need not take the loosest interpretation in all cases, and sometimes a more stringent interpretation may be more to the point, such as in the example above where an interpretation of T^2 as a g-operator restriction of $\bar{T}T$ did not involve the possibly superfluous adjustments. In any case, any interpretation of expressions in generalized operators that involves manipulations of multivalued maps and then a passage to an appropriate single-valued restriction we shall call the *multivalued approach*. The multivalued methods do not by themselves lead to very deep results, but they provide very necessary material which can then be treated by more sophisticated methods.

E. Elementary Theory of Subtractions

In the previous section we gave an interpretation of the formal product of several generalized operators; namely we take any g-operator version of certain multivalued combinations of the closures of the factors. We now describe certain more stringent interpretations that can reflect better the specific properties of the factors. Consider therefore a formal product $T_1T_2 \cdots T_N$. We first want to replace each factor T_i by $\pi_{J_i}T_i\pi_{G_i}$, where J_i and G_i are finite-dimensional subspaces of Φ_-. Let $\mathfrak{J}_i$ and $\mathfrak{G}_i$ be cofinal sets of finite-dimensional subspaces of Φ_- and define an order on $\mathfrak{J}_i \times \mathfrak{G}_i$ by $(J, G) \geq (J', G')$ if and only if $J \supset J'$ and $G \supset G'$. Let $\mathcal{F}_i$ be any cofinal subset of $\mathfrak{J}_i \times \mathfrak{G}_i$; then $(T_i)_{F_i} = \pi_{J_i}(T_i)\pi_{G_i} \to T$ weakly, where $F_i = (J_i, G_i) \in \mathcal{F}_i$. Consider now the product $(T_1)_{F_1} \cdots (T_N)_{F_N}$ and let $\mathcal{F}$ be a cofinal subset of $\mathcal{F}_1 \times \cdots \times \mathcal{F}_N$, where the last space is ordered componentwise. In general the product has no limit on $\mathcal{F}$ but we can modify each product by a term K_F, $F \in \mathcal{F}$, where K_F is a single-valued counter of some multivalued interpretation of $T_1 \cdots T_N$. We are then interested in the limit

$$\lim_{F \in \mathcal{F}} [(T_1)_{F_1} \cdots (T_N)_{F_N} + K_F]$$

and this affords an interpretation of the product if the limit exists. This interpretation of course depends on the choice of $\mathcal{F}$ and K_F but if these are obtained from considerations relating to the specific properties of the factors, the interpretations so obtained can be more limited than those obtained from the loosest multivalued methods and more relevant to the problem at hand. We shall call the above interpretation and modifications of it the *subtracted product*.

As an example let us compute T^2, where T is the symmetric discrete Lee model. Let $\mathfrak{G}$ be $\{G_n\}_{n=0,\cdots}$, where G_n is the finite-dimensional subspace of d generated by $\{e_0, e_1, \cdots, e_n\}$. We consider the expression $T\pi_{G_n}T + K_n$, where K_n is a symmetric single-valued counter of $\bar{T}T$, namely $K_n : f \rightsquigarrow (\kappa_n f_0; 0, 0, \cdots)$. We have

$$T\pi_{G_n}T + K_n : f \rightsquigarrow \Big(\Big(\lambda_0^2 + \sum_1^n \rho_k^2 + \kappa_n\Big)f_0 + \lambda_0 \sum_1^\infty \rho_k f_k + \sum_1^n \lambda_k \rho_k f_k; \cdots, \theta_{n,k}\lambda_k^2 f_k + \theta_{n,k}\lambda_k\rho_k f_0 + \rho_k\Big(\lambda_0 f_0 + \sum_1^\infty \rho_l f_l\Big), \cdots\Big),$$

where

$$\theta_{n,k} = 1 \quad \text{if} \quad k \leq n$$
$$= 0 \quad \text{if} \quad k > n.$$

Thus, if we pick $\kappa_n = \kappa - \sum_1^n \rho_k^2$, the limit converges and gives as an interpretation of T^2 the transformation

$$f \rightsquigarrow \Big((\lambda_0^2 + \kappa)f_0 + \lambda_0 \sum_1^\infty (\rho_k + \lambda_k\rho_k)f_k; \cdots, \lambda_k^2 f_k + \lambda_k\rho_k f_0 + \rho_k\Big(\lambda_0 f_0 + \sum_1^\infty \rho_l f_l\Big), \cdots\Big).$$

We now turn to certain constructions which we call *analytic subtractions*. Let $\{f_\alpha(\lambda)\}_{\alpha \in A}$ be a net of functions each analytic in λ in a fixed domain $\mathfrak{D}$. We say that the net *converges with subtractions* in $\mathfrak{D}$ if there is a net of polynomials $P_\alpha(\lambda)$ such that

$$\lim_{\alpha \in A} [f_\alpha(\lambda) + P_\alpha(\lambda)]$$

converges pointwise to an analytic function in $\mathfrak{D}$. We call the net P_α the *net of subtractions*. Let n_α be the order of P_α. If Q_α is another net of subtractions and m_α is the order of Q_α, then we say that $\{Q_\alpha\}$ *has fewer subtractions* than $\{P_\alpha\}$ if $m_\alpha \leq n_\alpha$ and $\lim m_\alpha/n_\alpha = 0$, where we interpret 0/0 as 0. If n_α is bounded, then we

say that P_α has a finite number of subtractions. We say that P_α is *canonical* if P_α is minus a first part of the Taylor expansion of f_α about some point λ_α; that is,

$$P_\alpha(\lambda) = -(f_\alpha(\lambda_\alpha) + (\lambda - \lambda_\alpha)f'_\alpha(\lambda_\alpha) + \cdots + (\lambda - \lambda_\alpha)^{n_\alpha}f_\alpha^{(n_\alpha)}(\lambda_\alpha)/n_\alpha!).$$

In most practical cases where subtracted limits are used, the weaker the subtraction the more desirable it is; in particular, if a finite number of subtractions can be used, then it is desirable to have n_α bounded by the least possible number.

One common way in which analytic subtractions are used is in interpreting $\int f_x(\lambda)\mu(dx)$, where $\lambda \rightsquigarrow f_x(\lambda)$ is a family of functions, each analytic in a fixed domain $\mathfrak{D}$, and x is a point in a measure space X equipped with a measure μ. If A is the set of all measurable subsets $a \subset X$ such that

$$f_a(\lambda) = \int_a f_x(\lambda)\mu(dx)$$

exists as an analytic function in $\mathfrak{D}$ and if $\bigcup_{a\in A} a = X$, then, even if $\int f_x(\lambda)\mu(dx)$ does not exist, one can interpret it as the subtracted limit of $f_a(\lambda)$. An example of the above is of course the Mittag-Leffler expansion for a meromorphic function. Let r_i be a sequence of complex numbers and let λ_i be a sequence of complex numbers with ∞ as the only accumulation point; then the sum $\sum_{i=1}^{\infty} r_i/(\lambda - \lambda_i)$ need not converge but one can pick a sequence of integers n_i such that

$$\sum_{i=1}^{\infty} (\lambda/\lambda_i)^{n_i} r_i/(\lambda - \lambda_i)$$

does converge and is in fact a canonical subtracted sum in the sense defined above. A similar interpretation can be assigned to the integral $g(\lambda)\int \mu(dx)/(\lambda - x)g(x)$, where $\mu(dx)$ is a measure on $\mathbb{R}$ and g is an entire function; this is a subtracted interpretation of $\int \mu(dx)/(\lambda - x)$ which need not converge.

As a matter of notation we shall use the tilde $\sim$ in the form

$$\tilde{\lim}, \quad \tilde{\sum}, \quad \text{and} \quad \tilde{\int}$$

to denote subtracted interpretations whenever we do not wish to explicitly exhibit the subtractions involved. Thus one interpretation of

$$\tilde{\sum_{i=1}^{\infty}} r_i/(\lambda - \lambda_i) \quad \text{is} \quad \sum_{i=1}^{\infty} (\lambda/\lambda_i)^{n_i} r_i/(\lambda - \lambda_i)$$

but of course others are also possible.

Analytic subtractions can likewise be applied whenever we are dealing with a product of generalized operators when at least one factor is an analytic generalized operator-valued function; for example, $STQ(\lambda)R$. We again introduce finite-dimensional projections and thus consider the product

$$S_{F_1}T_{F_2}Q(\lambda)_{F_3}R_{F_4},$$

but for the added term we take a polynomial $P_F(\lambda)$, that is, $P_F(\lambda) = \sum_{i=0}^{n_F} \lambda^i P_{i,F}$, where the $P_{i,F}$ are generalized operators. Thus, an interpretation of the product is

$$\lim_{F\in\mathcal{F}} [S_{F_1}T_{F_2}Q(\lambda)_{F_3}R_{F_4} + P_F(\lambda)].$$

As before, various procedures and modifications are possible, but these must be adapted to the specific situation at hand. We adopt the same terminology here as for subtracted limits of functions.

Analytic subtractions is one of the most powerful tools available, and its versatility will be demonstrated in subsequent examples.

F. Elementary Spectral Theory of Generalized Operators

In this section we give sense to the formal equations $(T - \lambda)\phi = 0$ and $(T - \lambda)\phi = \psi$ for a generalized operator T. We develop the theory by first giving several equivalent definitions of the generalized eigenvectors of a generalized operator. The first one in terms of limits is the most intuitive but the subsequent ones are more amenable to study and will eventually provide deeper insights into the structure of generalized operators.

Let T be a generalized operator on a rigged Hilbert space Φ. We want to generalize the notion of eigenvectors of an operator to that of generalized eigenvectors of a generalized operator. It is convenient to break this definition into two parts.

A generalized vector $\phi \in \Phi_+$ is said to belong to the *generalized null space* $N(T)$ of T if, given any $\mathfrak{W}^+$ neighborhood U of $0 \in \Phi_+$, there exists a smooth vector $f_U \in \Phi_-$ such that $\phi - f_U \in U$ and $Tf_U \in U$.

A generalized vector $\phi \in \Phi_+$ is said to belong to $\sum'(T)$ if, given any $\mathfrak{W}^+$ neighborhood U of $0 \in \Phi_+$, there exists a smooth vector $f_U \in \Phi_-$ and a complex number μ_U such that $\phi - f_U \in U$ and $\phi - \mu_U Tf_U \in U$.

A generalized vector $\phi \in \Phi_+$ is said to be a *generalized eigenvector* of T if either $\phi \in N(T)$ or $\phi \in \sum'(T)$. We define $\sum(T) = N(T) \cup \sum'(T)$. We note that if $\phi \in \sum(T)$ and $\lambda \in \mathbb{C}$, then $\lambda\phi \in \sum(T)$ and thus the set $\sum(T)$ is trivially too large in that for most purposes it is necessary to consider only a single representative from every ray $\{\lambda\phi; \lambda \neq 0\}$ and disregard the rest. As will become clear soon, the resulting set is still too

large and certain subsets will become relevant. The freedom in choosing both the ray representatives and these further subsets corresponds in part to the freedom in making certain arbitrary numerical assignments in conventional renormalization theory.

The $\mathfrak{W}^+$ neighborhoods of $0 \in \Phi_+$ form a directed set and so for $\phi \in \sum(T)$ the possible families $\{f_U\}$ form nets of smooth vectors which we call the *approximating nets of vectors*; likewise, if $\phi \in \sum'(T)$, the possible families of numbers $\{\mu_U\}$ form nets of complex numbers called the *approximating nets* of *spectral reciprocals*. Let $\phi \in \sum'(T)$ and $\{\mu_U\}$ be any approximating net of spectral reciprocals and let μ be any accumulation point of the net in $\mathbb{C} \cup \{\infty\}$. Since $\mathbb{C} \cup \{\infty\}$ is compact, we see that μ always exists. We also see that by passing to approximating subnets we can achieve μ to be $\lim \mu'_U$ for some other approximating net $\{\mu'_U\}$. From now on we assume that we have picked $\{f_U\}$ such that $\mu = \lim \mu_U$ exists in $\mathbb{C} \cup \{\infty\}$. The set of all such limits for all possible approximating nets we shall call the *reciprocal spectral range* of ϕ. The reciprocal spectral range is clearly a closed set.

We see that we can now rephrase the above definitions into the following: $\phi \in N(T)$ if and only if $0 \in \bar{T}\phi$; $\phi \in \sum'(T)$ and μ is in the reciprocal spectral range if and only if $\phi \in (\lim \sigma T)\phi$ as $\sigma \to \mu$, where $\sigma \to \mu$ by means of nets indexed by the $\mathfrak{W}^+$ neighborhoods of zero in Φ_+; but this implies that we can rephrase the condition further to read $\phi \in (\lim \sigma \bar{T})\phi$ as $\sigma \to \mu$. This last form admits of further analysis. If $\mu \neq 0$, $\mu \neq \infty$, we can rephrase it to $\phi \in \mu \bar{T}\phi$, which is equivalent to $(1/\mu)\phi \in \bar{T}\phi$. If $\mu = \infty$ and $f_U \to \phi$, $\phi - \mu_U T f_U \to 0$, then $1/\mu_U \phi - T f_U \to 0$ also and since $1/\mu_U \phi \to 0$ we see that $T f_U \to 0$ and $\phi \in N(T)$; thus the case $\mu = \infty$ does not lead to any new generalized eigenvectors. If $\mu = 0$, then we keep the condition $\phi \in (\lim \sigma \bar{T})\phi$ as $\sigma \to \mu$.

Before proceeding with the study of generalized eigenvectors, it becomes necessary to point out certain features. This is best done with an example, so consider the discrete Lee model again.

We have calculated $\bar{T}$ in the last section and we reproduce the results:

For $\rho \in d$,

$$\bar{T}\psi = \left\{\left(\lambda_0\psi_0 + \sum_1^\infty \rho_k\psi_k; \cdots, \lambda_k\psi_k + \psi_0\rho_k, \cdots\right)\right\}$$

and, for $\rho \notin d$,

$$\bar{T}\psi = \{(\alpha; \cdots, \lambda_k\psi_k + \psi_0\rho_k, \cdots) \mid \alpha \in \mathbb{C}\}.$$

Also its clear that $\lim \sigma \bar{T}$ as $\sigma \to 0$ is given by: for $\rho \in d$,

$$\lim_{\sigma \to 0} \sigma \bar{T} = 0,$$

for $\rho \notin d$,

$$\left(\lim_{\sigma \to 0} \sigma \bar{T}\right)\psi = \{(\alpha; 0, 0, \cdots) \mid \alpha \in \mathbb{C}\}.$$

From our rephrasing of the definitions the computation of $\sum(T)$ is now entirely straightforward. The results are the following. (We single out only a single representative from each ray.)

For $\rho \in d$ the only possible values of $\lambda = \mu^{-1}$ are first of all those which satisfy the equation

$$\lambda - \lambda_0 - \sum_1^\infty \rho_k^2/(\lambda - \lambda_k) = 0$$

and those λ_k for which $\rho_k = 0$. In the first case, the eigenvector is $(1; \cdots, \rho_k/(\lambda - \lambda_k), \cdots)$ and in the second case it is $(0; 0, \cdots, 0, 1, 0, \cdots)$, where the 1 occupies the kth place. Of course both situations may hold simultaneously, in which case any linear combination of the two eigenvectors is also an eigenvector with the same eigenvalue λ.

For $\rho \notin d$ any value of λ is allowed. If $\infty \neq \lambda \neq \lambda_k$, the eigenvector is $(1; \cdots, \rho_k/(\lambda - \lambda_k), \cdots)$. If $\lambda = \lambda_k$ and $\rho_k \neq 0$, the eigenvector is $(0; 0, \cdots, 0, 1, 0, \cdots)$, where the 1 is in the kth place. If $\lambda = \lambda_k$ and $\rho_k = 0$, any linear combination of the above two eigenvectors are allowed. If $\lambda = \infty$, the eigenvector is $(1; 0, 0, \cdots)$.

Now the case $\rho \in d$ is entirely understandable and the results coincide with the usual ones of bona-fide spectral analysis.

The case $\rho \notin d$ presents several difficulties. In the first place we seemingly have too many eigenvectors for there is no restriction on λ and any complex number will do. Now if $\rho \notin d$ but the sum

$$I(\lambda) = \sum_1^\infty \frac{\rho_k^2}{\lambda - \lambda_k} < \infty$$

for some λ, then this sum converges for all $\lambda \neq \lambda_k$, and a straightforward formal solution to the eigenvalue problem would yield the restriction

$$\lambda - \lambda_0 - \sum_1^\infty \frac{\rho_k^2}{\lambda - \lambda_k} = 0,$$

which is the restriction one would "normally" take. Our theory so far does not give any method for arriving at such restrictions.

Now if $I(\lambda)$ is divergent, consider the effect of introducing a cutoff h into $\rho: \rho'_k = \rho_k h_k$, where $0 < h_k \leq 1$ and where $I_h(\lambda) = \sum (\rho_k^2 h_k^2)/(\lambda - \lambda_k)$ is convergent. For simplicity, take ρ_k and λ_k real,

$\rho_k \neq 0$, and $\lambda_0 < \lambda_1 < \lambda_2 < \cdots$. As $h_k \to 1$ we see that the eigenvalues approach from above the values λ_k except for the lowest one which moves down to $-\infty$. The corresponding eigenvectors approach

$$\left(1, \cdots, \frac{\rho_k}{0}, \cdots, \frac{\rho_m}{\lambda_k - \lambda_m}, \cdots\right)$$

except for the lowest one which approaches $(1; 0, 0, \cdots)$.

Our second difficulty now is whether sense can be made out of singular behavior under cutoff limits.

We will eventually recognize that these problems are related but for now we briefly discuss them separately.

The restriction $\lambda - \lambda_0 - I(\lambda)$ for $\rho \notin d$ involves an infinite sum. An infinite sum is not really a sum and without some external reasoning we are not justified in taking any particular version of it, so we should not at this point be surprised that these sums admit of arbitrary definition. In the same vein, consider the case when $I(\lambda)$ is divergent but

$$I(\lambda) - I(0) = \lambda \sum \frac{\rho_k^2}{(\lambda - \lambda_k)\lambda_k}$$

is convergent, that is, when *subtracted sum* is appropriate. Such an interpretation of an infinite sum may come from physical considerations imposing possibly now the restriction

$$\lambda - \lambda_0 - \lambda \sum \frac{\rho_k^2}{(\lambda - \lambda_k)\lambda_k} = 0$$

but such an interpretation is again external.

We now consider the second problem; we have

$$\lim_{\lambda \to \lambda_k} \left(1; \cdots, \frac{\rho_k}{\lambda - \lambda_k}, \cdots, \frac{\rho_m}{\lambda - \lambda_m}, \cdots\right) = \left(1; \cdots, \frac{\rho_k}{0}, \cdots, \frac{\rho_m}{\lambda_k - \lambda_m}, \cdots\right),$$

but also

$$\lim_{\lambda \to \lambda_k} \frac{\lambda - \lambda_k}{\rho_k} \left(1, \cdots, \frac{\rho_k}{\lambda - \lambda_k}, \cdots, \frac{\rho_m}{\lambda - \lambda_m}, \cdots\right) = (0; 0, \cdots, 0, 1, 0, \cdots),$$

which is likewise a legitimate limit since we do not usually distinguish between multiples of the same vector. This last vector actually lies in $\sum(T)$ with $\lambda = \lambda_k$. The trouble here is that the original vector in the limit had to "diverge on a set of positive measure" and the only limit which can exist is one in which this divergence has been normalized. We note that in this case, even though the limit eigenvector ϕ is in fact smooth, we do not have $T\phi = \lambda_k\phi$, again pointing to the fact that for elementary spectral theory it is the closure $\bar{T}$ that plays the important role.

A way out of these difficulties will be hinted at in the last part of this section where we combine elementary spectral and function theory with certain considerations of analyticity. Before proceeding with this, we must, however, recast our definition of $\sum(T)$ into a still more useful form.

We now introduce the number $\lambda = 1/\mu$ called the *spectral value* or the *eigenvalue*, where we are considering $\mu \in \mathbb{C} \cup \{\infty\}$ to be in the reciprocal spectral range of some generalized eigenvector. For $\phi \in N(T)$, we shall take by definition $\lambda = 0$. We can now rephrase our definition of $\sum(T)$ as follows: If $\lambda \neq \infty$, then $\phi \in \sum(T)$ if and only if $0 \in (\lambda I - \bar{T})\phi$, and if $\lambda = \infty$, then $\phi \in \sum(T)$ if and only if $0 \in \lim (I - \mu\bar{T})\phi$ as $\mu \to 0$, where the limit is taken over nets converging to zero and indexed by the $\mathfrak{W}^+$ neighborhoods of zero in Φ_+.

The first part of this rephrasing is obvious; we prove the second. Suppose $0 \in \lim (I - \mu\bar{T})\phi$ as $\mu \to 0$; then by the nature of the indexing set for the convergent nets we have that $0 \in \lim (I - \mu T)\phi$ as $\mu \to 0$, but this is the same as saying there are approximating nets $f_U \to \phi$, $\mu_U \to 0$ such that $(I - \mu_U T)f_U \to 0$ and so $\phi - \mu_U T f_U \to 0$ and $\phi \in \sum'(T)$ with $\lambda = \infty$; conversely, if $\phi \in \sum'(T)$ with $\lambda = \infty$, then the above reasoning proceeds backwards to the conclusion $0 \in \lim (I - \mu\bar{T})\phi$ as $\mu \to 0$.

The final version of our definition is the following: $\phi \in \sum(T)$ and $\lambda \neq \infty$ if and only if $\phi \in (\lambda I - \bar{T})^{-1}0$; $\phi \in \sum(T)$ and $\lambda = \infty$ if and only if

$$\phi \in \lim_{\mu \to 0} (I - \mu\bar{T})^{-1}0 = \lim_{\lambda \to \infty} \lambda(\lambda I - \bar{T})^{-1}0,$$

where the limit is taken over convergent nets indexed by the $\mathfrak{W}^+$ neighborhood of zero in Φ_+. The second half of the above contention follows from the fact that $(\lim_N \mathfrak{Q}_\xi)^{-1} = \lim_N \mathfrak{Q}_\xi^{-1}$.

We now define the relations

$$\mathfrak{R}_\lambda(\bar{T}) = (\lambda I - \bar{T})^{-1},$$
$$\mathfrak{S}_\infty(\bar{T}) = \lim_{\lambda \to \infty} \lambda \mathfrak{R}_\lambda(\bar{T}),$$

where the limit is taken over converging nets indexed by the $\mathfrak{W}^+$ neighborhoods of zero in Φ_+; $\mathfrak{S}_\infty(\bar{T})$ is therefore closed as is $\mathfrak{R}_\lambda(\bar{T})$. The relation $\mathfrak{R}_\lambda$ is called the *resolvent relation*.

We also define

$$\sum_{\lambda}(T) = \mathcal{R}_{\lambda}(\bar{T})0, \quad \lambda \neq \infty,$$
$$\sum_{\infty}(T) = \mathcal{S}_{\infty}(\bar{T})0$$

and call $\sum_{\lambda}(T)$ the space of *generalized eigenvectors with eigenvalue* λ, or more briefly, the λ *multiplet* of T.

We have

$$\sum(T) = \bigcup_{\lambda \in \mathbb{C} \cup \{\infty\}} \sum_{\lambda}(T).$$

Just as $\phi \in \sum_{\lambda}(T)$ is in a certain sense a solution to the formal equation $T\phi = \lambda\phi$, so $\psi \in \mathcal{R}_{\lambda}\phi$ and $\psi \in \mathcal{S}_{\infty}\phi$ are in the same sense solutions to the formal equations $(\lambda - T)\psi = \phi$ and $(\infty - T)\psi = \phi$, respectively.

Let us return to the resolvent relations. By a *resolvent* of T we shall mean a family of g-operator versions $R_{\lambda}(T)$ of $\mathcal{R}_{\lambda}(\bar{T})$. For useful results we must impose on this family enough additional restrictions to recover as much as possible the properties of the usual resolvents of bona-fide operators. Here we consider only two such: hermiticity and analyticity. If T has a dual T^* as a generalized operator, then $R_{\lambda}(T)$ will be said to be *Hermitian* if there is a resolvent $R_{\lambda}(T^*)$ of T^* such that $R_{\lambda}(T) = R_{\bar{\lambda}}(T^*)^*$. In addition, we may want $R_{\lambda}(T)$ to be as analytic in λ as possible; this is not a well-defined notion here, but vaguely speaking what we want is that all other versions with larger domains of analyticity be in some sense less useful and also that the severity of the singularities at points of non-holomorphy be as small as possible. These questions will be dealt with in greater detail in the next section, but here let us see what these requirements suggest for our example of the discrete Lee model.

Let us again assume for simplicity that λ_0, λ_k, and ρ_k are real, that $\rho_k \neq 0$, that $\lambda_1 < \lambda_2 < \cdots < \lambda_k < \lambda_{k+1} < \cdots$, and that this sequence has no finite accumulation point. By a straightforward calculation we find that for $\lambda \neq \lambda_k$, $k \geq 1$,

$$\mathcal{R}_{\lambda}\psi = \left\{ \left(\alpha; \cdots, \frac{\psi_k + \alpha\rho_k}{\lambda - \lambda_k}, \cdots \right) \middle| \alpha \in \mathbb{C} \right\}$$

and that for $\lambda = \lambda_k$, $k \geq 1$,

$$\mathcal{R}_{\lambda_k}\psi = \left\{ \left(\frac{-\psi_k}{\rho_k}; \cdots, \gamma, \cdots, \psi_m - \frac{\psi_k}{\rho_k}\rho_m \Big/ (\lambda_k - \lambda_m), \cdots \right) \middle| \gamma \in \mathbb{C} \right\},$$

where the γ stands in the kth place and $m \neq k$.

Also

$$\mathcal{S}_{\infty}\psi = \{(\alpha; \psi_1, \psi_2, \cdots, \psi_k, \cdots) \mid \alpha \in \mathbb{C}\}.$$

The g-operator versions of these are

$$R_{\lambda}f = \left(\alpha(\lambda)f; \cdots, \frac{f_k + [\alpha(\lambda)f]\rho_k}{\lambda - \lambda_k}, \cdots \right) \qquad \lambda \neq \lambda_k, \quad k \geq 1,$$

$$R_{\lambda_k}f = \left(\frac{-f_k}{\rho_k}; \cdots, \gamma_k f, \cdots, f_m - \frac{f_k}{\rho_k}\rho_m \Big/ (\lambda_k - \lambda_m), \cdots \right), \quad k \geq 1,$$

$$S_{\infty}f = (\eta f; f_1, f_2, \cdots, f_k, \cdots),$$

where we have

$$\alpha(\lambda)f = \alpha(\lambda)_0 f_0 + \sum_1^{\infty} \alpha(\lambda)_k f_k,$$
$$\gamma_k f = \gamma_{k0} f_0 + \sum_1^{\infty} \gamma_{kl} f_l,$$
$$\eta f = \eta_0 f_0 + \sum_1^{\infty} \eta_k f_k$$

are all linear functionals on Φ_-.

Imposing Hermiticity now imposes the following restrictions which are easily calculated:

$$\alpha(\lambda)_k = \alpha(\lambda)_0 \frac{\rho_k}{\lambda - \lambda_k},$$
$$\alpha(\lambda)_0 = \overline{\alpha(\bar{\lambda})_0};$$
$$\gamma_{k0} = \frac{-1}{\rho_k}, \quad \gamma_{kl} = \frac{-1}{\rho_k}\frac{\rho_l}{\lambda_k - \lambda_l} \quad \text{for} \quad l \neq k,$$
$$\gamma_{kk} = \overline{\gamma_{kk}};$$
$$\eta_k = 0, \quad k \geq 1, \quad \eta_0 = \overline{\eta_0}.$$

It is useful to exhibit explicitly the Hermitian resolvents: for $\lambda \neq \lambda_k$, $k \geq 1$,

$$R_{\lambda}(T)f = \left(\alpha(\lambda)_0 f_0 + \alpha(\lambda)_0 \sum_1^{\infty} \frac{f_k\rho_k}{\lambda - \lambda_k}; \cdots, \frac{f_k}{\lambda - \lambda_k} + \alpha(\lambda)_0 \frac{f_0\rho_k}{\lambda - \lambda_k} + \alpha(\lambda)_0 \left(\sum_1^{\infty} \frac{f_l\rho_l}{\lambda - \lambda_l} \right) \frac{\rho_k}{\lambda - \lambda_k}, \cdots \right),$$

for $\lambda = \lambda_k$,

$$R_{\lambda_k}(T) = \left(\frac{-f_k}{\rho_k}; \cdots, \frac{-1}{\rho_k} \times \left(f_0 + \sum_{\substack{m=1 \\ m \neq k}}^{\infty} \frac{f_m\rho_m}{\lambda_k - \lambda_m} + \gamma_{kk} f_k \right), \cdots, \frac{f_m - (f_k/\rho_k)\rho_m}{\lambda_k - \lambda_m}, \cdots \right),$$

where the term involving the infinite sum stands in the kth place.

We are now left with the problem of determining the function $\alpha(\lambda)_0$ subject to the reflection condition $\alpha(\lambda)_0 = \overline{\alpha(\bar{\lambda})_0}$ and also the real numbers γ_{kk}. The analyticity properties of $R_\lambda(T)$ are determined by these choices. Now if $\alpha(\lambda)_0$ is analytic in a domain $\mathfrak{D}$, then so is R_λ except perhaps for the points λ_k and the point ∞. Let us determine under what conditions R_λ is analytic at $\lambda = \lambda_k$; certainly $\alpha(\lambda)_0$ must have a zero there,which we may want in any case in order to avoid a second-order pole in R_λ there; furthermore,

$$\frac{f_k}{\lambda - \lambda_k} + \alpha(\lambda)_0 \frac{\rho_k^2}{(\lambda - \lambda_k)^2} f_k$$

must be regular there,which implies that the zero must be of the form $-(1/\rho_k^2)(\lambda - \lambda_k)$. In addition we must have $\lim R_k(T) = R_{\lambda_k}(T)$ as $\lambda \to \lambda_k$ which determines γ_{kk}. It is convenient to look at $\tau(\lambda) = 1/\alpha(\lambda)$. Let us for example decide we want $R_\lambda(T)$ to be analytic at each λ_k; then $\tau(\lambda)$ must have a pole of the form $-\rho_k^2/(\lambda - \lambda_k)$. Assume in addition that $\tau(\lambda)$ is analytic at all other points. Thus, if $\sum_1^\infty \rho_k^2/(\lambda - \lambda_k)$ converges, then the most general form of $\tau(\lambda)$ is

$$\tau(\lambda) = E(\lambda) - \sum_1^\infty \frac{\rho_k^2}{\lambda - \lambda_k},$$

where $E(\lambda)$ is entire. Further detailed considerations will show that $E(\lambda) = \lambda + \sigma$, where σ is a real constant is especially singled out,and thus $\tau(\lambda) = \sigma + \lambda - \sum_1^\infty \rho_k^2/(\lambda - \lambda_k)$. Likewise, if $\sum_1^\infty \rho_k^2/(\lambda - \lambda_k)$ diverges but the subtracted sum $\lambda \sum_1^\infty \rho_k^2/(\lambda - \lambda_k)\lambda_k$ converges,the same considerations will single out the $\tau(\lambda)$ of the form $\sigma + \lambda - \lambda \sum_1^\infty \rho_k^2/(\lambda - \lambda_k)\lambda_k$.

Analyticity arguments therefore provide us with some of the external reasoning needed to interpret infinite sums and to produce the correct subtractions when necessary. The relevant eigenvalues are of course now determined by the poles in R_λ,which in the two cases above are located by the conditions

$$\sigma + \lambda - \sum_1^\infty \frac{\rho_k^2}{\lambda - \lambda_k} = 0$$

or

$$\sigma + \lambda - \lambda \sum_1^\infty \frac{\rho_k^2}{(\lambda - \lambda_k)\lambda_k} = 0,$$

whichever is appropriate. There are of course other possible choices for $\alpha(\lambda)_0$ and these may correspond to different interpretations of the model.

G. Elementary Orthogonal Methods

There is another approach to resolvents that complements the multivalued methods of picking a restriction of $\mathfrak{R}_\lambda$ and leads to more stringent results. Its application,however,requires an insight into the specific properties of a given generalized operator.

Let T be a generalized operator and suppose Φ is a direct sum $\Phi_1 \oplus \Phi_2$. Then,as was pointed out before, we can write T as a matrix

$$T = \begin{pmatrix} T_{11}, & T_{12} \\ T_{21}, & T_{22} \end{pmatrix}.$$

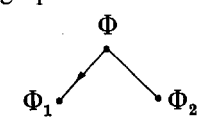
We now formally compute $(\lambda - T)^{-1}$ in terms of the matrix elements. Let

$$P_\lambda = (\lambda - T_{22})^{-1},$$
$$T_\lambda = T_{11} + T_{12} P_\lambda T_{21},$$
$$Q_\lambda = (\lambda - T_\lambda)^{-1};$$

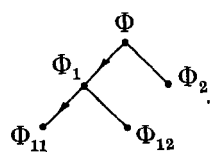
then one can easily show that

$$R_\lambda = (\lambda - T)^{-1} = \begin{pmatrix} Q_\lambda, & Q_\lambda T_{12} P_\lambda \\ P_\lambda T_{21} Q_\lambda, & P_\lambda + P_\lambda T_{21} Q_\lambda T_{12} P_\lambda \end{pmatrix}.$$

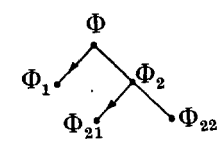
The actual procedure to follow is to compute P_λ and Q_λ by some method, such as the multivalued approach, and to compute T_λ and R_λ by analytic subtractions whenever appropriate. The above procedure, of course, may have been in turn applied to computing P_λ and Q_λ themselves so that we can get a (possibly infinite) nesting of this method. One can exhibit such nestings in terms of certain tree graphs. Note that in the above decomposition Φ_1 is distinguished from Φ_2 and we express the above decomposition by the graph



If furthermore Q_λ were computed by orthogonal methods by decomposing Φ_1 into $\Phi_{11} \oplus \Phi_{12}$ with Φ_{11} distinguished,we would express it by the graph



On the other hand



states that P_λ was computed by orthogonal decomposition of Φ_2 into $\Phi_{21} \oplus \Phi_{22}$ with Φ_{21} distinguished. It is clear now how to construct a graph for a general nesting and it should be possible to find rules for expressing R_λ, once the graph is given, in terms of the matrix entries such as was done for the simple decomposition above. The above procedure is one that is most closely related to the subtraction methods of conventional perturbation theory.

The use of orthogonal methods is of course predicated on the assumption that the direct sum decompositions of Φ are somehow naturally adapted to the generalized operator on hand.

As an example consider the discrete Lee model and let Φ be decomposed as $\mathbb{C} \oplus \Phi_2$ as in Sec. IIC. We then have

$$T_{11}: c \rightsquigarrow \lambda_0 c,$$
$$T_{12}: \{f_k\} \rightsquigarrow \sum \rho_k f_k,$$
$$T_{21}: c \rightsquigarrow \{c\rho_k\},$$
$$T_{22}: \{f_k\} \rightsquigarrow \{\lambda_k f_k\}.$$

We have

$$P_\lambda: \{f_k\} \rightsquigarrow \left\{\frac{f_k}{\lambda - \lambda_k}\right\}, \quad \lambda \neq \lambda_k.$$

Let F be any finite-dimensional subspace of Φ_{-2} and let it be generated by the orthonormal set $f^{(1)}, f^{(2)}, \cdots, f^{(N)}$. Consider $T_{12}\pi_F P_\lambda T_{21}$; we have

$$T_{12}\pi_F P_\lambda T_{21}: c \rightsquigarrow c \sum_{i=1}^{N} \sum_{k,l=1}^{\infty} \frac{\rho_l f_l^{(i)} \rho_k f_k^{(i)}}{\lambda - \lambda_k}.$$

Assume ∞ is the only accumulation point of the λ_k. If F belongs to a cofinal set, then eventually it contains any set of the form $e_1, e_2, \cdots, e_M$, $M \leq N$ and so we have eventually

$$T_{12}\pi_F P_\lambda T_{21}: c \rightsquigarrow c \times \left(\sum_{k=1}^{M} \frac{\rho_k^2}{\lambda - \lambda_k} + \sum_{i=M+1}^{N} \sum_{k,l=M+1}^{\infty} \frac{\rho_l f^{(i)} \rho_k f_k^{(i)}}{\lambda - \lambda_k} \right);$$

and if this is to have a subtracted limit in $\lambda \neq \lambda_k$, it can only be of the form

$$c \rightsquigarrow c\left(\tilde{\Sigma} \frac{\rho_k^2}{\lambda - \lambda_k}\right),$$

where

$$\tilde{\Sigma} \frac{\rho_k^2}{\lambda - \lambda_k} = E(\lambda) + \sum \left(\frac{\lambda}{\lambda_k}\right)^{n_k} \frac{\rho_k^2}{\lambda - \lambda_k}$$

and $E(\lambda)$ is entire. If a finite number of subtractions are sufficient, then the subtracted sum can be taken to be

$$P_{n-1}(\lambda) + \lambda^n \sum \frac{\rho_k^2}{(\lambda - \lambda_k)\lambda_k^n},$$

where n is taken to be the lowest possible integer and P_{n-1} is a polynomial of degree $n-1$. If we set $\alpha(\lambda) = (\lambda - \lambda_0 - \sum \rho_k^2/(\lambda - \lambda_k))^{-1}$, then we find that $Q_\lambda: c \rightsquigarrow \alpha(\lambda)c$ and that a computation of R_λ involves no further ambiguities and is given by

$$R_\lambda f = \left(\alpha(\lambda)\left(f_0 + \sum \frac{\rho_l f_l}{\lambda - \lambda_l}\right); \cdots, \frac{f_k}{\lambda - \lambda_k} + \alpha(\lambda)\left(f_0 + \sum \frac{\rho_l f_l}{\lambda - \lambda_l}\right)\frac{\rho_k}{\lambda - \lambda_k}, \cdots \right).$$

We shall obtain the same result by other methods in the next chapter.

H. Remarks

The discussion in this chapter has been very general and it introduces just enough assumptions to be able to define the important basic concepts. We are yet to determine what additional assumptions are needed to insure a successful program. We have made practically no topological assumptions concerning Φ_-, Φ_+, j, and T; this is a difficult problem but the following remarks are appropriate.

We have already spoken in Sec. IIA of the possibility that for specific applications certain choices of the rigging may be more natural than others. Now for the multivalued approach to the resolvent we feel it is in a sense natural to pick the riggings that exhibit all the possible generalized eigenvectors; we make this notion understandable by the following examples: Let $T = (1/i)d/dx$ on $\mathcal{S}(\mathbb{R}) \subset \mathcal{S}'(\mathbb{R})$; then the only generalized eigenvectors are $e^{i\lambda x}$, λ real; however, in going to $\mathcal{D}(\mathbb{R}) \subset \mathcal{D}'(\mathbb{R})$ by the procedure of Sec. IIA we acquire all the eigenvectors $e^{i\lambda x}$, λ complex. As another example let T be multiplication by the function x on $\mathcal{D}(\mathbb{R}) \subset \mathcal{D}'(\mathbb{R})$. The only eigenvectors are $\delta(x - \lambda)$, λ real, but by passing first to $\mathcal{S}(\mathbb{R}) \subset \mathcal{S}'(\mathbb{R})$ and then to $\mathcal{E}(\mathbb{R}) \subset \mathcal{E}'(\mathbb{R})$, where $\mathcal{E}(\mathbb{R}) \subset \mathcal{S}(\mathbb{R})$ is the Fourier transform of $\mathcal{D}(\mathbb{R})$ and $\mathcal{E}'(\mathbb{R})$ is the Fourier transform of $\mathcal{D}'(\mathbb{R})$, we obtain all the eigenvectors $\delta(x - \lambda)$, λ complex. Eventually, however, we expect in many cases to get no new further eigenvectors. We are reminded here of the regularity theorems for differential equations in which the introduction of the possibility that the solutions may lie in a larger space does not introduce new solutions.

The existence of an appropriately modified rigged Hilbert space which in some sense exhibits all of the possible eigenvectors of the original generalized operator T in the original rigged Hilbert space Φ must of course depend on certain regularity properties of the original matrix elements of T and when these

properties hold, one should be able to construct these natural rigged Hilbert spaces in which to exhibit T. We shall then say we have a *complete rigging* for T. What the appropriate regularity conditions are we have not determined but for generalized operators that arise from physical considerations we can expect no unusual pathologies and so these generalized operators form a class of examples by which the problem can be approached.

Now one may think originally that by choosing Φ_- and Φ_+ sufficiently well only the physically relevant eigenvectors will remain, as it happens for $(1/i)d/dx$ and more generally for the Gel'fand–Vilenkin riggings for self-adjoint and unitary bona fide operators. We, however, do not expect this to be a feasible program in general. It is very hard to see how the subtracted sum $\lambda \sum (\rho_k^2)/(\lambda - \lambda_k)\lambda_k$, which we met in the previous sections, can arise out of an appropriate choice of Φ_- and Φ_+ and appropriate topological restrictions. Analyticity provides an insight into this problem. We are again reminded of the theory of differential operators. Thus even though a solution to a differential eigenvalue problem $Lf = \lambda f$ may correspond to the Hilbert space spectrum of L only for certain values of λ, solutions exist for other values and have certain analyticity properties in λ. If L is sufficiently well behaved, then these analytic properties contain all the information needed to recover the Hilbert space structure of L. In our case, analyticity has to be imposed from the outside rather than existing *a priori* because we have allowed ourselves quite a lot of freedom in using weak topologies and multivalued maps; this freedom, however, was necessary to develop an elementary function theory that avoids ambiguous divergent expressions. In the end, we try to recover what we lost by imposing analyticity and other assumptions.

Whenever orthogonal methods are used, then one has to be able to choose riggings in such a way that Φ is appropriately a direct sum of other rigged Hilbert spaces. In a way this requirement can run counter to looking for complete riggings, as will become more apparent later, but is nevertheless a natural procedure.

Underlying the above procedures is of course the view that the analytic structure of the physically relevant mathematical objects for the case of infinitely many degrees of freedom is somehow similar to the analytic structure of similar objects in the theory of differential equations. Resolvent theory for generalized operators is an attempt to express this view in precise terms, and it is the task of the next chapter to develop these ideas in greater detail for the case of symmetric generalized operators.

III. RESOLVENT THEORY FOR SYMMETRIC GENERALIZED OPERATORS

Let T be a symmetric generalized operator on a rigged Hilbert space Φ. We shall here try to obtain insight into the structure of T by two methods. The first of these is to look for g-operator versions $R_\lambda(T) \subset \mathcal{R}_\lambda(\bar{T})$ and $S_\infty(T) \subset \mathcal{S}_\infty(\bar{T})$ which satisfy certain conditions to be listed below. These conditions should be considered as merely attempts and a given resolvent may or may not satisfy them or else satisfy them only partially. For reasons to become clear later we shall not endeavor to prove under what hypotheses these conditions can be met; furthermore, certain parts of this chapter will invoke extramathematical reasoning based on notions current in existing physical theories. What this method provides us with is a sieve by which we try to isolate the intrinsic structural properties of T from the ambiguities inherent in the multivalued approach of the previous section. Unfortunately, the sieve is too coarse to deal effectively with many problems but we have good reasons to believe that a refinement of these methods together with other approaches will soon provide tools for dealing with such problems.

The second approach is to use orthogonal methods as outlined in Sec. IIG. Of course this approach must use a decomposition of the rigged Hilbert space and directed sets of finite-dimensional subspaces of Φ_- that are somehow specifically adapted to T. We have not developed a theory of such relationships but will use what seems appropriate in a given situation.

A. Mathematical Renormalization

Let K be a positive generalized operator; then one can introduce a Hilbert space H_K which consists of the equivalence classes of K-Cauchy sequences of elements of Φ_-. Namely, a sequence $\{f_n\} \subset \Phi_-$ is called *K-Cauchy* if for any $\epsilon > 0$ there is an N such that $n, m \geq N$ implies $\langle f_n - f_m, K(f_n - f_m)\rangle < \epsilon$. Two such sequences $\{f_n\}$ and $\{g_n\}$ are *K-equivalent* $\{f_n\} \tilde{}_K \{g_n\}$ if $\lim \langle f_n - g_n, K(f_n - g_n)\rangle = 0$ as $n \to \infty$. The linear operations $\lambda\{f_n\} = \{\lambda f_n\}$ and $\{f_n\} + \{g_n\} = \{f_n + g_n\}$ do not conflict with the introduction of K equivalence and so the set of equivalence classes forms a well-defined linear space. The introduction of the inner product $(\{f_n\}, \{g_n\})_K = \lim \langle f_n, Kg_n\rangle$ as $n \to \infty$ is also well-defined and makes the set of equivalence classes of K-Cauchy sequences into a Hilbert space.

There is a canonical linear map $\Delta_K : \Phi_- \to H_K$ which assigns to $f \in \Phi_-$ the constant sequence $\Delta_K f = \{f, f, f, \cdots\}$. The map Δ_K is not necessarily an inclusion, for we may have $\Delta_K f = 0$, that is, $\langle f, Kf\rangle = 0$.

The subspace $\Delta_K\Phi_- \subset H_K$ is clearly dense for $\{f_n\} = \lim \Delta_K f_n$ as $n \to \infty$.

Let now $J \geq K \geq 0$ be two positive generalized operators. We see that every J-Cauchy sequence is also K-Cauchy and if $\{f_n\}\tilde{J}\{g_n\}$, then also $\{f_n\}\tilde{K}\{g_n\}$. Thus, there is a map $\rho_{KJ}: H_J \to H_K$ which takes $\{f_n\} \in H_J$ to $\rho_{KJ}\{f_n\} = \{f_n\} \in H_K$; ρ_{KJ} thus does nothing to the elements of the sequence but it reinterprets it in a different Hilbert space. We obviously have $\rho_{KJ}\Delta_J = \Delta_K$ and if $J \geq K \geq L \geq 0$, then we have $\rho_{LK}\rho_{KJ} = \rho_{LJ}$. Furthermore,we have

$$\begin{aligned}\|\rho_{KJ}\Delta_J f\|_J^2 &= \|\Delta_K f\|_K^2 \\ &= \langle f, Kf\rangle \leq \langle f, Jf\rangle = \|\Delta_J f\|_J^2,\end{aligned}$$

from which we can conclude that $\|\rho_{KJ}\| \leq 1$.

Introduce now the sesquilinear form $h, k \rightsquigarrow (\rho_{KJ}h, \rho_{KJ}k)_K$ on H_J; we have

$$|(\rho_{KJ}h, \rho_{KJ}k)_K| \leq \|\rho_{KJ}h\|_K \|\rho_{KJ}k\|_K \leq \|h\|_J \|k\|_J$$

and by a standard theorem we conclude that there is a bounded operator K_J in H_J such that

$$(\rho_{KJ}h, \rho_{KJ}k)_K = (h, K_J k)_J.$$

From this we deduce that

$$0 \leq (h, K_J h)_J = \|\rho_{KJ}h\|_K^2 \leq \|h\|_J^2$$

so that $0 \leq K_J \leq 1$. We likewise have

$$(h, K_J k)_J = (\rho_{KJ}h, \rho_{KJ}k)_K = (h, \rho_{KJ}^*\rho_{KJ}k)_J$$

so that

$$K_J = \rho_{KJ}^*\rho_{KJ}.$$

Define now the map $r_{JK}: \Delta_K\Phi_- \to H_J$ by

$$r_{JK}\Delta_K f = (K_J)^{\frac{1}{2}}\Delta_J f.$$

We have

$$\begin{aligned}\|K_J^{\frac{1}{2}}\Delta_J f\|_J^2 &= (K_J^{\frac{1}{2}}\Delta_J f, K_J^{\frac{1}{2}}\Delta_J f)_J \\ &= (\Delta_J f, K_J\Delta_J f)_J \\ &= (\rho_{KJ}\Delta_J f, \rho_{KJ}\Delta_J f)_K \\ &= (\Delta_K f, \Delta_K f) = \|\Delta_K f\|_K^2,\end{aligned}$$

so that r_{JK} is well defined and is in fact an isometry. By continuity we can extend r_{JK} to an isometry, again called r_{JK}, between H_K and the closure of the range of $K_J^{\frac{1}{2}}$:

$$r_{JK}: H_K \approx \overline{\text{range}\,(K_J^{\frac{1}{2}})} \subset H_J.$$

Now

$$\|r_{JK}\rho_{KJ}\Delta_J f\|_J^2 = \|r_{JK}\Delta_K f\|_J^2 = \|K_J^{\frac{1}{2}}\Delta_J f\|_J^2$$

and we have

$$r_{JK}\rho_{KJ} = K_J^{\frac{1}{2}}.$$

Of particular interest is the case when K_J is an orthogonal projection, which we shall denote by Q_{KJ}. If this be the case, we say that $J \geq K$ is a *projective pair*. We then have $K_J^{\frac{1}{2}} = Q_{KJ}^{\frac{1}{2}} = Q_{KJ} = K_J$ and so $r_{JK}\rho_{KJ} = \rho_{KJ}^*\rho_{KJ}$; but this implies that

$$r_{JK}\rho_{KJ}\Delta_J f = \rho_{KJ}^*\rho_{KJ}\Delta_J f \text{ or } r_{JK}\Delta_K f = \rho_{KJ}^*\Delta_K f$$

and we have that *for a projective pair*

$$r_{JK} = \rho_{KJ}^*.$$

Furthermore,

$$\begin{aligned}\|\rho_{KJ}\rho_{KJ}^*\Delta_K f\|_K^2 &= \|\rho_{KJ}r_{JK}\Delta_K f\|_K^2 \\ &= \|\rho_{KJ}Q_{KJ}^{\frac{1}{2}}\Delta_J f\|_K^2 \\ &= \|Q_{KJ}\Delta_J f\|_J^2 = \|\Delta_K f\|_K^2\end{aligned}$$

and we have *for a projective pair*

$$\rho_{KJ}\rho_{KJ}^* = 1_K.$$

Suppose now $J \geq K \geq L \geq 0$ and the pairs $J \geq K$ and $K \geq L$ are projective. We first show that $J \geq L$ is also projective. Indeed we have $\rho_{LK}\rho_{KJ} = \rho_{LJ}$ and so

$$L_J = \rho_{LJ}^*\rho_{LJ} = \rho_{KJ}\rho_{LK}^*\rho_{LK}\rho_{KJ} = \rho_{KJ}^*L_K\rho_{KJ};$$

consequently,

$$\begin{aligned}L_J^2 &= \rho_{KJ}^*L_K\rho_{KJ}\rho_{KJ}^*L_K\rho_{KJ} \\ &= \rho_{KJ}^*L_K 1_K L_K\rho_{KJ} = \rho_{KJ}^*L_K\rho_{KJ} = L_J,\end{aligned}$$

proving that L_J is a projection. By taking adjoints in the equality $\rho_{LK}\rho_{KJ} = \rho_{LJ}$ we now find the important transitivity relation

$$r_{JK}r_{KL} = r_{JL}.$$

We now prove a few simple results which we dignify to the status of lemmas for convenient future referral.

Lemma 1: Let $L_i \geq 0$, $i = 1, 2, \cdots$ be a countable family of positive generalized operators such that the weak sum $L = \sum L_i$ exists;then $\sum (L_i)_L = 1$, where the sum is strong.

Proof: For $f, g \in \Phi_-$ we have

$$\begin{aligned}(\Delta_L g, \Delta_L f)_L &= \langle g, Lf\rangle = \sum \langle g, L_i f\rangle \\ &= \sum (\Delta_{L_i} g, \Delta_{L_i} f)_{L_i} \\ &= \sum (\Delta_L g, (L_i)_L \Delta_L f)_L\end{aligned}$$

and so *on* $\Delta_L\Phi_-$ we have the weak sum $\sum (L_i)_L = 1$. Since $0 \leq (L_i)_L \leq 1$, we have that any partial sum $\sum_1^N (L_i)_L$ is positive and so,on $\Delta_L\Phi_-$, $0 \leq \sum_1^N (L_i)_L \leq 1$ and so by continuity $0 \leq \sum_1^N (L_i)_L \leq 1$ on H_L; the partial sums are therefore uniformly bounded in norm and we conclude that $\sum (L_i)_L = 1$ weakly on all of H_L.

To show that the sum is strong,we invoke the result of Hilbert space theory that a bounded converging weak sum of positive operators in fact converges strongly to the same limit■

Lemma 2: If on a Hilbert space H,we have $1 = \sum P_i$, where P_i, $i = 1, 2, \cdots$, are orthogonal projections; then $P_iP_j = 0$ for $i \neq j$, that is,we have an orthogonal decomposition of 1.

Proof: For $h \in P_iH$ we have $h = P_ih + \sum_{j\neq i} P_jh$ or $\sum_{j\neq i} P_jh = 0$;but each P_j is positive,so we have

$$0 \leq (h, P_jh) = -\left(h, \sum_{k\neq i,j} P_kh\right) \leq 0;$$

so $P_jh = 0$,showing that $P_jP_i = 0$ for $j \neq i$■

Combining the above two lemmas,we have immediately:

Lemma 3: Let $L_i \geq i = 1, 2, \cdots$ be a countable family of positive generalized operators such that the weak sum $L = \sum L_i$ exists and such that each pair $L \geq L_i$ is projective; then the sum $1 = \sum (L_i)_L$ is an orthogonal decomposition of 1■

Lemma 4: Let $L_i \geq 0$ $i = 1, 2, \cdots$ be a countable family of generalized operators such that the weak sum $L = \sum L_i$ exists. Introduce the sesquilinear forms $S_i(h, k)$ on H_L by $S_i(h, k) = (\rho_{L_iL}h, \rho_{L_iL}k)_{L_i} = (h, (L_i)_Lk)_L$ and let $S_i(h) = S_i(h, h)$ be the corresponding quadratic forms. Let $N_i = \{h \in H_L \mid S_i(h) = 0\}$; we call N_i the subspace of H_L degenerate with respect to L_i. We now claim that $(L_i)_L$ are all orthogonal projections if and only if the subspaces $N_i^\perp$ form an orthogonal decomposition of H_L.

Proof: The necessity was already established in Lemma 3 since if $(L_i)_L$ are all projections, then $1 = \sum (L_i)_L$ is an orthogonal decomposition and $N_i^\perp = (L_i)_LH_L$.

To prove sufficiency,let $H_L = \oplus N_i^\perp$ be on orthogonal decomposition and let π_i be the projection onto $N_i^\perp$. Then

$$\begin{aligned} S_i(h, k) &= S_i(h, \pi_ik + (1 - \pi_i)k) \\ &= S_i(h, \pi_ik) + S_i(h, (1 - \pi_i)k) \end{aligned}$$

but by the Schwartz inequality

$$|S_i(h, (1 - \pi_i)k)|^2 \leq S_i(h)S_i((1 - \pi_i)k) = 0$$

since $(1 - \pi_i)k \in N_i$. Therefore $S_i(h, k) = S_i(h, \pi_ik)$. Now by Lemma 1, $(h, k)_L = \sum S_j(h, k)$ so we have $(h, \pi_ik)_L = \sum_j S_j(h, \pi_ik)$ but,for $j \neq i$, $S_j(h, \pi_ik) = S_j(h, \pi_j\pi_ik) = 0$ since $\pi_j\pi_i = 0$ by assumption. Thus $(h, \pi_ik) = S_i(h, \pi_ik) = S_i(h, k) = (h, (L_i)_Lk)_L$ and we have $(L_i)_L = \pi_i$ an orthogonal projection■

A corollary to Lemma 4 is the following lemma:

Lemma 5: Let $J \geq K \geq 0$ be positive generalized operators; let $S_K(h) = (h, K_Jh)$; and let $N_K = \{h \in H_J \mid S_K(h) = 0\}$. We claim that K_J is a projection if and only if for $f \in N_K^\perp$ we have $S_K(f) = \|f\|_J^2$.

Proof: We have $J = (J - K) + K$ and $J - K \geq 0$, $K \geq 0$, so by Lemma 1 K_J is a projection if and only if $(J - K)_J$ is. Thus, by Lemma 4 $H_J = N_K^\perp \oplus N_{J-K}^\perp$ and so if $f \in N_K^\perp$, then $S_{J-K}(f) = 0$ and $\|f\|_J^2 = S_K(f) + S_{J-K}(f) = S_K(f)$■

Now let A be a directed set and $\{J_\alpha\}_{\alpha\in A}$ a family of positive generalized operators such that for $\beta \geq \alpha$, $J_\beta \geq J_\alpha$, and the pair is projective. Such a family we shall call a *projective net*. For the sake of notational ease we shall use just the symbol α as a subscript where otherwise we would have used the symbol J_α.

Introduce now the family of rigged Hilbert spaces $\Phi_\alpha = (H_\alpha, H_\alpha, \mathrm{id}_\alpha)$; then it is readily apparent that by defining $k_{-\beta\alpha} = r_{\beta\alpha} = k_{+\beta\alpha}$ we have an inductive family of rigged Hilbert spaces. Let $\Psi = \varinjlim \Phi_\alpha$ be the inductive limit of the family $\{\Phi_\alpha\}$ and likewise let $\Delta = \varinjlim \Delta_\alpha$ be the inductive limit of the family of maps $\Delta_\alpha: \Phi_- \to \Phi_{+\alpha} = H_\alpha$. The maps Δ_α and Δ will be called *dressing transformations*.

Define the canonical conclusion $k_\alpha: H_\alpha \to \Psi_0$ by the composition of canonical inclusions $j_-k_{-\alpha}: \Phi_{-\alpha} = H_\alpha \to \Psi_- \to \Psi_0$. We have $k_\beta r_{\beta\alpha} = k_\alpha$. The inclusion k_α is an isometry for

$$\|k_\alpha h_\alpha\|^2 = \langle k_{-\alpha}h_\alpha, jk_{-\alpha}h_\alpha\rangle = \|h_\alpha\|_\alpha^2$$

by inductive limit theory. Thus, each H_α can be identified as a closed subspace of Ψ_0; let E_α be the orthogonal projection in Ψ_0 onto this subspace.

The rigged Hilbert space Ψ is called the *renormalized rigged Hilbert space*; the elements of Ψ_0 are called the *renormalized vectors*, and the elements of Ψ_+ are called the *dressed vectors*.

The family $\{E_\alpha\}$ is called the *renormalized projective net*.

Now let $\mathcal{A}$ be a σ ring of subsets of some set X and suppose $\mathcal{A}$ is generated by a subring $\mathcal{A}^0$. We do not assume $\mathcal{A}^0$ is necessarily a σ ring. By a *complex measure μ on $\mathcal{A}$, finite on $\mathcal{A}^0$*,we mean a complex combination $\mu = m_+ - m_- + i(n_+ - n_-)$ of finite positive measures m_+, m_-, n_+, and n_- on $\mathcal{A}^0$. By standard measure theory we can extend the four finite positive measures to $\mathcal{A}$ but on $\mathcal{A}$ they need not be finite. We set $|\mu| = m_+ + m_- + n_+ + n_-$.

If J is a map assigning to each $A \in \mathcal{A}^0$ a generalized operator $J(A)$, then we say that J is a *generalized operator-valued measure on $\mathcal{A}$ finite on $\mathcal{A}^0$* if every matrix element is a complex measure on $\mathcal{A}$ finite on $\mathcal{A}^0$. Namely we require that for all $f, g \in \Phi_-$ the correspondence $A \rightsquigarrow \langle f, J(A)g\rangle$, $A \in \mathcal{A}^0$, define a complex combination of finite positive measures.

If J is as above, then we say that J is a *generalized spectral measure on $\mathcal{A}$* if in addition J is first positive on $\mathcal{A}^0$, $J(A) \geq 0$ for $A \in \mathcal{A}^0$, and if, secondly, whenever $A, B \in \mathcal{A}^0$, and $A \supset B$ the pair $J(A) \geq J(B)$ is projective.

We note that $\mathcal{A}^0$ is a directed set so that a generalized spectral measure on $\mathcal{A}$ is in fact a projective net. We therefore have the renormalized rigged Hilbert space Ψ and the renormalized projective net $\{E_A\}_{A\in\mathcal{A}^0}$. The main result of this section is the following:

Theorem 1: There exists a unique spectral measure $A \rightsquigarrow E(A)$ on $\mathcal{A}$ which extends the renormalized projective net $A \rightsquigarrow E_A$ on $\mathcal{A}^0$. That is, $E(A) = E_A$ for $A \in \mathcal{A}^0$ and furthermore:

(SM1) E is σ additive on $\mathcal{A}$;

(SM2) if $A, B \in \mathcal{A}$, then $E(A)E(B) = E(A \cap B)$;

(SM3) $\sup_{A\in\mathcal{A}} E(A) = 1$.

Proof: Since $\mathcal{A}^0$ is a generating subring for $\mathcal{A}$, then by standard extension theorems[12] we need only prove (SM1)-(SM3) on $\mathcal{A}^0$. To do so, we need to transfer results from some H_S to Ψ_0; we make use of the fact that for S and $T \in \mathcal{A}^0$ the maps k_S and r_{ST} are isometries and that $k_S r_{ST} = k_T$.

Thus, let $A_k \in \mathcal{A}^0$ be a disjoint sequence of sets and assume $A = \bigcup A_k \in \mathcal{A}^0$. Then we have $J(A) = \sum J(A_k)$ by hypothesis and by Lemma 1 we have, therefore, that in H_A, $1 = \sum Q_{A_k A}$; but by the above remarks this implies that $E_A = \sum E_{A_k}$ and since $A, A_k \in \mathcal{A}^0$ we have $E(A) = \sum E(A_k)$ and so E is σ additive on $\mathcal{A}^0$.

Suppose now that $A, B \in \mathcal{A}^0$; then since $A \cup B = (A\backslash B) \cup (A \cap B) \cup (B\backslash A)$ is a disjoint union, we have by hypothesis and by Lemma 1 that in $H_{A\cup B}$,

$$Q_{A\backslash B, A\cup B} + Q_{A\cap B, A\cup B} + Q_{B\backslash A, A\cup B} = 1$$

and similarly

$$Q_{A\backslash B, A\cup B} + Q_{B, A\cup B} = 1,$$

$$Q_{B\backslash A, A\cup B} + Q_{A, A\cup B} = 1,$$

so

$$\begin{aligned} Q_{B, A\cup B} Q_{A, A\cup B} &= (1 - Q_{A\backslash B, A\cup B})(1 - Q_{B\backslash A, A\cup B}) \\ &= 1 - Q_{A\backslash B, A\cup B} - Q_{B\backslash A, A\cup B} \end{aligned}$$

since $Q_{A\backslash B, A\cup B} Q_{B\backslash A, A\cup B} = 0$ by Lemma 3 and the first equation. By the first equation again we have

$$Q_{A, A\cup B} Q_{B, A\cup B} = Q_{A\cap B, A\cup B},$$

which implies that

$$E_A E_B = E_{A\cap B}$$

and since all three sets A, B, and $A \cap B$ are in $\mathcal{A}^0$, we have $E(A)E(B) = E(A \cap B)$ on $\mathcal{A}^0$.

Finally suppose $h \perp \sup_{A\in\mathcal{A}} E(A)\Psi_0$ and $h \neq 0$; then there is an $f \in \Psi_-$ such that $\|f - h\| < \|f\|$; but $f \in k_A H_A$ for some $A \in \mathcal{A}^0$ and so $\|f\| = \|E(A)f\| = \|E(A)(f - h)\| \leq \|f - h\| < \|f\|$, which is a contradiction; thus $h = 0$ and (SM3) is true ∎

The spectral measure E we shall call the *renormalized spectral measure.*

Finally a last mathematical notion: By a *Hilbert space with indefinite metric* we shall mean a Hilbert space together with a symmetric bounded operator η whose spectrum is contained in the points ± 1. We define $(f, g)_\eta = (f, \eta g)$. By a *rigged Hilbert space with indefinite metric* we shall mean a rigged Hilbert space Φ in which Φ_0 is a Hilbert space with indefinite metric.

B. The Resolvent Conditions

We now list the conditions on R_λ and S_∞ by which we hope to gain insight into the properties of T.

Condition 1 (Hermiticity): This condition was introduced in the last section and for symmetric generalized operators this now reads

$$R_\lambda(T)^* = R_{\bar\lambda}(T).$$

This equation really makes sense only if the domain of $R_{\bar\lambda}(T)$ is Φ_-; however, the domain can be forced to be smaller. Let $\Theta_\lambda \subset \Phi_-$ be the domain of R_λ; then our condition is

$$\langle g, R_\lambda(T)f\rangle = \overline{\langle f, R_{\bar\lambda}(T)g\rangle}$$

for all $f \in \Theta_\lambda$, $g \in \Theta_{\bar\lambda}$. This condition reduces to the above one in case $\Theta_\lambda = \Phi_- = \Theta_{\bar\lambda}$.

Condition 2 (Analyticity): This is a heuristic requirement. We should pick the domain of analyticity so that any version with a larger domain is somehow not as interesting and furthermore we should endeavor to have the severity of the singularities in R_λ be as mild as possible. For example, we could endeavor to have analyticity everywhere off the real axis or off the real axis in a neighborhood thereof.

Condition 3 (Positivity): We assume we have analyticity off the real axis in a neighborhood thereof.

Let the *imaginary part* $\operatorname{Im} R_\lambda$ of R_λ be defined as $(1/2i)[R_\lambda(T) - R_\lambda(T)^*]$. We now require that there be a neighborhood of the real singular points of R_λ such that for λ in the open *upper* half-plane and within this neighborhood $\operatorname{Im} R_\lambda$ be a *negative* generalized operator.

For bona fide operators $\operatorname{Im} R_\lambda$ is negative in the entire open upper half-plane but we cannot expect this for generalized operators. The truth of positivity for bona fide operators follows from the equation

$$\operatorname{Im} R_\lambda = \operatorname{Im} \int \frac{E(dx)}{\lambda - x} = \int \frac{(-\operatorname{Im} \lambda) E(dx)}{|\lambda - x|^2} = -\operatorname{Im} \lambda \int \frac{E(dx)}{|\lambda - x|^2},$$

which is negative for $\operatorname{Im} \lambda > 0$.

Condition 4 (The measure condition): The intuitive statement of this condition is that the imaginary part of $R_\lambda(T)$ along the real axis should be a generalized operator-valued measure.

We assume we have analyticity off the real axis in a neighborhood thereof. If l is a continuous positive real-valued function of compact support on $\mathbb{R}$, then there is a real number $\epsilon \neq 0$ such that the set supp $(l) + i\epsilon$ lies within the holomorphy domain of $R_\lambda(T)$. In this case we can construct the integral

$$J_\epsilon(l) = -\frac{1}{\pi} \int \operatorname{Im} R_{x+i\epsilon}(T) l(x)\, dx$$

and, again by analyticity, the matrix elements of $J_\epsilon(l)$ define complex-valued measures $\mu_{f,g,\epsilon}(dx)$ such that

$$\int \mu_{f,g,\epsilon}(dx) l(x) = \langle f, J_\epsilon(l) g \rangle = -\frac{1}{\pi} \int \langle f, \operatorname{Im} R_{x+i\epsilon} g \rangle l(x)\, dx, \qquad f, g \in \Phi_-.$$

This measure of course has to be thought of as being concentrated within the set of points x such that $x + i\epsilon$ is within the holomorphy domain of R_λ and by assumption, if ϵ is small enough, this set contains an arbitrary finite interval. The function l must of course have its support within this interval.

Consider now the measures $|\mu_{f,g,\epsilon}|\,(dx)$; then if

$$\lim_{\epsilon \downarrow 0} \int |\mu_{f,g,\epsilon}|\,(dx) l(x)$$

exists for all $f, g \in \Phi_-$, we have that $\lim J_\epsilon(l)$ as $\epsilon \downarrow 0$ exists and we call this limit $J(l)$.

The measure condition is now the following: We assume there is a σ ring $\mathcal{A}$ of subsets of $\mathbb{R}$ with a generating subring $\mathcal{A}^0$ of open sets and a generalized operator-valued measure $J(dx)$ on $\mathcal{A}$ finite on $\mathcal{A}^0$ such that whenever $J(l)$ exists l is measurable with respect to $\mathcal{A}$ and

$$\int J(dx) l(x) = J(l).$$

Instead of the resolvent R_λ we could have applied the same reasoning to a function $\alpha(\lambda)$ analytic off the real axis in a neighborhood thereof to arrive at the notion of $\alpha(\lambda)$ satisfying the measure condition. In this case there is no need to take matrix elements and we would have therefore introduced the complex measure μ_ϵ by $\mu_\epsilon(dx) = -(1/\pi) \operatorname{Im} \alpha(x + i\epsilon)\, dx$. The measure whose existence is asserted by the measure condition we shall call $\mu(\alpha, dx)$.

An example of a function satisfying neither positivity nor the measure condition is $1/\lambda^2$.

Condition 5 (Projectivity): Assume the measure condition is satisfied and assume further that there is a $C \in \mathcal{A}$ such that if we define $\mathcal{A}_C = \{A \in \mathcal{A} \mid A \subset C\} = \mathcal{A} \cap C$ and $\mathcal{A}^0_C = \{A \in \mathcal{A}^0 \mid A \subset C\}$, then $\mathcal{A}^0_C$ is a generating subring of $\mathcal{A}_C$ and the measure J is positive on $\mathcal{A}_C$. Projectivity now requires that J be a generalized spectral measure on $\mathcal{A}_C$; that is, if $A, B \in \mathcal{A}^0_C$, and $A \supset B$, then $J(A) \geq J(B)$ must be a projective pair. If this is satisfied, then we say we have *projectivity on C*.

If C is as above except that J is negative on $\mathcal{A}_C$, then by working with $-J$ we could again impose projectivity and we shall still say that J is projective on C.

Projectivity is a generalization of the notion of orthonormal decomposition. We now introduce another generalization which should be closely related to projectivity through the exact relationship we do not know at present.

Condition 6 (Decomposability): Assume we have a measurable space $(Z, \mathfrak{Z})$, where $\mathfrak{Z}$ is a σ ring of subsets of Z, which in addition carries two further structures: a real measurable function $z \rightsquigarrow \lambda(z)$ and a weakly measurable function $z \rightsquigarrow \psi_z \in \sum(T)$ such that $\psi_z \in \sum_{\lambda(z)}(T)$. What we mean for ψ_z to be weakly measurable is that all the functions of the form $z \rightsquigarrow \langle f, \psi_z \rangle$ for $f \in \Phi_-$ are measurable. This situation corresponds to the familiar practice in elementary quantum mechanics of introducing "auxiliary variables" into eigenvectors for the space $\{z \mid \lambda(z) = \lambda\}$, where λ is real is the space of auxiliary variables for eigenvalue λ.

Now let μ be a real measure on Z and consider the generalized operator-valued measure $K(dz)$ given by

$$K(dz) = |\psi_z\rangle\langle\psi_z|\,\mu(dz).$$

We assume now that R_λ satisfies the measure condition.

The condition of decomposability is now the following: there exist appropriate $(Z, \mathfrak{Z})$, $\lambda(z)$, ψ_z, and $\mu(dz)$ such that for $A \in \mathcal{A}^0$

$$J(A) = \int_{\{\lambda(z)\in A\}} K(dz) = \int_{\{\lambda(z)\in A\}} |\psi_z\rangle\langle\psi_z|\,\mu(dz).$$

If the above is true as for projectivity only for $A \in \mathcal{A}^0_C = \{A \in \mathcal{A}^0 \mid A \subset C \in \mathcal{A}\}$, then we say we have decomposability *on* C.

Let us discuss the two conditions of projectivity and decomposability. Both of these conditions allow us to reinterpret T as a bona fide operator in a new Hilbert space.

In the first place, assume we have projectivity on $C \in \mathcal{A}$ with J a positive measure on $\mathcal{A}_C$. Then we can construct the renormalized rigged Hilbert space $\Psi_C = {}_\rightarrow\lim \Phi_A$, where $A \in \mathcal{A}^0_C$. The measure J then becomes renormalized to a bona fide spectral measure E on C and we define $T_R \mid C$ to be the operator $\int_C \lambda E(d\lambda)$ and call it the *renormalized T in C*.

If J were negative on C, then using $-J$ we again repeat the above construction and keep the same designations.

Now let $C = C_1 \cup C_2$ be a disjoint union of $C_i \in \mathcal{A}$ such that J is projective on each one and is positive on the first and negative on the second. We can now introduce $\Psi_C = \Psi_{C_1} \oplus \Psi_{C_2}$, which again obviously carries a spectral measure E on C and defines a bona fide operator $T_R \mid C$. In this case, however, we make Ψ_C into a rigged Hilbert space with indefinite metric η by defining $\eta(h_1 \oplus h_2) = h_1 \oplus (-h_2)$. The spectral measure E is a spectral measure in the sense of spectral measures with indefinite metric; that is, we have $(f, E(A)g)_\eta = (E(A)f, g)_\eta$, which follows from the fact that η commutes with E.

If we have projectivity on all of $\mathbb{R}$ with J positive, we say that T is *renormalizable* and we define the *renormalized* T, T_R to be $T_R \mid \mathbb{R}$ as defined above.

Suppose now that R_λ is decomposable on $C \in \mathcal{A}$, and let $(Z_C, \mathfrak{Z})$ be the appropriate measure space. We can now introduce the Hilbert space $L^2(Z_C, |\mu|\,(dz))$ which will be called the *space of renormalized vectors*. The *dressing transformation* Δ_C will be defined as the map which to every $f \in \Phi_-$ assigns the complex valued function $z \rightsquigarrow \overline{\langle f, \psi_z\rangle}$ on Z_C. This function is of course not necessarily square integrable. One may like to construct a rigging $\Psi_{-C} \subset L^2(Z_C, |\mu|\,(dz)) \subset \Psi_{+C}$ such that $\Delta_C\Phi_- \subset \Psi_{+C}$ but there is no canonical way to do this.

We can now introduce the spectral measure F on Z_C in $L^2(Z_C)$ by setting $(F(W)h)(z) = \chi_W(z)h(z)$ for $h \in L^2$ and $W \in \mathfrak{Z}$. We can likewise introduce the spectral measure E on C in $L^2(Z_C)$ by $E(A) = F(\{z \mid \lambda(z) \in A\})$. This will be called the renormalized spectral measure of T on C and we define $T_R \mid C = \int_C \lambda E(d\lambda)$, which we call *the renormalized T in C*.

One finally makes the Hilbert space $L^2(Z_C)$ into a Hilbert space with indefinite metric by defining η to be multiplication by the Radon-Nikodym derivative $d\mu/d\,|\mu|$, which as is clear takes on the values ± 1 almost everywhere. Since F commutes with η, we see that F, and subsequently E, are in fact spectral measures in the sense of Hilbert spaces with indefinite metric.[13]

Condition 7 (*Well dressing*): This condition concerns the dressing transformation Δ. As we shall see from the heuristic discussion below, Δ can be thought of as a solution to the formal problem of finding a transformation that intertwines T_R and T; that is, $T_R\Delta = \Delta T$. This equation is purely formal since ΔT is not defined, for if $f \in \Phi_-$, then Tf is not necessarily in Φ_-, the domain of Δ. We want to make this formal consideration closer to the truth. More generally, if Δ formally intertwines T_R and T, then one could also consider it to formally intertwine E and J, or $1/(\lambda - T_R)$ and $R_\lambda(T)$ or any other such pair. Let us work with E and J to state the condition and then for any other pair the situation is entirely similar. Formally, we have $E(A)\Delta = \Delta J(A)$ or $E(A) = \Delta J(A)\Delta^{-1}$; we now want to use elementary function theory to interpret this equation. To do so, we consider each of $E(A)$, $J(A)$, and Δ as a linear relation between some appropriate spaces and then take an interpretation of the formal equation. The loosest sense of the formal equation is to consider $E(A) \in M(\Psi_+)$, $J(A) \in M(\Phi_+)$, and $\Delta \in M(\Phi_+, \Psi_+)$ and require

$$E(A) \subset \overline{\overline{\Delta}\overline{J(A)}\overline{\Delta}^{-1}}\overline{\overline{\Delta}\overline{J(A)}\overline{\Delta}^{-1}}.$$

However, one can take less loose interpretation such as not taking the closure of $J(A)$ in the above expression or even simply looking at $\bar{\Delta}J(A)\bar{\Delta}^{-1}$, which may make sense. Alternatively, one could consider Δ to be an element of $M(\Phi_-, \Psi_+)$ rather than $M(\Phi_+, \Psi_+)$.

In any case, if in some of the above senses the formal equation holds, then we say that J is *well dressed*. In practice one can start with the loosest

interpretation and then try to make it more stringent to see whether the choice of R_λ is in this way narrowed down but still interesting. Had we instead used the pair $1/(\lambda - T_R)$ and R_λ, then similar conditions would have defined the notion of R_λ being well dressed, and so forth.

The use of the pair $1/(\lambda - T_R)$ and R_λ can be more advantageous since one can use analytic subtraction methods on it, in which case it is better to use the formal equation in the form $[1/(\lambda - T_R)]\Delta = \Delta R_\lambda$. Now the left-hand side is often well defined already, for T_R is often concretely a multiplication operator and thus can be extended to Ψ_+ without much ambiguity. On the other hand ΔR_λ, in general, needs a subtracted interpretation which most often can be taken in the form $\lim_{F\in\mathcal{F}} \Delta\pi_F R_\lambda + P_F(\lambda)$.

To clarify the terms *dressed vector* and *renormalized vector*, we consider a decomposable resolvent and the corresponding function $z \rightsquigarrow \psi_z$. The function $z \rightsquigarrow \overline{\langle f, \psi_z\rangle}$ is heuristically the decomposition of f with respect to a "complete orthonormal set"; however, the set of generalized vectors $\{\psi_z\}$ is orthonormal only *by fiat* since, in general, the ψ_z will not lie in Φ_0 the "bare" Hilbert space nor is the inner product in Φ_0 the appropriate one to exhibit the structure of T. The resolvent is used here to determine which particular set of generalized vectors it is appropriate to make orthonormal by fiat and this is done by requiring $K(\{\lambda(z) \in A\}) = J(A)$. Now since $\psi_z \in \sum_{\lambda(z)}(T)$, we have the formal equation $T\psi_z = \lambda(z)\psi_z$ and so ΔTf is formally the function

$$z \rightsquigarrow \overline{\langle Tf, \psi_z\rangle} = \overline{\langle f, T\psi_z\rangle} = \lambda(z)\overline{\langle f, \psi_z\rangle} = (T_R\Delta f)(z)$$

since T_R is the operator of multiplication by the function $z \rightsquigarrow \lambda(z)$. Thus Δ is indeed formally a dressing transformation.

Consider now a function $h \in L^2(Z, |\mu|)$; then it need not be of the form Δf for $f \in \Phi_-$; however, formally it is of the form $\Delta\phi$ for $\phi \in \Phi_+$. Let

$$\phi = \int \psi_z h(z)\,|\mu|\,(dz);$$

then formally

$$(\Delta\phi)(z) = \int \overline{\langle \psi_w \psi_z\rangle} h(w)\,|\mu|\,(dw) = h(z)$$

since by the by fiat orthonormalization procedure $\overline{\langle \psi_w, \psi_z\rangle}$ should be a reproducing kernel with respect to $|\mu|\,(dw)$. The inner product $\overline{\langle \psi_w, \psi_z\rangle}$ is not defined and when expressed concretely in any given concrete rigged Hilbert space will involve meaningless divergent expressions. By a formal "renormalization" of these divergent expressions one can then consider h as a "renormalized" version of ϕ-hence the name *renormalized vector*. The name is appropriate from another point of view, for in the projective case the states in Ψ_0 arise via the process of introducing *new* pseudonorms on Φ_-, namely $f \rightsquigarrow \|\Delta_A f\|_A$.

What the above processes of renormalization have to do with the physical renormalization program is not here clear since the former are mathematical notions whereas the latter embodies many physical ideas; however, we feel that part of the physical renormalization program is embodied in the mathematical methods exhibited above.

Condition 8 (Nondegeneracy): This is a somewhat ad hoc condition and it requires that $R_\lambda f = 0$ imply $f = 0$ for all points of holomorphy of R_λ. For bona fide operators this is true since R_λ is the bounded inverse of an operator at all regular points.

Condition 9 (Ray limit): This condition is also somewhat ad hoc and it is based on the following observation concerning a self-adjoint bona fide operator A. The resolvent of A is given in terms of the spectral measure E by the integral

$$R_\lambda = \int \frac{E(dx)}{\lambda - x}.$$

We note that

$$|\lambda|\, e^{i\theta} R_{|\lambda| e^{i\theta}} = \int \frac{[1 - (xe^{i\theta}/|\lambda|)]E(dx)}{[\cos\theta - (x/|\lambda|)]^2 + \sin^2\theta}$$

has, when $\theta \neq 0$, π and $|\lambda| \to \infty$, the weak limit $\int E(dx) = 1$.

Our condition now is that along the rays $|\lambda|e^{i\theta}\theta \neq 0$, π, which eventually, as $|\lambda| \to \infty$, enter the holomorphy domain of R_λ, we have

$$\lim_{|\lambda|\to\infty} \lambda R_\lambda(T) = S_\infty(T),$$

where S_∞ is a symmetric g-operator version of $\mathcal{S}_\infty$ picked independently of the rays used in the limit above. Certain of these versions may be more natural than others, but as will be seen from the examples, even if $j \subset \mathcal{S}_\infty$, it is not always useful to pick this version; however, we could require that j be a version of S_∞, in which case if the ray limit holds for such an S_∞ we say we have a *normal* ray limit.

Having constructed the renormalized rigged Hilbert space Ψ, there arises the question of how one is to interpret physical processes within it. This question is complicated by the fact that the by-fiat procedure of orthonormalization destroys the detailed concrete structure of the generalized vectors ψ_z and it is this

structure that carries any additional physical information. This question will be looked into in the next section on several operators since often physical information is carried by a set of generalized operators rather than by a single one. For now we examine only single operators.

In the examples that follow we shall in the multivalued approach endeavor to pick resolvents satisfying all of the following conditions:

(1) hermiticity;
(2) analyticity;
(3) positivity with measure condition on $\mathbb{R}$ with $\mathcal{A}$ being the σ algebra $\mathcal{B}$ of Borel subsets of $\mathbb{R}$;
(4) projectivity on $\mathbb{R}$ or decomposability in some sense;
(5) well dressing of R_λ in some sense;
(6) nondegeneracy;
(7) normal ray limit.

We cannot hope to satisfy all of the above conditions in all cases but the obstruction to satisfying them should lead to important insights into the differences between bona fide and generalized operators.

We shall also apply orthogonal methods to supplement the multivalued approach. To be effective, we must at times nest the approach, that is, apply orthogonal methods to the calculation of Q_λ and P_λ. This nesting can continue indefinitely, thus arriving at something resembling an infinite continued fraction expansion of the resolvent. The combinatorial problems involved in such expansions are formidable and so we shall pursue only simple situations.

At various points in the following we shall refer to the *intrinsic structural properties* of T. This is not a well-defined notion but we mean by it the following: Since our approach is inductive and combines both mathematical and physical reasoning, any result we obtain is influenced by both a yet to be discovered mathematical theory and by a physical interpretation. Some of the results should, however, be less contaminated by extramathematical ideas than others and thus should appear in a cogent form in a proper deductive theory. At the moment, however, we can recognize such results only by an insight and when we feel we have such a result, we shall refer to it as an *intrinsic structural property*.

C. Examples

Example 15: Consider first the simple case of multiplication by the δ distribution on $\mathcal{D}(\mathbb{R}) \subset \mathcal{D}'(\mathbb{R})$,

$$(Tf)(x) = f(x)\delta(x) = f(0)\delta(x).$$

If $\tau \in \mathcal{D}'(\mathbb{R})$, we can have $f \to \tau$ in the sense of distributions in such a way that $f(0)$ approaches any complex number; thus we have

$$\bar{T}\tau = \{\alpha\delta(x) \mid \alpha \in \mathbb{C}\} \quad \text{for} \quad \tau \in \mathcal{D}',$$

and we find

$$\lambda \neq 0 : \mathcal{R}_\lambda \tau = \left\{\frac{1}{\lambda}\tau + \alpha\delta(x) \mid \alpha \in \mathbb{C}\right\},$$

$$\lambda = 0 : \mathcal{R}_0 \tau = \mathcal{D}' \quad \text{if} \quad \tau = \alpha\delta(x), \quad \alpha \in \mathbb{C},$$

$$= \phi \quad \text{otherwise},$$

$$\mathcal{S}_\infty \tau = \{\tau + \alpha\delta(x) \mid \alpha \in \mathbb{C}\}.$$

The Hermitian g-operator versions are easily computed to be

$$\lambda \neq 0 : R_\lambda f = \frac{1}{\lambda} f + \alpha(\lambda) f(0)\delta(x) \quad \text{where} \quad \alpha(\lambda) = \overline{\alpha(\bar{\lambda})},$$

$$\lambda = 0 : R_0 f = \phi \quad \text{if} \quad f \neq 0,$$

$$= 0 \quad \text{if} \quad f = 0,$$

$$S_\infty f = f + \beta f(0)\delta(x), \quad \beta \text{ real}.$$

Imposing analyticity, we require $\alpha(\lambda)$ to be real analytic with singularities on the real axis. We have

$$\operatorname{Im} R_\lambda f = -\frac{\operatorname{Im}\lambda}{|\lambda|^2} f + \operatorname{Im}\alpha(\lambda) f(0)\delta(x)$$

and for $\operatorname{Im} R_\lambda$ to be negative for $\operatorname{Im}\lambda > 0$ in a neighborhood of the singularities we must have $\operatorname{Im}\alpha(\lambda) \leq 0$ for $\operatorname{Im}\lambda > 0$ and in a neighborhood of the singularities.

By the measure condition with $\mathcal{A} = \mathcal{B}$ and positivity we have

$$J(A)f = \chi_A(0) f + \mu(\alpha, A) f(0)\delta(x),$$

where $\mu(\alpha, A) \geq 0$.

There are thus four possibilities for H_A:

(1) $\chi_A(0) = 0, \mu(\alpha, A) = 0$; hence $H_A = \{0\}, \Delta_A f = 0$;

(2) $\chi_A(0) = 0, \mu(\alpha, A) > 0$; hence $H_A = \mathbb{C}, \|c\|_A^2 = \mu(\alpha, A)\,|c|^2$, and $\Delta_A f = f(0)$;

(3) $\chi_A(0) = 1, \mu(\alpha, A) = 0$; hence $H_A = L^2(\mathbb{R}, dx)$, $\|h\|_A^2 = \|h\|_{L^2}^2$, and $\Delta_A f = f$;

(4) $\chi_A(0) = 1, \mu(\alpha, A) > 0$; hence

$$H_A = L^2(\mathbb{R}, dx) \oplus \mathbb{C},$$

$$\|h \oplus c\|_A^2 = \|h\|_{L^2}^2 + \mu(\alpha, A)\,|c|^2,$$

and $\Delta_A f = f \oplus f(0)$.

We see that by projectivity it is impossible for there to be two disjoint sets A_1 and A_2 such that $\mu(\alpha, A_1) > 0$ and $\mu(\alpha, A_2) > 0$. In that case, H_{A_1} will have a direct summand of the form $\mathbb{C}$ with the norm squared $\mu(\alpha, A_1)\,|\cdot|^2$, and $H_{A_1 \cup A_2}$ will also have a direct summand of the form $\mathbb{C}$ with norm squared

$$[\mu(\alpha, A_1) + \mu(\alpha, A_2)]\,|\cdot|^2.$$

Regardless of whether we have a direct summand of the form $L^2(\mathbb{R}, dx)$, the space $H_{A_1 \cup A_2}$ has no degenerate subspace with respect to $J(A_1)$ and so by Lemma 5 we must have $\|\Delta_{A_1} f\|^2_{A_1} = \|\Delta_{A_1 \cup A_2} f\|^2_{A_1 \cup A_2}$, which contradicts the hypothesis.

Thus, $\alpha(\lambda)$ must have only a single simple pole on the real axis:

$$\alpha(\lambda) = E(\lambda) + \frac{\beta_0}{\lambda - \lambda_0},$$

where $E(\lambda)$ is entire and $\beta_0 \geq 0$ by positivity. The normal ray limit implies $E(\lambda) = 0$ and $\beta_0 = \beta$.

For $\beta = 0$ we find that

$$\Psi_- = \Psi_0 = \Psi_+ = L^2(\mathbb{R}, dx), \quad \Delta f = f,$$

and $E(A) = \chi_A(0) I$. For $\beta > 0$, we find that

$$\Psi_- = \Psi_0 = \Psi_+ = L^2(\mathbb{R}, dx) \oplus \mathbb{C}$$

with $\|\cdot\|_0^2 = \|\cdot\|_{L^2}^2 + \beta |\cdot|^2$ and $\Delta f = f \oplus f(0)$. The renormalized spectral measure is

$$E(A) = \chi_A(0) P_{L^2} + \chi_A(\lambda_0) P_{\mathbb{C}}$$

where P_{L^2} and $P_{\mathbb{C}}$ are the canonical projections onto the respective direct summands.

The case $\beta = 0$ is not interesting since E is the spectral measure of the zero operator in L^2. We do not consider this case further.

For $\beta > 0$, E is the spectral measure of $\lambda_0 P_{\mathbb{C}}$ and thus T is reinterpreted as λ_0 times the projection onto the subspace generated by $0 \oplus 1/\sqrt{\beta}$. This is consistent with the formal picture of T, for formally $(T^2 f)(x) = \delta(0) f(0) \delta(x) = \delta(0)(Tf)(x)$ and so $T^2 = \delta(0) T$ and T is $\delta(0)$ times the projection onto the subspace generated by the δ function. We have "renormalized" $\delta(0)$ to λ_0 and made the δ function a new discrete state of norm $1/\sqrt{\beta}$.

The above resolvent is decomposable, for let $Z = \mathbb{N} \vee \mathbb{R}$ be the disjoint union of the set of integers $\mathbb{N} = \{1, 2, 3, \cdots\}$ and the real line $\mathbb{R}$; let $\mathfrak{Z}$ = the σ algebra of all subsets of Z. Let

$$\lambda(z) = 0 \quad \text{for} \quad z = n \in \mathbb{N},$$
$$= \lambda \quad \text{for} \quad z = \lambda \in \mathbb{R};$$

let

$$\psi_z = \psi_n \quad \text{for} \quad z = n \in \mathbb{N}, \quad \text{where } \psi_n \text{ is some orthonormal set in } L^2(\mathbb{R}, dx),$$
$$= \delta(x) \quad \text{for} \quad z = \lambda \in \mathbb{R};$$

and finally let

$$\mu(L) = \text{the number of points in } L \text{ for } L \subset \mathbb{N}$$
$$= \beta \chi_L(\lambda_0) \quad \text{for} \quad L \subset \mathbb{R}.$$

Then this is easily seen to afford a decomposition of J,

$$J(A) = \int_{\{\lambda(z) \in A\}} |\psi_z\rangle \langle \psi_z| \, \mu(dz).$$

The resolvent R_λ is well dressed for a computation of

$$P_\lambda = \overline{\overline{\Delta R_\lambda \Delta^{-1}} \vee \overline{\Delta \bar{R}_\lambda \Delta^{-1}}}$$

shows that

$$P_\lambda h \oplus c = \left\{ \frac{1}{\lambda} h \oplus \alpha \,\middle|\, \alpha \in \mathbb{C} \right\}$$

and so $1/(\lambda - T_R) \subset P_\lambda$ and R_λ is well dressed.

In connection with well dressing we further note that $Q_\lambda = \Delta R_\lambda \Delta^{-1}$ is given by

$$Q_\lambda h \oplus c = \left\{ \frac{1}{\lambda} h \oplus \alpha \,\middle|\, \alpha \in \mathbb{C} \right\} \quad \text{if } h \in \mathfrak{D} \text{ and } h(0) = 0,$$
$$= \phi \quad \text{otherwise}$$

and so Q_λ has a restriction which coincides with $1/(\lambda - T_R)$ on a dense subspace of Ψ_0 and thus R_λ is well dressed in even a more stringent sense.

In this example we, of course, could have violated some of the conditions; thus taking $\beta < 0$ we still would have had $\Psi_- = \Psi_0 = \Psi_+ + L^2(\mathbb{R}, dx) \oplus \mathbb{C}$ with $\|\cdot\|_0^2 = \|\cdot\|_{L^2}^2 + |\beta| \, |\cdot|^2$ and $\Delta f = f \oplus f(0)$ but we would now have had a rigged Hilbert space with indefinite metric with $\eta(h \oplus c) = h \oplus -c$. We shall not pursue the discussion of any other violations.

We note that the pole in R_λ at $\lambda = 0$ with residue j is present in all cases and that $J(A)$ is always a linear combination of j and the multiplication by δ. These are the intrinsic structural properties of T; the rest reflect certain arbitrary choices that can be made.

We shall not work out subsequent examples with as much explicit detail since much of the detail is entirely straightforward but often cumbersome to write out.

Example 16: As a slightly more complicated example consider $\mathbb{C} \oplus \mathfrak{D}(\mathbb{R}) \subset \mathbb{C} \oplus \mathfrak{D}'(\mathbb{R})$ and

$$T(f_0, f_1) = (-g f_1(0), -g f_0 \delta(x)).$$

We have

$$\bar{T}(\tau_0, \tau_1) = \{(\alpha, -g\tau_0 \delta(x)) \mid \alpha \in \mathbb{C}\},$$

where $(\tau_0, \tau_1) \in \mathbb{C} \oplus \mathfrak{D}'(\mathbb{R})$.

One easily finds

$$\lambda \neq 0 : \mathcal{R}_\lambda(\tau_0, \tau_1) = \left\{ \left(\alpha, \frac{1}{\lambda}(\tau_1 - g\alpha\delta) \right) \,\middle|\, \alpha \in \mathbb{C} \right\},$$
$$\lambda = 0 : \mathcal{R}_0(\tau_0, \tau_1) = \{(\tau_0, \psi) \mid \psi \in \mathfrak{D}'\} \text{ if } \tau_1 = g\tau_0 \delta,$$
$$= \varnothing \text{ otherwise},$$
$$\mathcal{S}_\infty(\tau_0, \tau_1) = \{(\alpha, \tau_1) \mid \alpha \in \mathbb{C}\}$$

and the g-operator versions of these are

$$\lambda \neq 0: R_\lambda(f_0, f_1) = \left(\alpha_\lambda(f_0, f_1), \frac{1}{\lambda}[f_1 - g\alpha_\lambda(f_0, f_1)\delta]\right),$$

$$\lambda = 0: R_\lambda(f_0, f_1) = \phi \quad \text{if} \quad (f_0, f_1) \neq (0, 0),$$
$$= (0, 0) \quad \text{if} \quad (f_0, f_1) = (0, 0),$$
$$S_\infty(f_0, f_1) = (\beta(f_0, f_1), f_1),$$

where α_λ and β are linear functionals $\Phi_- \to \mathbb{C}$.

The imposition of hermiticity requires that

$$\alpha_\lambda(f_0, f_1) = \alpha(\lambda)\left[f_0 - g\frac{1}{\lambda}f_1(0)\right], \quad \alpha(\lambda) = \overline{\alpha(\bar{\lambda})},$$

$$\beta(f_0, f_1) = \beta_0 f_0, \quad \beta_0 \text{ real}.$$

By analyticity we require $\alpha(\lambda)$ to be analytic off the real axis.

The ray limit implies that $\lambda\alpha(\lambda) \to \beta_0$ along rays $|\lambda|e^{i\theta} \neq 0, \pi$ as $|\lambda| \to \infty$. To have $j \subset S_\infty$, we must take $\beta_0 = 1$ and we do so.

Assume the measure condition; then we see that $\alpha(\lambda)$, $\lambda^{-1}\alpha(\lambda)$, and $\lambda^{-2}\alpha(\lambda)$ must all satisfy the measure condition and we set $\mu_k(A) = \mu(\lambda^k\alpha(\lambda), A)$ for $k = 0, -1, -2$. We have

$$J(A)(f_0, f_1) = (\mu_0(A)f_0 - g\mu_{-1}(A)f_1(0),$$
$$\chi_A(0)f_1 - g\mu_{-1}(A)f_0\delta + g^2\mu_{-2}(A)f(0)\delta).$$

Let $M(A)$ be the matrix

$$\begin{pmatrix} \mu_0(A), & -g\mu_{-1}(A) \\ -g\mu_{-1}(A), & g^2\mu_{-2}(A) \end{pmatrix}.$$

For $J(A)$ to be positive $M(A)$ must be a positive matrix on $\mathbb{C}^2$. There are three possibilities for $\rho_A = \text{rank } M(A)$, namely $\rho_A = 0, 1, 2$. Furthermore depending on whether $0 \in A$ or not, we do or do not have a direct summand of the form $L^2(\mathbb{R}, dx)$ in H_A. The presence or absence of this summand does not affect our discussion and so we proceed with the analysis of $M(A)$ assuming projectivity.

Suppose $\rho_A = 2$; then if A' is disjoint from A, we must have by positivity $\rho_{A \cup A'} = 2$ for $M(A \cup A') = M(A) + M(A')$. By Lemma 5, we must have $u^+M(A)u = u^+M(A \cup A')u$ for every 2-component vector $u = \binom{u_1}{u_2}$, where $u^+ = (\overline{u_1}, \overline{u_2})$; but this implies $u^+M(A')u = 0$ and since $M(A')$ is a symmetric matrix on $\mathbb{C}^2$, we have $M(A') = 0$.

A similar discussion also shows that if $\rho_A = 1$ and A' is disjoint from A' such that $\rho_{A \cup A'} = 1$, then likewise $M(A') = 0$.

The above two paragraphs imply that $\alpha(\lambda)$ can have at most two singularities at say λ_1 and λ_2 and these by positivity must be first-order poles with positive residues. Furthermore, unless $\alpha(0) = 0$, $R_\lambda(T)$ will have a higher-order pole at $\lambda = 0$ and thus not satisfy positivity. We have

$$\alpha(\lambda) = E(\lambda) + \frac{\rho_1}{\lambda - \lambda_1} + \frac{\rho_2}{\lambda - \lambda_2},$$

where $E(\lambda)$ is entire and ρ_1 and $\rho_2 \geq 0$.

The normal ray limit now implies $E(\lambda) = 0$ and $\rho_1 + \rho_2 = 1$. To have $\alpha(0) = 0$, we must have

$$(\rho_1/-\lambda_1) + (\rho_2/-\lambda_2) = 0,$$

which shows that $\lambda_1 \neq \lambda_2$ and $\lambda_1\lambda_2 < 0$, say $\lambda_1 < 0 < \lambda_2$.

We find

$$\rho_1 = \lambda_2/(\lambda_2 - \lambda_1), \quad \rho_2 = -\lambda_1/(\lambda_2 - \lambda_1).$$

Now let A_1 contain λ_1 but not λ_2 and let A_2 be disjoint from A_1 and contain λ_2 but not λ_1; set $A = A_1 \cup A_2$. After a computation:

$$M(A_1) = \frac{\lambda_2}{\lambda_2 - \lambda_1}\begin{pmatrix} 1, & -g/\lambda_1 \\ -g/\lambda_1, & g^2/\lambda_1^2 \end{pmatrix},$$

$$M(A_2) = -\frac{\lambda_1}{\lambda_2 - \lambda_1}\begin{pmatrix} 1, & -g/\lambda_2 \\ -g/\lambda_2, & g^2/\lambda_2^2 \end{pmatrix},$$

$$M(A) = \frac{1}{\lambda_2 - \lambda_1} \times \begin{pmatrix} \lambda_2 - \lambda_1, & -g(\lambda_2/\lambda_1 - \lambda_1/\lambda_2) \\ -g(\lambda_2/\lambda_1 - \lambda_1/\lambda_2), & g^2(\lambda_2/\lambda_1^2 - \lambda_1/\lambda_2^2) \end{pmatrix}.$$

By Lemma 4 we see that the null spaces of $M(A_1)$ and $M(A_2)$ must be orthogonal with respect to $M(A)$ so we must have

$$\left(\frac{g}{\lambda_1}, 1\right)\begin{pmatrix} \lambda_2 - \lambda_1, & -g(\lambda_2/\lambda_1 - \lambda_1/\lambda_2) \\ -g(\lambda_2/\lambda_1 - \lambda_1/\lambda_2), & g(\lambda_2/\lambda_1^2 - \lambda_1/\lambda_2^2) \end{pmatrix} \times \begin{pmatrix} g/\lambda_2 \\ 1 \end{pmatrix} = 0,$$

which is easily computed to be satisfied.

We now have the solution

$$\Psi_- = \Psi_0 = \Psi_+ = \mathbb{C}^2 \oplus L^2(\mathbb{R}),$$
$$(u \oplus h, v \oplus k)_0 = u^+Mv + (h, k)_{L^2},$$

where $M = M(A)$ exhibited above,

$$\Delta(f_0, f_1) = \begin{pmatrix} f_0 \\ f_1(0) \end{pmatrix} \oplus f_1$$

and

$$E(A) = \chi_A(-\lambda_1)P_- + \chi_A(\lambda_2)P_+ + \chi_A(0)P_{L^2},$$

where P_- is the projection onto the subspace generated by the vector $\binom{g/\lambda_2}{1} \oplus 0$, P_+ is the projection onto

the subspace generated by the vector $(^{-g/\lambda_1}_{1}) \oplus 0$, and P_{L^2} is the projection onto the direct summand $L^2(\mathbb{R})$. The renormalized T and T_R thus has a null space of infinite multiplicity and two singlets of energy λ_1 and λ_2, respectively.

Formally, if we solve $T(f_0, f_1) = \lambda(f_0, f_1)$, we find that for $(0, f)$ with $f(0) = 0$ we have $T(0, f) = 0$ and so $\lambda = 0$. Assuming formally $\lambda \neq 0$ yields the formal solutions $(1, \pm\delta(x)/[\delta(0)]^{\frac{1}{2}})$ with eigenvalues $\mp g[\delta(0)]^{\frac{1}{2}}$, thus "renormalizing" $\delta(0)$ to g^2/λ_0^2 and considering $(0, \delta(x))$ to be a new discrete state, we get two eigenvectors $(1, \pm g\delta(x)/\lambda_0)$ with eigenvalue $\mp\lambda_0$; these visibly correspond to our solutions above if we take $-\lambda_1 = \lambda_0 = \lambda_2$. We note, however, that our methods do not "renormalize" the expression $\delta(0)$ to the same value in each of the two eigenvectors. We do not know what additional criteria are needed to accomplish this.

We still have not discussed the case that α is nonsingular but in this case $\alpha(\lambda)$ is an entire function and the ray limit implies $\alpha(\lambda) \equiv 0$ so that $\Psi_- = \Psi_0 = \Psi_+ = L^2(\mathbb{R})$ and T_R is the zero operator.

Both of the above solutions are decomposable and well dressed.

For the intrinsic structural properties of T we may note the pole in R_λ at $\lambda = 0$ with residue j; the fact that J is always a linear combination of j, of the creation operator $C^+: (f_0, f_1) \rightsquigarrow (0, f_0\delta(x))$, and of the annihilation operator $C: (f_0, f_1) \rightsquigarrow (f_1(0), 0)$; and lastly the impossibility of making either λ_1 or λ_2 zero.

Example 17: Let us now do the discrete Lee model by the multivalued methods. For simplicity we assume $0 < \lambda_1 < \lambda_2 < \cdots$ with no finite accumulation point. We further assume $\rho_k \neq 0$, which in particular means that $\rho \notin d$ and we are in the nontrivial case. We restate the results for $R_\lambda(T)$ and $S_\infty(T)$, calling $\alpha(\lambda_0)$ by $\alpha(\lambda)$ now:

$$\text{for} \quad \lambda \neq \lambda_k, \quad k \geq 1,$$

$$R_\lambda f = \Big(\alpha(\lambda) f_0 + \alpha(\lambda) \sum_1^\infty \frac{f_k \rho_k}{\lambda - \lambda_k}\,; \cdots, \frac{f_k}{\lambda - \lambda_k} + \alpha(\lambda) \frac{f_0 \rho_k}{\lambda - \lambda_k} + \alpha(\lambda) \Big(\sum_1^\infty \frac{f_l \rho_l}{\lambda - \lambda_l} \Big) \frac{\rho_k}{\lambda - \lambda_k}, \cdots \Big),$$

where $\alpha(\lambda) = \overline{\alpha(\bar{\lambda})}$;

$$S_\infty f = (\eta f_0, f_1, f_2, \cdots),$$

where η is real.

For normal ray limits we must have $\eta = 1$.

We assume analyticity off the real axis in a neighborhood thereof and we assume the measure condition.

Let

$$\zeta f(\lambda) = f_0 + \sum \frac{\rho_k f_k}{\lambda - \lambda_k}$$

and let $\bar{f} = (\overline{f_0}, \overline{f_1}, \overline{f_2}, \cdots)$; then we find that

$$\langle g, R_\lambda f \rangle = \alpha(\lambda) \zeta \bar{g}(\lambda) \zeta f(\lambda) + \sum_1^\infty \frac{\overline{g_k} f_k}{\lambda - \lambda_k}$$

and so

$$\langle f, J(A) f \rangle = \mu(\langle f, R_\lambda f \rangle, A) = \mu(\alpha(\lambda) \zeta \bar{f}(\lambda) \zeta f(\lambda), A) + \sum \chi_A(\lambda_k) |f_k|^2.$$

The measure condition implies that $\alpha(\lambda) \zeta \bar{f}(\lambda) \zeta f(\lambda)$ must satisfy the measure condition, which, taking $f = (1, 0, 0, \cdots)$, means, in particular, that α satisfies the measure condition.

We shall at this point assume that $\alpha(\lambda)$ has nothing but first-order poles on the real axis; that is, that $\mu(\alpha, d\lambda)$ has no part which is absolutely continuous with respect to Lebesgue measure. This is based on the heuristic principle that the analyticity domain of R_λ be as large as possible while still keeping the result useful. This assumption will be justified *a posteriori* by the interesting and natural class of resolvents that are so obtained. We call the real poles of $\alpha(\lambda)$ by λ_ν, where $\nu \in N$ and N is a countable set. Let r_ν be the residue of $\alpha(\lambda)$ at $\lambda = \lambda_\nu$. We see that $\alpha(\lambda)$ must have a zero at $\lambda = \lambda_k$, $k \geq 1$, in order that $\alpha(\lambda) \zeta \bar{g}(\lambda) \zeta f(\lambda)$ not have a higher-order singularity there which would contradict the measure condition. We take $\alpha(\lambda) \sim (\lambda - \lambda_k)\beta_k$ at $\lambda = \lambda_k$.

A simple calculation reveals

$$\langle g, J(A) f \rangle = \sum_{\nu \in N} \chi_A(\lambda_\nu) r_\nu \zeta \bar{g}(\lambda_\nu) \zeta f(\lambda_\nu) + \sum_{k=1}^\infty \chi_A(\lambda_k) \bar{g}_k f_k (1 + \beta_k \rho_k^2).$$

This is explicitly a decomposition of J; for it we take for Z the disjoint union $N \vee \mathbb{N}$; for $\mathfrak{Z}$ the σ algebra of all subsets; for $\lambda(z)$ the function $\lambda(\nu) = \lambda_\nu$, $\lambda(k) = \lambda_k$; for ψ_z,

$$\psi_z = \Big(1, \cdots, \frac{\rho_k}{\lambda_\nu - \lambda_k}, \cdots\Big), \quad z = \nu \in N,$$

$$= (0; 0, \cdots, 0, 1, 0, \cdots), \quad z = k \in \mathbb{N},$$

where the 1 is in the kth place,

and for $\mu(dz)$ we take the point mass of weight r_ν at $\nu \in N$ and the point mass of weight $(1 + \beta_k \rho_k^2)$ at $k \in \mathbb{N}$; then clearly

$$J(A) = \int_{\lambda(z) \in A} |\psi_z\rangle \langle \psi_z| \, \mu(dz).$$

Let us designate an arbitrary function l on Z by $l = (l_\nu, l_k)$.

The dressing transformation is given by

$$(\Delta f)_\nu = f_0 + \sum \frac{\rho_k f_k}{\lambda_\nu - \lambda_k} = (\zeta f)(\lambda_\nu),$$

$$(\Delta f)_k = f_k.$$

The indefinite metric in $L^2(Z, |\mu|)$ is given by

$$\eta(h_\nu, h_k) = (\operatorname{sgn}(r_\nu)h_\nu, \operatorname{sgn}(1 + \beta_k\rho_k^2)h_k).$$

We cannot impose well dressing on R_λ unless we rig $L^2(Z)$, and although there is no canonical way of doing so, we pick the following: Let Ψ_- be the set of functions $(h_\nu, h_k) \in L^2$ which vanish at all but a finite set of points of nonzero weight. For Ψ_+ we pick the set of functions (ϕ_ν, ϕ_k) with *arbitrary* components. The pairing is of course

$$\langle f, \phi\rangle = \sum |r_\nu| \bar{f}_\nu\phi_\nu + \sum |(1 + \beta_k\rho_k^2)| \bar{f}_k\phi_k.$$

Strictly speaking, because of the possibility that $1 + \beta_k\rho_k^2 = 0$ for some k, one should use equivalence classes but for convenience we express everything in terms of representatives.

If we consider Δ to be in $M(\Phi_+, \Psi_+)$, then we find, for $\phi = (\phi_0; \cdots, \phi_k, \cdots) \in \Phi_+$,

$$\bar{\Delta}\phi = \{(\phi_\nu, \phi_k) \mid \phi_\nu \in \mathbb{C}\}.$$

A computation now shows that if $\alpha(\lambda) \not\equiv 0$, then R_λ is well dressed in the loosest sense but not in the sense $1/(\lambda - T_R) \subset \bar{\Delta}R_\lambda\bar{\Delta}^{-1}$ *unless* $\beta_k = -1/\rho_k^2$. Thus the resolvents that are analytic at $\lambda = \lambda_k$ are singled out from the others by being in a sense better dressed. We continue to study only this case.

If $\alpha(\lambda)$ were to have any zero other than at λ_k, then $R_\lambda(1; 0, 0, \cdots) = 0$ at these zeros and nondegeneracy rules this out. We are now thus confronted with the construction of a function which has zeros of the form $(-1/\rho_k^2)(\lambda - \lambda_k)$ at $\lambda = \lambda_k$ and only these. If we set $\tau(\lambda) = 1/\alpha(\lambda)$, then $\tau(\lambda)$ must be meromorphic with poles of the form $-\rho_k^2/(\lambda - \lambda_k)$ at $\lambda = \lambda_k$ and only these. The most general function of this form can be written as

$$\tau(\lambda) = E(\lambda) - \sum \left(\frac{\lambda}{\lambda_k}\right)^{n_k} \frac{\rho_k^2}{\lambda - \lambda_k},$$

where $E(\lambda)$ is entire and the n_k are integers picked so that the series converges to a meromorphic function.

There are now two possible situations: Either it is possible to take $n_k \leq n$ where n is a fixed integer or else it is impossible to do so. We first study the former case.

There are in addition two important subcases to consider:

(1) One can take $n_k = 0$;

(2) one can take $n_k \leq 1$ but not $n_k = 0$.

If we can take $n_k = 0$, then

$$\tau(\lambda) = E(\lambda) - \sum_{k=1}^{\infty} \frac{\rho_k^2}{\lambda - \lambda_k}.$$

Imposing the ray limit, we must have $\tau(\lambda)/\lambda \to 1/\eta$ as $|\lambda| \to \infty$ along $|\lambda| e^{i\theta}$, $\theta \neq 0, \pi$. Since

$$\frac{1}{\lambda}\sum \frac{\rho_k^2}{\lambda - \lambda_k} \to 0,$$

we must have $E(\lambda)/\lambda \to 1/\eta$, which implies that $E(\lambda) = (\lambda/\eta) + \sigma$, where σ is real. We see that we can take a normal ray limit $\eta = 1$ and then we have

$$\tau(\lambda) = \lambda + \sigma - \sum_{k=1}^{\infty} \frac{\rho_k^2}{\lambda - \lambda_k}.$$

This reproduces precisely the formal result if σ is taken to be $-\lambda_0$. Thus, in case $I(\lambda) = \sum \rho_k^2/(\lambda - \lambda_k)$ converges, our methods single out the formal solution up to a choice of σ. These solutions satisfy positivity in the whole upper half-plane and are analytic off the real axis everywhere except for $\lambda = \infty$. We further note that these solutions are not restricted only to the case of bona-fide operators T which would mean $\sum \rho_k^2 < \infty$ but also include generalized operators which are not bona fide since we only require that $\sum \rho_k^2/\lambda_k < \infty$.

To get an insight into the general case, consider the formal sum $I(\lambda) = \sum \rho_k^2/(\lambda - \lambda_k)$ even in the case when it is divergent. We note that the sum

$$I_n(\lambda) = \sum \left(\frac{\lambda}{\lambda_k}\right)^n \frac{\rho_k^2}{\lambda - \lambda_k}$$

is a subtracted sum obtained from $I(\lambda)$ by subtracting the first n terms of the Taylor expansion of $I(\lambda)$ about $\lambda = 0$:

$$\begin{aligned} I_n(\lambda) &= \sum \left(\frac{\lambda}{\lambda_k}\right)^n \frac{\rho_k^2}{\lambda - \lambda_k} \\ &= I(\lambda) - I(0) - \lambda I'(0) \\ &\quad + \cdots - \frac{\lambda^{n-1}}{n-1!} I^{(n-1)}(0) \end{aligned}.$$

so that in case we can take $n_k \leq n$ but not $n_k \leq n - 1$ we see that $\sum (\lambda/\lambda_k)^n \rho_k^2/(\lambda - \lambda_k)$ is obtained from $I(\lambda)$ by subtracting a polynomial of degree $n - 1$ *with divergent coefficients*. One could then consider a further adjustment by a polynomial of degree $n - 1$ with *finite* coefficients and so whenever we can take

$n_k \leq n$ but not $n_k \leq n-1$, we obtain a choice for $\tau(\lambda)$ of the form

$$\tau_n(\lambda) = \lambda + \sigma + P_{n-1}(\lambda) - \sum_1^\infty \left(\frac{\lambda}{\lambda_k}\right)^n \frac{\rho_k^2}{\lambda - \lambda_k}.$$

Here P_{n-1} is a real polynomial of degree $n-1$. Notice that if it is possible to take $n_k = n$, then it is not necessary to take *all* the n_k equal to n, for we could take any finite, and possibly some infinite, subset of them to be less; this, however, corresponds to a change of P_{n-1}; the presence of P_{n-1} is therefore reasonable since there is no *a priori* reason to "renormalize" the whole sum $\sum \rho_k^2/(\lambda - \lambda_k)$, for one could single out a convergent subsum and then "renormalize" the rest.

The above formal game is of course very familiar in theoretical physics where one adds counterterms to divergent formal expressions to make them converge and then allows for the possibility of further adjustments by finite counterterms of the same type as the infinite ones.

In case we cannot take $n_k \leq n$, then we see that $I(\lambda)$ needs an infinite number of subtractions and we do not have an argument, not even a formal one, for singling out a particular $E(\lambda)$. We do not pursue this case further; it corresponds to the so-called "non-renormalizable" case of conventional formal theory.

We note that in case $n = 1$, P_{n-1} must be a constant which we can absorb into σ; we have

$$\tau_1 = \lambda + \sigma - \sum_1^\infty \left(\frac{\lambda}{\lambda_k}\right) \frac{\rho_k^2}{\lambda - \lambda_k}.$$

The above expression would also be the formal solution *if we take* λ_0 *to be infinite*. Letting $\lambda_0 = -\sigma + \sum \rho_k^2/\lambda_k$ in T, a formal calculation will yield the τ_1 given above; thus in this case we can obtain the solution by adding an infinite counterterm to T, namely by adding

$$K: f \rightsquigarrow ((-\sigma + (\textstyle\sum \rho_k^2/\lambda_k) - \lambda_0) f_0; 0, 0, 0 \cdots)$$

to the original T. This counterterm is of the same form as an "adjustment of the vacuum self-energy." Whether τ_n for $n > 1$ can be achieved by suitable formal counterterms in T we do not know.

We note that τ_1 satisfies positivity in the whole upper half-plane, for we have

$$I_1 = \sum_1^\infty (\lambda/\lambda_k)\rho_k^2/(\lambda - \lambda_k)$$
$$= \sum_1^\infty (|\lambda|^2 - \lambda\lambda_k)\rho_k^2/\lambda_k |\lambda - \lambda_k|^2,$$

whose imaginary part is $-\operatorname{Im} \lambda \sum \rho_k^2/|\lambda - \lambda_k|^2$ and so $\operatorname{Im} \tau_1 = \operatorname{Im} \lambda(1 + \sum \rho_k^2/|\lambda - \lambda_k|^2) > 0$ for $\operatorname{Im} \lambda > 0$ and so $\operatorname{Im} \alpha(\lambda) = -\operatorname{Im} \tau_1(\lambda)/|\tau_1(\lambda)|^2 < 0$ for $\operatorname{Im} \lambda > 0$. This also implies that $\tau_1(\lambda)$ has no complex zeros and so $\alpha(\lambda)$ is analytic off the real axis. A computation also reveals that R_λ itself satisfies positivity in the whole upper half-plane and is analytic except for poles at $\lambda = \lambda_\nu$ where λ_ν are the zeros of $\tau_1(\lambda)$. Now $\tau_1(\lambda)$ is a Herglotz function[14] and so we have along the rays $|\lambda|e^{i\theta}$, $\theta \neq 0, \pi$, that $\tau_1(\lambda)/\lambda \to 1$ and so the normal ray limit is satisfied.

We therefore have that in the two distinguished subcases we are able to satisfy all of our requirements. These two subcases contain not only the bona fide discrete Lee models but also certain generalized discrete Lee models including the ones requiring an infinite formal adjustment of λ_0 but no further renormalizations.

Let us now return to the general case of $n_k \leq n$ but not $n_k \leq n-1$ where $n \geq 2$. Consider

$$I_n(\lambda)/\lambda^{n-1} = \lambda \sum \rho_k^2/\lambda_k^n(\lambda - \lambda_k) = \lambda H_n(\lambda),$$

where $H_n(\lambda) = \sum \rho_k^2/\lambda_k^n(\lambda - \lambda_k)$; $H_n(\lambda)$ is manifestly Herglotz and if along the rays $|\lambda|\, e^{i\theta}$, $\theta \neq 0, \pi \lambda H_n(\lambda)$ were to have a finite ray independent limit, then this limit would be[15] $\sum \rho_k^2/\lambda_k^n$, which, however, diverges by the choice of n. We must therefore conclude that for $n \geq 2$ the normal ray limit cannot be satisfied.

The next question we want to consider is whether one can pick P_{n-1} so that the residues r_ν of $\alpha(\lambda)$ at λ_ν are positive. This has a simple geometric interpretation: If we draw the graphs of $I_n(\lambda)$ and of $\lambda + \sigma + P_{n-1}(\lambda)$, the latter must intersect the former *from below*. The only possible unavoidable difficulty can occur in the region $\lambda < \lambda_1$. Now $I_n(\lambda)$ is monotonic in the region $-\infty < \lambda \leq 0$ and $I_n(\lambda) \to \pm\infty$ as $\lambda \downarrow -\infty$, depending on whether n is odd or even; furthermore, $I_n(\lambda) \to -\infty$ as $\lambda \uparrow \lambda_1$ and we see that it is indeed possible to pick P_{n-1} so that $r_\nu > 0$.

Finally, we want to know whether $\tau_n(\lambda)$ has any *complex* zeros. If it does, then R_λ will not be analytic off the real axis but will have poles at these locations. Now $\tau_n(\lambda)$ certainly *can* have complex zeros since for example if we look at the equation $\tau_n(i) = 0$ we see that for $n \geq 2$ we can always pick the real coefficients of P_{n-1} so as to satisfy the equation, showing that at least for some P_{n-1}, $\tau_n(\lambda)$ has complex zeros. We have, however, been unable to determine whether one can pick P_{n-1} so that $\tau_n(\lambda)$ has no complex zeros and furthermore if that be possible whether we can in addition satisfy positivity.

Let us now consider the effect of introducing a cut-off into $\rho: \rho_k' = \rho_k h_k$ such that $0 < h_k \leq 1$ and such that with the cutoff the sum $\sum (\rho_k')^2/(\lambda - \lambda_k)$ converges. We note that as the cutoff is removed, the function

$\tau(\lambda)$ does not converge to the corresponding function without cutoffs; only by subtracting from $\tau(\lambda)$ an appropriate entire function depending on the cutoff do we obtain a converging limit. This observation sheds light on the behavior under cutoffs that was noted in Sec. IIF; there these subtractions were not taken into account and the limiting situation there actually corresponds to the case $\alpha(\lambda) \equiv 0$ which we have not considered.

For another insight into this model let us apply analytic subtraction to the well-dressing equation $[1/(\lambda - T_R)]\Delta = \Delta T$. We take T_R to be a multiplication on Ψ_+ given by $T_R(\psi_\nu, \psi_k) = (\lambda_\nu \psi_\nu, \lambda_k \psi_k)$; then we find that the well-dressing requirement gives

$$\frac{1}{\lambda - \lambda_\nu}\zeta f(\lambda_\nu) = \alpha(\lambda)\zeta f(\lambda) + \sum \frac{\rho_k f_k}{(\lambda_\nu - \lambda_k)(\lambda - \lambda_k)} + \alpha(\lambda)\zeta f(\lambda)\tilde{\sum}\frac{\rho_k^2}{(\lambda - \lambda_k)(\lambda_\nu - \lambda_k)},$$

$$\frac{1}{\lambda - \lambda_k} f_k = \frac{1}{\lambda - \lambda_k} f_k + \alpha(\lambda)\frac{\rho_k}{\lambda - \lambda_k}\zeta f(\lambda) \quad \text{if} \quad 1 + \beta_k \rho_k^2 \neq 0.$$

So unless $\alpha(\lambda) \equiv 0$, we must have $\beta_k = -1/\rho_k^2$. Using the identity

$$\frac{1}{(\lambda_\nu - \lambda_k)(\lambda - \lambda_k)} = \frac{1}{\lambda - \lambda_\nu}\left(\frac{1}{\lambda_\nu - \lambda_k} - \frac{1}{\lambda - \lambda_k}\right),$$

we find that the first requirement leads to

$$\alpha(\lambda)\left(\lambda - \lambda_\nu + \tilde{\sum_k}\frac{\rho_k^2}{\lambda_\nu - \lambda_k} - \tilde{\sum}\frac{\rho_k^2}{\lambda - \lambda_k}\right) = 1.$$

For this to make sense we must have

$$-\lambda_\nu + \tilde{\sum}\frac{\rho_k^2}{\lambda_\nu - \lambda_k} = \sigma$$

equals a constant independent of ν; so

$$\alpha(\lambda) = \left(\lambda + \sigma - \tilde{\sum}\frac{\rho_k^2}{\lambda - \lambda_k}\right)^{-1}$$

and the λ_ν are determined by the equation

$$\lambda_\nu + \sigma - \tilde{\sum}\frac{\rho_k^2}{\lambda_\nu - \lambda_k} = 0$$

and this is precisely the answer we obtained by other methods. We see therefore that certain considerations of analyticity along with well dressing appropriately interpreted is all that is necessary for this model.

One begins to see from this example that the purely multivalued approach to the resolvent is at times a coarse method at arriving at the intrinsic structural properties of T, for we can arrive at the same results more quickly by methods using analytic subtractions. We believe that for the discrete Lee model the conclusion that $\tau(\lambda)$ must have a pole of the form $-\rho_k^2/(\lambda - \lambda_k)$ at $\lambda = \lambda_k$ is an expression of an intrinsic property; what other intrinsic properties have been exhibited in the above calculations is not very clear.

Example 18: Let us now consider a generalized operator T on $\mathbb{C} \oplus d \oplus d^2 \subset \mathbb{C} \oplus d' \oplus d^{2'}$ defined by

$$T(f_0, f_i, f_{ij}) = \left(\lambda_0 f_0 + \sum_{i=1}^{\infty}\rho_i f_i + \sum_{i,j=1}^{\infty}\sigma_{ij}f_{ij}, \lambda_i f_i + \rho_i f_0 + \sum_{j=1}^{\infty}\rho_j f_{ji}, \lambda_{ij}f_{ij} + \rho_i f_j + \sigma_{ij}f_0\right),$$

where λ_0, λ_i, λ_{ij}, ρ_i, and σ_{ij} are all real numbers. The multivalued approach is a bit crude here, for we have, whenever $\rho \notin d$,

$$\bar{T}(\phi_0, \phi_i, \phi_{ij}) = \{(\alpha, \beta_i, \lambda_{ij}\phi_{ij} + \rho_i\phi_i + \sigma_{ij}\phi_0) \mid \alpha, \beta_i \in \mathbb{C}\}$$

and a computation of $\mathcal{R}_\lambda$ will reveal that the possibilities for the resolvent are too numerous. We now apply orthogonal methods. Let $\Phi_{-1} = \mathbb{C} \oplus d$, $\Phi_{-2} = d^2$; then

$$T_{11}: (f_0, f_i) \rightsquigarrow (\lambda_0 f_0 + \sum \rho_i f_i, \lambda_i f_i + \rho_i f_0),$$
$$T_{12}: \{f_{ij}\} \rightsquigarrow (\sum \sigma_{ij} f_{ij}, \sum \rho_j f_{ji}),$$
$$T_{21}: (f_0, f_i) \rightsquigarrow \{\rho_i f_j + f_0 \sigma_{ij}\},$$
$$T_{22}: \{f_{ij}\} \rightsquigarrow \{\lambda_{ij} f_{ij}\},$$

so

$$P_\lambda: \{f_{ij}\} \rightsquigarrow \left\{\frac{f_{ij}}{\lambda - \lambda_{ij}}\right\}.$$

By analytic subtraction we find

$$T_{12}P_\lambda T_{21}: (f_0, f_i) \rightsquigarrow \left(f_0 \tilde{\sum_{i,j}}\frac{\sigma_{ij}^2}{\lambda - \lambda_{ij}} + \sum_{j=1}^{\infty} f_j \tilde{\sum_{i=1}^{\infty}}\frac{\sigma_{ij}\rho_i}{\lambda - \lambda_{ij}}, f_i \tilde{\sum_{l=1}^{\infty}}\frac{\rho_l^2}{\lambda - \lambda_{li}} + f_0 \tilde{\sum_{j=1}^{\infty}}\frac{\rho_j \sigma_{ji}}{\lambda - \lambda_{ji}}\right).$$

Let now

$$\lambda_0(\lambda) = \lambda_0 + \tilde{\sum_{i,j=1}^{\infty}}\frac{\sigma_{ij}^2}{\lambda - \lambda_{ij}},$$

$$\rho_i(\lambda) = \rho_i + \tilde{\sum_{l=1}^{\infty}}\frac{\rho_l \sigma_{li}}{\lambda - \lambda_{li}},$$

$$\lambda_i(\lambda) = \lambda_i + \tilde{\sum_{l=1}^{\infty}}\frac{\rho_l^2}{\lambda - \lambda_{li}}.$$

Then

$$T_\lambda:(f_0,f_i) \rightsquigarrow \left(\lambda_0(\lambda)f_0 + \sum_1^\infty \rho_i(\lambda)f_i,\ \lambda_i(\lambda)f_i + \rho_i(\lambda)f_0\right).$$

We see that Q_λ can be computed by orthogonal methods in precisely the same way as the discrete Lee model. Let

$$\alpha(\lambda) = \left(\lambda - \lambda_0(\lambda) - \sum_{i=1}^\infty \frac{\rho_i(\lambda)^2}{\lambda - \lambda_i(\lambda)}\right)^{-1},$$

$$\mathfrak{D}_\lambda(f_0,f_i) = f_0 + \sum_{i=1}^\infty \frac{\rho_i(\lambda)f_i}{\lambda - \lambda_i(\lambda)};$$

then

$$Q_\lambda:(f_0,f_i) \rightsquigarrow \left(\alpha(\lambda)\mathfrak{D}_\lambda(f_0,f_i),\ \frac{f_i}{\lambda - \lambda_i(\lambda)} + \frac{\rho_i(\lambda)}{\lambda - \lambda_i(\lambda)}\alpha(\lambda)\mathfrak{D}_\lambda(f_0,f_i)\right).$$

We can now compute R_λ without introducing any new subtracted sums.

Example 19: We now consider a rigged Fock space which is the direct sum $\oplus_{n=0}^\infty \Phi^{(n)}$, where $\Phi^{(0)}$ is $\mathbb{C} \subset \mathbb{C}$ and $\Phi^{(n)}$ is $d^n \subset d'^n$. Define the generalized operator

$$(Tf)^{(n)}_{i_1\cdots i_n} = \lambda_{i_1\cdots i_n} f^{(n)}_{i_1\cdots i_n} + \sum_{l=1}^\infty \rho_l f^{(n+1)}_{li_1\cdots i_n} + \rho_{i_1} f^{(n-1)}_{i_2\cdots i_n}.$$

Note that if we introduce the annihilation and creation operators

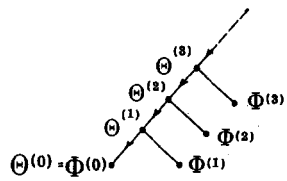
$$(\mathcal{A}_k^+ f)^{(n)}_{i_1\cdots i_n} = \delta_{i_1 k} f^{(n-1)}_{i_2\cdots i_n},$$
$$(\mathcal{A}_k f)^{(n)}_{i_1\cdots i_n} = f^{(n+1)}_{ki_1\cdots i_n},$$

then $T = \Lambda + \sum_{i=1}^\infty \rho_i(\mathcal{A}_i + \mathcal{A}_i^+)$, where Λ is the diagonal part of T.

One easily sees that if $\rho \notin d$, then $\bar{T}\psi = \Phi_+$ so that the multivalued approach is totally ineffective unless supplemented by additional analysis. For example, if we write $\Phi = \Phi_1 \oplus \Phi_2$, where

$$\Phi_1 = \bigoplus_{i=0}^N \Phi^{(i)} \quad \text{and} \quad \Phi_2 = \bigoplus_{i=N+1}^\infty \Phi^{(i)},$$

then $\bar{T}_{11} \oplus \bar{T}_{22}$ is strictly smaller than $\bar{T}$, and given a resolvent for T_{22} the problem can be treated by orthogonal methods. On the other hand, $\bar{T}_{22}\psi_2 = \Phi_{+2}$ and so the original problem is not entirely avoided. To treat this generalized operator adequately, one needs an infinite nesting of orthogonal decompositions to arrive at something like an infinite-continued fraction expansion of R_λ with possibly an infinite number of subtracted sums. Let $\Theta^{(n)} = \oplus_0^n \Phi^{(n)}$; then if we nest according to the graph



then the infinite continued fractions in R_λ can actually be computed. We give only $(R_\lambda)_{00}$ here:

$$(R_\lambda)_{00} = \cfrac{1}{\lambda - \lambda_0 - \sum_{i=1}^\infty \cfrac{\rho_i^2}{\lambda - \lambda_i - \sum_{j=1}^\infty \cfrac{\rho_j^2}{\lambda - \lambda_{ji} - \sum_{k=1}^\infty \cfrac{\rho_k^2}{\lambda - \lambda_{kji} - \cdots}}}}.$$

Example 20: We now study an example which exhibits a further difficulty: On $\mathfrak{D}(\mathbb{R}^2) \subset \mathfrak{D}'(\mathbb{R}^2)$ let T be the generalized operator of multiplication by $t(x)\delta(y)$ where t is a real C^∞ function:

$$(Tf)(x,y) = t(x)\delta(y)f(x,y) = t(x)f(x,0)\delta(y).$$

Now if f_U is an approximating net for the distribution τ, then it is possible to pick f_U in such a way that $f_U(x,0)$ approaches weakly in $\mathfrak{D}'(\mathbb{R})$ any given distribution σ. Thus we have

$$\bar{T}\tau = t(x)\mathfrak{D}'_x\delta(y) = \{t(x)\sigma(x)\delta(y) \mid \sigma \in \mathfrak{D}'(\mathbb{R})\}.$$

Straightforwardly solving for the resolvent, we get

$$\mathcal{R}_\lambda\tau = \frac{1}{\lambda}\tau + t(x)\mathfrak{D}'_x\delta(y), \quad \lambda \neq 0,$$
$$\mathcal{R}_0\tau = \mathfrak{D}'(\mathbb{R}^2) \quad \text{if} \quad \tau \in t(x)\mathfrak{D}'_x\delta(y),$$
$$= \varnothing \quad \text{otherwise},$$
$$\mathcal{S}_\infty\tau = \tau + t(x)\mathfrak{D}'_x\delta(y).$$

The g-operator versions of the above are

$$R_\lambda f = \frac{1}{\lambda} f + t(x)\rho_\lambda(x)(f)\delta(y), \quad \lambda \neq 0,$$
$$R_0 f = 0 \quad \text{if} \quad f = 0,$$
$$= \varnothing \quad \text{otherwise},$$
$$S_\infty f = f + t(x)\sigma(x)(f)\delta(y),$$

where $\rho_\lambda(x)(f)$ and $\sigma(x)(f)$ are in $\mathfrak{D}'$ and depend linearly on f.

Hermiticity requires that

$$\langle g(x, y), t(x)\rho_\lambda(x)(f)\delta(y)\rangle = \overline{\langle f(x, y)t(x)\rho_{\bar\lambda}(x)(g)\delta(y)\rangle}$$

or

$$\langle g(x, 0), t(x)\rho_\lambda(x)(f)\rangle = \overline{\langle f(x, 0), t(x)\rho_{\bar\lambda}(x)(g)\rangle}.$$

Now this requires $t(x)\rho_\lambda(x)(f)$ to be of the form

$$t(x)\int r_\lambda(x, z)f(z, 0)\, dz,$$

where $t(x)r_\lambda(x, z) \in \mathfrak{D}'_{xz}$; furthermore

$$t(x)r_\lambda(x, y) = t(y)\overline{r_{\bar\lambda}(y, x)}.$$

Similarly,

$$t(x)\sigma(x)(f) = t(x)\int s(x, z)f(z, 0)\, dz,$$

where $t(x)s(x, z) \in \mathfrak{D}'_{x,z}$ and

$$t(x)s(x, y) = t(y)\overline{s(y, x)}.$$

Now it is in fact fairly apparent that one can find appropriate r_λ and s to satisfy all the conditions so far expressed; however, the difficulty again is that there are too many such solutions. From the explicit form of T one can, however, write down what answer seems appropriate: One should like the renormalized Hilbert space to be say

$$L^2(\mathbb{R}^2) \oplus L^2(\text{supp } (t), \gamma\, dx),$$

where $\gamma > 0$ is a constant, and for T_R to be given by

$$T(f(x, y) \oplus g(x)) = 0 \oplus \beta t(x)g(x),$$

where $\beta \in \mathbb{R}$ is another constant. The dressing transformation is then given by

$$\Delta f = f(x, y) \oplus f(x, 0).$$

This choice corresponds to the choice

$$r_\lambda(x, y) = \frac{1}{t(x)}\chi_{\text{supp}\,(t)}(x)\frac{\gamma}{\lambda - \beta t(x)}\delta(x - y)$$

and

$$s(x, y) = \frac{1}{t(x)}\chi_{\text{supp}\,(t)}(x)\gamma\delta(x - y).$$

Formally since T is a multiplication operator, the eigenstates are $\psi_{ab} = \delta(x - a)\delta(y - b)$ and $T\psi_{ab} = t(a)\delta(b)\psi_{ab}$ so $T\psi_{ab} = 0$ for $b \neq 0$ or $t(a) = 0$ and $T\psi_{a0} = t(a)\delta(0)\psi_{a0} = \lambda_a\psi_{a0}$.

In the solution for r_λ and s presented above $\delta(0)$ has been "renormalized" to the number β and all generalized vectors of the form $\chi_{\text{supp}\,(t)}(x)h(x)\delta(y)$, where $h \in L^2[\text{supp } (t)]$ have been made into new bona fide Hilbert space states in $L^2(\text{supp } (t), \gamma\, dx)$.

We have now to extend our program to be able to do two new things: Show how the solutions in which r_λ is "diagonal," that is, $r_\lambda(x, y) = r_\lambda(x)\delta(x - y)$, are singled out; and then show how the solutions $r_\lambda(x) = [1/t(x)]\chi_{\text{supp}\,(t)}(x)\gamma/[\lambda - \beta t(x)]$ are singled out. We note that the second part is independent of the first since if we had chosen $r_\lambda(x) = [1/t(x)]\chi_{\text{supp}\,(t)}(x)1/[\lambda - u(x)]$, where $u(x)$ is *any* real measurable function, then we still would have a solution satisfying all our conditions. Notice that this solution corresponds to the case in which formally $\lambda_a = t(a)\delta(0)$ has been "renormalized" to $u(a)$ so that $\delta(0)$ has been "renormalized" to $u(a)/t(a)$, which renormalization, however, depends now on a. Unfortunately, we have as yet been unable, except for the comments to follow, to formulate effective conditions to pick out the "natural" answers.

One begins here to appreciate some of the difficulties of the multivalued approach, noticing that the new conditions cannot refer to $\mathcal{R}_\lambda$ and $\mathcal{S}_\infty$ alone (or equivalently to $\bar T$ alone); for in these relations the space $t(x)\mathfrak{D}'_x\delta(y)$ occurs and one can find many other C^∞ functions $w(x)$ such that $t(x)\mathfrak{D}'_x\delta(y) = w(x)\mathfrak{D}'_x\delta(y)$ and so the function $t(x)$ has become lost. Thus it is necessary to consider T itself along with $\bar T$ and to explore more thoroughly the relationship between a generalized operator and its closure.

We isolate one part of the difficulty by the following observation. Consider a direct sum of two rigged Hilbert spaces $\Phi = \Phi_1 \oplus \Phi_2$, and let T be a direct sum of two generalized operators $T = T_1 \oplus T_2$. Then clearly $\bar T = \bar T_1 \oplus \bar T_2$ and $\mathcal{R}_\lambda(\bar T) = \mathcal{R}_\lambda(\bar T_1) \oplus \mathcal{R}_\lambda(\bar T_2)$; however, it is *not* true that any g-operator version of $\mathcal{R}_\lambda$ is a direct sum $R_\lambda(T_1) \oplus R_\lambda(T_2)$ of g-operator versions of $\mathcal{R}_\lambda(\bar T_1)$ and $\mathcal{R}_\lambda(\bar T_2)$, respectively. The reason for this is easy to see for $\mathcal{R}_\lambda(\bar T_1)$ and $\mathcal{R}_\lambda(\bar T_2)$ are in general multivalued and this heuristically speaking is expressed by saying that $\mathcal{R}_\lambda(\bar T_i)f$ contain arbitrary parameters; now in picking a g-operator version of $\mathcal{R}_\lambda(\bar T)$ these arbitrary parameters must become functions on Φ_-; but as Φ_- is a direct sum $\Phi_{-1} \oplus \Phi_{-2}$ we see that an arbitrary parameter in $\mathcal{R}_\lambda(\bar T_1)$ can now become a function of $f \in \Phi_-$ which *depends on the part of* f *in* Φ_{-2} and this will violate the direct

sum decomposition. This leads to the possibility of choosing a nondiagonal $r_\lambda(x, y)$ above, for multiplication by $t(x)\delta(y)$ is formally a direct integral of multiplications by the δ function in $\mathcal{D}(\mathbb{R}) \subset \mathcal{D}'(\mathbb{R})$. We now introduce two conditions to partly overcome this problem.

Condition 10 (Weak diagonalizability): We say that a $\mathfrak{W}^+$ closed subspace L of Φ_+ is a *weak invariant subspace* of a generalized operator T if first of all the $\mathfrak{W}^+$ closure of $L \cap \Phi_-$ is L and furthermore if $T(L \cap \Phi_-) \subset L$. We see that Φ_+ and $\{0\}$ are always weak invariant subspaced and these may be the only ones. We say that a resolvent R_λ of T is *weakly diagonal* if whenever L is a weak invariant subspace of T, then it is also a weak invariant subspace of R_λ.

Notice that this condition forces the resolvent for a direct sum to be a direct sum of resolvents, for we have that $\Phi_{+1} \oplus 0$ and $0 \oplus \Phi_{+2}$ are weak invariant subspaces and therefore in picking an $R_\lambda \subset R_\lambda$ we are forced to take a resolvent of the form $R_\lambda(T_1) \oplus R_\lambda(T_2)$. This is the same result that is obtained if we compute R_λ by orthogonal methods applied to the decomposition $\Phi_1 \oplus \Phi_2$ since in this case $T_{12} = 0$ and $T_{21} = 0$.

For the case of multiplication by $t(x)\delta(y)$ if we choose the family of weak invariant subspaces $L_{ab} = \{\phi \in \mathcal{D}'(\mathbb{R}^2) \mid \text{supp } (\phi) \subset [a, b] \times \mathbb{R}\}$, then we find that $r_\lambda(x, y)$ must have its support in the set $\{x = y\}$ though it need not be of the form $r_\lambda(x) \times \delta(x - y)$.

The existence of weak invariant subspaces of T is very sensitive to the rigging. An example that illustrates this dramatically is that of the bona fide operator q of multiplication by x on either $\mathcal{D}(\mathbb{R}) \subset \mathcal{D}'(\mathbb{R})$ or $\mathcal{S}(\mathbb{R}) \subset \mathcal{S}'(\mathbb{R})$, where $\mathcal{S}(\mathbb{R})$ is the Fourier transform of $\mathcal{D}(\mathbb{R})$ and $\mathcal{S}'(\mathbb{R})$ is the Fourier transform of $\mathcal{D}'(\mathbb{R})$. In the first case, every subspace of $\mathcal{D}'$ having support in a closed interval is a weak invariant subspace of q; in the second case, because every element of $\mathcal{S}(\mathbb{R})$ is an entire function we see that the only weak invariant subspaces of q are $\{0\}$ and $\mathcal{S}'$. It is useful at this point to introduce a new notion: We say that a rigging is *diagonal* with respect to a generalized operator T if in some appropriate sense the set of weak invariant subspaces of T is as large as possible. A diagonal rigging like a complete rigging is one which is naturally adapted to the structure of T; furthermore a diagonal rigging is in a sense one which comes closest to the attempt at constructing a rigging which exhibits only the physically relevant generalized eigenvectors of T and should thus supply the most natural extension of the Gel'fand–Vilenkin formalism. Diagonal riggings and complete riggings stand at opposite extremes in this sense.

The above considerations touch upon the notion of *direct integrals*. We shall here give a rather unnatural but useful notion of a direct integral of rigged Hilbert spaces. Let Φ be a rigged Hilbert space; then we say Φ is a direct integral of rigged Hilbert spaces Φ_x, where $x \in X$ and X is a measure space with a measure μ if $\Phi_0 = \int^\oplus \Phi_{0x}\mu(dx)$ a direct integral of Hilbert spaces, Φ_- is a subspace of the space of vectors of Φ_0 of the form $\int^\oplus f_x\mu(dx)$, $f_x \in \Phi_{-x}$ and each $\phi \in \Phi_+$ can be expressed as $\int^\oplus \phi_x\mu(dx)$, $\phi_x \in \Phi_{+x}$. We now introduce the following:

Condition 11 (Direct Integral Condition): If Φ is a direct integral of Φ_x and T a generalized operator of the form $\int^\oplus T_x\mu(dx)$, that is,

$$T\int^\oplus f_x\mu(dx) = \int^\oplus (T_x f_x)\mu(dx),$$

where $T_x : \Phi_{-x} \to \Phi_{+x}$ is a generalized operator on Φ_x, then we require that $R_\lambda = \int^\oplus (R_\lambda)_x\mu(dx)$, where $(R_\lambda)_x$ is a resolvent of T_x.

If we view our example as a direct integral of multiplication by the δ function, $\int^\oplus M_{t(x)\delta}\,dx$, then we are forced to take $r_\lambda(x, y)$ to be $r_\lambda(x)\delta(x - y)$ with $r_\lambda(x) = [1/t(x)]\chi_{\text{supp }(t)}(x)1/[\lambda - u(x)]$ but still we do not single out $u(x) = \beta t(x)$.

Example 21: We now want to consider the $N\theta$ sector of the Lee model in n space dimensions, $n \geq 1$. We work with the rigged Hilbert space

$$\mathbb{C} \oplus \mathcal{D}(\mathbb{R}^n) \subset \mathbb{C} \oplus \mathcal{D}'(\mathbb{R}^n)$$

and with the Hamiltonian

$$T(f_0, f_1) = \left(-\int \frac{f_1(\mathbf{k})Q(\mathbf{k})d^nk}{[2\omega(k)]^{\frac{1}{2}}},\ \omega(k)f_1(\mathbf{k}) - \frac{f_0Q(\mathbf{k})}{[2\omega(\mathbf{k})]^{\frac{1}{2}}}\right).$$

Here $\omega(k) = (m^2 + k^2)^{\frac{1}{2}}$ and $Q(\mathbf{k})$ is a C^∞ cutoff function $0 \leq Q(\mathbf{k}) \leq q_0$, where q_0 is the bare coupling constant. We introduce the cutoff function in order to be able to compare our results with the conventional ones but of course we shall also be able to work in the case $Q(k) \equiv q_0$. We assume $Q \notin \mathcal{D}(\mathbb{R}^n)$, for otherwise $\bar{T}$ would be single valued and the results could not be extended to the case of no cutoff. One sees immediately that

$$\bar{T}\psi = \bar{T}(\psi_0, \psi_1) = \left\{\left(\alpha, \omega(k)\psi_1(k) - \frac{\psi_0Q(\mathbf{k})}{[2\omega(k)]^{\frac{1}{2}}}\right) \,\middle|\, \alpha \in \mathbb{C}\right\},$$

where $\psi_1 \in \mathcal{D}'(\mathbb{R}^n)$.

The resolvents are easily computable; they are:

$\lambda \notin [m, \infty) \subset \mathbb{R}$:

$$\mathcal{R}_\lambda \psi = \left\{ \left(\alpha, \frac{1}{\lambda - \omega(k)} \left(\psi_1(\mathbf{k}) - \alpha \frac{Q(\mathbf{k})}{[2\omega(k)]^{\frac{1}{2}}} \right) \right) \middle| \alpha \in \mathbb{C} \right\};$$

$\lambda \in (m, \infty)$:

$$\mathcal{R}_\lambda \psi = \left\{ \left(\alpha, \text{“} \frac{1}{\lambda - \omega(k)} \left(\psi_1(\mathbf{k}) - \alpha \frac{Q(\mathbf{k})}{[2\omega(k)]^{\frac{1}{2}}} \right) \text{”} + \beta(\hat{k}) \delta[\lambda - \omega(k)] \right) \middle| \alpha \in \mathbb{C}, \beta(\hat{k}) \in \mathcal{D}'(S^{n-1}) \right\},$$

where we have put the expression

$$\{1/[\lambda - \omega(k)]\}\{\psi(\mathbf{k}) - \alpha Q(\mathbf{k})\}/[2\omega(k)]^{\frac{1}{2}}$$

in quotes because this is ambiguous as a distribution but it is defined up to a term of the type $\beta(\hat{k})\delta[\lambda - \omega(k)]$. Here $\hat{k} = \mathbf{k}/|\mathbf{k}|$ and S^{n-1} is the $(n-1)$-dimensional sphere. Furthermore,

$\lambda = m$:

$$\mathcal{R}_\lambda \psi = \left\{ \left(\alpha, \text{“} \frac{1}{\lambda - \omega(k)} \left(\psi_1(\mathbf{k}) - \alpha \frac{Q(\mathbf{k})}{[2\omega(k)]^{\frac{1}{2}}} \right) \text{”} + \beta \delta(\mathbf{k}) \right) \middle| \alpha, \beta \in \mathbb{C} \right\},$$

where we have a similar meaning for the quotes as above. Finally

$$\mathcal{S}_\infty \psi = \{(\alpha, \psi_1(k)) \mid \alpha \in \mathbb{C}\}.$$

A g-operator version of $\mathcal{R}_\lambda$ for $\lambda \notin [m, \infty)$ is of the form

$$R_\lambda f = \left(\alpha_\lambda(f), \frac{1}{\lambda - \omega(k)} \left(f_1(\mathbf{k}) - \alpha_\lambda(f) \frac{Q(\mathbf{k})}{[2\omega(k)]^{\frac{1}{2}}} \right) \right),$$

where α_λ is a linear map $\Phi_- \to \mathbb{C}$. Hermiticity requires now that

$$\alpha_\lambda(f) = \alpha(\lambda) \left(f_0 - \int \frac{f_1(\mathbf{k}) Q(\mathbf{k}) d^n k}{[\lambda - \omega(k)][2\omega(k)]^{\frac{1}{2}}} \right),$$

where

$$\alpha(\lambda) = \overline{\alpha(\bar{\lambda})}.$$

Likewise

$$S_\infty f = (\alpha(f), f(\mathbf{k}))$$

and symmetry implies $\alpha(f) = \gamma f_0$, γ real.

Because of analyticity properties we shall not need versions of other relations.

It is convenient to introduce the following map:

$$\zeta : f \longrightarrow f_0 - \int \frac{f_1(\mathbf{k}) Q(\mathbf{k}) d^n k}{[\lambda - \omega(k)][2\omega(k)]^{\frac{1}{2}}} = (\zeta f)(\lambda),$$

where $\lambda \notin [m, \infty)$.

We assume $\alpha(\lambda)$ is analytic off the real axis in a neighborhood thereof. Let $\bar{f} = (\overline{f_0}, \overline{f_1(k)})$.

One easily finds for $\lambda \notin [m, \infty)$

$$\langle g, R_\lambda f \rangle = \alpha(\lambda)(\zeta \bar{g})(\lambda)(\zeta f)(\lambda) + \int \frac{\overline{g_1(\mathbf{k})} f_1(\mathbf{k}) d^n k}{\lambda - \omega(k)}.$$

We assume the measure condition, which means in general that $\alpha(\lambda)\zeta\bar{g}(\lambda)\zeta f(\lambda)$ must satisfy the measure condition and particular implies that $\alpha(\lambda)$ must satisfy the measure condition, for if we take $f = (1, 0)$, then $\langle f, R_\lambda f \rangle = \alpha(\lambda)$.

If we assume positivity on (m, ∞), then we can show that $\alpha(\lambda)$ cannot have a pole there; that is, $\alpha(\lambda)$ is not of the form $\beta/(\lambda - \lambda_0) + \gamma(\lambda)$ when $\mu(\gamma, d\lambda)$ is absolutely continuous with respect to Lebesgue measure in the vicinity of λ_0. This follows from the fact that $\alpha(\lambda)(\zeta \bar{f})(\lambda)(\zeta f)(\lambda)$ for $f_0 = 0$ is

$$\alpha(\lambda) \left(\int \overline{f_1(\mathbf{k})} Q(\mathbf{k}) \Big/ [\lambda - \omega(k)][2\omega(k)]^{\frac{1}{2}} \right) \times \left(\int f_1(\mathbf{k}) Q(\mathbf{k}) \Big/ [\lambda - \omega(k)][2\omega(k)]^{\frac{1}{2}} \right)$$

and it can be shown by taking $f_1(\mathbf{k})$ to be a sufficiently good approximation to $\delta(\omega(k) - \lambda_0)$ that the expression does not satisfy positivity.

We assume that $\mu(\alpha, d\lambda)$ is therefore absolutely continuous with respect to Lebesgue measure

$$\mu(\alpha, d\lambda) = (-1/\pi) \operatorname{Im} \alpha(\lambda)\, d\lambda,$$

where we write $\alpha(\lambda) = \operatorname{Re} \alpha(\lambda) + i \operatorname{Im} \alpha(\lambda)$ along (m, ∞).

We can now also show that by positivity $\operatorname{Im} \alpha(\lambda)$ cannot vanish on any interval $A \subset (m, \infty)$. Under the assumption $\operatorname{Im} \alpha(\lambda) \equiv 0$ on A we have for an interval $B \subset A$:

$$\langle f, J(B) f \rangle = \mu(\alpha(\lambda)\zeta\bar{f}(\lambda)\zeta f(\lambda), B) + \int_{\lambda \in B} \delta(\lambda - \omega(k)) |f_1(\mathbf{k})|\, d\lambda d^n k;$$

we see that the term proportional to $|f_0|^2$ does not contribute since its coefficient is $\mu(\alpha; B) = 0$. The terms linear in f_0 and $\bar{f}_0$ are

$$\int_{\lambda \in B} \delta(\lambda - \omega(k)) \operatorname{Re} \alpha(\lambda) \frac{f_1(\mathbf{k}) \bar{f}_0 + f_0 \overline{f_1(\mathbf{k})}}{[2\omega(k)]^{\frac{1}{2}}} Q(\mathbf{k})\, d\lambda d^n k.$$

Now unless $\operatorname{Re} \alpha(\lambda) = 0$ almost everywhere on B, we cannot satisfy positivity, for no matter what $f_1(\mathbf{k})$ is we can always choose f_0 so as to make the above expression sufficiently negative to make $\langle f, J(B) f \rangle$ negative. If $\operatorname{Re} \alpha(\lambda) = 0$ almost everywhere on B, then along with $\operatorname{Im} \alpha(\lambda) = 0$ and analyticity we conclude that $\alpha(\lambda) \equiv 0$, which is not an interesting solution and

we discount it. Positivity on (m, ∞) therefore forces $\alpha(\lambda)$ to have a cut at least along (m, ∞).

It is convenient to write $J(A)$ in terms of a matrix

$$J(A) = \begin{pmatrix} J_{00}, & J_{01}(\mathbf{k}) \\ J_{10}(\mathbf{k}), & J_{11}(\mathbf{k}_1, \mathbf{k}_2) \end{pmatrix};$$

that is,

$$\langle g, J(A)f\rangle = J_{00}\overline{g_0}f_0 + \overline{g_0}\int J_{01}(\mathbf{k})f_1(\mathbf{k})\,d\mathbf{k} + f_0\int J_{10}(\mathbf{k})\overline{g_1(\mathbf{k})}\,d\mathbf{k} + \int J_{11}(\mathbf{k}_1, \mathbf{k}_2)\overline{g_1(\mathbf{k}_1)}f_1(\mathbf{k}_2)\,d\mathbf{k}_1\,d\mathbf{k}_2.$$

An easy computation reveals

$$J_{00} = -\frac{1}{\pi}\int_A d\lambda\,\operatorname{Im}\alpha(\lambda),$$

$$J_{01}(\mathbf{k}) = \left(-\frac{1}{\pi}\int_A d\lambda\,\operatorname{Im}\alpha(\lambda)\frac{\mathcal{P}}{\lambda - \omega(k)} + \int_A d\lambda\,\operatorname{Re}\alpha(\lambda)\delta(\lambda - \omega(k))\right)\frac{Q(\mathbf{k})}{[2\omega(k)]^{\frac{1}{2}}},$$

$$J_{10}(\mathbf{k}) = J_{01}(\mathbf{k}),$$

$$J_{11}(\mathbf{k}_1, \mathbf{k}_2) = -\frac{1}{\pi}\int_A d\lambda\,\operatorname{Im}\alpha(\lambda)\frac{\mathcal{P}}{\lambda - \omega(k_1)}\frac{\mathcal{P}}{\lambda - \omega(k_2)} + \left[\int_A d\lambda\,\operatorname{Re}\alpha(\lambda)\left(\frac{\mathcal{P}}{\lambda - \omega(k_1)}\delta(\lambda - \omega(k_2)) + \delta(\lambda - \omega(k_1))\frac{\mathcal{P}}{\lambda - \omega(k_2)}\right) + \pi\int_A d\lambda\,\operatorname{Im}\alpha(\lambda)\delta(\lambda - \omega(k_1))\delta(\lambda - \omega(k_2))\right] \times \frac{Q(\mathbf{k}_1)Q(\mathbf{k}_2)}{[4\omega(k_1)\omega(k_1)]^{\frac{1}{2}}} + \int_A d\lambda\delta(\lambda - \omega(k_1))\delta(\mathbf{k}_1 - \mathbf{k}_2).$$

To deduce the consequences of imposing projectivity on J is very difficult, so here we will impose decomposability in (m, ∞). For $\lambda \in (m, \infty)$ the λ multiplet of T is degenerate over S^{n-1} and we introduce the variable $\hat{s} \in S^{n-1}$ to express this degeneracy. Let us therefore consider the family of generalized eigenvectors

$$\psi_{\lambda\hat{s}} = \left(\eta_{\lambda\hat{s}}, \eta_{\lambda\hat{s}}\frac{\mathcal{P}}{\lambda - \omega(k)}\frac{Q(\mathbf{k})}{[2\omega(k)]^{\frac{1}{2}}} + \beta_{\lambda\hat{s}}(\hat{k})\delta(\lambda - \omega(k))\right),$$

where we have made a choice of the ambiguous product of distribution in $\mathcal{R}_\lambda 0$ by choosing the principal value integral.

Let $K_{\lambda\hat{s}} = |\psi_{\lambda\hat{s}}\rangle\langle\psi_{\lambda\hat{s}}|$ and then, writing K in terms of a matrix of the same form as J, we get

$$K_{00} = |\eta_{\lambda\hat{s}}|^2,$$

$$K_{01}(\mathbf{k}) = |\eta_{\lambda\hat{s}}|^2\frac{\mathcal{P}}{\lambda - \omega(k)}\frac{Q(\mathbf{k})}{[2\omega(k)]^{\frac{1}{2}}} + \overline{\eta_{2\hat{s}}}\beta_{2\hat{s}}(\hat{k})\delta(\lambda - \omega(k))$$

$$K_{10} = K_{01},$$

$$K_{11}(\mathbf{k}_1, \mathbf{k}_2) = |\eta_{\lambda\hat{s}}|^2\frac{\mathcal{P}}{\lambda - \omega(k_1)}\frac{\mathcal{P}}{\lambda - \omega(k_2)}\frac{Q(\mathbf{k}_1)Q(\mathbf{k}_2)}{[4\omega(k_1)\omega(k_2)]^{\frac{1}{2}}} + \overline{\eta_{\lambda\hat{s}}}\frac{\mathcal{P}}{\lambda - \omega(k_1)}\frac{Q(\mathbf{k}_1)}{[2\omega(k_1)]^{\frac{1}{2}}}\beta_{\lambda\hat{s}}(\hat{k}_2)\delta(\lambda - \omega(k_2)) + \overline{\beta_{\lambda\hat{s}}(\hat{k}_1)}\delta(\lambda - \omega(k_1))\eta_{\lambda\hat{s}}\frac{\mathcal{P}}{\lambda - \omega(k_1)}\frac{Q(\mathbf{k}_2)}{[2\omega(k_2)]^{\frac{1}{2}}} + \overline{\beta_{\lambda\hat{s}}(\hat{k}_1)}\beta_{\lambda\hat{s}}(\hat{k}_2)\delta(\lambda - \omega(k_1))\delta(\lambda - \omega(k_2)).$$

Let $\Omega(d\hat{s})$ be the unit measure on S^{n-1} and let Ω be the volume of the unit sphere in $\mathbb{R}^n$ with respect to Lebesgue measure. We take $d\lambda\Omega(d\hat{s})$ to be the measure with respect to which the decomposition is to be performed.

We define a few auxiliary functions

$$\kappa(\lambda) = (\lambda^2 - m^2)^{\frac{1}{2}}, \quad \kappa \text{ is the inverse function to } \omega,$$

$$\|Q\|_\lambda^2 = \int |Q(\kappa(\lambda)\hat{k})|^2\,\Omega(d\hat{k}),$$

$$\hat{Q}_\lambda(\hat{k}) = \frac{Q(\kappa(\lambda)\hat{k})}{\|Q\|_\lambda}.$$

By a straightforward computation the requirements of decomposability become

(1) $$\int |\eta_{\lambda\hat{s}}|^2\,\Omega(d\hat{s}) = -\frac{1}{\pi}\operatorname{Im}\alpha(\lambda),$$

(2) $$\int \overline{\eta_{\lambda\hat{s}}}\beta_{\lambda\hat{s}}(\hat{k})\Omega(d\hat{s}) = \operatorname{Re}\alpha(\lambda)\frac{\|Q\|_\lambda\,\hat{Q}_\lambda(\hat{k})}{(2\lambda)^{\frac{1}{2}}},$$

(3) $$\int \overline{\beta_{\lambda\hat{s}}(\hat{k}_1)}\beta_{\lambda\hat{s}}(\hat{k}_2)\Omega(d\hat{s}) = \pi\operatorname{Im}\alpha(\lambda)\,\|Q\|_\lambda^2\frac{\hat{Q}_\lambda(\hat{k}_1)\hat{Q}_\lambda(\hat{k}_2)}{2\lambda} + \frac{1}{\Omega\kappa(\lambda)^{n-1}\lambda}\delta(\hat{k}_1; \hat{k}_2),$$

where $\delta(\hat{k}_1; \hat{k}_2)$ is the Dirac δ function on S^{n-1}:

$$\int \delta(\hat{k}_1; \hat{k}_2)f(\hat{k}_2)\Omega(d\hat{k}_2) = f(\hat{k}_1) \quad \text{for} \quad f \in C^\infty(S^{n-1}).$$

Let us introduce the following transformation for $f \in C^\infty(S^{n-1})$:

$$(B_\lambda f)(\hat{s}) = \beta_{\lambda\hat{s}}(f) = \int \beta_{\lambda\hat{s}}(\hat{k}) f(\hat{k}) \Omega(d\hat{k}).$$

We assume now, to be justified *a posteriori*, that B_λ in fact defines a bounded operator on $L^2(S^{n-1}, \Omega(d\hat{s}))$. Likewise, let I be the identity operator on $L^2(S^{n-1})$ and let $P_{\hat{Q}}$ be the orthogonal projection onto $\hat{Q}_\lambda(\hat{s}) \in L^2(S^{n-1})$.

The requirements of decomposability now become:

(1) If we define $\eta_\lambda : \hat{s} \rightsquigarrow \eta_{\lambda\hat{s}}$, then $\eta_\lambda \in L^2(S^{n-1})$ and

$$\|\eta_\lambda\|^2 = -\frac{1}{\pi} \operatorname{Im} \alpha(\lambda),$$

(2) $$B_\lambda^* \eta_\lambda = \operatorname{Re} \alpha(\lambda) \frac{\|Q\|_\lambda \hat{Q}_\lambda}{(2\lambda)^{\frac{1}{2}}},$$

(3) $$B_\lambda^* B_\lambda = |B_\lambda|^2 = \pi \operatorname{Im} \alpha(\lambda) \frac{\|Q\|_\lambda^2}{2\lambda} P_{\hat{Q}} + \frac{1}{\Omega\kappa(\lambda)^{n-1}\lambda} I.$$

For consistency since $|B_\lambda|^2 \geq 0$, we of course must have

(4) $$\pi \operatorname{Im} \alpha(\lambda) \frac{\|Q\|_\lambda^2}{2\lambda} \geq -\frac{1}{\Omega\kappa(\lambda)^{n-1}\lambda}.$$

For convenience we temporarily drop the subscript λ in certain expressions.

By the polar form of a bounded operator we have $B = W|B|$, where W is an isometry, so $B^* = |B| W^*$ and (2) becomes

$$|B| W^* \eta_\lambda = \operatorname{Re} \alpha(\lambda) \frac{\|Q\|_\lambda \hat{Q}_\lambda}{(2\lambda)^{\frac{1}{2}}}.$$

We further assume, under *a posteriori* justification, that W^* is an isometry and $|B|$ is invertible. We have

$$W^* \eta_\lambda = \operatorname{Re} \alpha(\lambda) \frac{\|Q\|_\lambda}{(2\lambda)^{\frac{1}{2}}} |B|^{-1} \hat{Q}_\lambda.$$

Since W^* is an isometry, the norm of the right-hand side must equal $\|\eta_\lambda\|$ and so we must have

$$\|\eta_\lambda\|^2 = [\operatorname{Re} \alpha(\lambda)]^2 \frac{\|Q\|_\lambda^2}{2\lambda} \||B|^{-1} \hat{Q}_\lambda\|^2.$$

Now

$$\| |B|^{-1} \hat{Q}\|^2 = (|B|^{-1} \hat{Q}, |B|^{-1} \hat{Q}) = (\hat{Q}, |B|^{-2} \hat{Q})$$

and from (3) $\hat{Q}$ is an eigenvector of $|B|^2$ with eigenvalue

$$\pi \operatorname{Im} \alpha(\lambda) \frac{\|Q\|_\lambda^2}{2\lambda} + \frac{1}{\Omega\kappa(\lambda)^{n-1}\lambda}$$

and so, since $\|\hat{Q}\| = 1$, $(\hat{Q}, |B|^{-2} \hat{Q})$ is the reciprocal of this number. Furthermore, $\|\eta_\lambda\|^2$ is given by (1) and so we have

$$-\frac{1}{\pi} \operatorname{Im} \alpha(\lambda) = [\operatorname{Re} \alpha(\lambda)]^2 \frac{\|Q\|_\lambda^2}{2\lambda} \times \left(\pi \operatorname{Im} \alpha(\lambda) \frac{\|Q\|_\lambda^2}{2\lambda} + \frac{1}{\Omega\kappa(\lambda)^{n-1}\lambda} \right)^{-1},$$

which simplifies to

$$\operatorname{Im} \alpha(\lambda) = -\tfrac{1}{2}\pi \Omega\kappa(\lambda)^{n-1} \|Q\|_\lambda^2 |\alpha(\lambda)|^2$$

and shows that along the cut (m, ∞) the imaginary part of $\alpha(\lambda)$ is proportional to the square of its absolute value. This is nothing more than the familiar unitarity relation.

Introducing $\tau(\lambda) = 1/\alpha(\lambda)$, we find the unitarity relation is equivalent to

$$\operatorname{Im} \tau(\lambda) = \tfrac{1}{2}\pi \Omega\kappa(\lambda)^{n-1} \|Q\|_\lambda^2.$$

If this relation is satisfied, then a quick check shows that the consistency condition (4) holds with a *strict* inequality for almost all λ and this implies that it was consistent to choose W^* to be an isometry and to have $|B|$ invertible. If we now take $B_\lambda = (|B_\lambda|^2)^{\frac{1}{2}}$, where $|B_\lambda|^2$ is given by the right-hand side of (3) and if we take

$$\eta_\lambda = \operatorname{Re} \alpha(\lambda) \|Q\|_\lambda B_\lambda^{-1} \frac{\hat{Q}}{(2\lambda)^{\frac{1}{2}}},$$

then decomposability is in fact satisfied. The unitarity relation therefore is equivalent to decomposability in (m, ∞).

Before proceeding to further examine $\alpha(\lambda)$ let us compute what may be interpretable as the scattering matrix of this model.

We take the following expression for the scattering states $\psi_{\lambda\hat{s}}^{(\pm)}$:

$$\psi_{\lambda\hat{s}}^{(\pm)} = \left(\eta_{\lambda\hat{s}}^{(\pm)}, \eta_{\lambda\hat{s}}^{(\pm)} \frac{1}{\lambda - \omega(k) \pm i\epsilon} \frac{Q(\mathbf{k})}{[2\omega(k)]^{\frac{1}{2}}} + \beta_\lambda^{(\pm)} \delta(\hat{s}; \hat{k}) \delta(\lambda - \omega(k)) \right).$$

These are precisely the scattering states in the usual solutions except now $Q(\mathbf{k})$ is not necessarily restricted in any way; we choose the states in this way since we do not know of any other possible definition.

We thus have

$$\beta_{\lambda\hat{s}}^{(\pm)}(\hat{k}) = \mp i\pi \eta_{\lambda\hat{s}}^{(\pm)} \frac{\|Q\|_\lambda \hat{Q}_\lambda(\hat{k})}{(2\lambda)^{\frac{1}{2}}} + \beta_\lambda^{(\pm)} \delta(\hat{s}; \hat{k}).$$

We now make another formal assumption, namely we assume the resolvent is decomposable with respect to the scattering states. Now while this is a reasonable

assumption when Q is such that the conventional physical interpretation is allowed,it need not be reasonable in the general case; however, we make this assumption not knowing of any other formal way to determine $\eta_\lambda^{(\pm)}$.

Assuming (1) to hold,we find by (2) that we must have

$$\int \overline{\eta_{\lambda\hat{s}}^{(\pm)}}\beta_{\lambda\hat{s}}^{(\pm)}(\hat{k})$$

$$= \mp\pi \int \overline{|\eta_{\lambda\hat{s}}^{(\pm)}|^2}\,\Omega(d\hat{s})\,\frac{\|Q\|_\lambda\,\hat{Q}_\lambda(\hat{k})}{(2\lambda)^{\frac{1}{2}}} + \beta_\lambda^{(\pm)}\overline{\eta_{\lambda\hat{k}}^{(\pm)}}$$

$$= \pm i\,\mathrm{Im}\,\alpha(\lambda)\,\frac{\|Q\|_\lambda\,\hat{Q}_\lambda(\hat{k})}{(2\lambda)^{\frac{1}{2}}} + \beta_\lambda^{(\pm)}\overline{\eta_{\lambda\hat{k}}^{(\pm)}}$$

$$= \mathrm{Re}\,\alpha(\lambda)\,\frac{\|Q\|_\lambda\,\hat{Q}_\lambda(\hat{k})}{(2\lambda)^{\frac{1}{2}}}.$$

So

$$[\mathrm{Re}\,\alpha(\lambda) \mp i\,\mathrm{Im}\,\alpha(\lambda)]\,\frac{\|Q\|_\lambda\,\hat{Q}_\lambda(\hat{k})}{(2\lambda)^{\frac{1}{2}}} = \overline{\beta_\lambda^{(\pm)}}\eta_{\lambda\hat{k}}^{(\pm)}.$$

Since $\alpha(\lambda)$ has a cut along (m, ∞) and since it satisfies $\alpha(\lambda) = \overline{\alpha(\bar{\lambda})}$,we see that $\mathrm{Re}\,\alpha(\lambda) \mp i\,\mathrm{Im}\,\alpha(\lambda) = \alpha(\lambda \mp i0)$, and we finally have

$$\eta_{\lambda\hat{s}}^{(\pm)} = \frac{1}{\overline{\beta_\lambda^{(\pm)}}}\,\alpha(\lambda \mp i0)\,\frac{\|Q\|_\lambda\,\hat{Q}_\lambda(\hat{s})}{(2\lambda)^{\frac{1}{2}}}.$$

The requirement (1): $\|\eta_\lambda^{(\pm)}\|^2 = (-1/\pi)\,\mathrm{Im}\,\alpha(\lambda)$ when combined with the unitarity relation yields after a quick computation

$$|\beta_\lambda^{(\pm)}|^2 = \frac{1}{\Omega\kappa(\lambda)^{n-1}\lambda}.$$

A computation of $|B_\lambda|^2$ now shows that (3) is in fact satisfied and so we take finally

$$\beta_\lambda^{(\pm)} = [\Omega\kappa(\lambda)^{n-1}\lambda]^{-\frac{1}{2}},$$

$$\eta_{\lambda\hat{s}}^{(\pm)} = \{\tfrac{1}{2}[\Omega\kappa(\lambda)^{n-1}]\}^{\frac{1}{2}}\alpha(\lambda \mp i0)\,\|Q\|_\lambda\,\hat{Q}_\lambda(\hat{s}).$$

The scattering matrix $\sigma_\lambda(\hat{s}, \hat{t})$ we now take to be the kernel which effects the transformation

$$\psi_{\lambda\hat{s}}^{(-)} = \int \sigma_\lambda(\hat{s}, \hat{t})\psi_{\lambda\hat{t}}^{(+)}\Omega(d\hat{t}).$$

This requirement consists of two parts coming from the different components of $\psi_{\lambda\hat{s}}^{(\pm)}$:

$$\eta_{\lambda\hat{s}}^{(-)} = \int \sigma_\lambda(\hat{s}, \hat{t})\eta_{\lambda\hat{t}}^{(+)}\Omega(d\hat{t}),$$

$$i\pi\eta_{\lambda\hat{s}}^{(-)}\,\frac{\|Q\|_\lambda\,\hat{Q}_\lambda(\hat{k})}{(2\lambda)^{\frac{1}{2}}} + [\Omega\kappa(\lambda)^{n-1}\lambda]^{-\frac{1}{2}}\delta(\hat{s}; \hat{k})$$

$$= \int \sigma_\lambda(\hat{s}, \hat{t})(-i\pi\eta_{\lambda\hat{t}}^{(+)})\Omega(d\hat{t})\,\frac{\|Q\|_\lambda\,\hat{Q}(\hat{k})}{(2\lambda)^{\frac{1}{2}}} + [\Omega\kappa(\lambda)^{n-1}\lambda]^{-\frac{1}{2}}\sigma_\lambda(\hat{s}, \hat{k}).$$

Assuming the first equation to hold, the second one can be solved for σ_λ to give

$$\sigma_\lambda(\hat{s}, \hat{t}) = \delta(\hat{s}; \hat{t}) + 2\pi i[\Omega\kappa(\lambda)^{n-1}\lambda]^{\frac{1}{2}}\eta_{\lambda\hat{s}}^{(-)}\,\frac{\|Q\|_\lambda\,\hat{Q}_\lambda(\hat{t})}{(2\lambda)^{\frac{1}{2}}}$$

or, substituting the expression for $\eta_\lambda^{(-)}$,

$$\sigma_\lambda(\hat{s}, \hat{t}) = \delta(\hat{s}; \hat{t}) + i\pi\Omega\kappa(\lambda)^{n-1} \times \alpha(\lambda + i0)\,\|Q\|_\lambda^2\,\hat{Q}(\hat{s})\hat{Q}(\hat{t}).$$

Substituting this final expression into the first requirement,one sees that it is in fact satisfied; use must be made though of the relations

$$\alpha(\lambda + i0)\alpha(\lambda - i0) = |\alpha(\lambda)|^2 = -\mathrm{Im}\,\alpha(\lambda)\,\frac{2}{\pi\Omega\kappa(\lambda)^{n-1}\,\|Q\|_\lambda^2},$$

$$\eta_\lambda^{(+)} - \eta_\lambda^{(-)} = i\,\mathrm{Im}\,\alpha(\lambda)\left(\frac{2}{\pi\Omega\kappa(\lambda)^{n-1}}\right)^{\frac{1}{2}}\|Q\|_\lambda\,\hat{Q}_\lambda.$$

Let us return to the function $\alpha(\lambda)$. Consider the region $(-\infty, m]$. We shall assume that $\alpha(\lambda)$ has only simple poles there; as for the discrete Lee model such an assumption invokes the heuristic principle that the analyticity domain of R_λ be as large as possible while still remaining interesting. This assumption has *a posteriori* justification by the class of solutions it singles out. Nondegeneracy requires that $\alpha(\lambda)$ not vanish at any regular point and we conclude therefore that $\tau(\lambda)$ should be regular off (m, ∞) along which it has a cut. Since we know the imaginary part of $\tau(\lambda)$ along (m, ∞),the most general form for this function is therefore

$$\tau(\lambda) = E(\lambda) - g(\lambda)\tfrac{1}{2}\Omega\int_m^\infty \frac{\kappa(\nu)^{n-1}\,\|Q\|_\nu^2}{(\lambda - \nu)g(\nu)}\,d\nu,$$

where $g(\nu)$ is an entire function picked to make the integral converge and $E(\lambda)$ is an entire function.

We see that $g(\lambda)$ can in fact be picked to be a power; for $\|Q\|_\lambda^2 \leq q_0^2$ and $\kappa(\lambda)^{n-1} \sim \lambda^{n-1}$ as $\lambda \to \infty$ so that one can take $g(\lambda) = \lambda^n$. This is the best possible choice unless Q is picked in such a way as to improve convergence. Let $n(Q)$ be the best possible choice for a given Q; thus $n(q_0) = n$. By the formal argument used for the discrete Lee model we single out the solutions

$$\tau(\lambda) = \lambda + \sigma + P_{n(Q)-1}(\lambda) - \tfrac{1}{2}\Omega\lambda^{n(Q)}\int_m^\infty \frac{\kappa(\nu)^{n-1}\,\|Q\|_\nu^2}{(\lambda - \nu)\nu^{n(Q)}}\,d\nu,$$

where P is a polynomial of degree $n(Q) - 1$.

The case $n(Q) = 0$ is the conventional solution and we do not discuss it any further except to note that our results coincide exactly with the conventional

ones. For no model is $n(q_0) = 0$ since this would imply zero space dimension.

The case $n(Q) = 1$ can be obtained as in the discrete Lee model by an infinite counterterm in T which in this case introduces an infinite "bare V particle mass." This solution satisfies positivity (absence of "ghosts"), normal ray limit, and analyticity in the whole plane save for the real physical singularities. We have $n(q_0) = 1$ in the 1-dimensional model and so the 1-dimensional Lee model has a cutoff free solution obtained by an infinite adjustment of the bare V particle mass.

The cases $n(Q) > 1$ have as in the discrete case certain distinctive features. Thus the normal ray limit cannot be realized and R_λ can have complex poles. As before the polynomial $P_{n(Q)-1}$ can be picked so that the pole residues of $\alpha(\lambda)$ on the real axis are positive but we do not know whether simultaneously, or at all, we can avoid $\alpha(\lambda)$ having complex poles.

If we apply orthogonal methods to the $N\theta$ sector of the Lee model, then as for the discrete Lee model we quickly obtain the same result as above. We take the decomposition $\Phi = \mathbb{C} \oplus \Phi_2$, where Φ_2 is $\mathfrak{D} \subset \mathfrak{D}'$. In computing T_λ we are faced with the analytic integral

$$\int \frac{|Q(\mathbf{k})|^2 \, d^n k}{2\omega(k)[\lambda - \omega(k)]} = \tfrac{1}{2}\Omega \int_m^\infty \frac{\kappa(\nu)^{n-1} \, \|Q\|_\nu^2}{(\lambda - \nu)} \, d\nu,$$

which, if we perform the minimum number of subtraction necessary, is given by

$$P_{n(Q)-1}(\lambda) + \tfrac{1}{2}\Omega\lambda^{n(Q)} \int_m^\infty \frac{\kappa(\nu)^{n-1} \, \|Q\|_\nu^2}{(\lambda - \nu)\nu^{n(Q)}} \, d\nu,$$

which then leads to the same result as before.

The outstanding intrinsic structural property of the N sector of the Lee model we believe to be the unitarity relation

$$\operatorname{Im} \tau(\lambda) = \tfrac{1}{2}\pi\Omega\kappa(\lambda)^{n-1} \, \|Q\|_\lambda^2.$$

What other intrinsic structural properties have become manifest is, as in the case of the discrete Lee model, not very clear.

D. Remarks

We now discuss in greater detail our views concerning the various conditions. As was pointed out at the end of the previous section, there are many generalized operators that are candidates for a particular physical problem; different candidates lead to different resolvent relations and thus whether the above conditions can or cannot be met depends on the candidate. The intrinsic structural properties of the physical generalized operator should of course manifest themselves most readily in a natural rigging such as the complete rigging or the diagonal rigging. However, even if we do not choose these riggings, the resolvent relations often still contain all the possible relevant information and it is thus natural to ask to what extent do the conditions reflect *candidate independent* properties.

Whether there exists a nontrivial Hermitian resolvent depends on the particular candidate. For example, consider the operator $T: f(x) \rightsquigarrow (1/i)f'(x)$ on either of two rigged Hilbert spaces (1) $\mathfrak{D}(0, \infty) \subset \mathfrak{D}'(0, \infty)$ and (2) $\mathfrak{D}(0, \infty) \subset L^2(0, \infty)$. One easily finds in case (1) that there is no problem in finding a Hermitian resolvent.

In case (2), however, we find that the only g-operator version R_λ of $\mathcal{R}_\lambda$ is

$$\begin{aligned} R_\lambda f &= \varnothing \quad \text{if} \quad f \neq 0, \\ &= 0 \quad \text{if} \quad f = 0. \end{aligned}$$

There is thus no nontrivial Hermitian resolvent.

The difficulty here is of course that T, which is a bona fide operator, is not essentially self-adjoint on $\mathfrak{D}(0, \infty)$. This fact, however, does not show itself in the hermiticity condition in case (1) even though hermiticity does imply a relationship between the behavior of $\mathcal{R}_\lambda$ in the upper and lower complex half-planes reminiscent of the theory of defects.

Because of this candidate dependence of hermiticity it is best to use this condition in a natural rigging.

In general one cannot expect analyticity everywhere off the real axis and this raises the problem of the physical meaning of any possible singularities off the real axis since these singularities will induce singularities in "unusual" places in various physically relevant objects such as the scattering matrix. This became apparent in the $N\theta$ sector of the Lee model where there arose the possibility of complex poles in the physical sheet of the scattering matrix. The understanding of these singularities we believe is best achieved by studying their structure in actual physical theories such as quantum electrodynamics or some experimentally accessible solid state phenomena. As such we have to wait for the development of more powerful methods to construct the resolvents of these theories.

The use of orthogonal methods determines some of the analyticity properties, thus eliminating some of the choices. This fact should provide an approach into more precise analyticity statements concerning the resolvents.

The positivity requirement reflects the probabilistic interpretation of quantum mechanics. If $A \rightsquigarrow J(A)$ is not a positive measure, then one could of course restrict oneself to a subset of $\mathbb{R}$ on which it is a positive

measure. In this case, one is disregarding certain singularities which are considered unphysical. If we pick an R_λ such that $A \rightsquigarrow J(A)$ is not a positive measure, we say that the resolvent has *ghosts*. Positivity may often be impossible to achieve and similar statements apply to it as to the analyticity requirement.

Some sort of requirement is needed to control the behavior near $\lambda = \infty$ and the ray limit is an attempt at such a control. For a bona fide self-adjoint operator the fact that the ray limit is I is an expression of the completeness of the set of eigenvectors; that is, the spectral projections constitute a resolution of the identity. The concept of completeness for generalized operators is elusive both mathematically and physically, for it is not clear whether the set of "bare" generalized vectors Φ_+ are all necessary or whether they are sufficient for the mathematical and physical structure of T. We say that the normal ray limits cannot be achieved in many cases. A detailed description of behavior near $\lambda = \infty$ should reflect the various subtractions made in constructing R_λ.

Both the nondegeneracy and the well-dressing conditions are somewhat ad hoc but are useful in limiting the choices. These two conditions also seem to touch upon a notion of completeness and may be useful in isolating such a notion. Well dressing in terms of analytic subtraction, however, may be more intrinsic.

The measure condition is of course very important as it stands at the foundation of any renormalization program. We have formulated it only with regard to the real singularities of R_λ but in view of the possibility that analyticity everywhere off the real axis is in general unattainable it may be useful to extend it in some way to the complex singularities to arrive at a generalized operator-valued measure on the complex plane. This again brings in the question of the physical role of "unusual" singularities.

The projectivity condition was introduced in order to effect a renormalization program. It generalizes the notion of orthonormal decomposition.

The condition of decomposability is closely related to that of projectivity but is more candidate dependent: The rigging may not exhibit the generalized eigenstates necessary for decomposition. For example if T is a bona fide bounded self-adjoint operator having a purely continuous spectrum on a Hilbert space $\mathcal{H}$, we can consider it as a generalized operator in the rigged Hilbert space $(\mathcal{H}, \mathcal{H}, \mathrm{id})$. In this case for λ in the resolvent set, $\mathcal{R}_\lambda$ coincides with the bona fide resolvent R_λ. By the spectral theorem R_λ is projective and J is the bona fide spectral measure. The resolvent, however, is not decomposable since $\Phi_+ = \mathcal{H}$ does not contain any generalized eigenstates other than 0: $\sum(T) = \{0\}$. If, however, T were supplied with an appropriate rigging, the resolvent would be decomposable. Decomposability is thus a useful condition only if one is working in a rigging which exhibits all the generalized eigenstates with real eigenvalues.

We find the state of the theory as presented in this section very unsatisfactory for several reasons. The process of arriving at a resolvent R_λ consists of a search among many possibilities for those choices that satisfy certain requirements. These requirements were based on analogy and do not necessarily reflect in a natural way the intrinsic structure of T. What we need is an explicit construction of explicit mathematical objects rather than a search. Orthogonal methods provide a step in this direction but are still not very well formulated. Thus the present theory can only be regarded as provisional, giving only certain tools by which the intrinsic structural properties of T can be found and later concretized into definite mathematical objects which can be exhibited by constructive procedures.

We believe that the following point of view should be adopted to attack this problem: The passage from T to J should be broken into two distinct stages; first there is the construction of the objects intrinsic to and uniquely determined by T, and second there is the choice of certain arbitrary elements which must be present in any physical renormalization theory. Thus in the discrete Lee model it appears that the residues $-1/\rho_\kappa^2$ of $\tau(\lambda)$ at λ_κ is an intrinsic property while the choice of $E(\lambda)$ must be influenced by extraneous reasoning; likewise in the $N\theta$ sector of the continuous Lee model the unitarity relation

$$\operatorname{Im} \tau(\lambda) = 1/2\pi\Omega\kappa(\lambda)^{n-1} \|Q\|_\lambda^2$$

also appears to be an intrinsic property. A most important question to ask now is the following: How can one derive the unitarity relation without going through the search procedure for the appropriate restrictions of the multivalued linear maps $\mathcal{R}_\lambda$? The knowledge of $\sum(T)$ and of the unitarity relation should be considered as the mathematical answer to the formal question: What are all the appropriately normalized solutions to the eigenvalue problem? For bona fide operators the answer to this question is embodied in a cogent form in a definite mathematical object, namely the resolvent. We have used this fact to try to construct a theory of generalized operators but because of the renormalization programs we now see that the resolvent is not really the appropriate object. The second important question is: What is the corresponding appropriate object for generalized operators?

In view of the above discussion and in view of the remarks to be made later in Sec. V we believe it is now profitable to modify the methodology somewhat and instead of trying to further refine the conditions on the resolvent we should try to approach the intrinsic properties directly now that we have some idea of what they are. This will be our main thrust in our future investigations into the structure of generalized operators.

The search for intrinsic objects is important in another respect, for they should provide an approach to generalized operators that are not symmetric; at present we do not have even a rudimentary theory for these.

IV. SEVERAL OPERATORS

In this section we would like to examine the question as to what extent can several generalized operators be treated simultaneously with respect to their structure; in particular, when can they be simultaneously renormalized to bona fide operators on a Hilbert space.

There is of course the problems of candidacy: Is there a rigging which is natural to all operators simultaneously, and likewise is there a rigging which is adequate to all operators simultaneously in the sense that no information about their structure is lost? Since we have not pursued these questions in any detail even for the case of a single operator, we shall not pursue them in any great detail here even though as we shall see they appear here even in stronger force.

There is here also a further complication: Given a set of generalized operators one may not want to renormalize all of them. For example, given a Hamiltonian T expressed in terms of a free field $\phi(\mathbf{x}, 0)$ at $t = 0$ one would want to renormalize T but the field $\phi(\mathbf{x}, 0)$ should again be re-expressed as a generalized operator. We shall call this the problem of *partial renormalization* of a set of generalized operators.

The problems dealt with in this section have great physical significance since the physical content of a physical system is most often carried by a set of operators satisfying various relations and conditions.

A. The Dressing Method

In this subsection we assume we are given a renormalizable generalized operator T on a rigged Hilbert space Φ. We denote by Ψ the renormalized rigged Hilbert space. We also assume that we are given in addition a set $\mathcal{S}$ of generalized operators on Φ. We now want to transfer the set $\mathcal{S}$ to Ψ to get a new set $\mathcal{S}_R$ of generalized operators on Ψ. In particular, in going from Φ to Ψ we would like to know if we can preserve various properties such as being bona fide, being symmetric, being unitary, being essentially self-adjoint, and so forth. Furthermore, if $\mathcal{S}$ is a unitary group or an algebra or in general comprises some algebraic system, then one would like to reconstruct again a similar algebraic system in Ψ.

The most elementary approach to this problem, which is the approach used in this section, is to use elementary function theory as it was used for the concept of well dressing. Formally if Δ is the dressing transformation, then $\mathcal{S}_R$ should be identified with $\Delta\mathcal{S}\Delta^{-1}$. More precisely, we consider Δ as an element of $M(\Phi_+, \Psi_+)$ and $\mathcal{S}$ as a subset of $M(\Phi_+)$; we then construct the set $\mathcal{S}_R \subset M(\Psi_+)$ by

$$\mathcal{S}_{\mathcal{R}} = \{\mathcal{S}_{\mathcal{R}} = \overline{\overline{\Delta S \Delta^{-1}} V \overline{\Delta S \Delta^{-1}}} \mid S \in \mathcal{S}\}.$$

This set constitutes the loosest possible interpretation of the formal set $\Delta\mathcal{S}\Delta^{-1}$. As in the case of well dressing a more stringent interpretation may also be effective and at times be even more to the point but even if more stringent interpretations are used, we shall still call the resulting set $\mathcal{S}_{\mathcal{R}}$ and consider these procedures as also part of the dressing method.

For each $S_{\mathcal{R}} \in \mathcal{S}_{\mathcal{R}}$ we now pick a g-operator version $S_R \subset S_{\mathcal{R}}$ and so get a set $\mathcal{S}_R$ of generalized operators in Ψ. This set is obviously not unique and to further restrict our choices we must impose additional constraints. Such constraints would require, for example, that $S \rightsquigarrow S_R$ preserve unitarity, essential self-adjointness, or any algebraic structure that the set $\mathcal{S}$ may have.

To see the features and problems of this method, we study a few examples.

Example 22: We take T to be the multiplication by the δ function on $\mathcal{D}(\mathbb{R}) \subset \mathcal{D}'(\mathbb{R})$. Let $\mathcal{S}$ be the unitary group $U(t)$ given by

$$(U(t)f)(x) = f(x + t) = f_t(x).$$

A computation of $U(t)$ either by the loosest method or by $U(t) = \overline{\Delta U(t) \Delta^{-1}}$ yields

$$U(t)h \oplus c = \{h_t \oplus \alpha \mid \alpha \in \mathbb{C}\}.$$

A g-operator version of $U(t)$ is therefore of the form

$$U(t)_R h \oplus c = h_t \oplus \alpha(t, h, c),$$

where $\alpha(t, h, c)$ is of the form $\alpha_1(t)h + \alpha_2(t)c$, the α_i being linear functionals on L^2 and $\mathbb{C}$, respectively. Imposing isometry, we must have $|\alpha_1(t)h + \alpha_2(t)c|^2 = |c|^2$ and since h is arbitrary, we must take $\alpha_1(t) = 0$ and $|\alpha_2(t)| = 1$. Imposing that $U(t)_R$ be a continuous unitary representation of $U(t)$, we get $\alpha_2(t) = e^{i\omega t}$,

where ω is real. Thus

$$U(t)_R h \oplus c = h_t \oplus e^{i\omega t} c$$

constitutes a reinterpretation of U in Ψ in terms of a unitary group.

Example 23: We still consider the same T as in Example 8, but now let $\mathcal{S}$ consist of the single operator S which is multiplication by $\delta_y = \delta(x-y)$. We find

$$\mathcal{S}\tau = \{\alpha\delta(x-y) \mid \alpha \in \mathbb{C}\},$$
$$S\bar{\Delta}^{-1}h \oplus c = \mathcal{S}h = \{\alpha\delta(x-y) \mid \alpha \in \mathbb{C}\}$$

and this is a closed relation; therefore,

$$\bar{\Delta}\overline{(S\bar{\Delta}^{-1})}h \oplus c = 0;$$

likewise

$$\bar{\Delta}\mathcal{S}\tau = \bar{\Delta}\{\alpha\delta(x-y) \mid \alpha \in \mathbb{C}\} = 0.$$

In this case we have $\mathcal{S}_{\mathcal{R}} = \{0\}$ and so $S_R = 0$.

We have here an example of a difficulty in *adequacy*; the renormalized Hilbert space Ψ is not adequate to express S as a generalized operator. Roughly speaking the elements of Ψ_- are not "smooth enough." There are two possible desirable constructions one can pursue: (1) pass to a new rigging of Ψ_0 possessing smoother states than Ψ_- in which S can be reinterpreted as a generalized operator, or (2) renormalize both T and S simultaneously in still a different Hilbert space. We shall carry through both of these programs in the Sec. IVC but for now we proceed with more examples in which $\mathcal{S}$ is a set of bona-fide operators.

Example 24: Consider now the discrete Lee model. We take the case when β_k is not necessarily $-1/\rho_k^2$. Let now $U(t)$ be the one-parameter group

$$(U(t)f)_n = e^{i\omega_n t} f_n, \quad \omega_n \in \mathbb{R}.$$

A computation of $\bar{\Delta}\overline{U(t)}\bar{\Delta}^{-1}$ gives

$$U(t)_{\mathcal{R}}(\phi_\nu, \phi_k) = \{(\psi_\nu, e^{i\omega_k t}\phi_k) \mid \psi_\nu \in \mathbb{C}\}.$$

Any unitary group of the form

$$(h_\nu, h_k) \rightsquigarrow ((W(t)h)_\nu, e^{i\omega_k t}h_k)$$

can be taken for $U(t)_R$, where $W(t)$ is unitary on $L^2(N, |\mu|)$. Thus, in case $\beta_k = -1/\rho_k^2$ the dressing method is so coarse that *any* unitary group in Ψ can serve as $U(t)_R$.

Example 25: On $\mathbb{C} \oplus \mathfrak{D}(\mathbb{R}) \subset \mathbb{C} \oplus \mathfrak{D}'(\mathbb{R})$ let T be given by $Tf_0 \oplus f_1(k) = 0 \oplus \alpha\delta(k)f_1(k)$. A simple computation now shows that we have a solution $\Psi = (\mathbb{C} \oplus L^2(\mathbb{R}) \oplus \mathbb{C}, \mathbb{C} \oplus L^2(\mathbb{R}) \oplus \mathbb{C}, \mathrm{id})$,

$$\|f_0 \oplus f_1 \oplus c\|^2 = |f_0|^2 + \|f_1\|^2 + \beta|c|^2, \beta > 0,$$
$$T_R f_0 \oplus f_1 \oplus c = 0 \oplus 0 \oplus \lambda_0 c,$$
$$\Delta f_0 \oplus f_1 = f_0 \oplus f_1 \oplus f_1(0).$$

Let $\mathcal{S}$ now consist of the "creation" operators $a^+(g): f_0 \oplus f_1(k) \rightsquigarrow 0 \oplus f_0 g(k)$ and "annihilation" operators $a(g): f_0 \oplus f_1(k) \rightsquigarrow (g, f_1) \oplus 0$, where $g \in \mathfrak{D}$, g real, and $(\cdot,\cdot)$ is the L^2 inner product. A computation of $\mathcal{S}_{\mathcal{R}} = \bar{\Delta}\bar{\mathcal{S}}\bar{\Delta}^{-1}$ now yields

$$a^+(g)_{\mathcal{R}} f_0 \oplus f_1 \oplus c = \{0 \oplus f_0 g(k) \oplus \alpha \mid \alpha \in \mathbb{C}\},$$
$$a(g)_{\mathcal{R}} f_0 \oplus f_1 \oplus c = \{(g, f_1) \oplus 0 \oplus \alpha \mid \alpha \in \mathbb{C}\}.$$

We now notice the following feature: If we pick $a^+(g)_R \subset a^+(g)_{\mathcal{R}}$ and $a(g)_R \subset a(g)_{\mathcal{R}}$ and insist that the two operators be formal adjoints of each other, we would get a smaller number of answers than if we had picked just any $a^+(g)_R \subset a^+(g)_{\mathcal{R}}$ and then taken $a(g)_R$ to be the formal adjoint. In other words, if $\mathcal{S}$ can be generated from a subset $\mathcal{G}$ by certain generating procedures, then, computing first $\mathcal{G}_R$ and then obtaining $\mathcal{S}_R$ by applying the same generating procedures, we would in general get a different answer than if we had computed $\mathcal{S}_R$ directly; this is so even when the set $\mathcal{S}_R$ must satisfy algebraic relations. Still a different answer would in general be obtained if we chose a different generating set $\mathcal{G}'$. This situation is in principle expected but the dressing method does not allow much further elucidation.

We shall not pursue the dressing method further since it is very elementary and can give only a very coarse view. It is helpful nevertheless in pointing out certain features of the theory of several operators.

B. Remarks toward a General Multiplicity theory

For a set of bounded bona fide operators the relevant multiplicity theory studies the von Neumann ring generated by this set. The spectral theorem in its greatest generality expresses a von Neumann ring as a direct integral of factors over the spectrum of its center. This theory has a strong algebraic character which however is intimately related to certain topological consideration. For the case of generalized operators we have to replace the study of $\mathfrak{B}(\mathcal{H})$, the set of bounded operators on a Hilbert space $\mathcal{H}$, by the study of appropriate subsets of $M(\Phi_+)$, the set of linear relations on a rigged Hilbert space Φ. There is little hope yet for achieving a mostly algebraic theory for even in the case of a single symmetric generalized operator T the passage from T to the generalized spectral measure J required analytic considerations.

Possibly when the intrinsic structure of generalized operators is well known, one could develop an algebraic theory of several operators in terms of the intrinsic objects, thus bystepping the analytic considerations needed to pass to the generalized spectral measure. We cannot pursue this possibility not having an intrinsic theory; however, if we assume that starting with a set of symmetric generalized operators we have somehow picked the appropriate generalized spectral measures, we can ask whether any form of a multiplicity theory can be constructed from this material. The rest of this subsection contains an investigation of this possibility.

A bounded self-adjoint operator S will be called a *hemiprojection* if the spectrum of S is contained in the points 0, $\frac{1}{2}$, and 1. Therefore, if S is a hemiprojection, then $S = E(1) + \frac{1}{2}E(\frac{1}{2})$, where $E(1)$ and $E(\frac{1}{2})$ are orthogonal projections and $E(1)E(\frac{1}{2}) = 0$; moreover, any operator of this form is hemiprojection.

Let now J and K be two positive generalized operators. We say that the unordered pair $\{J, K\}$ is *projective* if both J_{J+K} and K_{J+K} are hemiprojections. This is a generalization, as we shall see, of the notion of a projective pair when one of the generalized operators is larger than the other.

Let $\{J, K\}$ be a projective pair. We now have

$$\begin{aligned} J_{J+K} &= E_J(1) + \tfrac{1}{2}E_J(\tfrac{1}{2}), \\ K_{J+K} &= E_K(1) + \tfrac{1}{2}E_K(\tfrac{1}{2}), \\ 1 &= J_{J+K} + E_{J+K} = E_J(1) + \tfrac{1}{2}E_J(\tfrac{1}{2}) \\ &\quad + E_K(1) + \tfrac{1}{2}E_K(\tfrac{1}{2}). \end{aligned}$$

Multiplying this last expression for 1 by $E_J(1)$, $E_J(\frac{1}{2})$, $E_K(1)$, and $E_K(\frac{1}{2})$, respectively, we get a set of four equations which can be immediately expanded to eight by taking adjoints; we make use of four of these:

(1) $\frac{1}{2}E_J(\frac{1}{2}) = E_J(\frac{1}{2})E_K(1) + \frac{1}{2}E_J(\frac{1}{2})E_K(\frac{1}{2})$,

(2) $\frac{1}{2}E_K(\frac{1}{2}) = E_K(\frac{1}{2})E_J(1) + \frac{1}{2}E_K(\frac{1}{2})E_J(\frac{1}{2})$,

(3) $\frac{1}{2}E_J(\frac{1}{2}) = E_K(1)E_J(\frac{1}{2}) + \frac{1}{2}E_K(\frac{1}{2})E_J(\frac{1}{2})$,

(4) $E_J(1)E_K(1) + \frac{1}{2}E_J(\frac{1}{2})E_K(1) = 0$.

Multiplying (3) on the left by $E_K(1)$, we conclude that $E_K(1)E_J(\frac{1}{2}) = 0$, which combined with (4) yields $E_K(1)E_J(1) = 0$. A combination of (1) and (2) now gives $E_J(\frac{1}{2}) = E_K(\frac{1}{2})E_J(\frac{1}{2}) = E_K(\frac{1}{2})$ so that $E_K(\frac{1}{2}) = E_J(\frac{1}{2}) = E(\frac{1}{2})$, where the last equality is a definition. We now conclude

$$1 = E_J(1) + E_K(1) + E(\tfrac{1}{2})$$

and by Lemma 2 of Sec. III this is an orthonormal decomposition. Let us now prove the following useful fact:

Lemma 6: If $J \geq K$, then $\{J, K\}$ is a projective pair if and only if $J \geq K$ is a projective pair in the sense of Sec. III.

Proof: We first prove the necessity. Since $J \geq K$, we must have $J_{J+K} \geq K_{J+K}$ or $E_J(1) + \frac{1}{2}E(\frac{1}{2}) \geq E_K(1) + \frac{1}{2}E(\frac{1}{2})$, which is possible only if $E_K(1) = 0$ in which case $K_{J+K} = \frac{1}{2}E(\frac{1}{2})$. Now H_J and H_{J+K} are isomorphic since J-Cauchy and J-equivalence implies by virtue of $J \geq K$, respectively, $(J + K)$-Cauchy and $(J + K)$-equivalence. Thus, $\rho_{J,J+K}$ has a bounded inverse

$$\rho^{-1}_{J,J+K}: \Delta_J f \rightsquigarrow \Delta_{J+K} f \quad \text{with} \quad \|\rho^{-1}_{J,J+K}\| \leq 2.$$

We have $\rho_{KJ} = \rho_{K,J+K}\rho^{-1}_{J,J+K}$. Consider now the map $\sigma: H_J \to H_J$ given by $\rho_{J,J+K}E(\frac{1}{2})\rho^{-1}_{J,J+K}$; then clearly $\sigma^2 = \sigma$ and furthermore

$$\begin{aligned} &(\Delta_J g, \rho_{J,J+K}E(\tfrac{1}{2})\rho^{-1}_{J,J+K}\Delta_J f)_J \\ &\quad = (\rho_{J,J+K}\Delta_{J+K}g, \rho_{J,J+K}E(\tfrac{1}{2})\Delta_{J+K}f)_J \\ &\quad = (\Delta_{J+K}g, \tfrac{1}{2}E(\tfrac{1}{2})\Delta_{J+K}f)_{J+K} \\ &\quad = (\tfrac{1}{2}E(\tfrac{1}{2})\Delta_{J+K}g, \Delta_{J+K}f)_{J+K} \\ &\quad = (J_{J+K}E(\tfrac{1}{2})\Delta_{J+K}g, \Delta_{J+K}f)_{J+K} \\ &\quad = (\rho_{J,J+K}E(\tfrac{1}{2})\Delta_{J+K}g, \rho_{J,J+K}\Delta_{J+K}f)_J \\ &\quad = (\rho_{J,J+K}E(\tfrac{1}{2})\rho^{-1}_{J,J+K}\Delta_J g, \Delta_J f)_J \end{aligned}$$

shows that $\sigma^* = \sigma$ and thus σ is an orthogonal projection. We now have

$$\begin{aligned} \|\sigma\Delta_J f\|^2_J &= \|\rho_{J,J+K}E(\tfrac{1}{2})\rho^{-1}_{J,J+K}\Delta_J f\|^2_J \\ &= \|\rho_{J,J+K}E(\tfrac{1}{2})\Delta_{J+K}f\|^2_J \\ &= \|[E_J(1) + \tfrac{1}{2}E(\tfrac{1}{2})]E(\tfrac{1}{2})\Delta_{J+K}f\|^2_{J+K} \\ &= \|\tfrac{1}{2}E(\tfrac{1}{2})\Delta_{J+K}f\|^2_{J+K} \\ &= \|K_{J+K}\Delta_{J+K}f\|^2_{J+K} = \|\Delta_K f\|^2_K \end{aligned}$$

and so $K_J = \sigma$ and $J \geq K$ is a projective pair.

We now prove the converse in much less detail but in the same spirit.

If $J \geq K$ and K_J is a projection, then H_{J+K} can be identified with H_J by introducing a new norm $\|h\|_{J+K} = \|(1 + K_J)h\|_J$. In this way the subspace $K_J H_J$ becomes identified with a subspace $\mathcal{F}_{J+K} \subset H_{J+K}$. Let F_{J+K} be the projection onto this subspace; then it is clear that $K_{J+K} = \frac{1}{2}F_{J+K}$ and $J_{J+K} = (1 - F_{J+K}) + \frac{1}{2}F_{J+K}$ and so $\{J, K\}$ is a projective pair■

Introduce now the maps $J \wedge K: \Phi_- \to \Phi'_-$ and $J \vee K: \Phi_- \to \Phi'_-$ by the relations

$$\begin{aligned} \langle g, J \wedge Kf\rangle &= (\Delta_{J+K}g, \tfrac{1}{2}E(\tfrac{1}{2})\Delta_{J+K}f)_{J+K}, \\ J \vee K &= J + K - J \wedge K. \end{aligned}$$

We have

$$\langle g, J \vee Kf\rangle = (\Delta_{J+K}g, [E_J(1) + E_K(1) + \tfrac{1}{2}E(\tfrac{1}{2})]\Delta_{J+K}f)_{J+K}.$$

We shall assume that $J \wedge K$ and $J \vee K$ are in fact generalized operators; this is certainly true in all of the concrete rigged Hilbert spaces considered so far such as $d \subset d'$, $\mathfrak{D} \subset \mathfrak{D}'$, $\mathfrak{K} \subset \mathfrak{K}$, and direct sums of these.

One clearly has $J \geq J \wedge K$, $K \geq J \wedge K$, $J \leq J \vee K$, $K \leq J \vee K$ and furthermore all these pairs are projective. This latter contention is in fact quite clear but it can be proven in detail using the methods of the lemma above.

Finally if $J \geq K$ and $\{J, K\}$ is projective, then $J \vee K = J$, $J \wedge K = K$.

A set $\mathcal{S}$ of positive generalized operators is said to be *projective* if every unordered pair $\{J, K\}$, $J \in \mathcal{S}$, $K \in \mathcal{S}$, is a projective pair.

We say that a set $\mathcal{S}$ of positive generalized operators is *projectively closed* if whenever $J, K \in \mathcal{S}$ and $\{J, K\}$ is a projective pair, then $J \vee K \in \mathcal{S}$. By assumption the set of all positive generalized operators is projectively closed. The intersection of any number of projectively closed sets is projectively closed. The intersection of all projectively closed sets containing a given set $\mathcal{S}$ will be called the *projective closure* of $\mathcal{S}$ and denoted by $\overline{\mathcal{S}}^{\mathrm{pr}}$.

The set $\mathcal{S}$ will be called *spectral* if its projective closure is projective. Notice that if $\mathcal{S}$ is spectral, then its projective closure is directed for given J, $K \in \overline{\mathcal{S}}^{\mathrm{pr}}$ we have that $\{J, K\}$ is projective and $J \vee K \in \overline{\mathcal{S}}^{\mathrm{pr}}$ but $J \vee K \geq J$, $J \vee K \geq K$. By Lemma 6 we see that $\overline{\mathcal{S}}^{\mathrm{pr}}$ is a projective net and so defines a renormalized Hilbert space $\Psi_{\mathcal{S}} = \varinjlim \Phi_K$, where $\Phi_K = (H_K, H_K, \mathrm{id})$ and $K \in \overline{\mathcal{S}}^{\mathrm{pr}}$. The set $\mathcal{S}$ is now renormalized to a set of commuting projections in Ψ_0. We also have the dressing transformation $\Delta: \Phi_- \to \Psi_+$.

Let now $\mathcal{C}$ be a set of symmetric generalized operators. We assume each $T \in \mathcal{C}$ has a (positive) generalized spectral measure $J_T: A \rightsquigarrow J_T(A)$. We then say that the set $\mathcal{C}$ is *spectrally commutative* if the measures J_T can be picked such that the set $\mathcal{S} = \{J_T(A) \mid T \in \mathcal{C}, A \in \mathfrak{A}_T^0\}$ is spectral. A spectrally commutative set can be simultaneously renormalized and exhibited by means of a set of bona fide mutually commutative spectral measures $E_T: A \rightsquigarrow E_T(A)$.

We give two examples:

Example 26: On $\mathfrak{D}(\mathbb{R}) \subset \mathfrak{D}'(\mathbb{R})$ we consider the two generalized operators

$$(Tf)(x) = f(0)\delta(x),$$
$$(Sf)(x) = f(1)\delta(x-1).$$

This is the same as Example 23 with $y = 1$. We take the following two generalized spectral measures

$$J_T(A)f = \chi_A(0)f + \beta_1\chi_A(\lambda_1)\delta(x)f(0),$$
$$J_S(A)f = \chi_A(0)f + \beta_2\chi_A(\lambda_2)\delta(x-1)f(1),$$

where $\lambda_1, \lambda_2, \beta_1, \beta_2 \in \mathbb{R}$, and $\beta_i \geq 0$.

Either by direct apprehension or by a simple computation one sees that any pair $\{J_T(A), J_S(B)\}$ is projective, that

$$J_T(A) \vee J_S(B)f = \chi_{A\cup B}(0)f + \beta_1\chi_A(\lambda_1)\delta(x)f(0) + \beta_2\chi_B(\lambda_2)\delta(x-1)f(1)$$

and that the set of all generalized operators formed in this way is projective and projectively closed. Thus $\{T, S\}$ is spectrally commutative. We have $\Psi = (L^2(\mathbb{R}) \oplus \mathbb{C}^2, L^2(\mathbb{R}) \oplus \mathbb{C}^2, \mathrm{id})$ with

$$\|f \oplus a \oplus b\|^2 = \|f\|_2^2 + \beta_1|a|^2 + \beta_2|b|^2,$$
$$\Delta f = f \oplus f(0) \oplus f(1),\ T_R f \oplus a \oplus b = 0 \oplus \lambda_1 a \oplus 0,$$
$$S_R f \oplus a \oplus b = 0 \oplus 0 \oplus \lambda_2 b$$

and this accords precisely with the formal picture.

Example 27: We again consider $\mathfrak{D}(\mathbb{R}) \subset \mathfrak{D}'(\mathbb{R})$ and the following two generalized operators:

$$(Tf)(x) = f(0)\delta(x),$$
$$(Pf)(x) = (1/i)f'(x).$$

This corresponds to Example 22 but we have now taken the generator of the group U.

For the generalized spectral measures we take

$$J_T(A)f = \chi_A(0)f + \beta\chi_A(\lambda_0)f(0)\delta(x), \quad B \geq 0,\ \beta, \lambda_0 \in \mathbb{R},$$
$$J_P(A)f = \mathcal{F}^{-1}M_{\chi_A}\mathcal{F}f,$$

where $\mathcal{F}$ is the Fourier transform and M_{χ_A} is the multiplication by the characteristic function χ_A.

If we note that $\chi_A(0)f = \chi_A(0)\mathcal{F}^{-1}M_{\chi_{\mathbb{R}}}\mathcal{F}f$, then one can see that any pair $\{J_T(A), J_P(B)\}$ is projective, that

$$J_T(A) \vee J_P(B)f = \mathcal{F}^{-1}B_{\max(\chi_A(0),\chi_B)}\mathcal{F}f + \beta\chi_A(\lambda_0)f(0)\delta(x),$$

and that the set of these operators is projective and projectively closed.

We have $\Psi = (L^2(\mathbb{R}) \oplus \mathbb{C}, L^2(\mathbb{R}) \oplus \mathbb{C}, \mathrm{id})$ with

$$\|f \oplus c\|^2 = \|f\|_2^2 + \beta|c|^2,$$

$\Delta f = f \oplus f(0)$, $T_R f \oplus c = 0 \oplus \lambda_0 c$, and $P_R f \oplus c = (Pf) \oplus 0 = (1/i)f' \oplus 0$. Thus P_R generates the unitary group $U(t)f \oplus c = f(x+t) \oplus c$. This is one of the groups found in Example 22.

In case the set $\mathcal{C}$ is not spectrally commutative, then the set $\mathcal{S} = \{J_T(A) \mid A \in \mathfrak{A}_T^0 T \in \mathcal{C}\}$ is not spectral and therefore $\mathcal{C}$ cannot *a priori* be simultaneously renormalized by means of a commuting set of spectral measures. One would like to have, however, an analog of the reduction theory of rings of operators and in our case this now involves the following problems:

(1) The set $\mathcal{S}$ above is a candidate for renormalization to a set of not necessarily mutually commutative orthogonal projections. Is there any way of specifying to what extent a set of positive generalized operators is a candidate for renormalization to a set of orthogonal projections?

(2) What are the relationships between the renormalization and reduction programs?

(3) In the case of partial renormalization one can relax the condition that all the generalized operators be positive; in this case one renormalizes a certain set of positive generalized operators and wishes to retain a certain other set as a set of generalized operators on an appropriate rigging of the renormalized Hilbert space. How is the renormalization procedure modified in this case and how is the appropriate rigging to be effected? This again brings up the problems of adequacy.

These questions are interrelated and so the following discussion will not treat them in complete isolation.

Let us first approach the problem of adequacy. If we are given a set $\mathcal{C}$ of generalized operators and a spectral set $\mathcal{S}$ of positive generalized operators, then inductive limit theory provides us with a reinterpretation of each element S of $\mathcal{S}$ as an orthogonal projection S_R in Ψ_0 but it does not supply us with a reinterpretation of an element $T \in \mathcal{C}$. One of the problems connected with this is that the elements of $\Phi_{-S} = H_S$ of the rigged Hilbert space (H_S, H_S, id) and the elements of Ψ_- of the inductive limit may not be sufficiently "smooth" to reinterpret T. This was the situation encountered with Example 23. We want to pass to a different rigging of Ψ_0 and we here attempt to do this by constructing a different system of rigged Hilbert spaces Θ_S with $\Theta_{-S} \subset H_S$ but consisting of vectors "smoother" than an arbitrary vector of H_S.

Let $\Phi_S^{(0)} = \Phi_-$ and, having defined $\Phi_S^{(n)} \subset \Phi_-$, we define $\Phi_S^{(n+1)} \subset \Phi_-$ by

$$\Phi_S^{(n+1)} = \{f \in \Phi_S^{(n)} \mid Q_{ST}\Delta_T f \in \Delta_T \Phi_T^{(n)} \text{ for all } T \geq S\}$$

and let $\Theta_{-S} = \bigcap_{n=0}^{\infty} \Delta_S \Phi_S^{(n)}$.

We now show that, for $T \geq S$, $r_{TS}\Theta_{-S} \subset \Theta_{-T}$. For $n \geq 1$, let $f_S^{(n)} \in \Phi_S^{(n)}$ be such that $\Delta_S f_S^{(n)} = \zeta_S \in \Theta_{-S}$; we have

$$r_{TS}\zeta_S = r_{TS}\Delta_S f_S^{(n)} = Q_{ST}\Delta_T f_S^{(n)},$$

but by definition

$$Q_{ST}\Delta_T f_S^{(n)} = \Delta_T g_T^{(n-1)}, \quad g_T^{(n-1)} \in \Phi_T^{(n-1)},$$

so $r_{TS}\zeta_S = \Delta_T g_T^{n-1}$ and so

$$r_{TS}\zeta_S \in \bigcap_{n=0}^{\infty} \Delta_T \Phi_T^{(n)} = \Theta_{-T}.$$

Now Θ_{-S} need not be dense in H_S and the only time in which this procedure will work is when Θ_{-S} is dense in H_S for a cofinal subset $\mathcal{S}_\Theta$ of $\mathcal{S}$. Let us assume this is the case and we shall discuss this possibility later.

Let $\Upsilon_S \subset H_S$ be the space $\bigvee_{T \geq S} \rho_{ST}\Theta_{-T}$, that is, the linear span of the spaces $\rho_{ST}\Theta_{-T} \subset H_S$. We have $\rho_{ST}\Upsilon_T \subset \Upsilon_S$. For $\eta \in \Upsilon_S$, $\zeta \in \Theta_{-S}$ set $\langle \eta, \zeta \rangle_S = (\eta, \zeta)_S$, the last inner product taken in H_S. Thus, Θ_{-S} can be considered as a space Ξ_S of antilinear functionals on Υ_S. Let Υ_{+S} be the space of Cauchy sequences of Ξ_S with respect to the pairing of Υ_S with Θ_{-S} and let j_S be the canonical inclusion $\Theta_{-S} \subset \Upsilon_{+S}$. We endow Υ_{+S} with the weak$^+$ topology with respect to its pairing with Θ_{-S}. We now show that $r_{TS}: \Theta_{-S} \to \Theta_{-T}$ is continuous with respect to the topologies obtained from the Υ_{+S}, Υ_{+T} by restriction. Indeed for $\Delta_T g \in \Upsilon_T$ and $\Delta_S f \in \Theta_{-S}$ we have

$$\begin{aligned}\langle \Delta_T g, r_{TS}\Delta_S f \rangle &= (\Delta_T g, Q_{ST}\Delta_T f)_T \\ &= (\rho_{ST}\Delta_T g, \rho_{ST}\Delta_T f)_S = (\Delta_S g, \Delta_S f)_S,\end{aligned}$$

but $\Delta_S g \in \Upsilon_S$ so r_{TS} is in fact continuous in the stated topologies. We can therefore extend r_{TS} to a map, again called r_{TS}, from Υ_{+S} to Υ_{+T}. By continuity we still have the transitivity relations $r_{TS}r_{SP} = r_{TP}$.

The triple $(\Theta_{-S}, \Upsilon_{+S}, j_S)$ is not necessarily a rigged Hilbert space, the only difficulty being that there may be elements in Υ_{+S} which vanish on all elements of Θ_{-S}. One can introduce the space Θ_{+S} which is the space of equivalence classes of Υ_{+S} with respect to the equivalence relation that η is equivalent to ζ, $\eta, \zeta \in \Upsilon_{+S}$ if and only if $\eta - \zeta$ vanishes on all elements of Θ_{-S}. Since no subspace of a Hilbert space has an element orthogonal to this given subspace, we see that $\Theta_S = (\Theta_{-S}, \Theta_{+S}, j_S)$ is a rigged Hilbert space and in case $S \in \mathcal{S}_\Theta$, Θ_S is a rigging of H_S. The difficulty with introducing Θ_{+S} is that the map r_{TS} does not map equivalence classes into equivalence classes. Thus, $\{\Theta_S\}$ is not an inductive system of rigged Hilbert spaces. This is, however, only an apparent impasse, for a detailed examination of the inductive limit construction shows that the triple $(\Theta_{-S}, \Upsilon_{+S}, j_S)$ can be used with appropriate modifications. We now give these modifications.

As before we call the map $r_{TS}: \Theta_{-S} \to \Theta_{-T}$ by k_{-TS} and the extension $r_{TS}: \Upsilon_{+S} \to \Upsilon_{+T}$ by k_{+TS}.

We note that (IF4) is satisfied, for if $f \in \Theta_{-S}$, $\phi \in \Upsilon_{+S}$, then there is a sequence $\phi_n \in \Xi_S$ which converges weakly to ϕ. We have, for $T \geq S$, $\langle f, \phi_n \rangle_S = \langle r_{TS} f, r_{TS} \phi_n \rangle_T$, which by continuity of r_{TS} implies that $\langle f, \phi \rangle_S = \langle k_{-TS} f, k_{+TS} \phi \rangle_T$.

Property (IF5) is now modified to the following: The map $(k_{-TS})': \Upsilon_{+T} \to \Theta_{-S'}$ defined by

$$\langle f, (k_{-TS})' \phi \rangle_S = \langle k_{-TS} f, \phi \rangle_T$$

maps Υ_{+T} into Θ^{ω}_{+S}. We prove this: Let $\phi_n \in \Xi_S$ be a sequence converging weakly to ϕ; then

$$\langle k_{-TS} f, \phi \rangle_T = \lim \langle k_{-TS} f, \phi_n \rangle_T = \lim (r_{TS} f, \phi_n)_T = \lim (f, \rho_{ST} \phi_n)_S;$$

but $\rho_{ST} \phi_n \in \Upsilon_S$ so that $f \rightsquigarrow (f, \rho_{ST} \phi_n)_S$ is a functional in Θ_{+S} and since the above limit exists, we have that $f \rightsquigarrow \langle k_{-TS} f, \phi \rangle$ is a functional in Θ^{ω}_{+S}.

We now want to define a rigged Hilbert space Λ which is in some generalized sense an inductive limit of the system $\{(\Theta_{-S}, \Upsilon_{+S}, j_S)\}$ for $S \in \mathcal{S}_\Theta$.

For Λ_- we take as before $\varinjlim \Theta_{-S}$, an inductive limit of complex vector spaces. As before, the canonical maps $k_{-S}: \Theta_{-S} \to \Lambda_-$ are inclusions.

Again as before, we take for Λ_+ that subspace of Λ'_- each element of which when restricted to $k_{-S}\Theta_{-S}$ is an element of Θ^{ω}_{+S}. The canonical restrictions $r_S: \Lambda_+ \to \Theta^{\omega}_{+S}$ are therefore again defined.

The definition of $k_{+S}: \Upsilon_{+S} \to \Lambda_+$ proceeds as before but we now do not have that k_{+S} is an inclusion. What we do have is that k_{+S} does not annihilate any nonzero element of Ξ_S, for if $\phi \in \Xi_S$ and $k_{+S}\phi = 0$, we have, for all $f \in \Theta_{-S}$, $\langle k_{-S} f, k_{+S} \phi \rangle = (f, \phi)_S = 0$ and so $\phi = 0$.

The linear maps l_S and the linear map j are defined as before. These maps are in fact inclusions by the result of the previous paragraph.

As before, the quadratic form $f \rightsquigarrow \langle f, jf \rangle$ on Λ_- is positive definite and this also follows by the above mentioned property of k_{+S}.

The proof that $\Lambda_0 \subset \Lambda_+$ is the same as before and concludes the construction of Λ as a rigged Hilbert space■

The next question is: In what way can Λ carry a generalized operator T if $T \notin \mathcal{S}$? If $\zeta \in \Theta_{-S}$ and f is a representative of ζ in Φ_-, then Tf need not necessarily be in any sense an element of Υ_{+S}; however, one can ask whether this situation holds in the limit in the following sense: Let $l, l' \in \Lambda_-$; then each has a representative ζ, ζ' in $\Theta_{-S}, \Theta_{-S'}$ respectively; for $R \geq S$, $R \geq S'$ let $r_{RS}\zeta$ and $r_{RS}\zeta'$ have representatives f and f' in Φ_-, respectively. One can now ask whether the $\lim_R \langle f', Tf \rangle$ exists and whether it is independent of all the choices involved. If this be the case, then we call this limit $\langle l', T_\Lambda l \rangle$ and T then defines a map $\Lambda_- \to \Lambda'_-$. In case $T_\Lambda \Lambda_- \subset \Lambda_+$, then T_Λ is a generalized operator on Λ and we say that T is *expressible* in Λ. We note that each $S \in \mathcal{S}_\Theta$ is expressible and $S_\Lambda = E_S \mid \Lambda_-$.

We give some examples.

Example 28: Let $\mathcal{S}$ be given by the generalized spectral measure for the multiplication by the δ distribution $\mathcal{S} = \{J(A)\}_{A \in \mathfrak{B}}$, where

$$J(A)f = \chi_A(0) f + \beta \chi_A(\lambda_0) f(0) \delta(x), \quad \beta > 0.$$

Eventually both $\chi_A(0)$ and $\chi_A(\lambda_0)$ are 1 and so we consider this case. We have $H_A = L^2(\mathbb{R}, dx) \oplus \mathbb{C}$,

$$\|f \oplus c\|^2_A = \|f\|^2 + \beta |c|^2$$

and $\Delta_A f = f \oplus f(0)$. If $B \supset A$, then one easily sees that $Q_{AB} = I$ and so $\Phi^{(n)}_A = \Phi_-$ and thus $\Theta_{-A} = \{f \oplus f(0) \mid f \in \mathcal{D}\}$ and consequently $\Upsilon_A = \Theta_{-A}$. We now have that $\Upsilon_{+A} = \Theta_{+A} = \mathcal{D}' \oplus \mathbb{C}$ modulo, the equivalence relation $\sim$ by which $\phi \oplus a \sim \tau \oplus b$ if $\phi - \tau = \beta(b - a)\delta$; j_A is given by defining $j_A f \oplus f(0)$ to be the equivalence class of $f \oplus f(0)$ in Υ_{+A}. The rigged Hilbert space Λ is therefore identical with any one of the triplets $(\Theta_{-A}, \Theta_{+A}, j_A)$. The Hilbert space Λ_0 is indeed $\Psi_0 = L^2(\mathbb{R}, dx) \oplus \mathbb{C}$, where $h \oplus c$ is taken as an element of Λ_+ by considering its equivalence class in $\mathcal{D}' \oplus \mathbb{C}$; this inclusion is well defined, for there is only one representative $\phi \oplus a$ of this class in which $\phi \in L^2$.

Now let $l = f \oplus f(0)$, $l' = g \oplus g(0)$, both being elements of Λ_- and let T be multiplication by $\delta(x - y)$, $y \neq 0$. Then for A such that $\chi_A(0) = 1 = \chi_A(\lambda_0)$, f is the unique representative of l and g is the unique representative of l' and $\langle g, Tf \rangle = \overline{g(y)} f(y)$, which is independent of A and in fact $g \rightsquigarrow \overline{g(y)} f(y)$ defines an element of Λ_+ given by the equivalence class of $f(y)\delta(x - y) \oplus 0$. Thus T is in fact expressible in Λ with $T_\Lambda f \oplus f(0)$ being the equivalence class of $f(y)\delta(x - y) \oplus 0$.

Suppose now that $y = 0$; then $\langle g, Tf \rangle = \overline{g(0)} f(0)$ and, as before, T_Λ exists with $T_\Lambda f \oplus f(0)$ being the equivalence class of $f(0)\delta(x) \oplus 0$. Note now, however, that this equivalence class also contains the representative $0 \oplus (1/\beta) f(0)$ so $T_\Lambda f \oplus f(0) \in \Lambda_0$ and T_Λ is a densely defined bona fide operator; this operator is bounded on its domain and its continuous extension to Λ_0 is $1/\beta$ times the projection onto the direct summand $\mathbb{C}$.

Now J is in fact the generalized spectral measure of T and so it is of interest to compare T_Λ with T_R. We note that $T_\Lambda = T_R$ if and only if $1/\beta = \lambda_0$. Is there any way to understand this relation? We recall that according to the formal picture T is $\delta(0)$ times the projection onto the subspace generated by the delta function. The renormalization procedure "renormalizes" $\delta(0)$ to λ_0 and makes the δ function into a new discrete state of norm $1/\sqrt{\beta}$. Now the formal norm of $\delta(x)$ is

$$\|\delta\| = \left(\int \delta(x)^2\, dx\right)^{\frac{1}{2}} = [\delta(0)]^{\frac{1}{2}}$$

and so the $\delta(0)$ here is "renormalized" to $1/\beta$. The condition $1/\beta = \lambda_0$ therefore equates these two "renormalizations" of $\delta(0)$. The condition $T_\Lambda = T_R$ relates the "renormalizations" of divergent expressions occurring in formal inner products with the ones occurring in the formal eigenvalue problem. This can be seen to be in a sense true in general, for the construction of T_Λ is related to the inner products in the Hilbert spaces H_A which are related to formal inner product calculations while T_R is by construction more sensitive to the formal eigenvalue problem. Note that for $\lambda_0 < 0$ we would be forced to introduce an indefinite metric $\beta < 0$ if we were to insist on the equality $T_\Lambda = T_R$. The precise relationship of T_Λ to T_R is of course a very intriguing question which we further explore in another example below.

As a final observation we note that if we were to renormalize T_Λ for $y = 1$, then we would get the results of Example 25 where now $(T_\Lambda)_R$ corresponds to S_R of that example and the T_R of that example corresponds to expressing the bona fide projection $f \oplus f(0) \rightsquigarrow 0 \oplus \lambda_0 f(0)$ on the new rigged Hilbert space Λ' constructed from Λ in the same way that Λ was constructed from Φ. One can thus break the renormalization procedure into several stages of partial renormalization.

Example 29: Let Φ be the rigged Hilbert space $\mathcal{D}(\mathbb{R}) \subset \mathcal{D}'(\mathbb{R})$ and let S be the set of projections which are multiplications by $\chi_A(x)$ where A is a closed interval. H_A is of course $L^2(A, dx)$ and $\Delta_A f = \chi_A f$. We have $\Phi_A^{(1)} = \{f \in \mathcal{D}(\mathbb{R}) \mid Q_{AB}\chi_B f \in \chi_B \mathcal{D}$ for all $B \supset A\}$, but $Q_{AB}\chi_B f = \chi_A f$ and so $\Phi_A^{(1)} = \mathcal{D}(A)$; therefore, $\Phi_A^{(n)} = \mathcal{D}(A)$ and so $\Theta_{-A} = \mathcal{D}(A)$. Clearly, Υ_A is the set of all complex-valued functions on A that can be extended to a C^∞ function on a neighborhood of A, Υ_{+A} is then $\{\tau \in \mathcal{D}'(\mathbb{R}) \mid \text{supp } (\tau) \subset A\}$ and Θ_{+A} is $\mathcal{D}'(\mathbb{R}) \cap \mathcal{D}'(A)$. The rigged Hilbert space Λ is again $\mathcal{D}(\mathbb{R}) \subset \mathcal{D}'(\mathbb{R})$ and so $\Lambda = \Phi$. If now $l, l' \in \Lambda_-$, then these have representatives f and f' in some $\mathcal{D}(A) \subset \Phi_-$ and these representatives are unique; thus, if T is any generalized operator, then $\langle f', Tf\rangle$ is eventually independent of A and in fact defines a generalized operator in Λ which is identical with T.

The above two examples illustrate that the rigged Hilbert space Λ possesses very natural and elegant properties; however, a note of caution must be added here. The question of rigging is of great importance here: One must be able to pass to a rigging in which there exists a cofinal set S_Θ and which is otherwise adequate for the expressibility of other generalized operators. Candidacy problems therefore begin to play an even more important role here than for the theory of a single generalized operator.

Example 30: We here want to examine the consequences of imposing, up to domain considerations, the equality $T_\Lambda = T_R$ on Example 18 of Sec. III. We are working on $\mathcal{D}(\mathbb{R}^2) \subset \mathcal{D}'(\mathbb{R}^2)$ with T being multiplication by $t(x)\delta(y)$ where $t \in C^\infty$. In Sec. III we found that, for $\lambda \neq 0$,

$$R_\lambda f = \frac{1}{\lambda} f + t(x) \int r_\lambda(x, z) f(z, 0)\, dz \delta(y),$$

where $t(x) r_\lambda(x, y) = t(y)\overline{r_{\bar{\lambda}}(y, x)}$ and weak diagonalizability requires r_λ to have support on the set $\{x = y\}$. Assume for the sake of this example that $J(\mathbb{R})$ exists so that $\Psi = (H_{\mathbb{R}}, H_{\mathbb{R}}, \text{id})$; if r_λ is then sufficiently well behaved, $H_{\mathbb{R}}$ will be given by $H_{\mathbb{R}} = L^2(\mathbb{R}^2) \oplus H_\kappa$, where H_κ is the completion of C^∞ [supp (t)] with respect to the inner product

$$(f, g)_\kappa = \int \overline{f(x)} \kappa(x, y) g(y)\, dx\, dy,$$

where κ is a positive kernel with support in $\{x = y\} \cap$ [supp (t) × supp (t)] and where by definition $f \in C^\infty$ [supp (t)] if it has a C^∞ extension to a neighborhood of supp (t). The dressing transformation will then be given by $\Delta f = f \oplus f(x, 0)$. The construction of Λ is now straightforward and we clearly have $\Lambda_- = \{f \oplus f(x, 0) \in \mathcal{D}(\mathbb{R}^2) \oplus C^\infty \text{ (supp } (t))\} = \Delta\Phi_-$ and Λ_+ is given by the sequential weak$^+$ completion of this with respect to the inner product $(\cdot, \cdot)_\kappa$. Now T_R is given by $T_R = 0 \oplus T_\kappa$, where T_κ is obtained from $t(x) r_\lambda(x, y)$ by the renormalization procedure of Sec. III; on the other hand, T_Λ is given by

$$T_\Lambda f \oplus f(x, 0) = t(x)\delta(y) f \oplus 0,$$

which will equal T_R if and only if

$$\int g(x, 0) t(x) f(x, 0)\, dx = (g(x, 0), T_\kappa f(x, 0))_\kappa$$

for all $g \in \mathcal{D}(\mathbb{R}^2)$. This equality greatly limits our

choices for r_λ and brings us closer to the "desirable" answers of Sec. III. Among the allowed choices is now of course $r_\lambda(x, y) = [1/t(x)]\chi_{\mathrm{supp}(t)}(x)\beta\delta(x-y)/[\lambda - \beta t(x)]$, where $\beta > 0$ but other choices are also possible such as allowing β to depend on x. Notice that the β in the numerator of the above expression corresponds to a "renormalization" of a formal divergent inner product while the β which stands in front of $t(x)$ corresponds to a "renormalization" of a formal eigenvalue. The condition $T_\Lambda = T_R$ thus again relates the two formal considerations but whether such identifications are in fact desirable is not here clear.

Example 31: On $\mathcal{D}(\mathbb{R}) \subset \mathcal{D}'(\mathbb{R})$ let $\mathcal{S}$ consist of the single generalized operator S which is multiplication by the δ function. We have immediately $H_S = \mathbb{C}$, $\Delta_S f = f(0)$, $\Theta_{-S} = \mathbb{C}$, $\Upsilon_S = \mathbb{C}$ and so $\Lambda = (\mathbb{C}, \mathbb{C}, \mathrm{id})$. If $l, l' \in \Lambda_-$, then any $f, f' \in \Phi_-$ with $f(0) = l$, $f'(0) = l'$ are the respective representatives and if T is any generalized operator, $\langle f', Tf\rangle$ is independent of the choice of representatives if and only if T is proportional to S. This is an extreme example in which no operator other than one in the linear span of $\mathcal{S}$ is expressible in Λ; the source of the difficulty is moreover not in the choice of rigging. We have not therefore completely solved the problem of adequacy.

Concerning the construction of Λ some comments are now in order. In the first place the construction of Θ_{-S} and subsequently of Λ makes no use of Φ_+ except indirectly through the use of the set $\mathcal{S}$; thus the construction does not entirely reflect the properties of the original rigged Hilbert space Φ. We do not believe this to be a major drawback since Φ_+ is usually no larger than the $\mathfrak{W}^+$ sequential completion of Φ_- with respect to the sesquilinear form $f, g \rightsquigarrow \langle f, jg\rangle$ and so Φ_+ is closely related to Φ_-. The second comment is that the construction of Λ involves only the set $\mathcal{S}$ and makes no reference to $\mathcal{C}$ so that to have $T \in \mathcal{C}$ expressible in Λ is still partly an accident. As Example 31 points out, this problem of adequacy still remains even after the problem of constructing a sufficiently "smooth" rigging of Ψ_0 has been solved.

We now turn to some other considerations. Let $\mathcal{S}$ be a set of positive generalized operators which is to be a candidate for renormalization to a set of orthogonal projections. What we mean for $\mathcal{S}$ to be renormalizable is that there is a spectral set $\mathfrak{T}$ such that for all $J \in \mathcal{S}$ there is a $P \in \overline{\mathfrak{T}}^{\mathrm{pr}}$ such that $P \geq J$ and the pair is projective. We first note that if now P is *any* element of $\overline{\mathfrak{T}}^{\mathrm{pr}}$ such that $P \geq J$, then this pair is projective, for since $\mathfrak{T}$ is spectral there is now an element $S \in \overline{\mathfrak{T}}^{\mathrm{pr}}$ such that $S \geq P$, $S \geq J$ and both pairs are projective. Since $J \leq P$, we have $J_S \leq P_S$, and since both are projections, we have $J_S P_S = J_S$; therefore

$$\|\rho_{JS} r_{SP} \Delta_P f\|_J = \|\rho_{JS} P_S \Delta_S f\|_J = \|J_S P_S \Delta_S f\|_S = \|J_S \Delta_S f\|_S = \|\Delta_J f\|_J$$

and we conclude that $\rho_{JS} r_{SP} = \rho_{JP}$ and finally

$$J_P = \rho^*_{JP}\rho_{JP} = r^*_{SP}\rho_{JS}\rho_{JS} r_{SP} = \rho_{PS} J_S \rho^*_{PS},$$

which implies that

$$J_P^2 = \rho_{PS} J_S \rho^*_{PS}\rho_{PS} J_S \rho^*_{PS} = \rho_{PS} J_S P_S J_S \rho^*_{PS} = \rho_{PS} J_S \rho^*_{PS} = J_P$$

and $P \geq J$ is projective. We can now renormalize every element $J \in \mathcal{S}$ to an orthogonal projection J_R in $\Psi_{0\mathfrak{T}}$, where $\Psi_{\mathfrak{T}}$ is the renormalized Hilbert space of the spectral set $\mathfrak{T}$; namely, we set $J_R \Psi_0 = k_P J_P H_P$ for any $P \geq J$, $P \in \overline{\mathfrak{T}}^{\mathrm{pr}}$. We thus obtain a set $\mathcal{S}_R$ of orthogonal projections on Ψ_0 and we can ask in what way is the set $\mathcal{S}_R$ unique. A partial answer is provided by the following:

Theorem 2: Let the set $\mathcal{S}$ be renormalized by two spectral sets $\mathfrak{T}$ and $\mathfrak{T}'$ to the respective sets of orthogonal projections $\mathcal{S}_R$ on $\Psi_{0\mathfrak{T}}$ and $\mathcal{S}_{R'}$ on $\Psi_{0\mathfrak{T}'}$. Let S and S' be the closed linear spans of $\mathcal{S}_R\Psi_{0\mathfrak{T}}$ and $\mathcal{S}_{R'}\Psi_{0\mathfrak{T}'}$, respectively and furthermore if $\mathcal{F} \subset \mathcal{S}$ is any *finite* subset of $\mathcal{S}$, let $S_{\mathcal{F}}$ and $S'_{\mathcal{F}}$ be the closed linear spans of $\mathcal{F}_R\Psi_0$ and $\mathcal{F}_{R'}\Psi_0$, respectively.

There exists a densely defined operator $W: S \to S'$ satisfying the following properties:

(1) The domain of W is the linear span of $\bigcup_{\mathcal{F}\subset\mathcal{S}} S_{\mathcal{F}}$;
(2) W is bounded on each $S_{\mathcal{F}}$;
(3) W is an isometry on each $J_R S$;
(4) W intertwines J_R and $J_{R'}$; that is, $WJ_R \supset J_{R'}W$.

Proof: We first define W on a dense subspace $D_{\mathcal{F}}$ of each $S_{\mathcal{F}}$. Let $\mathcal{F} = \{J_i\}_{i=1,\cdots,N}$ and pick $P \in \overline{\mathfrak{T}}^{\mathrm{pr}}$ and $P' \in \overline{\mathfrak{T}}'^{\mathrm{pr}}$ such that $P \geq J_i$, $P' \geq J_i$. Let

$$h = \sum_{i=1}^{N} (J_i)_R k_P \Delta_P f_i;$$

then we define Wh by

$$Wh = \sum_{i=1}^{N} (J_i)_{R'} k_{P'} \Delta_{P'} f_i.$$

To show that W is well defined, let $0 = \sum (J_i)_R k_P \Delta_P f_i$; then for all $g \in \Phi_-$ we have

$$0 = (k_P \Delta_P g, \sum (J_i)_R k_P \Delta_P f_i) = (\Delta_P g, \sum (J_i)_P \Delta_P f_i) = \langle g, \sum J_i f_i\rangle$$

and therefore $\sum J_i f_i = 0$; a similar argument now shows that

$$\sum (J_i)_{R'} k_{P'} \Delta_{P'} f_i = 0$$

and so W is well-defined. Now for $J \in \mathcal{S}$ we have

$$\|J_R k_P \Delta_P f\| = \|J_P \Delta_P f\|_P = \langle f, Jf \rangle$$
$$= \|J_{P'} \Delta_{P'} f\|_{P'} = \|J_{R'} k_{P'} \Delta_{P'} f\|$$

and so W can be extended to an isometry on each $J_R S$. From this it follows that W can be extended to a bounded operator on each S since the bound of W as defined on $D_{\mathcal{F}}$ now depends only on the angular relationships among the closed spaces $\{J_{R'} S\}_{J \in \mathcal{F}}$ and $\{J_{R'} S'\}_{J \in \mathcal{F}}$. The other contentions of the theorem are now obvious■

The above theorem is quite simple but it provides important insights. One should note for example that W need not be an isometry nor need it be bounded. To construct examples of such, we first note the following:

Lemma 7: In $(\mathcal{H}, \mathcal{K}, \mathrm{id})$ if $\mathcal{S}$ is any set of orthogonal projections such that whenever $J, K \in \mathcal{S}$, $J \neq K$, we have $J \cap K = 0$, then $\mathcal{S}$ is spectral.

Proof: Let $\mathcal{F} = \{F_i\}_{i=1,\cdots,N}$ and $\mathcal{F}' = \{F_i'\}_{i=1,\cdots,N'}$ be two finite subsets of $\mathcal{S}$ and let $F = \sum F_i$, $F' = \sum F_i'$. we show that $\{F, F'\}$ is a projective pair. Consider the eigenvalue equation

$$Ff = \lambda(F + F')f,$$

where $f \in (\bigcup F_i \cup \bigcup F_i')\mathcal{H}$. We have

$$\sum_{i=1}^{N} (1 - \lambda) F_i f + \sum_{i=1}^{N'} (-\lambda) F_i' f = 0.$$

Let $\{G_i\}_{i=1,\cdots,M}$ be $\mathcal{F} \cap \mathcal{F}'$ and let $\hat{\sum}$ denote a summation which leaves out the G_i. Then we have

$$\hat{\sum} (1 - \lambda) F_i f + \hat{\sum} (-\lambda) F_i' f + \sum_{i=1}^{M} (1 - 2\lambda) G_i f = 0$$

and so by hypothesis we must have

$$(1 - \lambda) F_i f = 0, \quad F_i \neq G_j,$$
$$(-\lambda) F_i' f = 0, \quad F_i' \neq G_j,$$
$$(1 - 2\lambda) G_i f = 0,$$

and so if $f \neq 0$, we must have either $\lambda = 1$, $\lambda = 0$, $\lambda = \frac{1}{2}$, and $\{F, F'\}$ is a projective pair. This shows that $\bar{\mathcal{S}}^{\mathrm{pr}}$ is given by $\{\sum_{F_i \in \mathcal{F}} F_i\}_{\mathcal{F} \subset \mathcal{S}}$, where $\mathcal{F}$ is any finite subset of $\mathcal{S}$ and thus $\mathcal{S}$ is spectral■

Example 32: Consider $(\mathcal{H}, \mathcal{H}, \mathrm{id})$ and let $\{h_\alpha\}_{\alpha \in A}$ be any linearly independent set of vectors of norm one. Let $\mathcal{S} = \{|h_\alpha\rangle \langle h_\alpha|\}_{\alpha \in A}$; then $\mathcal{S}$ is spectral by the above lemma. Let $\mathcal{T}$ be the spectral set consisting of the singleton $\{1\}$ and let $\mathcal{T}' = \mathcal{S}$. Let $\{e_\alpha\}_{\alpha \in A}$ be an orthonormal set in an abstract Hilbert space of dimension cardinality of A; then W is, up to domain considerations, equivalent to the operator $f \rightsquigarrow \sum e_\alpha(h_\alpha, f)$ for $f \in S$, which clearly need not be an isometry nor be bounded.

It would be instructive to know to what extent one could prove the converse of the Theorem 2, namely if $\mathcal{S}$ is renormalized by a spectral set $\mathcal{T}$ to a set of orthogonal projections and if V is an operator on S taking values in a Hilbert space $\mathcal{K}$ and satisfying conclusions (1)–(3) of the Theorem 2, does there exist a spectral set $\mathcal{T}'$ renormalizing $\mathcal{S}$ to a set of orthogonal projections such that V is equivalent to the operator W whose existence is asserted by the theorem? At present we can only give a partial answer; namely, if V is bounded and invertible on Ψ_0 and is an isometry on each $J_R \Psi_0$, then the answer is yes, for we need only then take $\mathcal{T}'$ to be the set of all generalized operators given by the sesquilinear forms $f, g \rightsquigarrow (Vk_P \Delta_P g, Vk_P \Delta_P f)$ for $P \in \mathcal{T}$. This example covers sufficiently many cases to show that the set $\mathcal{T}$ is far from unique.

Let now $\mathcal{S}$ be spectral and so we can take $\mathcal{T} = \mathcal{S}$; suppose now that $\mathcal{T}'$ also renormalizes $\mathcal{S}$ to a set of *commuting* projections; then it is readily seen by its very construction that W is an isometry. Thus, a spectral set can be renormalized to a commuting set of projections in essentially a unique manner. Conversely, if $\mathcal{S}$ is renormalized by $\mathcal{T}$ to a set of commuting projections, then $\mathcal{S}$ is spectral. To see this, one needs only show the following easily proved fact: If $P \geq J$, $P \geq K$ are projective pairs and J_P and K_P commute, then $\{J, K\}$ is a projective pair and $J \vee K$ is given by $\langle g, J \vee Kf \rangle = (\Delta_P g, J_P \cup K_P \Delta_P f)_P$. A spectral set therefore has an essentially unique distinguished method of renormalization, namely, to a set of commuting projections. Other methods of renormalization are of course still possible.

If $\mathcal{S}$ is not spectral, then there is *a priori* no distinguished method of renormalization and the number of essentially different methods is in fact very large. The problems of renormalization and reduction therefore are not separate. The problem of ascertaining when a set $\mathcal{S}$ can be renormalized to a set of projections is therefore quite delicate and requires deeper investigations.

Let now $\mathcal{S}$ be an arbitrary set of generalized operators and let $\mathcal{T}$ be spectral. Assume that $\mathcal{S}$ is expressible in $\Lambda_{\mathcal{T}}$. We must now again realize that there is no *a priori* distinguished $\mathcal{T}$ and in general there is no way of separating the renormalization and reduction theories.

The above considerations have a bearing on the physical renormalization program, for if $\mathcal{C}$ is not a spectrally commutative set, then even if somehow we picked the generalized spectral measures $\{J_T\}_{T\in\mathcal{C}}$, part of the renormalization program now involves the essentially arbitrary choice of the "angles" among the renormalized spectral measures; even in the case of a spectrally commutative set one may be forced by physical consideration to renormalize the spectral measures to a noncommutative set.

A typical physical situation is the following: We are given a set $\mathcal{G}$ of generalized operators which are the formal generators of the one-parameter subgroups of some physical group such as the Poincaré group; $\mathcal{G}$ must of course be renormalized to a set of bona fide self-adjoint operators. We are also given another set $\mathcal{S}$ of generalized operators that we want to retain as a set of generalized operators. Typically $\mathcal{G}$ is given in terms of creation and annihilation operators and $\mathcal{S}$ is a set of free fields at a fixed time. The renormalization program now is the following: $\mathcal{G}_R$ is obtained by two steps; first, we must choose appropriate generalized spectral measures for each element of $\mathcal{G}$ and, second, we must find a spectral set $\mathcal{T}$ which renormalized the generalized spectral measures to a set of bona fide spectral measures the "angles" among which are such as to obtain a representation $\mathcal{G}_R$ of the Lie algebra of the physical group. We conjecture that if the second step is possible at all, then as for the spectrally commutative case it can be carried out in an essentially unique way. Lastly, we must endeavor to have $\mathcal{S}$ expressible in $\Lambda_{\mathcal{T}}$ to obtain a set of generalized operators $\mathcal{S}_\Lambda$.

The above of course is merely an outline but such an outline was in fact a prime goal of this section. It remains for future investigations to show to what extent this program can be carried out in practice.

V. CONCLUDING REMARKS

Historically functional analytic methods were introduced in order to deal with qualitative features of differential and integral equations; however, in the process of doing so certain concrete features of the original problems become lost. Thus, the fact that the original equations could be solved for concrete functions becomes lost in the study of abstract operators on abstract topological linear spaces. However, one has to reintroduce a certain amount of concreteness in order to deal with certain problems; thus, in studying the decomposition of operators with respect to eigenvectors one often needs more than the spectral measure (part of the purely abstract theory); it then becomes necessary to study features of the eigenvectors themselves. To do so, one needs spaces to carry them and this leads to the introduction of auxiliary vector spaces such as Φ_- and Φ_+ of rigged Hilbert spaces. We have been trying in the previous sections to see to what extent this formalism can handle problems outside of differential equations but which are closely related to them. We envisage a unified theory of differential equations and the so-to-speak generalized differential equations (generalized operators) seeking that such a theory would not only solve hitherto intractable problems but shed more light on differential equations per se by placing them in a wider context. The methods we have employed in developing an appropriate formalism have a large dose of extramathematical reasoning coming mostly from physics; thus we have in this work only a certain outline of a purely mathematical theory.

We believe that the reintroduction of a certain amount of concreteness into the abstract theory of operators is crucial. The abstract methods introduced to study the qualitative features of concrete differential and integral operators have often abstracted too much to permit simple analyses. A differential equation may have solutions which do not lie in the function space set up to study it, and information which these solutions provide either gets lost or else finds its way into the admissible solutions in such a complex way as to be almost inextricable. Thus, in analytic Hilbert space perturbation theory $T(\lambda) = T_0 + \lambda V$ one can often find situations in which the perturbation series for the normalized eigenvectors is not analytic at $\lambda = 0$ even though the concrete differential equation has solutions as concrete functions which are analytic at $\lambda = 0$. In such cases, it is often expedient to study these solutions even though in the end one wants to obtain information only about the Hilbert space behavior of $T(\lambda)$. One would like to have sufficient concreteness to be able to deal with such situations but at the same time have a sufficiently abstract formalism to study the qualitative features of large classes of problems. In this work we have introduced the notion of a rigged Hilbert space at the basis of such a formalism. We again point out that our notion is weaker than that normally used and this is necessarily so, for the usual notion so far has rarely provided results significantly different from those of bona fide spectral theory. One of the surprising insights of our investigations is the importance of what we call candidacy problems. We have not pursued these but there is a definite mathematical program connected with them, namely the following:

(1) Define precisely what is meant by a change of rigging and a change of candidate.

(2) Define precisely what is meant by the various natural riggings (such as the complete and the diagonal rigging).

(3) Prove regularity theorems which determine when various natural riggings exist.

(4) Discover procedures for constructing the natural riggings when these do exist.

(5) Find a natural way of expressing statements that are true "up to a change of rigging."

A natural question that often arises is what additional topological assumptions should one impose on Φ_- and Φ_+. In our view one natural *a priori* assumption is to require Φ_+ to be the $\mathfrak{W}^+$ sequential completion of $j\Phi_-$. This is a generalization of the sequential completeness of a Hilbert space. In this way, starting with any pre-Hilbert space V with the inner product $(\cdot, \cdot)$, one can construct a canonical rigged Hilbert space Φ by taking Φ_- to be V and Φ_+ to be the sequential weak completion of V with respect to the inner product. Any consideration of further topological properties of Φ_- and Φ_+ we believe has to be related to the structures of any given generalized operators and pursued within the mathematical program introduced above.

To disclose the structure of generalized operators we have used analogy with bona fide operators; namely, we assumed there is an analytic manifold of "solutions," given by R_λ, at the singular points of which lie the physically relevant structures. We have realized in the end that this approach is an attempt at two things at once, namely, a mathematical theory and a physical interpretation. For the program to be entirely successful, we must separate these two aspects and begin to decontaminate the mathematics from the physics. The mathematical aspects are the more imperative, a view which we shall further substantiate below. Because of this it is important to recognize the intrinsic properties which are independent of the physical renormalization program.

At the other end of the physical renormalization program one arrives at a set of generalized operator-valued measures and one is once again in the domain of pure mathematics. Section IVC indicates that there is a definite multiplicity theory of mathematical renormalization which can be pursued as an independent mathematical discipline. This theory still needs further material to be gained from physical considerations, for we have not developed in full detail the renormalization program for generalized operators containing "unusual" singularities in the resolvents. The mathematical and physical nature of these has yet to be clarified.

Underlying the whole program of investigating generalized operators is a definite conception of the physical processes that generalized operators represent. Our views on this matter, which are somewhat unconventional, are the following:

The introduction of well-defined mathematical objects corresponding to physical phenomena is in general not very difficult in two forms, what may be called the differential and the integral form. Most theoretical difficulties come from trying to relate the two forms to each other by a definite theory. Thus from an "infinitesimal" mechanical model of continuous media one arrives at the differential equations governing propagation and diffusion processes. There is an extensive body of functional analytic knowledge allowing us to study the structure of such equations and to relate them to the solutions, the integral form of the phenomena. Of course, a theory attempting to relate differential to integral forms may be frustrated by various "paradoxes" whose appearance may signal the existence of qualitatively new phenomena; thus the Klein–Gordon equation contains the Klein paradox which obstructs the quantum mechanical interpretation of the solutions and is finally resolved by the presence of pair creation.

With the experimental development of elementary particle physics it has become clear that in addition to propagation and diffusion phenomena one has to consider the processes of emission and absorption. The formalism of creation and annihilation operators has enabled one to express the differential form of a mathematical theory, namely the introduction of generalized operators. The mathematical structure of these has, however, not been developed; what confronts one immediately is the difficulty that the formal Hamiltonians, say, map vectors of Hilbert space out of Hilbert space. Now there are two possible attitudes toward this fact: (1) the formal Hamiltonian is suspect and one must alter it drastically to conform to quantum mechanics; (2) the view that the inner product in the rigged Hilbert space in which the Hamiltonian is defined carries quantum mechanical probabilistic information is suspect. Most theorists to date have taken the first viewpoint; we propose the second. Taking this point of view, the introduction of the Schrödinger equation

$$i\frac{\partial}{\partial t}\psi(t) = T\psi(t)$$

is seen to involve false pretenses. In the first place, the equation is practically meaningless in that there are in

general no solutions other than $\psi = 0$ and in the second place, since the inner product in Φ_0 does not carry quantum mechanical probabilistic information, there is no point in introducing the equation in the first place. Conventional perturbation and renormalization programs also suffer from these false pretenses since they assume one is starting from some known "bare" quantum mechanical system. In contrast to the case of propagation phenomena where physical reasoning leads directly to *equations to be solved*, physical reasoning in the case of emission and absorption does not lead directly to a set of equations to be solved. However, the reasoning leading to the formal Hamiltonian is sufficiently simple and compelling that one should take it in good faith. The formal Schrödinger equation also should contain some truth and one must now ask what procedure for generalized operators corresponds mathematically as closely as possible to finding the solutions of a differential equation. The answer to this question is the goal of the intrinsic structure theory. One can now see just what role the methods of the previous sections, especially Sec. III, play; they straddle the mathematical and physical frameworks in that having some idea of the final physical situation we use the information to obtain insight into the mathematical structure theory. Eventually, however, the resolvent methods as we have formulated them, especially in the multivalued approach, would have to be given up for more intrinsic methods; for, just as not every differential equation has a well-defined physical interpretation, not every generalized operator should answer to physical principles and we are endeavoring to have a theory of generalized operators in general. It is imperative, however, that the provisional methods which we do employ not do any violence to the mathematical objects involved, for these are not suspect; we are willing, however, to do some violence if necessary to certain physically related notions obtained from bona fide operator theory. We envisage a general theory of absorption and emission, such as a detailed theory of electrodynamics, which stands prior to quantum mechanics. Such a theory may in fact contain "paradoxes" forcing us to change some physical views and it is also for this reason that a purely mathematical theory is imperative.

Finally the theory of generalized operators may be considered as a form of calculus in infinitely many dimensions. There exist other infinite-dimensional calculi; namely the "functional methods." These are all based on the following formal considerations. One formally simultaneously diagonalizes a commuting set of real quantum field amplitudes $\phi(x)$, $x \in \mathbb{R}^n$; $[\phi(x), \phi(y)] = 0$, $x \neq y$. One then introduces a complete set of simultaneous eigenvectors $\phi(x)\Psi_f = f(x)\Psi_f$, where the vector Ψ_f is determined by the function $x \rightsquigarrow f(x)$. If Φ is any other vector, then one can pass to the "representation" $(\Phi, \Psi_f) = \Phi(f)$ which is a functional on the space of functions $x \rightsquigarrow f(x)$. The field $\phi(x)$ now acts as a multiplication operator multiplying $\Phi(f)$ by the evaluation functional $ev_x: f \rightsquigarrow f(x) = ev_x(f)$; thus $(\phi(x)\Phi)(f) = ev_x(f)\Phi(f) = f(x)\Phi(f)$. One then introduces the canonical conjugate amplitude $\pi(x)$; $[\phi(x), \pi(y)] = i\delta(x - y)$ which in the functional "representation" is taken to be the functional derivative $\pi(x) = (1/i)\delta/\delta f(x)$. The above approach is often used by physicists because a large class of Hamiltonians in quantum-field theory is formally given by the forms

$$T = \int \pi(x)^2 \, dx + V(\phi).$$

The "calculus" presented above, however, contains serious mathematical difficulties and there have been numerous attempts to put some form of it on a sound basis but few of these have met with any satisfying success.

One particular difficulty that one notices immediately is that for many of the functionals one wants to consider the higher-functional derivatives such as $\delta^2/\delta f(x)^2$ are not well defined. For example

$$[\delta^2/\delta f(x)^2] \int f(y)^3 \, dy = 6\delta(0) f(x)$$

and one meets the meaningless expression $\delta(0)$. This is reminiscent of the difficulty in formal calculations of generalized operators; however, this difficulty with the functional approach has in practice proved more intractable. Can generalized operator theory help us understand this intractability? The following remarks may be of relevance: Let $\phi(x)$ be a free neutral scalar boson field amplitude defined in Fock space; then formally $\phi(x)$ has eigenvectors $\phi(x)\Psi_\tau = \tau(x)\Psi_\tau$, where $\tau(x)$ is a formal symbol that one manipulates as though it were a function. These eigenvectors can be given explicitly in a concise manner by means of a generating functional; we define

$$G_\tau(f) = \exp\left(\int f(x)\tau(x) \, dx - \tfrac{1}{2}\int f(x)^2 \, dx\right);$$

then the n-particle component of Ψ_τ is given by

$$(\Psi_\tau)_n(x_1, \cdots, x_n) = \frac{1}{(n!)^{\frac{1}{2}}} \frac{\delta}{\delta f(x_1)} \cdots \frac{\delta}{\delta f(x_n)} G_\tau(f)|_{f=0}.$$

So, in particular, $(\Psi'_\tau)_1(x) = \tau(x)$ and we have chosen the normalization such that $(\Psi'_\tau)_0 = 1$. If we now rig the 1-particle subspace by $\Phi_- \subset \Phi_+$, then τ should be in Φ_+. An explicit calculation of $(\Psi'_\tau)_n(x_1, \cdots, x_n)$ shows that this is a sum of tensor products of factors either of the form $\tau(x_i)$ or $\delta(x_i, x_j)$, $i \neq j$, where $\delta(x, y)$ is a formal reproducing kernel

$$\int \delta(x, y) f(y)\, dy = f(x).$$

If we rig the multiparticle subspaces appropriately, these formal expressions will in fact be concrete generalized functions and Ψ'_τ will be in fact a generalized eigenvector. The conditions which these requirements place on Φ_+ are, however, quite liberal, so liberal in fact that there is no generalized function space which contains all possible τ. We have shown therefore that $\phi(x)$ *has no complete rigging*; no rigging of Fock space exhibits all the possible generalized eigenvectors of the bose field. The duals of function spaces therefore do not provide us with a universal infinite-dimensional independent variable but it is the belief in the existence of such a variable, which is naively taken to be a function, that underlies many of the formal methods of the functional approach as used by physicists. Whether in some sense universal infinite-dimensional variables exist is an intriguing question; we know that no dual of a function space provides it. The development of an ambitious multiplicity theory for generalized operators may hopefully shed further light on these problems and possibly facilitate the development of infinite-dimensional calculus.

ACKNOWLEDGMENT

I am grateful to Professor A. S. Wightman for his interest, help, and encouragement.

* Part of this work was done at Princeton University under an NSF Graduate Fellowship and under Putnam Fellowship. I am thankful for this financial support.

[1] A. S. Wightman, Phys. Rev. **101**, 860 (1956).

[2] R. F. Streater and A. S. Wightman, *PCT, Spin and Statistics, and All That* (Benjamin, New York, 1964).

[3] R. Jost, *The General Theory of Quantized Fields* (American Mathematical Society, Providence, R.I., 1965).

[4] A. M. Jaffe, Phys. Rev. **158**, 1454 (1967).

[5] K. O. Friedrichs, *Perturbation of Spectra in Hilbert Space* (American Mathematical Society, Providence, R.I., 1965).

[6] J. Glimm, Commun. Math. Phys. **5**, 343 (1967).

[7] J. Glimm, Commun. Math. Phys. **6**, 61 (1967).

[8] J. Glimm, Commun. Math. Phys. **10**, 1 (1968).

[9] J. Glimm and A. M. Jaffe, Commun. Math. Phys. **11**, 9 (1968).

[10] J. Glimm and A. M. Jaffe, Phys. Rev. **176**, 1945 (1968).

[11] I. M. Gel'fand and N. Ya. Vilenkin, *Obobschennie Funktsii* (Moscow, 1961), Vol. 4, and also in the English translation *Generalized Functions* (Academic, New York, 1964), Vol. 4.

[12] S. K. Berberian, *Notes on Spectral Theory* (Van Nostrand, New York, 1966); the relevant theorems are found in Chap. 4.

[13] We note that, in both the case of projectivity and the case of decomposability, whenever an indefinite metric is introduced, then T_R is *both* symmetric and η symmetric; that is, we have both $(f, T_R g) = (T_R f, g)$ and $(f, T_R g)_\eta = (T_R f, g)_\eta$ for f and g in the domain of T_R. To have a renormalized T which is η symmetric but not symmetric requires the renormalized Hilbert space to have contributions from singularities of R_λ that lie off the real axis; we have not pursued this problem.

[14] J. A. Shohat and J. D. Tomarkin, *The Problem of Moments* (American Mathematical Society, Providence, R.I., 1943); we are here making use of Lemma 2.1, p. 23.

[15] See Ref. 14, Lemma 2.2, p. 24.

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 11, NUMBER 12 DECEMBER 1970

S Matrix in the Heisenberg Representation

EDITH BORIE*†
University of North Carolina, Chapel Hill, North Carolina 27514

(Received 25 May 1970)

The elements of the S matrix are calculated directly from an operator formalism, using the method of Yang and Feldman. This method has the advantage of providing a simple and direct justification of the Feynman rules for gauge fields, which express the contribution to the S matrix from a diagram containing closed loops in terms of sums over lowest-order physical amplitudes (tree amplitudes) in which all external lines are on the mass shell and have physical polarizations. This guarantees unitarity. A condensed notation due to DeWitt is used. First, the one- and two-loop contributions to the amplitude for production of a single quantum, and the amplitudes for pair production and scattering of a single particle by a classical background field are calculated in the absence of an invariance group. Noncausal chains (loops of cylically connected advanced or retarded Green's functions) never appear at any stage of the calculation, thus giving the decomposition into sums over tree amplitudes directly. This result is then generalized in an obvious way to the case in which an invariance group is present. The amplitudes are expressed in terms of a noncovariant propagator which propagates only physical (transverse) quanta. Rewriting these expressions in terms of covariant propagators leads to the formal appearance of "fictitious quanta," which compensate the nonphysical modes carried by these propagators. All results are in agreement with those obtained by other methods.

1. INTRODUCTION

In his study of the radiative corrections to gravity, Feynman[1] proposed that the requirements of unitarity could be met by the use of a technique for calculating these corrections which is not obviously equivalent to the usual Feynman rules. This method is based on the decomposition of diagrams containing closed loops into sums over tree amplitudes (lowest-order physical scattering amplitudes) in which all external lines are on the mass shell and have physical transversality properties. This decomposition was accomplished by removing noncausal chains (loops of cyclically connected advanced or retarded Green's functions) from all closed loops, and has the effect of eliminating the longitudinal components of the propagator. When only single closed loops are involved, these amplitudes group themselves in an obvious way into "Feynman baskets," i.e., groups of tree diagrams, each of which represents a complete physical process. This decomposition guarantees unitarity. In addition, the tree theorem of Feynman, which states that the sum of all tree diagrams for a given physical process is gauge invariant, can be used to show that the radiative corrections are gauge invariant. However, there are difficulties in extending this procedure to the case in which two or more closed loops are present.

The aim of this paper is to attempt to put the justification of these rules on a firmer theoretical footing by showing that they can be derived in a straightforward manner from operator field theory by using a generalization of the method of Yang and Feldman.[2] Although the results agree with those obtained previously by other methods,[3–5] this derivation has the advantage of maintaining a closer connection with conventional field theory. Noncausal chains never appear, and all amplitudes are expressed in terms of propagators for transverse quanta only. The treatment of more than one closed loop is straightforward, although tedious.

A notation due to DeWitt[3,6] is used, which is sufficiently general to embrace all boson field theories, yet also condensed enough to reduce the analysis to manageable proportions.

We treat small disturbances on a classical background, which serves as a reference point about which quantum fluctuations are assumed to take place. This background plays an important role in the analysis of virtual processes. By varying the background field, we can reproduce the effect of individual quanta on a variety of fundamental processes, including the laws of propagation, and all radiative corrections. In addition, the use of a background field eliminates the need to introduce external sources, thus avoiding difficulties when a non-Abelian invariance group is present.

In Sec. 2, the notation is briefly described, and certain preliminary notions are introduced. Asymptotic fields and the corresponding creation and annihilation operators are defined. With the aid of these operators, it is possible to define "vacua" in the remote past and future, relative to the background field, and to construct the incoming and outgoing states which determine the S matrix.

The Yang–Feldman method and its application to the calculation of some elements of the S matrix are

described in Sec. 3. For simplicity, the presence of an invariance group is ignored in order to focus attention on the role of the Yang–Feldman method in yielding automatically the splitting of normal Feynman graphs into Feynman baskets. An iterative solution of the operator field equations gives an expansion of the outgoing fields in terms of the incoming fields, and from this one obtains an expansion of the creation and annihilation operators for outgoing states in terms of the corresponding operators for incoming states. The defining equation for the "relative vacua" is used to determine the "vacuum" in the remote future as an expansion in terms of the incoming states. We then determine the outgoing states as superpositions of incoming states, and use the orthonormality properties of the incoming states to calculate the elements of the S matrix. The amplitudes for one- and two-quantum production and for scattering of a single quantum by the background field are explicitly calculated. These amplitudes have the form of sums over tree amplitudes, and the results obtained in lowest order are in agreement with those obtained by DeWitt,[3] Fadeev and Popov,[4] and Mandelstam.[5] The two-loop contribution to the amplitude is calculated, and is found to be expressible in terms of a functional derivative of the vacuum-to-vacuum amplitude, provided that a certain extra term, which was also used by DeWitt, is added to the field equations. The relationship of the need for this extra term to the removal of the noncausal chains is discussed.

These results are generalized in an obvious way for the case in which an invariance group is present in Sec. 4. The procedure is the same as before, but it is now necessary to introduce two distinct propagators, both of which are defined relative to the background field. One is manifestly covariant, but propagates nonphysical as well as physical quanta; the other propagates physical quanta only, but lacks manifest covariance. Only the latter propagator enters into the calculation of the S matrix, in accordance with the tree theorem. When the amplitudes are re-expressed in terms of the covariant propagator, it is necessary to compensate the nonphysical modes by introducing "fictitious quanta" which couple to real quanta through asymmetric vertices which vanish when the invariance group is Abelian, and which appear only in closed loops.

2. NOTATION, ASYMPTOTIC FIELDS, DEFINITION OF THE S MATRIX

In what follows, a number of definitions and results from Ref. 3 will be used. The most important of these will be introduced as needed or given in Appendix A, but the reader is referred to the above-mentioned reference for a more complete discussion.

The field variables are assumed to be real and are denoted by $\boldsymbol{\phi}^i$. Letters from the middle of the Greek alphabet are used to denote space–time indices, while letters from the beginning of the Greek alphabet are reserved for group indices. Primes are used to distinguish different points of space–time, and also appear on associated indices in order to avoid explicit appearances of the space–time coordinates. In many cases, the primes are suppressed, and the indices i, j, etc., do double duty as discrete labels for the field components and as continuous labels over the points of space–time. The summation convention for repeated indices is extended to include integrations over the space–time variables. Thus, expressions such as M_{ij} are really elements of continuous matrices, and the symbol δ_{ij} involves a 4-dimensional δ function.

Functional differentiation with respect to the field variables is denoted by a comma followed by one or more Latin indices. Thus, if S is the action functional for the system, the field equations could be written as

$$S_{,i} = 0. \tag{2.1}$$

For local theories, $S_{,ij}$ plays the role of a linear differential operator with variable coefficients. The higher functional derivatives are known as bare vertex functions. They describe the basic interaction between finite disturbances and vanish for linear theories. Because of the commutativity of functional differentiation, the bare vertices $S_{,ijk\ldots}$ are completely symmetric in their indices, and $S_{,ij}$ corresponds to a self-adjoint linear operator.

The asymptotic forms of finite disturbances on a classical background are defined by

$$\begin{aligned}\boldsymbol{\phi}^{\pm i} &= \boldsymbol{\phi}^i - G^{\pm ij}(S_{,j} - S_{,jk}\boldsymbol{\phi}^k) \\ &= \boldsymbol{\phi}^i - G^{\pm ij}[(2!)^{-1}S_{,jkl}\boldsymbol{\phi}^k\boldsymbol{\phi}^l \\ &\quad + (3!)^{-1}S_{,jklm}\boldsymbol{\phi}^k\boldsymbol{\phi}^l\boldsymbol{\phi}^m + \cdots],\end{aligned} \tag{2.2}$$

where the $G^{\pm}$ are the advanced and retarded Green's functions of the operator $S_{,ij}$ (or of an associated nonsingular operator F_{ij}, defined in Ref. 3, if an invariance group is present) and the second part of the equation is obtained from the first by a formal expansion of the action about the background field φ. The asymptotic fields satisfy the equation for infinitesimal disturbances,

$$S_{,ij}\boldsymbol{\phi}^{\pm i} = 0, \tag{2.3}$$

and iteration of Eq. (2.2) gives a formal solution to the field equations

$$\begin{aligned} 0 &= S_{,i}[\varphi + \boldsymbol{\phi}] \\ &= S_{,ij}\boldsymbol{\phi}^j + (2!)^{-1}S_{,ijk}\boldsymbol{\phi}^j\boldsymbol{\phi}^k \\ &\quad + (3!)^{-1}S_{,ijkl}\boldsymbol{\phi}^j\boldsymbol{\phi}^k\boldsymbol{\phi}^l + \cdots, \end{aligned} \tag{2.4}$$

where the functional derivatives are evaluated at φ, which is a solution of $S_{,i}[\varphi] = 0$.

If the background field vanishes (flat empty space-time), the most general form for the asymptotic fields is given by

$$\boldsymbol{\phi}^{\pm i} = u^i_A \mathbf{a}^{\pm}_A + u^{*i}_A \mathbf{a}^{\pm *}_A + {}^{\circ}R^i_\alpha \boldsymbol{\zeta}^{+\alpha} \tag{2.5}$$

where the asymptotic wavefunctions u^i_A satisfy

$${}^{\circ}S_{,ij}u^i_A = 0 \tag{2.6}$$

and, in the presence of an invariance group, the supplementary conditions

$${}^{\circ}R_{i\alpha}u^i_A = 0, \quad {}^{\circ}F_{ij}u^j_A = 0. \tag{2.7}$$

Here the capital Latin indices are used as schematic labels for the states of the corresponding quanta. Explicit forms for the u^i_A for the case of the Yang–Mills and gravitational fields are given in Ref. 3. These functions also satisfy the orthonormality relations

$$\begin{aligned} -i\int_\Sigma u^i_A(s_\circ)^\mu_{ij}u^j_B\, d\Sigma_\mu &= -i\int_\Sigma u^i_A(f_\circ)^\mu_{ij}u^j_B\, d\Sigma_\mu = 0, \\ -i\int_\Sigma u^{*i}_A(s_\circ)^\mu_{ij}u^j_B\, d\Sigma_\mu &= -i\int_\Sigma u^{*i}_A(f_\circ)^\mu_{ij}u^j_B\, d\Sigma_\mu = \delta_{AB}, \\ -i\int_\Sigma u^i_A(s_\circ)^\mu_{ij}(R_\circ)^j_\alpha\, d\Sigma_\mu &= 0, \end{aligned} \tag{2.8}$$

where the hypersurface Σ is asymptotically spacelike, but otherwise arbitrary, and the quantities $s_\circ$ and $f_\circ$ are the Green's theorem operators associated with the operators ${}^{\circ}S_{,ij}$ and ${}^{\circ}F_{ij}$, respectively.

When the background field does not vanish, this form is generalized to

$$\boldsymbol{\phi}^{\pm i} = f^{\pm i}_A \boldsymbol{\alpha}^{\pm}_A + f^{\pm *i}_A \boldsymbol{\alpha}^{\pm *}_A + R^i_\alpha \boldsymbol{\zeta}^{\pm\alpha}, \tag{2.9}$$

where

$$f^{\pm i}_A = (1 + G^{\pm}_0 X^{\pm})^i_j u^j_A, \tag{2.10}$$

with the $G^{\pm}_0$ being the advanced and retarded Green's functions in the absence of a background field, and where

$$X^{\pm} = (1 - UG^{\pm}_0)^{-1}U \tag{2.11a}$$

and

$$U_{ij} = F_{ij} - {}^{\circ}F_{ij}. \tag{2.11b}$$

These quantities satisfy a long list of identities which are given in Appendix A.

The creation and annihilation operators $\boldsymbol{\alpha}^{\pm *}$ and $\boldsymbol{\alpha}^{\pm}$ are based on a separation of the total field into a classical background φ and a quantum remainder $\boldsymbol{\phi}$. The classical background is assumed to contain a finite amount of "energy", and hence it superposes linearly with $\boldsymbol{\phi}$ in the remote past and future and disperses ultimately to a state of infinite weakness. We may then write

$$\boldsymbol{\varphi}^{\pm i} = \varphi^{\pm i} + \boldsymbol{\phi}^{\pm i} \tag{2.12}$$

with $\boldsymbol{\phi}^{\pm i}$ given by Eq. (2.9).

We observe that we can write the commutator in the form

$$[\boldsymbol{\phi}^{\pm i}, \boldsymbol{\phi}^{\pm j}] = i\tilde{G}^{ij} = f^{\pm i}_A f^{\pm *j}_A - f^{\pm *i}_A f^{\pm j}_A, \tag{2.13}$$

which follows immediately from Eqs. (2.9), (2.10), and (A8i), the commutation relations obeyed by the $\boldsymbol{\alpha}^{\pm}$, and from

$$f^{\pm} = (1 + G^{\pm}_0 X^{\pm})(1 + G^{\mp}_0 X^{\mp})^{-1} f^{\mp} = f^{\mp} \pm \tilde{G}U, \tag{2.14}$$

which is proved by making use of Eqs. (A8i) and (2.10).

The functions $f^{\pm}$ are the basis functions for classical waves, and they satisfy a set of orthonormality relations which are similar to Eqs. (2.8) and which are given explicitly in Eq. (10.5) of Ref. 3.

The relative vacua $|0, \pm\infty\rangle$ are defined by

$$\boldsymbol{\alpha}^{\pm}_A |0, \pm\infty\rangle = 0. \tag{2.15}$$

It should be stressed that the states $|0, \pm\infty\rangle$ are functionals of the classical background. Because the background is capable of producing or absorbing any number of quanta in individual elementary processes, the two states are not identical.

Next we define incoming and outgoing states

$$|A_1 \cdots A_n, \pm\infty\rangle = \boldsymbol{\alpha}^{\pm *}_{A_1} \cdots \boldsymbol{\alpha}^{\pm *}_{A_n} |0, \pm\infty\rangle. \tag{2.16}$$

If the possibility of stable composite structures is ignored, these form two complete orthogonal bases in the physical Hilbert space.

The S matrix can now be defined relative to the background field, with elements given by the amplitudes

$$\langle A'_1 \cdots A'_n, +\infty \mid A_1 \cdots A_m, -\infty\rangle. \tag{2.17}$$

3. USE OF THE YANG–FELDMAN METHOD FOR THE CALCULATION OF SCATTERING AMPLITUDES

The basic idea of the Yang–Feldman method is to obtain an iterative solution to the field equations in terms of the asymptotic fields and to use this solution to calculate the outgoing states as superpositions of

the incoming states. The orthonormality properties of the incoming states make it possible to obtain the elements of the S matrix simply by taking the appropriate inner products and substituting our previously calculated expansions.

In this chapter, we calculate the amplitudes for one- and two-quantum production and for scattering of a quantum by the background field. For the moment, the presence of an invariance group is ignored in order to focus attention on the problem of the decomposition into Feynman baskets.

First, we write the outgoing fields ϕ^+ as an expansion in terms of the incoming fields ϕ^-, using Eq. (1.2). This gives

$$\begin{aligned}\phi^{+i} &= \phi^{-i} + (G^{-ij} - G^{+ij}) \\ &\quad \times [(2!)^{-1}S_{,jkl}\phi^k\phi^l \\ &\quad + (3!)^{-1}S_{,jklm}\phi^k\phi^l\phi^m + \cdots] \\ &= \phi^{-i} - \tilde{G}^{ij}\{(2!)^{-1}S_{,jkl}\phi^{-k}\phi^{-l} \\ &\quad + \tfrac{1}{2}[(2!)^{-1}S_{,jkl}G^{-ki}S_{,imn} + (3!)^{-1}S_{,jlmn}] \\ &\quad \times (\phi^{-l}\phi^{-m}\phi^{-n} + \phi^{-m}\phi^{-n}\phi^{-l}) + \cdots\},\end{aligned} \tag{3.1}$$

the second line following from the first by iteration.

Using Eqs. (2.12), (2.13), (2.15), and (3.1) and the orthonormality properties of the functions $f^{\pm}$, we obtain

$$\begin{aligned}\alpha_A^+ &= -i\int_\Sigma f_A^{+*i}f_{ij}^\mu\phi^{+j}\,d\Sigma_\mu \\ &= -i\int_\Sigma f_A^{+*i}f_{ij}^\mu\{f_B^{-j}\alpha_B^- + f_B^{-*j}\alpha_B^{-*} \\ &\quad - \tilde{G}^{jk}[(2!)^{-1}S_{,klm}\phi^l\phi^m + \cdots]\}\,d\Sigma_\mu \\ &= -i\int_\Sigma f_A^{+*i}f_{ij}^\mu[f_B^{+j}\alpha_B^- + f_B^{+*j}\alpha_B^{-*} \\ &\quad - \tilde{G}^{jk}(U_{kl}(u_B^l\alpha_B^- + u_B^{*l}\alpha_B^{-*}) \\ &\quad + (2!)^{-1}S_{,klm}\phi^l\phi^m + \cdots)]\,d\Sigma_\mu \\ &= \alpha_A^- + if_A^{+*k}[U_{kl}(u_B^l\alpha_B^- + u_B^{*l}\alpha_B^{-*}) \\ &\quad + (2!)^{-1}S_{,klm}\phi^{-l}\phi^{-m} + \cdots] \\ &= (A + a)_{AB}\alpha_B^- + (B + b)_{AB}\alpha_B^{-*} \\ &\quad + \beta_A + D_{ABC}\alpha_B^{-*}\alpha_C^- + E_{ABC}\alpha_B^-\alpha_C^- \\ &\quad + F_{ABC}\alpha_B^{-*}\alpha_C^{-*} + G_{ABCD}\alpha_B^{-*}\alpha_C^{-*}\alpha_D^- \\ &\quad + H_{ABCD}\alpha_B^{-*}\alpha_C^-\alpha_D^- \\ &\quad + I_{ABCD}\alpha_B^-\alpha_C^-\alpha_D^- + J_{ABCD}\alpha_B^{-*}\alpha_C^{-*}\alpha_D^{-*} + \cdots,\end{aligned} \tag{3.2}$$

where U_{ij} is given by Eq. (2.11b) and where

$$\begin{aligned}A_{AB} &= \delta_{AB} + iu_A^{*i}X_{ij}^-u_B^j, \quad B_{AB} = iu_A^{*i}X_{ij}^-u_B^{*j}, \\ a_{AB} &= if_A^{+*j}[(2!)^{-1}S_{,jkl}G^{-ki}S_{,imn} + (3!)^{-1}S_{,jlmn}] \\ &\quad \times (f_B^{-l}f_C^{-m}f_C^{-*n} + f_C^{-l}f_C^{-*m}f_B^{-n} \\ &\quad + f_C^{-*l}f_C^{-m}f_B^{-n}), \\ b_{AB} &= if_A^{+*j}[(2!)^{-1}S_{,jkl}G^{-ki}S_{,imn} + (3!)^{-1}S_{,jlmn}] \\ &\quad \times (f_B^{-*l}f_C^{-m}f_C^{-*n} + f_C^{-l}f_C^{-*m}f_B^{-*n} \\ &\quad + f_C^{-*l}f_C^{-m}f_B^{-*n}), \\ \beta_A &\simeq \tfrac{1}{2}if_A^{+*j}S_{,jkl}f_B^{-k}f_B^{-*l} \\ &= -\tfrac{1}{2}f_A^{+*j}S_{,jkl}(1 + G_0^-X^-)_m^k \\ &\quad \times G_0^{(+)mn}(1 + X^+G_0^+)_n^l, \\ D_{ABC} &\simeq \tfrac{1}{2}if_A^{+*j}S_{,jkl}(f_B^{-*k}f_C^{-l} + f_C^{-k}f_B^{-*l}), \\ E_{ABC} &\simeq \tfrac{1}{2}if_A^{+*j}S_{,jkl}f_B^{-k}f_C^{-l}, \\ F_{ABC} &\simeq \tfrac{1}{2}if_A^{+*j}S_{,jkl}f_B^{-*k}f_C^{-*l}, \\ G_{ABCD} &\simeq if_A^{+*j}[(2!)^{-1}S_{,jkl}G^{-ki}S_{,imn} + (3!)^{-1}S_{,jlmn}] \\ &\quad \times (f_B^{-*l}f_C^{-*m}f_D^{-n} + f_C^{-*l}f_B^{-*m}f_D^{-n} \\ &\quad + f_D^{-l}f_B^{-*m}f_C^{-*n}), \\ H_{ABCD} &\simeq if_A^{+*j}[(2!)^{-1}S_{,jkl}G^{-ki}S_{,mn} + (3!)^{-1}S_{,jlmn}] \\ &\quad \times (f_B^{-*l}f_C^{-m}f_D^{-n} + f_C^{-l}f_B^{-*m}f_D^{-n} \\ &\quad + f_D^{-l}f_B^{-*m}f_C^{-n}), \\ I_{ABCD} &\simeq \tfrac{1}{2}if_A^{+*j}[(2!)^{-1}S_{,jkl}G^{-ki}S_{,imn} + (3!)^{-1}S_{,jlmn}] \\ &\quad \times (f_B^{-l}f_C^{-m}f_D^{-n} + f_D^{-l}f_B^{-m}f_C^{-n}), \\ J_{ABCD} &\simeq \tfrac{1}{2}if_A^{+*j}[(2!)^{-1}S_{,jkl}G^{-ki}S_{,imn} + (3!)^{-1}S_{,jlmn}] \\ &\quad \times (f_B^{-*l}f_C^{-*m}f_D^{-*n} + f_D^{-*l}f_B^{-*m}f_C^{-*n}).\end{aligned} \tag{3.3}$$

It can be shown that Eq. (3.2) is consistent with the commutation rules for creation and annihilation operators for free fields. If the expansion (3.2) and its conjugate are inserted into the expressions for the commutators of α^+ and α^{+*}, tedious algebra[7] shows that the commutation rules are satisfied provided the α^- and α^{-*} satisfy them.

Next, we expand the outgoing "vacuum" in terms of the incoming states:

$$\begin{aligned}|0, +\infty\rangle &= C_0\,|0, -\infty\rangle \\ &\quad + \sum_{n=1}^{\infty}(n!)^{-1}C_{A_1\cdots A_n}\,|A_1\cdots A_n, -\infty\rangle.\end{aligned} \tag{3.4}$$

Substitution of this expansion and Eq. (3.2) into the defining equation (2.14) for the state $|0, +\infty\rangle$ gives conditions which determine the coefficients $C_{A_1\cdots A_n}$. Once these have been calculated, we obtain the elements of the S matrix by straightforward computation of inner products between incoming and

outgoing states. The conditions which we obtain are

$$\begin{aligned}0 &= \alpha_A^+ |0, +\infty\rangle \\ &= \sum_{n=0}^{\infty}(n!)^{-1}[(\beta_A C_{A_1\cdots A_n} + A_{AB}C_{BA_1\cdots A_n} \\ &\quad + E_{ABC}C_{BCA_1\cdots A_n} + I_{ABCD}C_{BCDA_1\cdots A_n} + \cdots) \\ &\quad \times |A_1\cdots A_n, -\infty\rangle \\ &\quad + (B_{AB}C_{A_1\cdots A_n} + D_{ABC}C_{CA_1\cdots A_n} \\ &\quad + H_{ABCD}C_{CDA_1\cdots A_n} + \cdots)|BA_1\cdots A_n, -\infty\rangle \\ &\quad + (F_{ABC}C_{A_1\cdots A_n} + G_{ABCD}C_{DA_1\cdots A_n} + \cdots) \\ &\quad \times |BCA_1\cdots A_n, -\infty\rangle \\ &\quad + (J_{ABCD}C_{A_1\cdots A_n} + \cdots) \\ &\quad \times |BCDA_1\cdots A_n, -\infty\rangle + \cdots]. \end{aligned} \tag{3.5}$$

An iterative method is used to obtain an approximate solution to the system of equations given in (3.5). First, only the zero-order terms[8] are kept, giving

$$A_{AB}C^0{}_{BA_1\cdots A_n} + \sum_{i=1}^{n} B_{AA_i}C^0{}_{A_1\cdots A_{i-1}A_{i+1}\cdots A_n} = 0$$

for each n. (3.6)

These can be solved exactly, with the result

$$\begin{aligned}&C^0{}_{BA_1\cdots A_{2n}} = 0, \\ &C^0{}_{BA_1\cdots A_{2n+1}} = \frac{(-1)^{n+1}}{2^n n!}\sum_{\substack{\text{perm}\\ 1\cdots 2n+1}}(A^{-1}B)_{BA_1} \\ &\qquad \times (A^{-1}B)_{A_2A_3}\cdots(A^{-1}B)_{A_{2n}A_{2n+1}}C_0, \end{aligned} \tag{3.7}$$

where

$$\begin{aligned}(A^{-1}B)_{AB} &= iu_A^{*j}(1 - X^-G_0^{(+)})^{-1}{}_j{}^k X^-_{kl}u_B^{*l} \\ &= iu_A^{*j}X^*_{jk}u_B^{*k} \end{aligned} \tag{3.8}$$

is symmetric, so that the $C_{A_1\cdots A_n}$ are totally symmetric in their indices, as would be expected.

These equations do not determine the vacuum-to-vacuum amplitude

$$\langle 0, +\infty \,|\, 0, \infty\rangle = C_0^* \equiv e^{iW[\phi]}, \tag{3.9}$$

since the coefficients are determined only up to the constant C_0 by Eqs. (2.5). However, only the vacuum-to-vacuum *probability* is really of interest, the phase of the amplitude being physically irrelevant. This can be obtained from the normalization condition on the state $|0, +\infty\rangle$:

$$\begin{aligned}1 &= \langle 0, +\infty \,|\, 0, +\infty\rangle = \sum_{n=0}^{\infty}(n!)^{-1}C^*_{A_1\cdots A_n}C_{A_1\cdots A_n} \\ &\simeq |C_0|^2\, e^{\frac{1}{2}\,\mathrm{Tr}\log[1-(A^{-1}B)^*(A^{-1}B)]} \\ &\simeq |C_0|^2\, e^{\frac{1}{2}\log\det(1-G^{(-)}{}_0XG^{(+)}{}_0X^*)}, \end{aligned} \tag{3.10}$$

which is in agreement with previous results.[9]

The zero-order expressions are next inserted back into equations (3.5), and first-order terms are retained. Details of some of the calculations are given in Appendix B. The results to first order are

$$\begin{aligned}C_A &\simeq \tfrac{1}{2}f_A^{*j}S_{,jkl}G^{(-)*kl}C_0, \\ C_{AB} &\simeq C^0{}_{AB} = -(A^{-1}B)_{AB}C_0, \\ C_{ABC} &\simeq -(A^{-1}B)_{AB}C_C - (A^{-1}B)_{AC}C_B \\ &\quad - (A^{-1}B)_{BC}C_A - if_A^{*j}S_{,jkl}f_B^{*k}f_C^{*l}C_0, \\ C_{ABCD} &\simeq C^0{}_{ABCD}, \end{aligned} \tag{3.11}$$

where

$$f_A^i = (1 + G_0X)^i{}_j u_A^j. \tag{3.12}$$

The first-order results are now inserted into Eqs. (3.5), and second-order terms are kept. The calculation is similar to the first-order calculation (and is sketched for the case of C_{AB} in Appendix B), with the result that

$$\begin{aligned}C_{AB} \simeq\ & -(A^{-1}B)_{AB}C_0 + C_AC_BC_0^{-1} \\ & - \tfrac{1}{2}f_A^{*j}S_{,jkl}G^{(-)*ki}S_{,imn}G^{(-)*ml}f_B^{*n}C_0 \\ & + f_A^{*j}[(2!)^{-1}S_{,jkl}G^{*ki}S_{,imn} + (3!)^{-1}S_{,jlmn}] \\ & \times [G^{(-)*mn}f_B^{*l} + (G^{(-)*ml} + G^{(-)*lm})f_B^{*n}]C_0. \end{aligned} \tag{3.13}$$

The correction to C_{ABCD} is given by a rather lengthy expression in Ref. 7.

The third-order correction to C_A is found in a similar manner by a very lengthy but straightforward computation and is also given in Ref. 7.

This procedure can in principle be continued to give all the coefficients $C_{A_1\cdots A_n}$ to arbitrarily high order. However, the number of terms one has to work with increases rapidly with increasing order, and the calculations become extremely lengthy. Those given here or in Ref. 7 are sufficient for the computation of the amplitudes for production of two quanta and for scattering of a single particle by the background field to second order (corresponding to a single closed loop in the radiative corrections) and the single-quantum production amplitude to third order (corresponding to two closed loops).

The calculation of the single-quantum production amplitude to lowest order is given in detail in Appendix C as an example of how the calculations go. The result is

$$\langle A, +\infty \,|\, 0, -\infty\rangle = -\tfrac{1}{2}C_0^*\bar{f}_A^j S_{,jkl}G^{(+)kl}, \tag{3.14}$$

where $\bar{f} = (1 + G_0X)u^*$. This can be written in the form

$$\langle A, +\infty| \ 0, -\infty\rangle = ie^{iW}f_A^j W_{,j},$$

with

$$W_{,j} \simeq \tfrac{1}{2} i S_{,jkl} G^{(+)kl}, \tag{3.15}$$

where the functional $W[\varphi]$ is defined in Eq. (3.8). This is in agreement with previous results [cf. Eq. (17.2) of Ref. 3].

The third-order contribution to the single-quantum production amplitude is given by

$$\begin{aligned}
&\langle A, \infty \mid 0, -\infty \rangle \\
&\simeq -\tfrac{1}{2} C_0^* \bar{f}^j_A S_{,jkl} G^{(+)kl} \\
&\quad - \tfrac{1}{8} i C_0^* \bar{f}^j_A S_{,jkl} G^{ki} S_{,imn} G^{lp} S_{,pab} G^{(+)mn} G^{(+)ab} \\
&\quad - \tfrac{1}{4} i C_0^* \bar{f}^j_A S_{,jkl} G^{ki} S_{,imn} G^{np} S_{,pab} G^{(+)ml} G^{(+)ab} \\
&\quad - \tfrac{1}{4} i C_0^* \bar{f}^j_A S_{,jkl} G^{-ki} S_{,imn} G^{np} S_{,pab} G^{(+)lm} G^{(+)ab} \\
&\quad - \tfrac{1}{4} i C_0^* \bar{f}^j_A S_{,jlmn} G^{lp} S_{,pab} G^{(+)mn} G^{(+)ab} \\
&\quad - \tfrac{1}{4} i C_0^* \bar{f}^j_A S_{,jkl} G^{ki} S_{,imnp} G^{(+)ml} G^{(+)np} \\
&\quad - \tfrac{1}{4} i C_0^* \bar{f}^j_A S_{,jkl} G^{-ki} S_{,imnp} G^{(+)lm} G^{(+)np} \\
&\quad - \tfrac{1}{8} i C_0^* \bar{f}^j_A S_{,jklmn} G^{(+)kl} G^{(+)mn} \\
&\quad - \tfrac{1}{6} i C_0^* \bar{f}^j_A S_{,jlmn} G^{np} S_{,pab} G^{(+)am} G^{(+)bl} \\
&\quad - \tfrac{1}{6} i C_0^* \bar{f}^j_A S_{,jlmn} G^{-np} S_{,pab} G^{(+)ma} (G^{(+)lb} + G^{(+)bl}) \\
&\quad - \tfrac{1}{4} i C_0^* \bar{f}^j_A S_{,jkl} G^{ki} S_{,imn} G^{np} S_{,pab} G^{(+)am} G^{(+)bl} \\
&\quad - \tfrac{1}{4} i C_0^* \bar{f}^j_A S_{,jkl} G^{ki} S_{,imn} G^{-np} S_{,pab} G^{(+)ma} G^{(+)bl} \\
&\quad - \tfrac{1}{4} i C_0^* \bar{f}^j_A A_{,jkl} G^{-ki} S_{,imn} G^{-np} S_{,pab} \\
&\quad \times (G^{(+)ma} + G^{(+)am}) G^{(+)lb} \\
&\quad - \tfrac{1}{4} i C_0^* \bar{f}^j_A S_{,jkl} G^{-ki} S_{,imn} G^{lp} S_{,pab} G^{(+)am} G^{(+)bn}.
\end{aligned} \tag{3.16}$$

This is almost, but not quite, expressible in the form $ie^{iW} \bar{f}^j_A W_{,j}$. What is missing is a term having the following structure:

$$-\tfrac{1}{12} i C_0^* \bar{f}^j_A S_{,jkl} G^{-ki} S_{,imn} G^{-np} S_{,pab} G^{-bl} \tilde{G}^{am}. \tag{3.17}$$

Such a term would come from adding to the right-hand side of the operator field equations (2.4) a term of the form

$$\tfrac{1}{12} S_{,ijk} G^{-jl} S_{,lmn} G^{-np} S_{,pab} G^{-bk} \tilde{G}^{ma} \tag{3.18}$$

and a corresponding term to the definition of the asymptotic fields (2.2). If this term is added, we obtain the following expression for $W_{,j}$:

$$\begin{aligned}
W_{,j} \simeq\ & \tfrac{1}{2} i S_{,jkl} G^{(+)kl} - \tfrac{1}{8} (S_{,lmn} G^{lp} S_{,pab})_{,j} \\
& - \tfrac{1}{8} (S_{,klmn} G^{(+)kl} G^{(+)mn})_{,j} \\
& - \tfrac{1}{12} (S_{,lmn} G^{np} S_{,pab} G^{(+)ma} G^{(+)lb} \\
& + S_{,lmn} G^{-np} S_{,pab} (G^{(+)am} + G^{(+)ma}) G^{(+)lb})_{,j}.
\end{aligned} \tag{3.19}$$

This same extra term had to be added to the field equations by DeWitt in order to be able to write them in the form

$$0 = T\{S_{,i}[\varphi + \phi] - i(\ln \Delta[\varphi + \phi])_{,i}\}. \tag{3.20}$$

It is also obtained when one explicitly removes the noncausal chains from the normal Feynman graphs. Since one expects that particle production amplitudes are obtained by functional differentiation of the vacuum diagrams and since the removal of the noncausal chains does not affect this relationship, it is perhaps not too surprising to find that it must be added in this case also. The expression for $W[\varphi]$ obtained by functional integration of Eq. (3.19) is the same as the corresponding expression obtained by DeWitt. It is given in diagram form, following the conventions of that paper in Fig. 1. However, DeWitt showed that still another term $Y_{(2)}$ must be added to the field equations in order (i) to complete the decomposition into Feynman baskets and (ii) to guarantee invariance of the theory under a change of variables. This term does not come automatically either from his calculation or from the present work. There is no difficulty in including it in the operator field equations of the Yang–Feldman formalism, and there are good consistency arguments for doing so. However, it cannot at the present time be said to emerge at a fundamental level, e.g., from a canonical formalism.

We now proceed to the calculation of the two-quantum production amplitude and the amplitude for scattering of a single quantum by the background field. After a rather lengthy but straightforward calculation, we obtain, for pair production,

$$\begin{aligned}
&\langle AA', +\infty \mid 0, -\infty \rangle \\
&\simeq i C_0^* u^{*p}_A X_{pr} u^{*r}_{A'} + \tfrac{1}{4} C_0^* \bar{f}^j_A S_{,jkl} G^{(+)kl} \bar{f}^i_{A'} S_{,imn} G^{(+)mn} \\
&\quad - \tfrac{1}{2} C_0^* \bar{f}^j_A (S_{,jkl} G^{ki} S_{,imn} + S_{,jlmn}) G^{(+)mn} \bar{f}^l_{A'} \\
&\quad - \tfrac{1}{2} C_0^* \bar{f}^j_A \bar{f}^n_{A'} S_{,jkl} G^{ki} S_{,imn} (G^{(+)ml} + G^{(+)lm}) \\
&\quad - \tfrac{1}{2} C_0^* \bar{f}^j_A \bar{f}^n_{A'} S_{,jkl} G^{(+)ki} S_{,imn} G^{(+)ml},
\end{aligned} \tag{3.21}$$

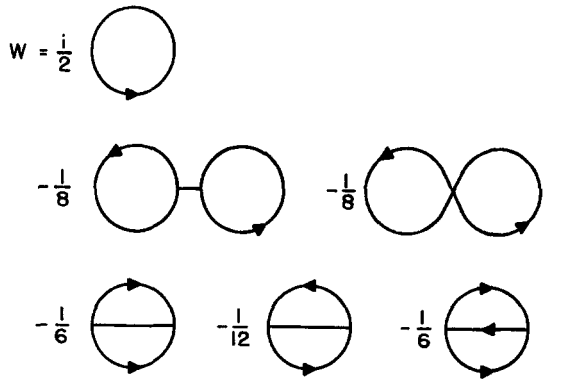


FIG. 1. Vacuum diagram in absence of an invariance group. Lines bearing arrows represent quanta on the mass shell.

which agrees with previous results and which can be written in the form

$$\langle AA', +\infty \mid 0, -\infty\rangle = ie^{iW}u^{*p}_{A}X_{pr}u^{*r}_{A'} + f^{i}_{A}f^{j}_{A'}(e^{iW})_{,ij}, \quad (3.22)$$

with W given approximately by (3.15).

Similarly, we obtain, for scattering,

$$\begin{aligned}\langle A, +\infty \mid A', -\infty\rangle &\simeq C_0^*(\delta_{AA'} + iu^{*p}_{A}X_{pr}u^{r}_{A'}) \\ &+ \tfrac{1}{4}C_0^*\bar{f}^{j}_{A}S_{,jkl}G^{(+)kl}f^{i}_{A'}S_{,imn}G^{(+)mn} \\ &- \tfrac{1}{2}C_0^*\bar{f}^{j}_{A}f^{l}_{A'}(S_{,jkl}G^{ki}S_{,imn} + S_{,jlmn})G^{(+)mn} \\ &- \tfrac{1}{2}C_0^*\bar{f}^{j}_{A}f^{n}_{A'}S_{,jkl}G^{ki}S_{,imn}(G^{(+)ml} + G^{(+)lm}) \\ &- \tfrac{1}{2}C_0^*\bar{f}^{j}_{A}f^{n}_{A'}S_{,jkl}G^{(+)ki}S_{,imn}G^{(+)ml}, \quad (3.23)\end{aligned}$$

which also agrees with previous results. It is seen that the radiative corrections for scattering differ from those for pair production only in the modification of the external line wavefunctions to suit the process being considered.

The formal results given above apply to any boson field theory. They have been checked, in lowest order, for the case of scalar fields with self-coupling by explicit computation of the relevant amplitudes.[10] The result obtained by calculating the amplitude as given in Eq. (3.22) is the same as that obtained by applying the usual Feynman rules.

4. GENERALIZATION TO THE CASE IN WHICH AN INVARIANCE GROUP IS PRESENT

It is not difficult to generalize the previous results to the case in which an invariance group is present. Equation (3.1) is unaffected, and Eq. (3.2) is modified only in the fact that the functions $iG_0^{(+)}$ which appear in the expressions for β_A, a_{AB}, and b_{AB} are replaced by their projections into the physical subspace $i\mathfrak{G}_0^{(+)}$ (see Appendix A). This occurs because only products of the form $uu^\dagger$ appear in any of the calculations. Repeating the previous calculations, but using Eqs. (A10) instead of Eqs. (A8), we obtain, to second order,

$$\begin{aligned}C_A &\simeq -\tfrac{1}{2}\bar{\mathfrak{f}}^{*j}_{+A}S_{,jkl}\mathfrak{G}^{(+)*kl}C_0 \\ C_{AB} &\simeq -(A^{-1}B)_{AB}C_0 + C_AC_BC_0^{-1} \\ &+ \tfrac{1}{2}C_0\bar{\mathfrak{f}}^{*j}_{+A}S_{,jkl}\mathfrak{G}^{(-)*ki}S_{,imn}\mathfrak{G}^{(+)*lm}\bar{\mathfrak{f}}^{*n}_{-B} \\ &- \bar{\mathfrak{f}}^{*j}_{+A}[(2!)^{-1}S_{,jkl}\mathfrak{G}_-^{*ki}S_{,imn} + (3!)^{-1}S_{,jlmn}] \\ &\times [\mathfrak{G}^{(+)*mn}\bar{\mathfrak{f}}^{*l}_{-B} \\ &+ (\mathfrak{G}^{(+)*ml} + \mathfrak{G}^{(+)*lm})\bar{\mathfrak{f}}^{*n}_{-B}]C_0, \\ C_{ABC} &\simeq -(A^{-1}B)_{AB}C_C - (A^{-1}B)_{AC}C_B \\ &- (A^{-1}B)_{BC}C_A - i\bar{\mathfrak{f}}^{*j}_{+A}S_{,jkl}\bar{\mathfrak{f}}^{*k}_{-B}\bar{\mathfrak{f}}^{*l}_{-C}, \quad (4.1)\end{aligned}$$

with

$$(A^{-1}B)_{AB} = iu^{*p}_{A}\mathfrak{X}^{*}_{-pr}u^{*r}_{B} \quad (4.2)$$

and

$$\bar{\mathfrak{f}}_{\pm} \equiv (1 + \mathfrak{G}_{0\pm}\mathfrak{X}_{\pm})u. \quad (4.3)$$

These two functions differ only by a gauge transformation, and $u^\dagger(\mathfrak{X}_+ - \mathfrak{X}_-)u^* = 0$, so that the plus and minus signs are physically irrelevant.

To lowest order, the amplitude for single-quantum production is given by

$$\langle A, +\infty \mid 0, -\infty\rangle \simeq -\tfrac{1}{2}C_0^*\bar{\mathfrak{f}}^{j}_{+A}S_{,jkl}\mathfrak{G}^{(+)kl}, \quad (4.4)$$

which differs from the result for the case in which no group was present only in the replacement of $G^{(+)}$ by $\mathfrak{G}^{(+)}$ and of the function f by its generalization $\bar{\mathfrak{f}}_{\pm}$. As has been shown by DeWitt, this is also equal to

$$\begin{aligned}&\tfrac{1}{2}C_0^*\bar{\mathfrak{f}}^{j}_{+A}[S_{,jkl}G^{(+)kl} - R^{k}_{\alpha,j}R_{k\beta}(\hat{G}^{(+)\alpha\beta} + \hat{G}^{(+)\beta\alpha})] \\ &\quad = -\tfrac{1}{2}C_0^*\bar{\mathfrak{f}}^{j}_{+A}[S_{,jkl}G^{(+)kl} - (V_{(\alpha j)\beta} + V_{(\beta j)\alpha})\hat{G}^{(+)\alpha\beta}], \quad (4.5)\end{aligned}$$

where $\hat{G}^{(+)}$ is the propagator for fictitious quanta and where

$$V_{(\alpha j)\beta} = R^{k}_{\alpha,j}R_{k\beta}. \quad (4.6)$$

We obtain the two-quantum production amplitude in a similar manner. The result is

$$\begin{aligned}\langle AA', +\infty \mid 0, -\infty\rangle &= iC_0^*u^{*p}_{A}\mathfrak{X}_{+pr}u^{*r}_{A'} \\ &+ \tfrac{1}{4}C_0^*\bar{\mathfrak{f}}^{j}_{+A}S_{,jkl}\mathfrak{G}^{(+)kl}\bar{\mathfrak{f}}^{i}_{+A'}S_{,imn}\mathfrak{G}^{(+)mn} \\ &- C_0^*\bar{\mathfrak{f}}^{j}_{+A}[(2!)^{-1}S_{,jkl}\mathfrak{G}_-^{ki}S_{,imn} + (3!)^{-1}S_{,jlmn}] \\ &\times (\mathfrak{G}^{(+)mn}\bar{\mathfrak{f}}^{l}_{+A'}(\mathfrak{G}^{(+)ml} + \mathfrak{G}^{(+)lm})\bar{\mathfrak{f}}^{n}_{+A'}) \\ &- \tfrac{1}{2}C_0^*\bar{\mathfrak{f}}^{j}_{+A}S_{,jkl}\mathfrak{G}^{(+)ki}S_{,imn}\mathfrak{G}^{(+)ml}\bar{\mathfrak{f}}^{n}_{+A'}. \quad (4.7)\end{aligned}$$

Again, this differs from (3.21) only in the replacement of $G^{(+)}$ by $\mathfrak{G}^{(+)}$ and of f by $\bar{\mathfrak{f}}_{\pm}$. This result was also obtained by DeWitt. The re-expression of (4.7) in terms of the covariant propagators and the resulting terms involving the fictitious quanta are given in Ref. 3.

This calculation verifies explicitly that, at least in lowest order, the elements of the S matrix in the presence of an invariance group can be obtained from the corresponding expressions in the absence of a group simply by replacing the propagators $G^{(+)}$ by $\mathfrak{G}^{(+)}$ everywhere and using suitably generalized external line wavefunctions. The calculation also indicates that this should be true for higher orders as well. Since the amplitudes have already been decomposed into Feynman baskets, the radiative corrections are automatically expressed in terms of sums over tree amplitudes. Hence the results are automatically unitary and, by the tree theorem, also gauge invariant. The use of physical gravitons in the

sums over the tree amplitudes introduces a non-relativistic element into the theory. However, the expressions can always be converted to manifestly covariant form, provided that one introduces the fictitious quanta to compensate the unphysical modes which are carried by the covariant propagator. The necessity for doing this arises in a completely natural way when one calculates the amplitudes with the Yang–Feldman method.

ACKNOWLEDGMENT

The author would like to express her appreciation to B. S. DeWitt for many helpful discussions.

APPENDIX A

For the convenience of the reader, we list here a number of useful identities which are given in Ref. 3 and which are needed in the present calculation.

By (2.11) we have

$$U_{ij} = F_{ij} - {}^{0}F_{ij},$$

and

$$X^{\pm} = (1 - UG_0^{\pm})^{-1}U = (1 + X^{\pm}G_0^{\pm})U.$$

From the self-adjointness of ${}^{0}F_{ij}$ it follows that $G_0^{+ij} = G_0^{-ji}$ and hence that

$$X_{ij}^{+} = X_{ji}^{-}. \tag{A1}$$

We define the positive energy function $G_0^{(+)}$ by

$$iG_0^{(+)ij} = u_A^i u_A^{*j} + R_{0\alpha}^i v_a^{\alpha}(N^{-1})^{ab}v_b^{*\beta}\bar{R}_{0\beta}^j + \bar{R}_{0\alpha}^i v_a^{\alpha}(N^{-1})^{ba}v_b^{*\beta}R_{0\beta}^j \tag{A2}$$

and the negative energy function $G_0^{(-)}$ by

$$G_0^{(-)ij} = G_0^{(+)*ij} = -G_0^{(+)ji}. \tag{A3}$$

It is convenient to introduce a special symbol for the projection of the positive energy function into the physical subspace (this is necessary only if an invariance group is present), which we define by

$$i\mathfrak{G}_0^{(+)ij} = u_A^i u_A^{*j}. \tag{A4}$$

We note that, in the absence of an invariance group, $G_0^{(+)} = \mathfrak{G}_0^{(+)}$.

In quantum theory, a dominant role is played by the Feynman propagator, which for zero background fields is defined by

$$G_0^{ij} = \begin{cases} -G_0^{(+)ij} & i \text{ to the future of } j \\ +G_0^{(-)ij} & j \text{ to the future of } i \end{cases} = G_0^{\pm ij} \mp G_0^{(\pm)ij} = G_0^{ji}. \tag{A5}$$

Just as we defined $X^{\pm}$, we have

$$X = (1 - UG_0)^{-1}U = (1 + XG_0)U, \tag{A6}$$

with $X_{ij} = X_{ji}$.

The values for the retarded and advanced Green's functions and for the Feynman propagator for a non-zero background field are given in terms of their values for zero background field by

$$G = G_0(1 - UG_0)^{-1} = G_0 + G_0XG_0,$$
$$G^{\pm} = G_0^{\pm}(1 - UG_0^{\pm})^{-1} = G_0^{\pm} + G_0^{\pm}X^{\pm}G_0^{\pm}. \tag{A7}$$

These quantities satisfy a number of identities which may be obtained by straightforward manipulation of the defining equations. We list a few of them here:

$$X = (1 \pm X^{\pm}G_0^{(\pm)})^{-1}X^{\pm}, \tag{A8a}$$
$$(1 \pm X^{\pm}G_0^{(\pm)})^{-1} = 1 \mp XG_0^{(\pm)}, \tag{A8b}$$
$$1 + XG_0 = (1 \mp XG_0^{(\pm)})(1 + X^{\pm}G_0^{\pm}), \tag{A8c}$$
$$\tilde{G} = G^{+} - G^{-} = (1 + G_0^{\pm}X^{\pm})\tilde{G}_0(1 + X^{\pm}G_0^{\pm}), \tag{A8d}$$

if $G = G^{\pm} \mp G^{(\pm)}$, then

$$G^{(\pm)} = (1 + G_0^{\pm}X^{\pm})G_0^{(\pm)}(1 + XG_0) = (1 + G_0X)G_0^{(\pm)}(1 + X^{\pm}G_0^{(\pm)}), \tag{A8e}$$
$$X^{+} - X^{-} = U\tilde{G}U = X^{+}\tilde{G}_0X^{-} = X^{-}\tilde{G}_0X^{+}, \tag{A8f}$$
$$X - X^{*} = X(G_0^{(-)} - G_0^{(+)})X^{*}, \tag{A8g}$$
$$X^{*} = [1 + X(G_0^{(-)} - G_0^{(+)})]^{-1}X, \tag{A8h}$$
$$(1 + G_0^{\pm}X^{\pm})(1 + G_0^{\mp}X^{\mp})^{-1} = 1 + (1 + G_0^{\pm}X^{\pm})(G_0^{\pm} - G_0^{\mp})U = 1 \pm \tilde{G}(1 + X^{\mp}G_0^{\mp})^{-1}U, \tag{A8i}$$
$$G^{(+)ij} = G^{(-)ij}. \tag{A8j}$$

In the theory of the S matrix, the function G plays the role of propagator of field quanta. If an invariance group is present, this function propagates non-physical as well as physical quanta. It is useful to introduce alternative functions which propagate real quanta only. They are defined by

$$\mathfrak{G}_{0\pm}^{ij} = G_0^{\pm ij} \mp \mathfrak{G}_0^{(\pm)ij}. \tag{A9}$$

These quantities satisfy a list of identities similar to those of Eqs. (A8). They are

$$\mathfrak{G}_{0\pm}^{ij} = \mathfrak{G}_{0\mp}^{ji}, \tag{A10a}$$
$$\mathfrak{X}_{\pm} = (1 - U\mathfrak{G}_{0\pm})^{-1}U = (1 \pm X^{\pm}\mathfrak{G}_0^{(\pm)})^{-1}X^{\pm}, \tag{A10b}$$
$$\mathfrak{X}_{+ij} = \mathfrak{X}_{-ji}, \tag{A10c}$$
$$(1 \pm X^{\pm}\mathfrak{G}_0^{(\pm)})^{-1} = 1 \mp \mathfrak{X}_{\pm}\mathfrak{G}_0^{(\pm)}, \tag{A10d}$$
$$1 + \mathfrak{X}_{\pm}\mathfrak{G}_{0\pm} = (1 \mp \mathfrak{X}_{\pm}\mathfrak{G}_0^{(\pm)})(1 + X^{\pm}G_0^{\pm}), \tag{A10e}$$
$$\mathfrak{G}_{\pm} = \mathfrak{G}_{0\pm}(1 - U\mathfrak{G}_{0\pm})^{-1} = G^{\pm} \mp \mathfrak{G}^{(\pm)}, \tag{A10f}$$

whence

$$\mathfrak{G}^{(\pm)} = (1 + G_0^{\pm}X^{\pm})\mathfrak{G}_0^{(\pm)}(1 + \mathfrak{X}_{\pm}\mathfrak{G}_{0\pm}) = (1 + \mathfrak{G}_{0\pm}\mathfrak{X}_{\pm})\mathfrak{G}_0^{(\pm)}(1 + X^{\pm}G_0^{\pm}).$$

APPENDIX B

An explicit calculation of the coefficient C_A and brief sketch of the calculations of C_{AB} and C_{ABC} are given here.

From Eq. (3.5) we deduce

$$0 = \beta_A C_0 + (A + a)_{AB} C_B + E_{ABC} C_{BC} + I_{ABCD} C_{BCD} + \cdots, \quad \text{(B1a)}$$

$$0 = \beta_A C_B + (A + a)_{AC} C_{CB} + E_{ABC} C_{CDB} + I_{ACDE} C_{CDEB} + (B + b)_{AB} C_0 + D_{ABC} C_C + H_{ABCD} C_{CD} + \cdots, \quad \text{(B1b)}$$

$$0 = \beta_A C_{BC} + (A + a)_{AD} C_{DBC} + E_{ADE} C_{DEBC} + I_{ADEF} C_{DEFBC} + (B + b)_{AB} C_C + (B + b)_{AC} C_B + D_{ABD} C_{DC} + D_{ACD} C_{DB} + H_{ABDE} C_{DEB} + (F_{ABC} + F_{ACB}) C_0 + (G_{ABCD} + G_{ACBD}) C_0 + \cdots, \quad \text{(B1c)}$$

etc.

Then to first order from Eq. (B1a) we have

$$C_A \simeq (\beta_B - E_{BCD}(A^{-1}B)_{CD}) C_0, \quad \text{(B2a)}$$

and from (B1c)

$$C_{ABC} \simeq -(A^{-1}B)_{AB} C_C - (A^{-1}B)_{AC} C_B + A^{-1}_{AD}(\beta_D - E_{DEF}(A^{-1}B)_{EF}) C_0 (A^{-1}B)_{BC} + A^{-1}_{AD}[2F_{DBC} - D_{DBE}(A^{-1}B)_{EC} - D_{DCE}(A^{-1}B)_{EB} + 2E_{DEF}(A^{-1}B)_{BE}(A^{-1}B)_{CF}] \quad \text{(B2b)}$$

and nothing new to this order for C_{AB}. Using Eq. (A8a), we find that

$$\begin{aligned} A^{-1} &= (1 + iu^\dagger X^- u)^{-1} \\ &= 1 - iu^\dagger X^- u - iu^\dagger X^- G_0^{(+)} X^- u \\ &\quad - iu^\dagger X^- G_0^{(+)} X^- u - \cdots \\ &= 1 - iu^\dagger X^- (1 - G_0^{(+)} X^-)^{-1} u \\ &= 1 - iu^\dagger X^* u. \end{aligned} \quad \text{(B3)}$$

Therefore,

$$\begin{aligned} A^{-1}_{AB} f_B^{+*j} &= (\delta_{AB} - iu^*_{pA} X^*_{pr} u^r_B) u_B^{*k} (1 + X^- G_0^-)^j_k \\ &= u_A^{*p} (1 + X^* G_0^*)^j_p = f_A^{*j}, \end{aligned} \quad \text{(B4)}$$

where Eq. (A8c) has been used in obtaining the second line from the first.

Substituting the explicit forms given in Eq. (3.3) into (B2a) and using Eqs. (B4) and (A8c), (A8e), we obtain for C_A

$$\begin{aligned} C_A &\simeq -\tfrac{1}{2} i A^{-1}_{AB} f_B^{+*j} S_{,jkl} (1 + G_0^- X^-)^k_m \\ &\quad \times (iG_0^{(+)mn} + iG_0^{(+)mp} X^*_{pr} G_0^{(+)nr})(1 + X^+ G_0^+)^l_n C_0 \\ &= \tfrac{1}{2} f_A^{*j} S_{,jkl} (1 + G_0^- X^-)^k_m G_0^{(+)mp} \\ &\quad \times (1 - X^* G_0^{(-)})^n_p (1 + X^+ G_0^+)^l_n C_0 \\ &= \tfrac{1}{2} f_A^{*j} S_{,jkl} G^{(-)*kl} C_0. \end{aligned} \quad \text{(B5)}$$

In calculating C_{ABC}, we make the same kind of substitutions and go through the same manipulations. This is only necessary for the last term of (B2b) (the one in square brackets), since the other one that is not already calculated is easily recognized to be equal to $-(A^{-1}B)_{BC} C_A$.

The calculation of C_{AB} is similar. We give here the equation which determines it to second order and the first and last steps of the calculations:

Vanishing of the coefficient of $|B, \infty\rangle$ in Eq. (3.5) implies that

$$0 = \beta_A C_B + (A + a)_{AC} C_{CB} + (B + b)_{AB} C_0 + E_{ACD} C_{CDB} + I_{ACDE} C_{CDEB} + D_{ABC} C_C + H_{ABCD} C_{DC} + \cdots, \quad \text{(B6)}$$

so that

$$\begin{aligned} C_{AB} \simeq\ & -(A^{-1}B)_{AB} C_0 \\ & - A^{-1}_{AD}(\beta_D C_B + D_{DBC} C_C + E_{DCE} C_{CEB}) \\ & - A^{-1}_{AD} \Big(b_{DB} - a_{DC}(A^{-1}B)_{CB} \\ & - H_{DBCE}(A^{-1}B)_{CE} + \tfrac{1}{2} I_{DCEF} \\ & \times \sum_{\substack{\text{perm} \\ CEF}} (A^{-1}B)_{CE} (A^{-1}B)_{FB} \Big) C_0 \\ =\ & -(A^{-1}B)_{AB} C_0 + C_A C_B C_0^{-1} \\ & + \tfrac{1}{2} f_A^{*j} S_{,jkl} G^{(-)*ki} S_{,imn} \\ & \times [G^{(-)*mn} f_B^{*l} + G^{(-)*lm} f_B^{*n}] C_0 \\ & + f_A^{*j} [(2!)^{-1} S_{,jkl} G^{-ki} S_{,imn} + (3!)^{-1} S_{,jlmn}] \\ & \times [G^{(-)*mn} f_B^{*l} + (G^{(-)*lm} + G^{(-)*ml}) f_B^{*n}] C_0. \end{aligned} \quad \text{(B7)}$$

APPENDIX C

We present here a detailed calculation of the single-quantum production amplitude to lowest order.

By Eqs. (3.2) and (3.4), we have for single-quantum production

$$\begin{aligned} &\langle A, +\infty \,|\, 0, -\infty \rangle \\ &\quad = \langle 0, +\infty | \, \alpha_A^+ \, |0, -\infty\rangle \\ &\quad = \beta_A C_0^* + B_{AB} C_B^* + (2!)^{-1} F_{ABC} (C_{BC}^* + C_{CB}^*) \\ &\qquad + (3!)^{-1} J_{ABCD} \sum_{\text{perm}} C_{BCD}^* + \cdots. \end{aligned} \quad \text{(C1)}$$

Using the explicit expressions for the various coefficients, as given by (3.3), (3.10), and (3.12), and

Eqs. (A8b), (A8c), and (A8e), we obtain, in first order,

$$
\begin{aligned}
&\langle A, +\infty \,|\, 0, -\infty\rangle \\
&\simeq \beta_A C_0^* + B_{AB} C_B^* + F_{ABC}(A^{-1}B)^*_{BC} C_0^* \\
&= C_0^*[-\tfrac{1}{2} f_A^{+*j} S_{,jkl}(1 + G_0^- X^-)_m^k G_0^{(+)mn}(1 + X^+ G_0^+)_n^l \\
&\quad - \tfrac{1}{2} u_A^{*p} X_{pr}^- G_0^{(+)mr}(1 + XG_0)_m^j S_{,jkl} G^{(-)kl} \\
&\quad + \tfrac{1}{2} f_A^{+*j} S_{,jkl}(1 + G_0^- X^-)_m^k \\
&\quad \times G_0^{(+)pm} X_{pr} G_0^{(+)rn}(1 + X^+ G_0^+)_n^l] \\
&= -\tfrac{1}{2} C_0^* u_A^{*p}[(1 + X^- G_0^-)_p^j S_{,jkl}(1 + G_0^- X^-)_m^k \\
&\quad \times (1 + G_0^{(-)} X)_r^m G_0^{(+)rn}(1 + X^+ G_0^+)_n^l \\
&\quad + X_{pr}^- G_0^{(-)rm}(1 + XG_0)_m^j S_{,jkl} G^{(+)kl}] \\
&= -\tfrac{1}{2} C_0^* u_A^{*p}[(1 + X^- G_0^-)_p^j \\
&\quad + X_{pr}^- G_0^{(-)rm}(1 + XG_0)_m^j] S_{,jkl} G^{(+)kl} \\
&= -\tfrac{1}{2} C_0^* u_A^{*p}(1 + XG_0)_p^j S_{,jkl} G^{(+)kl} \\
&= -\tfrac{1}{2} C_0^* \bar{f}_A^j S_{,jkl} G^{(+)kl}. \qquad \text{(C2)}
\end{aligned}
$$

* Present Address: National Bureau of Standards, Washington, D.C. 20234.

† National Research Council Postdoctoral Research Associate, 1968–1970.

[1] R. P. Feynman, Acta Phys. Polon. **24**, 697 (1963).

[2] C. N. Yang and D. Feldman, Phys. Rev. **79**, 972 (1950).

[3] B. S. DeWitt, Phys. Rev. **162**, 1195, 1239 (1967).

[4] L. D. Faddeev and V. N. Popov, Phys. Letters **25B**, 29 (1967).

[5] S. Mandelstam, Phys. Rev. **175**, 1580, 1604 (1968).

[6] B. S. DeWitt, *Dynamical Theory of Groups and Fields* (Gordon and Breach, New York, 1965).

[7] E. F. Borie, thesis, University of North Carolina, 1968 (unpublished).

[8] By zero order, we mean the absence of any bare vertex functions. A term is said to be first order if it contains exactly one vertex function $S_{,ijk}$, second order if it contains two such functions or one vertex function $S_{,ijkl}$, and so on.

[9] Compare this result

$$(e^{-2\,\mathrm{Im}\,W} = |C_0|^2 = \exp\{-\tfrac{1}{2}\log\det[1 - (A^{-1}B)^*(A^{-1}B)]\})$$

with Eq. (16.20) of Ref. 3, noting that $A^{-1}B = iu\dagger X^* u^* = i\Lambda\dagger$.

[10] E. Borie and A. Cadez (to be published).