JOURNAL OF
MATHEMATICAL
PHYSICS

VOLUME 4, NUMBER 6 JUNE 1963

Propagator Ghosts and Regge Poles

H. M. FRIED
Courant Institute of Mathematical Sciences, New York University, New York, New York
AND
J. G. TAYLOR
Institute for Advanced Study, Princeton, New Jersey
(Received 30 November 1962)

In the approximation of considering only vertex contributions to the propagator, we determine sufficient conditions on the two-particle scattering amplitude which does not generate ghosts in the propagator. The Regge type of behavior does not satisfy these conditions, and will give rise to ghosts unless cancellations occur coming from the higher inelastic contributions to the vertex-function unitarity condition, which have been neglected in our calculation. The double-pole behavior discussed by Frye is found to be the only two-particle scattering-amplitude behavior which will have no propagator ghosts, under certain general conditions.

ONE of the difficulties common to all existing field-theory approximation procedures is the occurrence of propagator ghosts.[1] Symanzik[2] has recently given a discussion of this point, and his conclusions may be rephrased as a set of necessary conditions on the asymptotic behavior of each of the absorptive parts entering into the propagator unitarity sum, in order that no propagator ghosts occur; these conditions essentially require the vanishing of the | vertex function $|^2$ and each of the | decay amplitudes $|^2$ at infinite energy. This situation does not occur in each order of perturbation theory, nor in a large class of nonperturbative approximation schemes described recently by one of us,[3] nor in a simple generalization of axiomatic perturbation theory described recently by the other.[4] Thus it is interesting to determine at least some sufficiency conditions in order that an approximation method in field theory does not generate propagator ghosts, and for this we require some mechanism that suitably damps out the vertex function and decay amplitudes at high energies. We consider only the vertex contribution to the propagator here.

Because the vertex function is related to the S-wave scattering amplitude through an Omnès solution to the vertex unitarity equation, it is a natural question to ask if the Regge-pole hypothesis provides such a mechanism for achieving the damping necessary at high energies. We think it sufficiently interesting to remark that, for the simplest type of Regge behavior of the scattering amplitude, this is almost but not quite the case, and one must require cancellations involving the higher inelastic contributions to the vertex unitarity relation. However, the more general behavior recently discussed by Frye,[5] in which the important asymptotic behavior of the scattering amplitude is controlled by multiple poles or cuts in the complex angular-

[1] G. Feldman, Proc. Roy. Soc. (London) **A223** 112, (1954); P. Redmond [Phys. Rev. **112**, 1404 (1958)] has described an *ad hoc* method for removing these ghosts from the perturbation approximations.

[2] K. Symanzik, J. Math. Phys. **1**, 249 (1960). See also H. Lehmann, K. Symanzik, and W. Zimmermann, Nuovo Cimento **2**, 425 (1955).

[3] J. G. Taylor, "On the Field Equations III. Non-Perturbative Approximation Method," (I. A. S. preprint).

[4] H. M. Fried, "Axiomatic Perturbation Theory II. Partial Sum Formalism," (C. I. M. S. preprint).

[5] G. Frye, Phys. Rev. Letters **8**, 494 (1962). Our scattering amplitude is normalized to be half as large as Frye's.

momentum plane, does provide such a mechanism for a range of the Regge parameters $\alpha'(0)$, $\beta(0)$.

The vertex contribution to the propagator of a scalar, self-interacting boson field takes the form

$$\Delta_F'(s) = \frac{1}{m^2 - s} + \int_{4m^2}^{\infty} \frac{ds' \Sigma(s') \, |\Gamma(s')|^2}{(s' - s)(s' - m^2)^2}, \quad (1)$$

where $\Sigma(s)$ is a real, positive, two-particle phase space factor, and $\Gamma(s)$ represents the vertex function (more precisely, Γ is a form factor: vertex function times Klein–Gordon amputated propagator). In a renormalizable theory, Σ typically has the asymptotic form $\Sigma \sim s$ for large s, and the ghost situation arises when $\Gamma(s)$ is approximated by any method for which[6]

$$\int^{\infty} \frac{ds}{s} \, |\Gamma(s)|^2 < \infty \quad (2)$$

is not satisfied.

The vertex unitarity equation reads[7]

$$\Gamma(s) - \Gamma^*(s) = i[\Gamma(s) A_0^*(s) + \Gamma^*(s) A_0(s) + H(s)], \quad (3)$$

where $A_0(s)$ is the S-wave scattering amplitude and $H(s)$ denotes the real functions obtained by summing over all the contributing inelastic and decay amplitudes. Equation (3) and the statement that $\Gamma(s)$ is analytic in the cut plane (with no essential singularities at infinity) imply the Omnès solution[8]

$$\Gamma(s) = \exp\,[u(s) - u(m^2)] \left\{ g + \frac{(s - m^2)}{2\pi} \times \int_{9m^2}^{\infty} \frac{ds' H(s')}{(s' - s - i\epsilon)(s' - m^2)} \cdot \frac{e^{-[\Delta(s') - \Delta(m^2)]}}{|1 + iA_0(s')|} \right\}, \quad (4)$$

which satisfies the boundary condition $\Gamma(m^2) = g$; here $u(s) = 1/\pi \int_{4m^2}^{\infty} ds' \, \delta(s') \times (s' - s - i\epsilon)^{-1} = \Delta(s) + i\delta(s)$, and $2i\,\delta(s) = \ln\,[1 + iA_0(s)/1 - iA_0^*(s)]$. In order to discuss the asymptotic form of $\Gamma(s)$ it is clearly necessary to make an assumption about $H(s)$; but the form of (4) suggests that the important asymptotic behavior of $\Gamma(s)$ is governed by the factor $\exp\,[u(s) - u(m^2)]$, and hence by the S-wave scattering. With reasonable assumptions about the form of $H(s)$, this would be the situation for certain classes of high-energy scattering behavior; e.g., if the double-pole model discussed below is correct, the presence of $H(s)$ is not pertinent to this discussion if, asymptotically, $H(s) \sim s^{-\gamma}$, $\gamma > 0$, or if $H(s) \sim (\ln s)^{-\gamma}$, $\gamma > \frac{3}{2}$; etc.

[6] Equation (2) implies that the wavefunction renormalization constant is positive.

[7] We use the same normalization as M. Goldberger, *Introduction to the Theory and Application of Dispersion Relations*, edited by DeWitt and Omnès, (Hermann & Cie., Paris, 1960).

[8] R. Omnès, Nuovo Cimento **8**, 316 (1958).

The S-wave scattering amplitude obeys the unitarity equation

$$A_0(s) - A_0^*(s) = 2i(|A_0(s)|^2 + |F(s)|^2), \quad (5)$$

where $|F(s)|^2$ denotes the real, positive contribution of all the inelastic production processes. With $A_0(s) = \rho(s) e^{i\zeta(s)}$, ρ and ζ real, (5) may be rewritten in the form

$$\rho = \tfrac{1}{2} (\sin \zeta)\, \xi_*, \qquad \xi_* = 1 \pm (1 - |G|^2)^{\frac{1}{2}},$$
$$|F|^2 = \tfrac{1}{4} |G|^2 \sin^2 \zeta, \quad (6)$$

which makes explicit the limitations on the strength of the inelastic terms, i.e., $|G|^2 \leq 1$. From (6) it is easy to see that

$$\sigma_{\text{tot}}^{(0)}(s)/\sigma_{\text{el}}^{(0)}(s) = 2/\xi_*(s), \quad (7)$$

where the superscripts (0) denote the S-wave cross sections. Below the inelastic threshold, at $s = 9m^2$ in the simple scalar theory, this ratio must be unity, indicating that we are on the "ξ_+ branch," with $|G|^2 = 0$. As the threshold is passed, the scattering amplitude cannot change discontinuously, which implies that we are still on the ξ_+ branch with $|G|^2$ increasing from zero. If $|G|^2$ increases to its maximum value at some point s_0, $\xi_+(s_0) = 1$, and we have the "diffraction limit," $\sigma_{\text{tot}}^{(0)}(s_0)/\sigma_{\text{el}}^{(0)}(s_0) = 2$. At this point it is possible for the scattering amplitude to pass smoothly to the ξ_- branch; and in fact this must be the situation if the potential theory Regge behavior occurs, since the ratio of (7) then diverges logarithmically at high energy.

Only the real part of the exponential factor of (4) is pertinent to the convergence of (2). For large s this becomes[9]

[9] This statement requires some justification. A discussion of such limiting behavior has been given by G. Frye and R. Warnock ["Analysis of Partial Wave Dispersion Relations," University of Washington preprint.] We give here a simple treatment for a suitable class of functions $\delta(s)$. To within a constant, the exponential factor of (4) may be written as $(1/\pi) I(s)$, with

$$I(s) = s \int_{4m^2}^{\infty} \frac{ds'}{(s' - s)} \frac{\delta(s')}{s'}.$$

We assume that

$$\int_{4m^2}^{\infty} ds' \, \delta(s')/s'$$

is divergent, since otherwise no damping is obtained. We consider the functions $\delta(s)$ which are bounded with bounded derivatives for $s > 8m^2$, and denote the bounds of $\delta(s)$ and $\delta'(s)$ by δ and δ_1; this situation arises when only a finite number of elastic resonances occur. The higher derivatives of $\delta(s)$ may be unbounded at higher particle thresholds, as has been discussed by A. J. Dragt and R. Karplus ["Analyticity and Unitarity of General Transition Thresholds," Berkeley preprint.] The singularities occuring in $\delta(s)$ at the threshold for the production of n particles were shown there to have the form $[s - (nm)^2]^{3n-5/2}$ if n is even and $[s - (nm)^2]^{3n-5/2} \times \ln\,(n^2m^2 - s)$ if n is odd, and these do not cause the derivative

$$\Delta(s) - \Delta(m^2) \sim -\frac{1}{\pi}\int^s \frac{ds'}{s'}\,\delta(s'),$$
$$\delta(s) = \tan^{-1}\left[\frac{\sin\zeta(s)\cos\zeta(s)}{2/\xi_+(s) - \sin^2\zeta(s)}\right]. \tag{8}$$

It is interesting to note that the arctan can jump to a different branch only if, as s' passes through s_r, $\zeta(s_r) = \frac{1}{2}\pi$ and $\xi_+(s_r) \to \xi_+(s_r) = 2$. At such a point, the inelastic cross section must vanish and we would have a pure elastic resonance. If this actually occurred we would pick up a term

$$-(1/\pi)\int^s ds'/s' \times \tfrac{1}{2}\pi$$

in (8), and gain a convergence factor $s^{-\frac{1}{2}}$ in the asymptotic expression for $\Gamma(s)$. However, because the inelastic amplitudes must be strictly zero at s_r this method of obtaining convergence is highly doubtful. Further, such an elastic resonance would have the same quantum numbers as one of the three particles defining the vertex, and such resonances are not seen experimentally.

It is clear that what is desired is an asymptotic behavior of ζ and $\sigma_{\text{tot}}^{(0)}/\sigma_{\text{el}}^{(0)}$, such that the integral of (8) diverges, with a positive sign, for large s. These quantities depend critically on the model behavior of the scattering amplitude at high energies. For a Regge-type amplitude, where $\alpha(t)$ denotes the vacuum trajectory, $\alpha(0) = 1$, one expects the dominant behavior

$$A(s, t) \sim \beta(t)s^{\alpha(t)}[\ln s]^{p-1}\{i + \operatorname{ctn}\tfrac{1}{2}[\pi\alpha(t)]\}, \tag{9}$$

where $p = 1$ corresponds to a first-order pole in the scattering amplitude considered as a function of complex angular momentum; $p = 2$ corresponds to a double pole; etc. Knowledge of the asymptotic phase ζ hinges on the ratio of the S-wave projections of both the real and imaginary parts of (9), and these have been related by a symmetrized dispersion relation.[10] In Frye's model, where

$$\operatorname{Im} A(s, t) \sim \beta(t)e^{\alpha(t)\phi(s)},$$

the desired phase relation is obscure (although symmetrization may be imposed; see below); however, it turns out that all the essential properties of Frye's model are simply reproduced by the double-pole model (e.g., one finds $\sigma_{\text{tot}} \sim \sigma_{\text{el}} \sim \ln s$).

The S-wave projections of (9) have the form[11]

$$\operatorname{Re} A_0(s) \sim \tfrac{1}{2}\pi[\beta(0)/\alpha'(0)]\,[\ln s]^{p-3},$$
$$\operatorname{Im} A_0(s) \sim [\beta(0)/\alpha'(0)]\,[\ln s]^{p-2},$$

and hence

$$\tan\zeta \sim (2/\pi)\ln s, \qquad \zeta \to \tfrac{1}{2}\pi. \tag{10}$$

The cross-section ratio is sensitive to p with the dependence

$$\sigma_{\text{tot}}^{(0)}/\sigma_{\text{el}}^{(0)} \sim [\alpha'(0)/\beta(0)]\,[\ln s]^{2-p}. \tag{11}$$

For the potential-theory Regge case, $p = 1$, we obtain $\delta(s) \sim (\ln s)^{-2}$, and the integral of (8) converges. However, for the double-pole model we have $\sigma_{\text{tot}}^{(0)}/\sigma_{\text{el}}^{(0)} \sim \alpha'(0)/\beta(0) = R$, the integral of (8) diverges according as

$$\Delta(s) - \Delta(m^2) \sim -\tfrac{1}{2}[1/(R-1)]\ln\ln s,$$

and a factor

$$[\ln s]^{-1/2(R-1)}$$

is introduced into the asymptotic form of Γ. We have already described the special case for which $R = 1$, since this corresponds to an elastic resonance at infinity.

The integral of (2) will now converge provided that $R < 2$. It is interesting to note that the value $R = 2$ is just the ratio which defines Frye's model, where, for the sum of all partial waves, $\sigma_{\text{tot}}/\sigma_{\text{el}} \sim 4$. (In comparison, the double-pole model gives $\sigma_{\text{tot}}/\sigma_{\text{el}} \sim 2R$.) The physical difference in interpretation here, for the situation $2 > R > 1$, is that each inelastic wave does not contribute its maximum possible value.

More generally, we may discuss the desired behavior of ζ and ξ_+. The required damping of Γ will be obtained if $\delta(s)$ converges to a positive constant as

---

$\delta'(s)$ to be unbounded for large s. It is convenient to write $I(s)$ in the form

$$I(s) = -\int_{4m^2}^{\frac{1}{2}s}\frac{ds'}{s'}\,\delta(s') + \int_{4m^2}^{\frac{1}{2}s}\frac{ds'}{s'-s}\,\delta(s') + s\int_{\frac{1}{2}s}^{\infty}\frac{ds'}{(s'-s)}\,\frac{\delta(s')}{s'} = I_1 + I_2 + I_3,$$

where we assume that I_1 is divergent as $s \to \infty$, and will show that I_2 and I_3 are bounded functions of s. This is immediate for I_2, since $|I_2| < \delta[1 - (8m^2/s)]$ for all $s > 8m^2$. If the integration variable in I_3 is changed to $x = s'/s$, the range of integration may be split into the two intervals $(\frac{1}{2}, \frac{3}{2})$ and $(\frac{3}{2}, \infty)$. The latter range gives a contribution bounded by

$$\delta\int_{\frac{3}{2}}^{\infty}\frac{dx}{(x-1)x};$$

the former gives a contribution which may be written in the form

$$\int_{\frac{1}{2}}^{\frac{3}{2}}\frac{[\delta(xs) - \delta(s)]}{xs(s-1)}\,dx + \delta(s)\int_{\frac{1}{2}}^{\frac{3}{2}}\frac{dx}{x(x-1)}.$$

The first of these terms is bounded by δ_1, and the second by a finite constant times δ, for sufficiently large s; hence I_3 is also bounded.

[10] S. Frautschi, M. Gell-Mann, and F. Zachariasen, Phys. Rev. **126**, 2204, (1962).

[11] We use the technique of integration by parts to define the asymptotic expansion of these projections, as in reference 5.

$s \to \infty$, while a ghost will occur if it converges to a negative constant for large s. If $\delta(s)$ vanishes for large s, and also

$$\int^{\infty} \frac{ds}{s} \exp\left[-\frac{2}{\pi}\int^{s} \frac{ds'}{s'}\,\delta(s')\right] < \infty\,,$$

then asymptotically the function

$$\gamma(s) = \exp\left[\frac{2}{\pi}\int^{s} ds'\,\delta(s')/s'\right]$$

must be of the form $\prod_{r=1}^{n} (\ln_r s)^{\alpha_r}$, where

$\alpha_1, \cdots \alpha_{n-1} = 1, \quad \alpha_n > 1;$ and $\ln_r s \equiv \ln(\ln_{r-1} s)$.

For large s, this implies

$$\delta(s) \sim \left(\frac{\alpha_1}{\ln s} + \frac{\alpha_2}{\ln s \ln_2 s} + \cdots\right), \qquad (12)$$

with $\alpha_1 \geq 1$. This is the most general form of $\delta(s)$ in the class of functions with expansions in formal multiple power series in $(\ln_r s)^{-1}$. Thus (12) is a sufficient condition, in a general form, on any approximation scheme which is to avoid propagator ghosts. If, following Frye, we may assume that Im $A(s, t) \sim \beta(t) \exp[\alpha(t)\phi(s)]$ for large s, then we find, by manipulations similar to those used in reaching (10) and (11), that the only form of $\phi(s)$ consistent with (12) is $\phi(s) = \ln s + \ln_2 s$; this is Frye's model with, however, $R < 2$.

It may well be that the propagator ghosts which are generated by the Regge single-pole behavior of the scattering amplitude will be removed when the inelastic contributions of (4) are included. The present discussion indicates that, with the neglect of these terms, the Regge single-pole behavior is not consistent with a local, Lorentz-invariant field theory with positive definite metric; the double-Regge-pole behavior is. This situation can only be made worse if inelastic contributions are included on the right-hand side of (1), since these extra positive terms may introduce new propagator ghosts but can never cancel any portion of the vertex contributions.

ACKNOWLEDGMENTS

It is a pleasure to thank Dr. G. Frye for several informative discussions, and to thank Dr. P. DeCelles for an enjoyable conversation relevant to this material.

Divergence Conditions and the Gravitational Field Equations*

HEINZ R. PAGELS†
Institute of Theoretical Physics, Department of Physics, Stanford University, Stanford, California
(Received 11 June 1962)

It is shown that in addition to the four divergence laws $T^{\alpha\beta}{}_{|\beta} = 0$, stress tensors satisfying the Einstein field equations must also satisfy the additional divergence conditions $(T^{\gamma\alpha|\beta} - T^{\gamma\beta|\alpha})_{|\gamma} = 0$. Conversely, under quite general assumptions these ten conditions on the source, $T_{\alpha\beta}$, dictate the metric–matter connection proposed by Einstein.

I. INTRODUCTION

IN the conventional derivation of the gravitational field equations, one begins with the Lagrangian density $L = (-g)^{\frac{1}{2}}(R + 4\lambda - \kappa T)$, and demands that the action integral be stationary with respect to variations in the $g_{\alpha\beta}$. As is well known, one arrives at Einstein's coupling of the metrical field to the matter field

$$R_{\alpha\beta} - \tfrac{1}{2}g_{\alpha\beta}R + \lambda g_{\alpha\beta} = \kappa T_{\alpha\beta}. \tag{1}$$

In virtue of these field equations, the contracted Bianchi identities imply that the divergence laws $T^{\alpha\beta}{}_{|\beta} = 0$ are automatically satisfied, this fact being a prime motivation for the choice of the field equations.

One might then ask: are the divergence laws $T^{\alpha\beta}{}_{|\beta} = 0$ the only conditions on the matter tensor $T_{\alpha\beta}$ implied by the explicit structure of $R_{\alpha\beta}$ and the relation (1)? The answer to this question turns out to be negative. Indeed, we find that in addition to the four equations

$$T^{\alpha\beta}{}_{|\beta} = 0, \tag{2}$$

the matter tensor must also satisfy the six equations

$$(T^{\gamma\alpha|\beta} - T^{\gamma\beta|\alpha})_{|\gamma} = 0. \tag{3}$$

In a flat space these last equations are trivially satisfied because of the conservation laws

$$\partial T^{\gamma}{}_{\alpha}/\partial x^{\gamma} = 0, \tag{4}$$

and the commutative nature of differentiation

$$(T^{\gamma}{}_{\alpha|\beta} - T^{\gamma}{}_{\beta|\alpha})_{|\gamma} = \frac{\partial^2 T^{\gamma}{}_{\alpha}}{\partial x^{\gamma}\,\partial x^{\beta}} - \frac{\partial^2 T^{\gamma}{}_{\beta}}{\partial x^{\gamma}\,\partial x^{\alpha}} = \frac{\partial^2 T^{\gamma}{}_{\alpha}}{\partial x^{\beta}\,\partial x^{\gamma}} - \frac{\partial^2 T^{\gamma}{}_{\beta}}{\partial x^{\alpha}\,\partial x^{\gamma}} = 0. \tag{5}$$

However in a curved space, Eqs. (3) are found to be satisfied as a consequence of the Einstein equations.

In the case of a flat space, we have freedom to perform a linear coordinate transformation without altering the physical content of the equations. As a consequence, we can derive the divergence laws (4), which we write symbolically as $\nabla\cdot\mathbf{T} = 0$, and which express the conservation of energy and momentum. In the more general case of a curved space, the freedom to perform a coordinate transformation reduces the ten components of $T_{\alpha\beta}$ to six independent components.

The second set of Eqs. (3), which we might write symbolically as $\nabla\cdot\nabla\times\mathbf{T} = 0$, imposes an additional six constraints on the tensor $T_{\alpha\beta}$, and as we shall see, serves to identify it as the source of the metrical field. What physical invariance principle is associated with this divergence condition is not clear, for, as remarked previously, these divergence conditions are trivially satisfied in a flat space. It should, however, be emphasized that $\nabla\cdot\nabla\times\mathbf{T} = 0$ is decidedly nontrivial in a space for which $R_{\alpha\gamma\delta\beta} \neq 0$ and their validity is in fact intimately associated with the Einstein field equations. The fact that these additional divergence conditions constitute six equations, and, in the case of flat space, follow from the conservation of energy–momentum, suggests a formal similarity with the conservation of angular momentum.

II. PROOF OF $\nabla\cdot\nabla\times\mathbf{T} = 0$

We shall now establish the following theorem:

Theorem I: If $T_{\alpha\beta}$ is related to the metric by $\kappa T_{\alpha\beta} = R_{\alpha\beta} - \frac{1}{2}g_{\alpha\beta}R + \lambda g_{\alpha\beta}$, where κ and λ are constants, then $T^{\alpha\beta}{}_{|\beta} = 0$ and $(T^{\gamma\alpha|\beta} - T^{\gamma\beta|\alpha})_{|\gamma} = 0$. The conditions $T^{\alpha\beta}{}_{|\beta} = 0$ follow from the contracted Bianchi identities, as is well known. To establish the second identity, it is instructive to inspect the formal properties of the tensor

$$\Lambda_{\alpha\beta\gamma} = T_{\alpha\beta|\gamma} - T_{\alpha\gamma|\beta}, \tag{6}$$

* Supported in part by the U. S. Air Force, through the Air Force Office of Scientific Research, Contract AF 49(638)-388.
† NSF Cooperative Fellow, September 1961–June 1962.

which has the algebraic symmetries

$$\Lambda_{\alpha\beta\gamma} + \Lambda_{\alpha\gamma\beta} = 0, \qquad \Lambda_{\alpha\beta\gamma} + \Lambda_{\beta\gamma\alpha} + \Lambda_{\gamma\alpha\beta} = 0. \tag{7}$$

The vector which we can form by contraction,

$$\Lambda_\beta = \Lambda^\alpha{}_{\beta\alpha} = T_\beta{}^\alpha{}_{|\alpha} - \partial T/\partial x^\beta, \qquad T = T^\alpha{}_\alpha, \tag{8}$$

has the form $\Lambda_\beta = -\partial T/\partial x^\beta$ since $T_\beta{}^\alpha{}_{|\alpha} = 0$ and therefore the curl vanishes:

$$\Lambda_{\gamma|\beta} - \Lambda_{\beta|\gamma} = 0. \tag{9}$$

Next we make use of the fundamental identity

$$T_{\alpha\beta|\gamma|\delta} - T_{\alpha\beta|\delta|\gamma} = T_{\sigma\beta}R^\sigma{}_{\alpha\gamma\delta} + T_{\alpha\sigma}R^\sigma{}_{\beta\gamma\delta}, \tag{10}$$

which expresses the noncommutativity of covariant differentiation. On contracting this expression, we have

$$T_\alpha{}^\beta{}_{|\beta|\delta} - T_\alpha{}^\beta{}_{|\delta|\beta} = -T_\alpha{}^\beta{}_{|\delta|\beta} = T_\sigma{}^\beta R^\sigma{}_{\alpha\beta\delta} + T_{\alpha\sigma}R^\sigma{}_\delta, \tag{11}$$

where we have put $R^\sigma{}_\delta = R^{\sigma\beta}{}_{\beta\sigma}$. From (11) and the definition of $\Lambda_{\gamma\alpha\beta}$, we find

$$\Lambda^\beta{}_{\alpha\delta|\beta} = \Lambda_{\alpha|\delta} - \Lambda_{\delta|\alpha} + T_{\delta\sigma}R^\sigma{}_\alpha - R_{\delta\sigma}T^\sigma{}_\alpha - T^{\sigma\beta}(R_{\sigma\alpha\beta\delta} - R_{\sigma\delta\beta\alpha}). \tag{12}$$

But $T^{\sigma\beta}(R_{\sigma\alpha\beta\delta} - R_{\sigma\delta\beta\alpha}) = T^{\sigma\beta}(R_{\sigma\alpha\beta\delta} + R_{\sigma\delta\alpha\beta}) = T^{\sigma\beta}R_{\sigma\beta\alpha\delta} = 0$ in virtue of the symmetries of $T_{\alpha\beta}$ and $R_{\alpha\beta\gamma\delta}$. Consequently, using (9), (12) becomes

$$\Lambda^\beta{}_{\alpha\delta|\beta} = T_{\delta\sigma}R^\sigma{}_\alpha - R_{\delta\sigma}T^\sigma{}_\alpha, \tag{13}$$

so that $(T^{\beta\alpha|\delta} - T^{\beta\delta|\alpha})_{|\beta} = 0$ follows from the Einstein equations since (1) implies that $R_{\alpha\beta}$ and $T_{\alpha\beta}$ commute. Hence we have Theorem I.

III. THE EINSTEIN EQUATIONS

It is interesting and natural to ask if the ten conditions on the matter tensor $\nabla\cdot\mathbf{T} = 0$ and $\nabla\cdot\nabla\times\mathbf{T} = 0$, imply the Einstein equations (1). We find that the Einstein equations and the ten divergence conditions are not in general coimplicative without the introduction of an additional assumption.

The Assumption of Locality: A relation between a matter tensor $T_{\alpha\beta}$ and a metric tensor $g_{\alpha\beta}$ is called local if $T_{\alpha\beta}$ at a point P can be expressed in terms of the metric tensor in an arbitrarily small neighborhood of P. We now assume that all physically meaningful matter tensors are locally related to the geometry.

In this context one should recall a theorem due to Cartan,[1] which states that if a symmetric tensor $T_{\alpha\beta}(x)$ depends only on the metric up to the second derivative of the metric at the space–time point x, is linear in the second derivative and has vanishing divergence, $\nabla\cdot\mathbf{T} = 0$, then it is related to the metric tensor by the Einstein equations (1) for a general Riemannian geometry. We however shall make no condition on the specific nature of the relation between $T_{\alpha\beta}$ and $g_{\alpha\beta}$, for this is the formal role that the six conditions $\nabla\cdot\nabla\times\mathbf{T} = 0$ play. We shall consequently prove:

Theorem II: *If we demand that in every Riemannian geometry the symmetric stress-energy tensor $T_{\alpha\beta}$ must satisfy the divergence laws $\nabla\cdot\mathbf{T} = 0$, $\nabla\cdot\nabla\times\mathbf{T} = 0$ and the relation between $T_{\alpha\beta}$ and the metric satisfies the assumption of locality, then this relation must be the system of Einstein equations, $\kappa T_{\alpha\beta} = R_{\alpha\beta} - \frac{1}{2}g_{\alpha\beta}R + \lambda g_{\alpha\beta}$, where λ and κ are constants.*

We have seen that the six divergence conditions $\nabla\cdot\nabla\times\mathbf{T} = 0$ and the condition $\nabla\cdot\mathbf{T} = 0$ imply that $T_{\alpha\beta}$ and $R_{\alpha\beta}$ commute, or equivalently,

$$G_\alpha{}^\gamma T_\gamma{}^\beta - T_\alpha{}^\gamma G_\gamma{}^\beta = 0, \tag{14}$$

where we have introduced $G_{\alpha\beta} = R_{\alpha\beta} - \frac{1}{2}g_{\alpha\beta}R$. The $G_\alpha{}^\beta$ satisfy the contracted Bianchi identities

$$G_\alpha{}^\beta{}_{|\beta} = 0. \tag{15}$$

At any given point x, we may assume without loss of generality that $G_\alpha{}^\beta$ has distinct and nonzero eigenvalues, where we mean by eigenvalues those numbers λ which solve the determinental condition, $\det|G_\alpha{}^\beta - \lambda\delta_\alpha{}^\beta| = 0$, with $G_\alpha{}^\beta$ evaluated at x. We now introduce a local Minkowski coordinate system at x so that $G_\alpha{}^\beta$ is diagonal in this coordinate system and has the values g_α along the main diagonal. Then (14) implies that we can write

$$T_\alpha{}^\beta = v^0 g_\alpha{}^\beta + v^1 G_\alpha{}^\beta + v^2 G_\alpha{}^\gamma G_\gamma{}^\beta + v^3 G_\alpha{}^\gamma G_\gamma{}^\delta G_\delta{}^\beta, \tag{16}$$

where the v^i are scalars. This we see from the fact that (14) implies that at x, the coordinate transformation that diagonalizes $G_\alpha{}^\beta$ with values g_α along the main diagonal, also diagonalizes $T_\alpha{}^\beta$ with values t_α along the main diagonal. From Lagrange's interpolation formula we then know that there exists a polynomial $P(y) = v^0 + v^1 y + v^2 y^2 + v^3 y^3$ such that $P(y)$ takes the values t_α at the points $y = g_\alpha$, or $t_\alpha = v^0 + v^1 g_\alpha + v^2(g_\alpha)^2 + v^3(g_\alpha)^3$, which is Eq. (16) in this particular coordinate system. But since (16) is a tensor equation, it is true in all coordinate systems.

We next impose the four divergence conditions $\nabla\cdot\mathbf{T} = 0$ on (16):

[1] E. Cartan, J. Math. 87, 141 (1922).

$$(v^0 g_\alpha{}^\beta + v^1 G_\alpha{}^\beta + v^2 G_\alpha{}^\gamma G_\gamma{}^\beta + v^3 G_\alpha{}^\gamma G_\gamma{}^\delta G_\delta{}^\beta)_{|\beta} = 0, \quad (17)$$

which are four partial differential equations for the four unknown scalars $v^i(x)$. It is not in general true that we must pick $v^2 = v^3 = 0$ and v^0, v^1 constants unless we invoke the assumption of locality, which from (16), we see implies that the $v^i(x)$ depend only on the $g_{\alpha\beta}$ in the neighborhood of x since $T_{\alpha\beta}(x)$ is to depend only on $g_{\alpha\beta}$ in the neighborhood of x.

We now ask what are the solutions $v^i(x)$ of (17) which are consistent with locality? We write (17) as

$$\sum_{i=0}^{3} \left(A_{\alpha i} v^i + B_\alpha{}^\beta{}_i \frac{\partial v^i}{\partial x^\beta} \right) = 0, \quad (18)$$

where $A_{\alpha 0} = A_{\alpha 1} = 0$,

$$A_{\alpha 2} = G_\alpha{}^\gamma{}_{|\beta} G_\gamma{}^\beta, \qquad A_{\alpha 3} = (G_\alpha{}^\gamma G_\gamma{}^\delta)_{|\beta} G_\delta{}^\beta,$$
$$B_\alpha{}^\beta{}_0 = g_\alpha{}^\beta, \qquad B_\alpha{}^\beta{}_1 = G_\alpha{}^\beta, \quad (19)$$
$$B_\alpha{}^\beta{}_2 = G_\alpha{}^\gamma G_\gamma{}^\beta, \qquad B_\alpha{}^\beta{}_3 = G_\alpha{}^\gamma G_\gamma{}^\delta G_\delta{}^\beta,$$

and we have used the fact that $G_\alpha{}^\beta{}_{|\beta} = 0$. The coefficients, $A_{\alpha i}$, $B_\alpha{}^\beta{}_i$, of Eq. (18) are specified once the geometry is given. Equation (18) is a first-order partial differential equation for the $v^i(x)$ so that we may appeal to an initial-value formulation in our inquiry concerning solutions.

We may choose the geometry so that the 4×4 matrix $B_\alpha{}^0{}_i$ is nonsingular, and hence solve (18) for $\partial v^i/\partial x^0$:

$$\frac{\partial v^i}{\partial x^0} = F^i\left(A, B, v^i, \frac{\partial v^i}{\partial x^1}, \frac{\partial v^i}{\partial x^2}, \frac{\partial v^i}{\partial x^3}\right). \quad (20)$$

Now, if on the hypersurface $x_0 = 0$ we specify the initial values $v^i(x_1, x_2, x_3, x_0 = 0)$, then by the initial-value formalism we can construct a solution $v^i(x_1, x_2, x_3, x_0)$. On some other hypersurface defined by $x_0 = a > 0$, this solution takes the values $v^i(x_1, x_2, x_3, x_0 = a)$, and at the space–time point P for which $x_1 = b_1$, $x_2 = b_2$, $x_3 = b_3$, $x_0 = b > a > 0$, the solution takes the value $v^i(b_1, b_2, b_3, b)$.

Now let us slightly change the geometry in the region of space–time R between the hypersurfaces $x_0 = 0$ and $x_0 = a$, by giving the metric a small perturbation, $g_{\alpha\beta} \to g_{\alpha\beta} + \delta g_{\alpha\beta}$. This perturbation is chosen so that it vanishes outside of the region R; in particular, the geometry in the neighborhood of P is unchanged. Let us assume that changing the geometry in R induces a change in the solutions $v^i(x) \to v^i(x) + \delta v^i(x)$ in R. If we now specified the *same* initial values $v^i(x_1, x_2, x_3, x_0 = 0)$ on the hypersurface $x_0 = 0$, the solution will now in general take a new value $\bar{v}^i(x_1, x_2, x_3, x_0 = a)$ on hypersurface $x_0 = a$ since the geometry in R is altered. By the initial-value formalism, however, the value of the solution at P depends on the value of the solution on the hypersurface $x_0 = a$ and the geometry in the space–time region between $x_0 = a$ and $x_0 = b > a$. Since the geometry in this region is assumed not to have changed, $v^i(b_1, b_2, b_3, b)$ will depend only on the value of the solution at $x_0 = a$, and since this has changed, $v^i(b_1, b_2, b_3, b)$ will *in general* be different. This, however, violates the assumption of locality for we see that $v^i(b_1, b_2, b_3, b)$ depends on more than the geometry in the neighborhood of P; in fact it depends on the geometry in R. Consequently, we must have that $\delta v^i(x) = 0$ in R for an arbitrary $\delta g_{\alpha\beta}$ in R. What does this imply about the possible solutions $v^i(x)$ in R?

For a small change in the metric in R there will be a small change in $G_\alpha{}^\beta$ and $G_\alpha{}^\beta{}_{|\gamma}$:

$$g_{\alpha\beta} \to g_{\alpha\beta} + \delta g_{\alpha\beta},$$
$$G_\alpha{}^\beta \to G_\alpha{}^\beta + \delta G_\alpha{}^\beta, \quad (21)$$
$$G_\alpha{}^\beta{}_{|\gamma} \to G_\alpha{}^\beta{}_{|\gamma} + \delta G_\alpha{}^\beta{}_{|\gamma},$$

so that the coefficients $A_{\alpha i}$ and $B_\alpha{}^\beta{}_i$ are also changed;

$$A_{\alpha i} \to A_{\alpha i} + \delta A_{\alpha i},$$
$$B_\alpha{}^\beta{}_i \to B_\alpha{}^\beta{}_i + \delta B_\alpha{}^\beta{}_i. \quad (22)$$

From the differential equation (18), we see that in R,

$$\sum_{i=0}^{3} \left(\delta A_{\alpha i} v^i + \delta B_\alpha{}^\beta{}_i \frac{\partial v^i}{\partial x^\beta} \right) = -\sum_{i=0}^{3} \left(A_{\alpha i}\, \delta v^i + B_\alpha{}^\beta{}_i\, \delta \frac{\partial v^i}{\partial x^\beta} \right). \quad (23)$$

We have previously seen that the assumption of locality requires that, in general, we must have $\delta v^i = 0$ in R, and hence also $\delta(\partial v^i/\partial x^\beta) = 0$ in R. Applying this result to Eq. (23) we have

$$\sum_{i=0}^{3} \left(\delta A_{\alpha i} v^i + \delta B_\alpha{}^\beta{}_i \frac{\partial v^i}{\partial x^\beta} \right) = 0. \quad (24)$$

Now consider any space–time point W in R, and chose the variation in the metric so that at the point W

$$\delta g_{\alpha\beta} = 0, \qquad \delta(\partial g_{\alpha\beta}/\partial x^\gamma) = 0,$$
$$\delta(\partial^2 g_{\alpha\beta}/\partial x^\gamma\, \partial x^\sigma) = 0, \quad (25)$$

and hence $\delta G_\alpha{}^\beta = 0$ since $\delta G_\alpha{}^\beta$ is formed only from the expressions (25). From the expressions (19) for the coefficients $A_{\alpha i}$ and $B_\alpha{}^\beta{}_i$ we have that at W

$$\delta B_{\alpha}{}^{\beta}{}_{i} = 0, \qquad \delta A_{\alpha 0} = 0, \qquad \delta A_{\alpha 1} = 0,$$
$$\delta A_{\alpha 2} = G_{\gamma}{}^{\beta}\,\delta G_{\alpha}{}^{\gamma}{}_{|\beta},$$
$$\delta A_{\alpha 3} = G_{\delta}{}^{\beta}\,\delta(G_{\alpha}{}^{\gamma}G_{\gamma}{}^{\delta})_{|\beta}$$
$$= G_{\delta}{}^{\beta}G_{\gamma}{}^{\delta}\,\delta G_{\alpha}{}^{\gamma}{}_{|\beta} + G_{\delta}{}^{\beta}G_{\alpha}{}^{\gamma}\,\delta G_{\gamma}{}^{\delta}{}_{|\beta}. \tag{26}$$

For convenience, we shall introduce Riemannian coordinates in R and set the origin of these coordinates at W. It is a well-known property of Riemannian coordinates that at the origin of coordinates, the Christoffel symbols all vanish, $\Gamma^{\delta}_{\alpha\beta} = 0$, and in virtue of (25), $\delta\Gamma^{\delta}_{\alpha\beta} = 0$. Moreover, the covariant derivative has the same value as the ordinary derivative at the origin so that one finds for $\delta G_{\alpha}{}^{\gamma}{}_{|\beta}$ at W explicitly,

$$\delta G_{\alpha}{}^{\gamma}{}_{|\beta} = \frac{g^{\gamma\delta}g^{\sigma\tau}}{2}\,\delta\Big(\frac{\partial^3 g_{\alpha\tau}}{\partial x^{\delta}\,\partial x^{\sigma}\,\partial x^{\beta}} + \frac{\partial^3 g_{\sigma\beta}}{\partial x^{\delta}\,\partial x^{\alpha}\,\partial x^{\tau}}$$
$$- \frac{\partial^3 g_{\alpha\beta}}{\partial x^{\delta}\,\partial x^{\sigma}\,\partial x^{\tau}} - \frac{\partial^3 g_{\sigma\tau}}{\partial x^{\alpha}\,\partial x^{\beta}\,\partial x^{\delta}}\Big)$$
$$+ g_{\alpha\beta}g^{\gamma\delta}g^{\epsilon\lambda}g^{\sigma\tau}\,\delta\Big(\frac{\partial^3 g_{\epsilon\tau}}{\partial x^{\delta}\,\partial x^{\sigma}\,\partial x^{\lambda}} + \frac{\partial^3 g_{\sigma\lambda}}{\partial x^{\delta}\,\partial x^{\epsilon}\,\partial x^{\tau}}$$
$$- \frac{\partial^3 g_{\epsilon\lambda}}{\partial x^{\delta}\,\partial x^{\sigma}\,\partial x^{\tau}} - \frac{\partial^3 g_{\sigma\tau}}{\partial x^{\epsilon}\,\partial x^{\lambda}\,\partial x^{\delta}}\Big). \tag{27}$$

Let us choose $\delta(\partial^3 g_{\alpha\tau}/\partial x^{\delta}\partial x^{\beta}\partial x^{\lambda})$ at W so that $\delta A_{\alpha 2} = G_{\gamma}{}^{\beta}\delta G_{\alpha}{}^{\gamma}{}_{|\beta} = 0$ for our geometry. This can be done so that $\delta A_{\alpha\beta} \neq 0$ at W as we see from the expressions (26):

$$\delta A_{\alpha 2} = G_{\gamma}{}^{\beta}\,\delta G_{\alpha}{}^{\gamma}{}_{|\beta} = 0,$$
$$\delta A_{\alpha 3} = G_{\delta}{}^{\beta}G_{\gamma}{}^{\delta}\,\delta G_{\alpha}{}^{\gamma}{}_{|\beta} + G_{\alpha}{}^{\gamma}\,\delta A_{\gamma 2} \tag{28}$$
$$= G_{\delta}{}^{\beta}G_{\gamma}{}^{\delta}\,\delta G_{\alpha}{}^{\gamma}{}_{|\beta} \neq 0.$$

For example, if we pick our geometry and our variation at W so that all $\delta G_{\alpha}{}^{\gamma}{}_{|\beta}$ vanish except $\delta G_0{}^{1}{}_{|1} = -G_0{}^{0}$ and $\delta G_0{}^{0}{}_{|0} = G_1{}^{1} \neq G_0{}^{0}$, and have $G_{\alpha}{}^{\beta}$ diagonal at W, then one finds that $\delta A_{\alpha 2} = 0$, $\delta A_{\alpha 3} \neq 0$. Using Eqs. (28) and (26) for the variations, Eq. (24) now reads $v^3\delta A_{\alpha 3} = 0$, since all other terms vanish at W and hence $v^3 = 0$ at W. But W was chosen arbitrarily, so that we must have $v^3(x) = 0$ throughout R.

Using the fact that $v^3(x) = 0$, we can iterate the above demonstration, again choosing the variations so that (25) holds at W. However this time we can choose $\delta(\partial^3 g_{\alpha\tau}/\partial x^{\delta}\partial x^{\beta}\partial x^{\lambda})$ so that $\delta A_{\alpha 2} \neq 0$ at W, so that the differential equation (24) reads $v^2\delta A_{\alpha 2} = 0$, and hence we must have $v^2(x) = 0$ throughout R.

With the knowledge that $v^3(x) = v^2(x) = 0$, we now pick our variations so that $\delta g_{\alpha\beta} = 0$, $\delta(\partial g_{\alpha\beta}/\partial x^{\gamma}) = 0$ at W, but $\delta(\partial^2 g_{\alpha\beta}/\partial x^{\tau}\partial x^{\gamma}) \neq 0$ at W so that $\delta G_{\alpha}{}^{\beta} \neq 0$. Equation (24) then implies that $(\partial v^1/\partial x^{\beta})\,\delta G^{\beta}{}_{\alpha} = 0$ at W so that for a nonsingular matrix $\delta G^{\beta}{}_{\alpha}$ we must have $\partial v^1/\partial x^{\beta} = 0$ in R, and therefore the only possibility is $v^1(x) =$ constant. Finally, since $v^3(x) = v^2(x) = 0$, and $v^1(x) =$ constant, Eq. (23) reads $(\partial v^0/\partial x^{\beta})\delta g_{\alpha\beta} = 0$, and for nonsingular $\delta g_{\alpha\beta}$ at W, we must have $\partial\gamma^0/\partial x^{\beta} = 0$ in R, and hence $v^0(x) =$ constant.

We have shown that the assumption of locality implies that the only possible choice of scalars $v^i(x)$ is $v^3 = v^2 = 0$, and v^1 and v^2 constants. Apply this result to (16) and let $\lambda = v^0 \mid v^1$ and $\kappa = 1/v^1$; we then have the Einstein equations

$$\kappa T_{\alpha}{}^{\beta} = G_{\alpha}{}^{\beta} + \lambda g_{\alpha}{}^{\beta} \tag{29}$$

as the only relation between $T_{\alpha\beta}$ and the metric-satisfying locality, and $\nabla\cdot\mathbf{T} = 0$, $\nabla\cdot\nabla\times\mathbf{T} = 0$, for an arbitrary Riemannian space. This concludes the proof of theorem II.

ACKNOWLEDGMENT

The author wishes to extend his gratitude to Professor M. Schiffer of the Stanford Mathematics Department for many illuminating discussions, for his penetrating criticisms, and for pointing out the necessity of the assumption of locality.

Fermi Normal Coordinates and Some Basic Concepts in Differential Geometry*

F. K. MANASSE† AND C. W. MISNER‡
Palmer Physical Laboratory, Princeton University, Princeton, New Jersey
(Received 31 May 1962)

Fermi coordinates, where the metric is rectangular and has vanishing first derivatives at each point of a curve, are constructed in a particular way about a geodesic. This determines an expansion of the metric in powers of proper distance normal to the geodesic, of which the second-order terms are explicitly computed here in terms of the curvature tensor at the corresponding point on the base geodesic. These terms determine the lowest-order effects of a gravitational field which can be measured locally by a freely falling observer. An example is provided in the Schwarzschild metric. This discussion of Fermi Normal Coordinate provides numerous examples of the use of the modern, coordinate-free concept of a vector and of computations which are simplified by introducing a vector instead of its components. The ideas of contravariant vector and Lie Bracket, as well as the equation of geodesic deviation, are reviewed before being applied.

I. INTRODUCTION

IN 1922 Fermi showed[1] that, given any curve in a Riemannian manifold, it is possible to introduce coordinates near this curve in such a way that the Christoffel symbols vanish along the curve, leaving the metric there rectangular. Several developments of this idea followed. One was a generalization[2] of the theorem to a manifold with a symmetric affine connection $\Gamma_{\mu\nu}{}^{\sigma}$, but without necessarily assuming any metric structure. A second development was an inquiry which showed that in general no coordinates exist for which $\Gamma_{\mu\nu}{}^{\sigma} = 0$ on surfaces of dimension greater than one, and which developed criteria for the special situations where this was possible.[3] A third variation of Fermi's idea is the set of coordinates based on an arbitrary curve which Synge[4] calls Fermi coordinates. Here one allows a few nonzero Christoffel symbols, although retaining a rectangular metric, for the advantage of making the curve become an axis of the coordinate system. These coordinates, as Synge shows, form a nonrotating system in a natural physical sense for a (not necessarily freely falling) observer in a gravitational field.

In this paper we consider not a modification or generalization of Fermi's idea, but a specialization and particularization of it. We specialize to the case where the curve in question is a geodesic, and we choose a particular set out of the many coordinate systems which satisfy his $\Gamma_{\mu\nu}{}^{\sigma} = 0$ condition along this geodesic. The resulting coordinates we call *Fermi normal coordinates* because of an analogy to that particular choice of the many coordinates satisfying $\Gamma_{\mu\nu}{}^{\sigma} = 0$ at a single point, called Riemann normal coordinates,[5] which in addition gives the series expansion[6]

$$ds^2 = \{\eta_{\mu\nu} + \tfrac{1}{3}R_{\alpha\mu\beta\nu}x^{\alpha}x^{\beta} + O[(x)^3]\}\, dx^{\mu}\, dx^{\nu}. \qquad (1)$$

The primary mathematical contribution of this paper is to compute the quadratic terms of a corresponding expansion in Fermi normal coordinates [see Eqs. (66)]. In this case the expansion parameter is the geodesic distance normal to the given geodesic; the expansion is valid for a limited region of space, and for *all time*. Thus, Fermi normal coordinates provide a standardized way in which a freely falling observer can report observations and local experiments. In particular, the quadratic terms of the metric, which we compute in terms of the curvature,

* Based in part on a Ph.D. Thesis by F. K. Manasse, Princeton University, 1961.

† Communications Development Training Program Fellow from Bell Telephone Laboratories.

‡ Alfred P. Sloan Research Fellow.

[1] E. Fermi, Atti Acad. Naz. Lincei Rend. Cl. Sci. Fiz. Mat. Nat. **31**, 21, 51 (1922).

[2] L. P. Eisenhart, *Non-Riemannian Geometry*, (American Mathematical Society Colloquium Publications, New York, New York, 1927), Sec. 25. The Fermi normal coordinates developed in the present paper are also defined in (symmetric) affine spaces, and all our results which can be stated in affine spaces are valid there. The proofs are obtained by replacing every set of orthonormal vectors by a set of linearly independent vectors.

[3] L. O'Raifeartaigh, Proc. Roy. Irish Acad. **A59**, 2 (1958).

[4] J. L. Synge, *Relativity, The General Theory* (North-Holland Publishing Company, Amsterdam, 1960).

[5] See, for example, L. P. Eisenhart, *Riemannian Geometry*, (Princeton University Press, Princeton, New Jersey, 1926).

[6] Here $\eta_{\mu\nu} = \text{diag}\,(-1, 1, 1, 1)$ is the Lorentz metric. We shall use Greek indices for space–time (μ, ν, etc. = 0, 1, 2, 3), while Latin indices give components along spatial axes (i, j, etc. = 1, 2, 3). Our sign conventions for the curvature tensor are

$$R_{\mu'\alpha\beta} = \partial_{\alpha}\Gamma_{\mu'\beta} - \partial_{\beta}\Gamma_{\mu'\alpha} - (\Gamma_{\mu'}{}^{\sigma}{}_{\alpha}\Gamma_{\sigma'\beta} - \Gamma_{\mu'}{}^{\sigma}{}_{\beta}\Gamma_{\sigma'\alpha}),$$

and

$$R_{\mu\nu} = R_{\mu}{}^{\alpha}{}_{\alpha\nu}.$$

The Riemann tensor convention corresponds to Cartan's definition (reference 13) of the curvature forms $O_{\mu'}{}^{\nu'} = \frac{1}{2}R_{\mu'}{}^{\nu'}{}_{\alpha\beta}\, dx^{\alpha}\, dx^{\beta}$ in terms of the connection forms

$$\omega_{\mu'}{}^{\nu'} = \Gamma_{\mu'}{}^{\nu'}{}_{\alpha}\, dx^{\alpha},\ \theta_{\mu'}{}^{\nu'} = d\omega_{\mu'}{}^{\nu'} - \alpha_{\mu'}{}^{\sigma'}\omega_{\sigma'}{}^{\nu'},$$

which definition is also valid in orthogonal (or other nonholonomic) frames.

determine the effects of gravitational field gradients upon experiments done in a freely falling elevator.

The procedure for constructing Fermi normal coordinates, which is given in Sec. II, is a variation of the standard procedure for constructing Riemann normal coordinates. It is also a special case of the procedures used by Levi-Civita[7] or Synge[4] to construct (inequivalent versions of) Fermi coordinates about an arbitrary, nongeodesic, curve. The present paper is very closely related to Levi-Civita's, since it discusses some of the same topics, but in inverse order. Levi-Civita, in the paper in question, developed for the first time the equation of geodesic deviation and used Fermi coordinates as a technique for simplifying this equation to display its properties more clearly. In contrast, our primary interest is here in the Fermi coordinates, but we shall use the equation of geodesic deviation as a device for studying the properties of Fermi coordinates and for computing the metric tensor in these coordinates.

The major part of the present paper is devoted to studying properties of the Fermi normal coordinates constructed in Sec. II. In Sec. IV we show that this construction leads to a nonsingular coordinate system in a neighborhood of the given geodesic, and in Sec. V we show that these coordinates satisfy the Fermi conditions

$$g_{\mu\nu}|_G = \eta_{\mu\nu}, \tag{2}$$

$$\Gamma_\mu{}^\alpha{}_\nu|_G = 0, \tag{3}$$

along the given geodesic G. In these discussions, as well as in later examples, it is useful to have unambiguous ways of indicating a vector without specifying a coordinate system, and of displaying its components in different coordinate systems without confusion. These notations, based on the idea of a tangent vector as a differentiation, are reviewed in Sec. III. This idea of a vector is also used in Sec. VI where we review the equation of geodesic deviation in order to see precisely what vector satisfies it. Then, in Sec. VII, we note that certain vectors occurring in the construction of Fermi normal coordinates must satisfy the equation of geodesic deviation; using this fact we evaluate the quadratic (curvature) terms in the expansion of the metric analogous to Eq. (1). Finally, Sec. VIII is an example, where, starting from the Schwarzschild metric in standard Schwarzschild coordinates, we evaluate the metric in Fermi normal coordinates surrounding a radial timelike geodesic. This represents this metric in a rest frame of a particle of negligible mass freely falling toward a large central mass. In the following paper,[8] this serves as a starting point from which to compute the metric surrounding a finite but small mass falling radially toward a large central mass, a particular case of the two-body problem in general relativity.

The paragraphs of mathematical "review" (Secs. III and VI), although they contain nothing new or original, are not considered by the authors as the least important part of this paper. Most physicists, even those very familiar with general relativity, continue to use the same definition of a vector as did Einstein, in spite of the considerable progress by mathematicians in the intervening half century. A particularly careful statement of this definition by Synge and Schild[9] gives a precise meaning to the sentence "The quantities v^α are components of a contravariant vector" without finding it worth the trouble to write a sentence of the form "a contravariant vector is a ..." The end of this sentence is, in fact, either rather unhelpful[10] or rather long[11] when it merely elucidates the transformation law definition. The transformation law outlook on geometry was an attempt to broaden[12] the Erlanger Programm viewpoint: (a geometry is characterized by invariance under a group of transformations) without repudiating it completely. The more geometrical approach to geometry, based on an intuition rooted in the classical studies of curves and surfaces in Euclidean three space, was hampered for a time because its most powerful computational techniques[13] employed elements which were defined only by their intuitive significance. As a consequence, many demonstrations were clear only to mathematicians with sufficient intuition.[13] This difficulty was eliminated by Chevalley[14] who gave new definitions of tangent vectors and differentials, providing them with a

[7] T. Levi-Civita, Math. Ann. 97, 291 (1926).

[8] F. K. Manasse, J. Math. Phys. 4, 746 (1963) (following paper).

[9] J. L. Synge and A. Schild, *Tensor Calculus* (University of Toronto Press, Toronto, 1952), Sec. 1.3.

[10] T. Y. Thomas, *The Differential Invariants of Generalized Spaces* (Cambridge University Press, New York, 1934,) p. 30.

[11] N. Steenrod, *The Topology of Fibre Bundles* (Princeton University Press, Princeton, New Jersey, 1951), Sec. 6.4.

[12] O. Veblen and J. H. C. Whitehead, *The Foundations of Differential Geometry*, (Cambridge University Press, New York, 1932, reprinted 1953), Sec. 16.

[13] E. Cartan, *Leçons sur la géométrie des espaces de Riemann* (Gauthier-Villars, Paris, 1951).

[14] C. Chevalley, *Theory of Lie Groups*. (Princeton University Press, Princeton, New Jersey, 1946), p. 77. A definition suitable for differentiable, rather than analytic, manifolds can be found in reference 15 or in H. Flanders, Trans. Am. Math. Soc. 75, 311 (1953). A definition of differentiable manifold which parallels Chevalley's for the analytic case is given by de Rham, *Variétes Différentiables* (Hermann et Cie., Paris, 1955).

clear formal structure compatible with their intuitive significance. These definitions are now also available in introductory texts.[15]

In Sec. III and subsequently we use this definition of a contravariant vector as a tangent to a curve. Since a curve is easily thought of in a coordinate-independent way as a moving point $P(t)$ in the manifold, this approach can avoid all mention of coordinates in defining a vector. A theoretical physicist should not be surprised that the tangent to a curve $P(t)$ is conceived of as the generator of infinitesimal translations along the curve and, hence, denoted by $\partial/\partial t$.

II. CONSTRUCTION OF FERMI NORMAL COORDINATES

Conditions (2) and (3), which Fermi coordinates satisfy, state that to the maximum extent possible one desires space in the neighborhood of some given geodesic G to look like flat space in rectangular coordinates. As motivation for the construction which will be given in this section, we suggest that a plausible way to try to achieve this is to use as many "straight" lines (geodesics) as possible in laying out the coordinates. What follows now is merely a recipe which purports to construct a coordinate system; the proof that it does so (i.e., that the coordinates constructed by this recipe are non-singular) is deferred to Sec. IV. That Eqs. (2) and (3) are satisfied is not shown until Sec. V.

In order to uniquely specify a set of Fermi normal coordinates it is necessary to choose arbitrarily a point P_0 to be the origin, and an orthonormal set of vectors $\mathbf{e}_0$, $\mathbf{e}_1$, $\mathbf{e}_2$, and $\mathbf{e}_3$ at P_0 to fix the coordinate axes there. The first step in the construction is then to solve the geodesic equation and obtain that unique geodesic G which starts at P_0 with tangent $\mathbf{e}_0$ there. We will describe the geodesic G by the equation

$$P = h(\tau). \tag{4}$$

The condition that G "starts at P_0" is just

$$P_0 = h(0), \tag{5}$$

and does not imply that we refuse to consider negative values of τ.

Because G is a geodesic, its tangent at any two points on G is related by parallel displacement along G. At P_0, the tangent was $\mathbf{e}_0$, which we now call $\mathbf{e}_0(0)$, while $\mathbf{e}_0(\tau)$ will mean the tangent to G at $P = h(\tau)$. Similarly, we can define $\mathbf{e}_i(\tau)$ for $i = 1, 2, 3$ as vectors at $h(\tau)$ obtained by parallel displacement

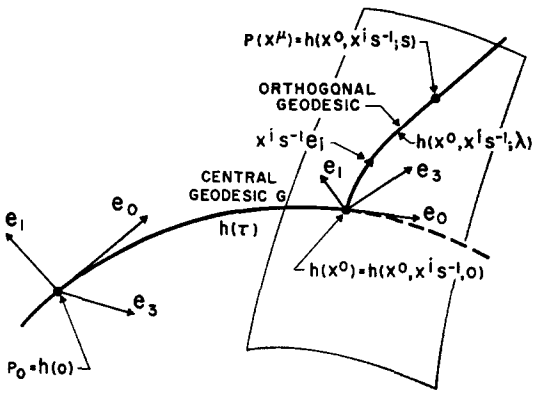

FIG. 1. Fermi normal coordinates are determined by a reference point P_0 and an orthonormal reference frame $\mathbf{e}_\mu$ there. The time axis G of the coordinates is the geodesic $h(\tau)$ tangent to $\mathbf{e}_0$ at P_0. The point $P(x^\mu)$, with given Fermi normal coordinates x^μ, is found by first following G for a proper time $\tau = x^0$, and then following a certain orthogonal geodesic at a proper distance $s = [(x^1)^2 + (x^2)^2 + (x^3)^2]^{\frac{1}{2}}$. This second spacelike geodesic $h(x^0, x^i s^{-1}, \lambda)$, is chosen by requiring that for $\lambda = 0$, where it crosses G, its tangent has direction cosines $x^i s^{-1}$ relative to the base vectors $\mathbf{e}_i$ carried by parallel transport along G from P_0.

along G of the vectors $\mathbf{e}_i = \mathbf{e}_i(0)$ given at $P_0 = h(0)$.

The prescription for locating the point P whose Fermi normal coordinates are x^μ can now be given. We assume for simplicity that $\mathbf{e}_0$ is a timelike vector and $\mathbf{e}_i$ are spacelike. Then given x^μ, we construct at the point $h(x^0)$ [i.e., along G at $\tau = x^0$] the unit vector

$$\mathbf{v} = \alpha^i \mathbf{e}_i(x^0), \tag{6}$$

whose (spatial) direction cosines α^i are

$$\alpha^i = x^i/s, \tag{7}$$

with

$$s^2 = (x^i)^2 \equiv (x^1)^2 + (x^2)^2 + (x^3)^2. \tag{8}$$

There is then a unique geodesic

$$P = h(x^0; \alpha^i; \lambda), \tag{9}$$

with path parameter λ which starts at $h(x^0)$ and is tangent to $\mathbf{v}$ there. The point $P(x^\mu)$ with Fermi normal coordinates x^μ is found by proceeding along this geodesic a proper distance s, i.e.,

$$P(x^\mu) = h(x^0; \alpha^i; s). \tag{10}$$

This prescription is summarized in Fig. (1). When points P are represented by their coordinate values $y^{\mu'}(P)$ in some coordinate system[16] in which the

[15] T. J. Willmore, *An Introduction to Differential Geometry* (Clarendon Press, Oxford, England, 1959), Chap. 6, Sec. 2.

[16] Since we think of the metric or any other tensor as an object which is independent of our choice of coordinate system, we prefer that the indication of the particular coordinate system to which a set of tensor components $g_{\mu\nu}$ refers be placed on the component (index) part of the symbol rather than on the tensor part. Thus $g_{\mu\nu}$ and $g_{\mu'\nu'}$ are components of the same metric tensor in two coordinate systems, while, should the occasion arise, $g_{\mu\nu}$ and $g'_{\mu\nu}$ might represent two different metrics in a single coordinate system. See also the transformation laws of Eqs. (40) and (76).

metric components $g_{\mu'\nu'}(y^\alpha)$ are known, the constructions which we have prescribed lead to the relationship (10) with the point $P(x^\mu)$ given in terms of its coordinates $y^{\nu'}$ as $y^{\nu'}(x^\mu)$. These functions $y^{\nu'}(x^\mu)$ specify the coordinate transformation between the arbitrary coordinates $y^{\nu'}$ and the Fermi normal coordinates x^μ. We have not needed to mention these arbitrary initial coordinates $y^{\mu'}$ while prescribing the construction of the Fermi normal coordinates x^μ, and have avoided doing so to emphasize the fact that the point $P(x^\mu)$ corresponding to given values of x^μ is independent of the coordinate system $y^{\mu'}$ in which the computations may have been performed.

III. TANGENT VECTORS AND LIE BRACKETS

In the preceding construction of Fermi normal coordinates, the vectors which appeared were all used as tangents to curves. We want to recall here that *all* contravariant vectors can be thought of as tangents to curves and identified with the derivative with respect to the corresponding curve parameter.

Given a curve $y^\mu(t)$ in some coordinate system y^μ, the tangent vector

$$t^\mu = dy^\mu/dt \tag{11}$$

is clearly a contravariant vector, and can be used to compute derivatives $\partial/\partial t$ along the curve $y^\mu(t)$ by the rule

$$\frac{\partial f}{\partial t} \equiv \frac{df(y^\mu(t))}{dt} = \frac{dy^\mu}{dt}\frac{\partial f}{\partial y^\mu} \equiv t^\mu f_{,\mu}. \tag{12}$$

Conversely, given a contravariant vector field[17] $t^\mu(y^\alpha)$, we can solve the ordinary differential equations

$$dy^\mu/dt = t^\mu(y^\alpha(t)) \tag{11}$$

to obtain curves $y^\mu(t)$ with tangents t^μ.

The advantage of thinking of contravariant vectors t^μ as tangents to curves is that this helps us find a concrete mathematical object we can identify with the abstract vector $\mathbf{t}$ whose components t^μ appear in our computations. This object is the operation of differentiation along the curve whose tangent is $\mathbf{t}$. That is, we write

$$\mathbf{t} \equiv \partial/\partial t. \tag{13}$$

The right-hand side of this identification is an operation which can be described in a *coordinate-independent way. The tangent,* $\mathbf{t}$ *or* $\partial/\partial t$*, to a curve* $P(t)$ *is the operation on scalar functions* $f(P)$ *defined by*

$$\partial f(P)/\partial t \equiv df(P(t))/dt, \tag{14}$$

i.e., by inserting the equation of the curve and taking an ordinary derivative. Since the operator $\partial/\partial t$ is applicable to all scalar functions, it can be applied in particular to those four scalar functions $y^\mu(P)$ we may be using as coordinates:

$$\partial y^\mu(P)/\partial t \equiv dy^\mu/dt = t^\mu. \tag{15}$$

In this way one can recover the components t^μ from the vector $\partial/\partial t$. Conversely, writing Eq. (12) in the form

$$\mathbf{t} \equiv \partial/\partial t = t^\mu\, \partial/\partial y^\mu, \tag{16}$$

we construct the contravariant vector $\partial/\partial t$ from a knowledge of its component t^μ. Equation (16) shows $\partial/\partial t$ as a linear combination, with coefficients t^μ, of four contravariant base vectors $\partial/\partial y^\mu$. These base vectors are tangents to the coordinate lines, e.g., $\partial/\partial y^0$ is the tangent $\partial/\partial t$ to the curve $y^i = \text{const}$, $y^0 = t$. We have frequent use for Eqs. (15) and (16) in what follows. In particular, Eq. (16) provides a method of displaying the components of a vector which simultaneously reminds us what coordinate system is being used and is, in this respect, superior to a statement of the form $\mathbf{t} = (t^0, t^1, t^2, t^3)$. We also find it convenient to be able to designate the components of a vector in several different ways, and thus write

$$(\partial/\partial t)^\mu = (\mathbf{t})^\mu = t^\mu = \partial y^\mu/\partial t. \tag{17}$$

Although we represent contravariant vectors $\mathbf{t}$ by the partial derivative symbol $\partial/\partial t$, it is not always possible to think of several vectors simultaneously as having the properties of standard partial derivatives.[18] In particular, consider the commutator of two tangent vector *fields* $\mathbf{u} \equiv \partial/\partial u$ and $\mathbf{v} \equiv \partial/\partial v$:

$$[\mathbf{u}, \mathbf{v}]f \equiv \frac{\partial}{\partial u}\left(\frac{\partial f}{\partial v}\right) - \frac{\partial}{\partial v}\left(\frac{\partial f}{\partial u}\right). \tag{18}$$

Since $\mathbf{v}$ is a field, $\partial f/\partial v$ is a function and can be subsequently differentiated along a curve tangent to $\mathbf{u}$. Thus the right-hand side of Eq. (18) is well defined, and evidently does not depend on the coordinates used to evaluate it. If we do pick a coordinate system, e.g., $\partial f/\partial v = v^\mu\, \partial f/\partial y^\mu$, Eq. (18)

[17] A vector t^μ given only at a point, or along a curve, etc., can always and in many ways be considered part of a vector field by arbitrarily defining $t^\mu(y^\alpha)$ at other points.

[18] A single vector $\partial/\partial t$ differs from a partial derivative by the possibility of vanishing; e.g., the tangent to a constant curve $P(t) = P_0$ is the zero vector $(\partial/\partial t) = 0$, since $\partial f/\partial t = df(P_0)/dt = 0$ for all functions f. However, in regions where $\partial/\partial t \neq 0$, coordinates can be introduced so that $\partial/\partial t = \partial/\partial y^0$ is a conventional partial derivative.

reads

$$[\mathbf{u}, \mathbf{v}]f = \left(u^\nu \frac{\partial v^\mu}{\partial y^\nu} - v^\nu \frac{\partial u^\mu}{\partial y^\nu}\right) \frac{\partial f}{\partial y^\mu}, \tag{19}$$

or

$$[\mathbf{u}, \mathbf{v}] = (v^\mu{}_{,\nu} u^\nu - u^\mu{}_{,\nu} v^\nu) \frac{\partial}{\partial y^\mu}. \tag{20}$$

As a linear combination of the base vectors $\partial/\partial y^\mu$, the object $[\mathbf{u}, \mathbf{v}]$ evidently is itself a contravariant vector, called the *Lie Bracket* of $\mathbf{u}$ and $\mathbf{v}$. Its components are displayed in Eq. (20). In case (as in this paper) a covariant derivative is defined, Eq. (20) can be rewritten as

$$[\mathbf{u}, \mathbf{v}]^\mu = v^\mu{}_{;\nu} u^\nu - u^\mu{}_{;\nu} v^\nu, \tag{21}$$

since the symmetry of the $\Gamma^\alpha_{\mu\nu} = \Gamma^\alpha_{\nu\mu}$ lets them cancel here in any case.

We most often wish to use Eq. (21) in the case where we know $[\mathbf{u}, \mathbf{v}] = 0$. This is true whenever $\mathbf{u}$ and $\mathbf{v}$ can be thought of as tangents to coordinate lines in a surface. That is, let $P(u, v)$ be the equation of a "surface" parameterized by u, v, and let $\mathbf{u} = \partial/\partial u$ be the tangent to lines of constant v in this surface, and similarly $\mathbf{v} = \partial/\partial v$ is tangent to lines of constant u. Then $[\mathbf{u}, \mathbf{v}]$ can be evaluated from Eq. (18) by setting $f = f(P(u, v))$ on the right-hand side. The derivatives are then standard partial derivatives which commute, so $[\mathbf{u}, \mathbf{v}] = 0$.[19] [This derivation requires only that $P(u, v)$ be a differentiable point-valued function; it actually represents a *two*-dimensional surface only if $\mathbf{u}$ and $\mathbf{v}$ are linearly independent vectors.]

IV. REGULARITY OF FERMI NORMAL COORDINATES

According to the construction of Sec. II, Fermi normal coordinates are specified in terms of the solution

$$P = h(\tau; \alpha^i; \lambda) \tag{22}$$

of the geodesic equation describing a geodesic which begins ($\lambda = 0$) at the point

$$h(\tau; \alpha^i; 0) = h(\tau) \tag{23a}$$

on the central geodesic G, and whose tangent there is

$$(\partial/\partial\lambda)_{\lambda=0} = \alpha^i \mathbf{e}_i(\tau). \tag{23b}$$

As used in Sec. II, the parameters α^i satisfied $(\alpha^i)^2 = 1$, but we ignore this condition now and consider all values of the α^i. We first prove that

$$h(\tau; s\alpha^i; \lambda) = h(\tau; \alpha^i; s\lambda) \tag{24}$$

holds for all s (by rescaling the path parameter λ) so that Eq. (10) defining Fermi normal coordinates may be replaced by

$$P(x^\mu) = h(x^0; x^i; 1). \tag{25}$$

This form allows us to verify more easily the differentiability of the inverse relationship, $x^\mu(P)$, i.e., of the coordinate functions.

We regard $h(\tau)$ and $h(\tau; \alpha^i; \lambda)$ as the point-valued function of one and five real variables, respectively, computed without regard to any interpretation placed on their real-number arguments. (In contrast, common usage for real-valued functions dictates that $f(x^\alpha)$ and $f(y^{\alpha'})$ mean different functions of their four real arguments so as to represent the same function of points $f(P)$ in two different coordinate systems.) Then, to prove Eq. (24) we rewrite Eq. (22) in some arbitrary regular coordinate system $y^{\mu'}(P)$ as

$$y^{\mu'} = h^{\mu'}(\tau; \alpha^i; \lambda). \tag{26}$$

The functions $h^{\mu'}$ are simply the *unique* solutions of the differential equations

$$\frac{d^2 h^{\mu'}}{d\lambda^2} + [\Gamma_{\alpha'}{}^{\mu'}{}_{\beta'}]_{\nu'=h'} \frac{dh^{\alpha'}}{d\lambda} \frac{dh^{\beta'}}{d\lambda} = 0, \tag{27}$$

which satisfy the initial conditions

$$h^{\mu'}(\tau; \alpha^i; 0) = h^{\mu'}(\tau), \tag{28}$$

and

$$\frac{dh^{\mu'}(\tau; \alpha^i; 0)}{d\lambda} \equiv \left(\frac{\partial}{\partial\lambda}\right)^{\mu'}_{\lambda=0} = \alpha^i (\mathbf{e}_i(\tau))^{\mu'}, \tag{29}$$

where $y^{\mu'} = h^{\mu'}(\tau)$ is the central geodesic G. After remarking that the differential equation (27) is unchanged upon replacing λ by $s\lambda$, we prove Eq. (24) by verifying that, as function of λ, $h^{\mu'}(\tau; s\alpha^i; \lambda)$ and $h^{\mu'}(\tau; \alpha^i; s\lambda)$ not only satisfy the same differential equation (27), but also the same initial condition. For, each reduces to $h^{\mu'}(\tau)$ for $\lambda = 0$ and has a first derivative $s\alpha^i(\mathbf{e}_i(\tau))^{\mu'}$ at $\lambda = 0$. Thus, by the Uniqueness Theorem[20] for solutions of differential equations we have

$$h^{\mu'}(\tau; s\alpha^i; \lambda) = h^{\mu'}(\tau; \alpha^i; s\lambda), \tag{24a}$$

[19] To see that the Lie Bracket does not *always* vanish, an example suffices. For the unit vectors $\mathbf{e}_\theta = \partial/\partial\theta$ and $\mathbf{e}_\varphi = (\sin\theta)^{-1}\partial/\partial_\varphi$ on the unit sphere, compute from Eq. (20)

$$[\mathbf{e}_\theta, \mathbf{e}_\varphi] = -\cot\theta \mathbf{e}_\varphi \neq 0.$$

[20] See, for example, F. J. Murray and K. S. Miller, *Existence Theorems*, (New York University Press, New York, 1954), Chap. 2, Theorems 1, 3; Chap. 3, Theorem 2; Chap. 5, Theorem 6. A discussion of the properties of geodesics from which we have borrowed much is found in H. Seifert and W. Threlfall, *Variationsrechnung im Grossen* (B. G. Teubner, Leipzig, 1938), footnote 20, p. 97.

which represents Eq. (24) in the $y^{\mu'}$ coordinate system.

The definition of Fermi normal coordinates in Eq. (25) gives now the transformation law

$$y^{\mu'}(x^\alpha) = h^{\mu'}(x^0; x^i; 1). \tag{30}$$

By a standard theorem,[20] the solutions of ordinary differential equations are differentiable functions of the initial conditions, so we have established the differentiability of $y^{\mu'}(x^\alpha)$. To show the existence of a differentiable inverse relation $x^\mu(y^\alpha)$, representing the coordinate functions $x^\mu(P)$, we must show that the Jacobian $|\partial y^{\mu'}/\partial x^\alpha|$ does not vanish.[20]

The condition of a nonvanishing Jacobian is precisely the condition that the coordinate axes do not collapse, i.e., that the vectors $\partial/\partial x^\mu$ be linearly independent. For, when we form the components of, say, $\partial/\partial x^0$ in the $y^{\mu'}$ frame, they are $\partial y^{\mu'}/\partial x^0$, so that the determinant formed from the components of the four vectors $\partial/\partial x^\alpha$ is

$$J \equiv \det(\partial y^{\mu'}/\partial x^\alpha), \tag{31}$$

and $J \neq 0$ is equivalent to the linear independence of these vectors. We prove $J \neq 0$ by showing that, along the central geodesic G,

$$(\partial/\partial x^\mu)_G = \mathbf{e}_\mu(\tau). \tag{32}$$

Then, since $\mathbf{e}_\mu(\tau)$ are orthonormal vectors, they are linearly independent and $J \neq 0$ on G. By continuity, then, we have $J \neq 0$ in some neighborhood of G.

The basic fact we need in order to prove the equation $(\partial/\partial x^\mu)_G = \mathbf{e}_\mu(\tau)$ in the preceding argument is the description in Fermi normal coordinates of the geodesics entering their construction. This is also the basis from which we will compute all other properties of Fermi normal coordinates. Consider then the curve $P(\lambda)$ defined in Fermi normal coordinates by

$$\begin{aligned} x^0 &= \tau = \text{const}, \\ x^i &= \alpha^i\lambda, \qquad \alpha^i = \text{const}. \end{aligned} \tag{33}$$

According to Eqs. (25) and (24), this curve is given by

$$P(\lambda) = h(\tau; \alpha^i\lambda; 1) = h(\tau; \alpha^i; \lambda), \tag{34}$$

and is, therefore, that geodesic whose tangent $\partial/\partial\lambda$ is given by Eq. (23b) at the point $x^i = 0$, $x^0 = \tau$ corresponding to $\lambda = 0$. But the components of $\partial/\partial\lambda$ can be computed from Eqs. (33), and are $(\partial/\partial\lambda)^0 = 0$ and $(\partial/\partial\lambda)^i = \alpha^i$, so

$$\partial/\partial\lambda = \alpha^i\,\partial/\partial x^i. \tag{35}$$

Comparing this with Eq. (23b) gives

$$(\partial/\partial x^i)_{x^i=0} = \mathbf{e}_i(x^0), \tag{32a}$$

since the α^i are arbitrary. Similarly, from Eqs. (25), (24), and (23a) we see that the curve $P(\tau)$ defined by

$$\begin{aligned} x^0 &= \tau, \\ x^i &= 0, \end{aligned} \tag{36}$$

is given by

$$P(\tau) = h(\tau; 0; 1) = h(\tau; 0; 0) = h(\tau), \tag{36b}$$

and is the central geodesic G whose tangent $\partial/\partial\tau$ is $\mathbf{e}_0(\tau)$. But the components of the tangent $(\partial/\partial\tau)$ are easily computed from Eq. (36) and give

$$(\partial/\partial x^0)_{x^i=0} = \mathbf{e}_0(x^0). \tag{32b}$$

To recapitulate, the question of the Jacobian or of the linear independence of the $\partial/\partial x^\mu$, reduces by Eq. (32) to the linear independence of the $\mathbf{e}_\mu(\tau)$. But $\mathbf{e}_\mu(\tau)$ are orthonormal, since they are defined by parallel displacement of the orthonormal vectors $\mathbf{e}_\mu(0)$, and parallel displacement preserves inner products,

$$\mathbf{e}_\mu(\tau)\cdot\mathbf{e}_\nu(\tau) = \eta\mu\nu. \tag{37}$$

V. THE FERMI CONDITIONS

We have actually already proven that in Fermi normal coordinates the metric is rectangular on G. For by definition, the metric components are the matrix of inner products of the base vectors, i.e.,

$$g_{\mu\nu}(x^\alpha) = (\partial/\partial x^\mu)\cdot(\partial/\partial x^\nu), \tag{38}$$

so Eqs. (32) and (37) give

$$g_{\mu\nu}(x^0; 0) \equiv g_{\mu\nu}|_G = \eta_{\mu\nu}. \tag{39}$$

It may, nevertheless, be instructive to see this equation arise by applying the tensor transformation law to the metric components $g_{\alpha'\beta'}(y^{\sigma'})$ of some original coordinate system $y^{\sigma'}$:

$$g_{\mu\nu}|_G = g_{\alpha'\beta'}\left(\frac{\partial y^{\alpha'}}{\partial x^\mu}\right)_G\left(\frac{\partial y^{\beta'}}{\partial x^\nu}\right)_G = \mathbf{e}_\mu\cdot\mathbf{e}_\nu = \eta_{\mu\nu}. \tag{40}$$

In the central equality here we recalled that $(\partial y^{\alpha'}/\partial x^\mu)_G$ are the components of $(\partial/\partial x^\mu)_G = \mathbf{e}_\mu(x^0)$ in the $y^{\alpha'}$ frame.

In order to show that $\Gamma_\mu{}^\alpha{}_\nu|_G = 0$ holds in Fermi normal coordinates, we begin by considering the consequences of the fact that the curve $x^0 = \tau$, $x^i = \alpha^i\lambda$, satisfies the geodesic equation

$$\frac{d^2x^\mu}{d\lambda^2} + \Gamma_\alpha{}^\mu{}_\beta\frac{dx^\alpha}{d\lambda}\frac{dx^\beta}{d\lambda} = 0. \tag{41}$$

Substitution gives

$$\Gamma_{i}{}^{\mu}{}_{j}(x^0 = \tau; x^i = \alpha^i\lambda)\alpha^i\alpha^j = 0, \tag{42a}$$

which holds in particular for $\lambda = 0$:

$$\Gamma_{i}{}^{\mu}{}_{j}(x^0 = \tau; x^i = 0)\alpha^i\alpha^j = 0. \tag{42b}$$

But now the $\Gamma_{ij}{}^{\mu}$ are independent of α^i and symmetric in i and j, while the α^i are arbitrary, so

$$\Gamma_{i}{}^{\mu}{}_{j}|_G = 0. \tag{43}$$

To show that the other Christoffel symbols vanish, we recall that each of the vectors $\mathbf{e}_\mu(\tau)$ satisfies the equation of parallel displacement along the central geodesic $x^0 = \tau$, $x^i = 0$. In general, to parallel displace a vector $\mathbf{u}$ along a curve $P(t)$, one solves the differential equation (linear)

$$u^{\mu}{}_{;\nu}t^\nu = du^\mu/dt + u^\alpha\Gamma_{\alpha}{}^{\mu}{}_{\beta}t^\beta. \tag{44}$$

We may take $\mathbf{u}$ to be any of the vectors $\mathbf{e}_\sigma(\tau) = (\partial/\partial x^\sigma)_G$ whose components in Fermi normal coordinates are therefore $(\mathbf{e}_\sigma)^\mu = \delta_\sigma{}^\mu$, and $\mathbf{t}$ becomes $\partial/\partial\tau = \mathbf{e}_0 = \delta_0{}^\beta(\partial/\partial x^\beta)_G$. Thus, from Eq. (44) we obtain

$$\Gamma_{\sigma}{}^{\mu}{}_{0}|_G = 0. \tag{45}$$

Combining Eqs. (43) and (45) gives the second of the Fermi conditions,

$$\Gamma_{\alpha}{}^{\mu}{}_{\beta}|_G = 0. \tag{46}$$

Since this implies $(\partial g_{\alpha\beta}/\partial x^\mu)_G = 0$, we have evaluated the first two terms in a Taylor expansion of the Fermi normal metric. The quadratic terms, which require us to evaluate $(\partial^2 g_{\alpha\beta}/\partial x^i\partial x^j)_G$, is computed in Sec. VII, after a diversion to review the computational technique we will use.

VI. EQUATION OF GEODESIC DEVIATION[7,21]

The construction of Fermi normal coordinates involves families of geodesics. Let us consider only a one-parameter family of geodesics for the present, say $P(n, s)$, where for each fixed value at $n = n_0$, $P(n_0, s)$ satisfies the geodesic differential equation with s as path parameter. The tangent vector $\mathbf{s} = \partial/\partial s$ can then be thought of as the generator of infinitesimal translations along geodesic n, while $\mathbf{n} = \partial/\partial n$ is the generator of infinitesimal translations along a curve $P(n, s_0)$ connecting corresponding points (same value of s) on adjacent geodesics. Along a fixed geodesic, $\mathbf{n}$ cannot vary arbitrarily, since the adjacent geodesic can be determined by only two points lying on it. These restrictions on $\mathbf{n}$ are expressed by the equation of geodesic deviation which is a differential equation satisfied by $\mathbf{n}$ along each geodesic, i.e., as a function of s.

To derive the equation of geodesic deviation, we begin with the geodesic equation in the form[9]

$$\delta\mathbf{s}/\delta s = 0, \tag{47}$$

where $\delta/\delta s$ is the covariant derivative along s. This is, of course, just an abbreviation for

$$s^{\mu}{}_{;\nu}s^\nu \equiv ds^\mu/ds + s^\alpha\Gamma_{\alpha}{}^{\mu}{}_{\beta}s^\beta = 0, \tag{48}$$

but the more compact notation lets us outline the derivation without computations. Since Eq. (47) holds for all values of n, we may differentiate it to obtain

$$\frac{\delta}{\delta n}\left(\frac{\delta\mathbf{s}}{\delta s}\right) = 0. \tag{49}$$

As the difference of two geodesic equations, this should be an equation for the difference vector $\mathbf{n}$, that is, $\mathbf{n}$ should appear differentiated, rather than as a derivative. The relationship which achieves this is Eq. (21) which can be written

$$0 = [\mathbf{n}, \mathbf{s}] \equiv \delta\mathbf{s}/\delta n - \delta\mathbf{n}/\delta s. \tag{50}$$

[The Lie bracket $[\mathbf{n}, \mathbf{s}]$ vanishes since n and s parameterize the surface $P(n, s)$.] Before this relation can be employed in Eq. (49), however, the covariant derivatives must be written in the opposite order,

$$0 = \frac{\delta}{\delta n}\left(\frac{\delta\mathbf{s}}{\delta s}\right) = \frac{\delta}{\delta s}\left(\frac{\delta\mathbf{s}}{\delta n}\right) + \left\{\frac{\delta}{\delta n}\frac{\delta}{\delta s} - \frac{\delta}{\delta s}\frac{\delta}{\delta n}\right\}\mathbf{s}, \tag{51}$$

or

$$\delta^2\mathbf{n}/\delta s^2 + [\delta/\delta n, \delta/\delta s]\mathbf{s} = 0. \tag{52}$$

When the commutator[22] of covariant derivatives here is expressed in terms of the curvature tensor, this Eq. (52) is the equation of geodesic deviation. The computation is

$$\begin{aligned}\{[\delta/\delta n, \delta/\delta s]\mathbf{s}\}^\mu &= (s^{\mu}{}_{;\alpha}s^\alpha)_{;\beta}n^\beta - (s^{\mu}{}_{;\beta}n^\beta)_{;\alpha}s^\alpha \\ &= (s^{\mu}{}_{;\alpha\beta} - s^{\mu}{}_{;\beta\alpha})s^\alpha s^\beta + s^{\mu}{}_{;\nu}(s^\nu{}_{;\sigma}n^\sigma - n^\nu{}_{;\sigma}s^\sigma) \\ &= s^\sigma R_{\sigma}{}^{\mu}{}_{\beta\alpha}n^\beta s^\alpha + s^{\mu}{}_{;\nu}[\mathbf{n}, \mathbf{s}]^\nu \\ &= s^\alpha R_{\alpha}{}^{\mu}{}_{\nu\beta}s^\beta n^\nu,\end{aligned} \tag{53}$$

[21] For physical applications see F. A. E. Pirani, Acta Phys. Polon. **15**, 389 (1956); Phys. Rev. **105**, 1089 (1957); and J. Weber, *General Relativity and Gravitational Waves* (Interscience Publishers, Inc., New York, 1961), Chap. 8.

[22] The entire derivation may be regarded as a process of evaluating the commutators which relate Eq. (49) in the curious form

$$\frac{\delta}{\delta n}\frac{\delta}{\delta s}\left(\frac{\partial}{\partial s}\right) = 0$$

to an equation whose leading term is

$$\frac{\delta^2 n}{\delta s^2} = \frac{\delta}{\delta s}\frac{\delta}{\delta s}\left(\frac{\partial}{\partial n}\right).$$

where we again use $[\mathbf{n}, \mathbf{s}] = 0$. Thus, we find for the equation of geodesic deviation

$$\delta^2 n^\mu/\delta s^2 + (s^\alpha R_\alpha{}^\mu{}_{\nu\beta} s^\beta) n^\nu = 0. \tag{54}$$

Let us briefly outline the principal properties of this equation. It is a second-order, *linear*, ordinary differential equation for $\mathbf{n}$ as a function of s. In addition to the trivial solution $\mathbf{n} = 0$, it also obviously has the solution $\mathbf{n} = \mathbf{s}$. A perhaps not so obvious solution which can, however, be easily verified is $\mathbf{n} = s\mathbf{s}$. The solution $\mathbf{n} = (A's + B')\mathbf{s}$ (where the adjacent geodesics coincide with the original one but are parameterized differently) clearly satisfies

$$\mathbf{n}\cdot\mathbf{s} = As + B, \tag{55}$$

where A and B are constants [$A = A'(\mathbf{s}\cdot\mathbf{s})$ while $\mathbf{s}\cdot\mathbf{s} = \text{const}$ according to the geodesic equations]. We can further show that *every* solution $\mathbf{n}$ satisfies Eq. (55) by using the product rule of covariant differentiation and the geodesic equation $\delta s_\mu/\delta s = 0$ to compute

$$\frac{\partial^2}{\partial s^2}(s_\mu n^\mu) = \frac{\delta^2}{\delta s^2}(s_\mu n^\mu) = \frac{\delta}{\delta s}\left(s_\mu \frac{\delta n^\mu}{\delta s}\right) = s_\mu \frac{\delta^2 n^\mu}{\delta s^2}$$
$$= -s^\alpha s^\mu R_{\alpha\mu\nu\beta} n^\nu s^\beta = 0. \tag{56}$$

The constant A in Eq. (55) is related to the normalization of the geodesic parameter, for by using the Lie bracket relation $[\mathbf{n}, \mathbf{s}] = 0$, we find

$$A = \frac{\delta}{\delta s}(s_\mu n^\mu) = s_\mu \frac{\delta n^\mu}{\delta s} = s_\mu \frac{\delta s^\mu}{\delta n} = \frac{1}{2}\frac{\partial}{\partial n}(s_\mu s^\mu). \tag{57}$$

Unless $\mathbf{s}\cdot\mathbf{s} = 0$, we can always modify any solution $\mathbf{n}$ of the geodesic deviation equation, adding terms of the form $(As + B)\mathbf{s}$, to obtain a solution satisfying $\mathbf{n}\cdot\mathbf{s} = 0$. This modification corresponds to a linear change in the parameterization of the adjacent geodesic, which is of course consistent with the geodesic equation. According to Eq. (57) the condition $\mathbf{n}\cdot\mathbf{s} = 0$ is consistent with the standard normalization $\mathbf{s}\cdot\mathbf{s} = \pm 1$ for geodesic parameters.

VII. QUADRATIC TERMS IN THE FERMI METRIC

A power-series expansion of the metric in Fermi normal coordinates is determined by the derivatives $g_{\mu\nu,ij\cdots k}|_G$. The linear terms $g_{\mu\nu,i}|_G x^i$ were shown to vanish in Sec. V, where we found that on the central geodesic G all the Christoffel symbols vanish. In this section we will compute the quadratic terms, $\frac{1}{2} g_{\mu\nu,ij}|_G x^i x^j$, by first computing $\Gamma_\mu{}^\alpha{}_{\nu,\beta}|_G$.

Since the equation $\Gamma_\mu{}^\alpha{}_\nu = 0$ holds for all x^0 at $x = 0$, it may be differentiated with respect to x^0 to give

$$\Gamma_\mu{}^\alpha{}_{\nu,0}|_G = 0. \tag{58a}$$

Also using $\Gamma_\mu{}^\alpha{}_\nu|_G = 0$, we note that on G the definition of the Riemann tensor[6] reduces to two terms, and, in particular, from Eq. (58a) we find

$$\Gamma_\mu{}^\alpha{}_{0,\nu}|_G = R_\mu{}^\alpha{}_{\nu 0}|_G. \tag{58b}$$

The remaining derivatives of the affine connection are

$$\Gamma_i{}^\mu{}_{j,k}|_G = -\tfrac{1}{3}(R_i{}^\mu{}_{jk} + R_j{}^\mu{}_{ik})|_G, \tag{58c}$$

as we now show by use of the equation of geodesic deviation. Note that this last equation implies a symmetry

$$\Gamma_i{}^\mu{}_{j,k}|_G + \Gamma_j{}^\mu{}_{k,i}|_G + \Gamma_k{}^\mu{}_{i,j}|_G = 0, \tag{58d}$$

peculiar to these coordinates which follows from the corresponding "triple symmetry" of the Riemann tensor.

The family of geodesics $P(\lambda) = h(\tau; \alpha^i; \lambda)$ used in constructing Fermi normal coordinates provides us with four vectors, $\partial/\partial\tau$ and $\partial/\partial\alpha^i$, which (since they generate displacements between adjacent geodesics) must each satisfy the equation of geodesic deviation as functions of λ for fixed τ, α^i [In contrast, $P(\tau) = h(\tau; \alpha^i; \lambda)$ is not a geodesic unless $\lambda\alpha^i = 0$, so we have no *family* of geodesics with tangents $\partial/\partial\tau$, and neither $\partial/\partial\alpha^i$ nor $\partial/\partial\lambda$ satisfies the equation of geodesic deviation as a function of τ, even for $\lambda\alpha^i = 0$.] Although there are moderate amounts of computation involved in what follows now, the basic idea is quite simple. In the geodesic deviation equation ("$\delta^2\mathbf{n} + R\mathbf{n} = 0$") we insert known solutions, $\mathbf{n} = \partial/\partial\tau$ or $\partial/\partial\alpha^i$. At the point $\lambda = 0$ (i.e., on G where $\Gamma_\mu{}^\alpha{}_\nu = 0$), the second covariant derivative term will reduce to the derivative of a Christoffel symbol evaluated on G, and the only other term in the equation will be the curvature term, so we will obtain a formula "$\partial\Gamma = R$", i.e., Eqs. (58).

The family of geodesics $P = h(\tau; \alpha^i; \lambda)$ is described in Fermi normal coordinates by the equations

$$x^0 = \tau, \qquad x^i = \alpha^i\lambda. \tag{59}$$

The components of the deviation vectors $\mathbf{n}$, computed using Eq. (17), are then

$$\partial/\partial\tau = \partial/\partial x^0 \equiv \delta_0{}^\mu\, \partial/\partial x^\mu, \tag{60}$$

$$\partial/\partial\alpha^i = \lambda\, \partial/\partial x^i \equiv \lambda\delta_i{}^\mu\, \partial/\partial x^\mu. \tag{61}$$

Similarly, the vector $\mathbf{s}$ tangent to the geodesic is

$$\partial/\partial\lambda = \alpha^i\, \partial/\partial x^i. \tag{62}$$

We use the components displayed here in the geodesic deviation equation (54) which in detail

becomes

$$\frac{d^2n^\mu}{ds^2} + 2\frac{dn^\sigma}{ds}\Gamma_\sigma{}^\mu{}_\alpha s^\alpha + n^\sigma s^\alpha s^\beta R_\alpha{}^\mu{}_{\sigma\beta}$$
$$+ n^\sigma s^\alpha s^\beta(\Gamma_\sigma{}^\mu{}_{\alpha,\beta} + \Gamma_\sigma{}^\tau{}_\alpha\Gamma_\tau{}^\mu{}_\beta - \Gamma_\sigma{}^\mu{}_\tau\Gamma_\alpha{}^\tau{}_\beta) = 0. \quad (63)$$

The case $\mathbf{n} = \partial/\partial\tau$ merely leads to some of Eqs. (58b) again, so we treat only the cases $\mathbf{n} = \partial/\partial\alpha^i$. Then dn^σ/ds becomes $d(\lambda\delta_i{}^\sigma)/d\lambda = \delta_i{}^\sigma$, and

$$d^2n^\mu/d\lambda^2 = 0. \quad (63a)$$

But, since $n^\mu = \lambda\delta_i{}^\mu$ vanishes for $\lambda = 0$, Eq. (63) is trivial on G unless we divide through by λ before setting $\lambda = 0$. In order to accomplish this, the second term in Eq. (63) can be expanded in powers of λ;

$$2\delta_i{}^\sigma\Gamma_\sigma{}^\mu{}_j\alpha^j = 2\Gamma_i{}^\mu{}_j|_G\,\alpha^j + 2\lambda\left(\frac{\partial}{\partial\lambda}\Gamma_i{}^\mu{}_j\alpha^j\right)\Big|_G$$
$$= 2\lambda\Gamma_i{}^\mu{}_{j,k}|_G\,\alpha^j\alpha^k + O(\lambda^2). \quad (64a)$$

Then, at $\lambda = 0$ where $\Gamma_\mu{}^\alpha{}_\nu = 0$, we obtain from Eq. (63)

$$(3\Gamma_i{}^\mu{}_{j,k} + R_j{}^\mu{}_{ik})|_G\,\alpha^j\alpha^k = 0, \quad (64b)$$

or

$$(\Gamma_i{}^\mu{}_{j,k} + \Gamma_i{}^\mu{}_{k,j})|_G = -\tfrac{1}{3}(R_j{}^\mu{}_{ik} + R_k{}^\mu{}_{ij})|_G. \quad (64c)$$

This equation can be solved for $\Gamma_i{}^\mu{}_{j,k}|_G$ by adding to it one cyclic permutation,

$$(\Gamma_j{}^\mu{}_{k,i} + \Gamma_j{}^\mu{}_{i,k})|_G = -\tfrac{1}{3}(R_k{}^\mu{}_{ji} + R_i{}^\mu{}_{jk})|_G, \quad (64d)$$

and subtracting another,

$$(\Gamma_k{}^\mu{}_{i,j} + \Gamma_k{}^\mu{}_{j,i})|_G = -\tfrac{1}{3}(R_i{}^\mu{}_{kj} + R_j{}^\mu{}_{ik})|_G. \quad (64e)$$

The result, after using the symmetry of the connection $\Gamma_\mu{}^\alpha{}_\nu = \Gamma_\nu{}^\alpha{}_\mu$, is just Eq. (58c).

From the definition of Christoffel symbols,

$$g_{\mu\nu,\alpha} = g_{\mu\sigma}\Gamma_\nu{}^\sigma{}_\alpha + g_{\sigma\nu}\Gamma_\mu{}^\sigma{}_\alpha, \quad (64f)$$

we find by differentiation that

$$g_{\mu\nu,\alpha\beta}|_G = \eta_{\mu\sigma}\Gamma_\nu{}^\sigma{}_{\alpha,\beta}|_G + \eta_{\sigma\nu}\Gamma_\mu{}^\sigma{}_{\alpha,\beta}|_G. \quad (64g)$$

Thus, Eqs. (58) imply that

$$g_{\mu\nu,0\alpha}|_G = 0, \quad (64h)$$

and that, for $g_{\mu\nu,ij}|_G$, we have

$$g_{00,ij}|_G = 2R_{0i0j}|_G, \quad (65a)$$
$$g_{0k,ij}|_G = \tfrac{2}{3}(R_{0jki} + R_{0ikj})|_G, \quad (65b)$$
$$g_{lm,ij}|_G = \tfrac{1}{3}(R_{iljm} + R_{imjl})|_G. \quad (65c)$$

To summarize all the information we have obtained about the metric in Fermi normal coordinates, we can write the Taylor series

$$g_{00} = -1 + R_{0l0m}|_G\,x^lx^m + \cdots, \quad (66a)$$
$$g_{0i} = 0 + \tfrac{2}{3}R_{0lim}|_G\,x^lx^m + \cdots, \quad (66b)$$
$$g_{ij} = \delta_{ij} + \tfrac{1}{3}R_{iljm}|_G\,x^lx^m + \cdots. \quad (66c)$$

Here the dependence of the metric on the spatial coordinates x^i is shown explicitly, while its dependence on x^0 is contained entirely in the curvature components which are evaluated at $x^i = 0$ for each x^0.

VIII. AN EXAMPLE

We compute the Schwarzschild metric to quadratic order, as in Eq. (66), in Fermi normal coordinates surrounding a radial geodesic. In Schwarzschild coordinates, which we will call $y^{\mu'}$ or T, R, Θ, Φ, the metric components $g_{\mu'\nu'}$ are displayed in the form

$$ds^2 = g_{\mu'\nu'}\,dy^{\mu'}\,dy^{\nu'}$$
$$= -X\,dT^2 + X^{-1}\,dR^2 + R^2\,d\Theta^2 + R^2\sin^2\Theta\,d\Phi^2, \quad (67)$$

where

$$X = 1 - 2M/R. \quad (68)$$

To find the equations of a radial geodesic, $T(t)$, $R(t)$, with Θ and Φ constant, one may replace the geodesic equations by two first integrals; one is the normalization of proper time

$$1 = XT'^2 - X^{-1}R'^2, \quad (69)$$

and the other is a dimensionless energy parameter

$$k = XT', \quad (70)$$

which yields $R(t)$ by quadratures, and expresses k of the metric. (The primes here indicate derivatives with respect to *proper* time t along this geodesic.) Eliminating T' gives

$$k^2 = X + R'^2 = 1 - 2M/R_0, \quad (71)$$

which yields $R(t)$ by quadratures, and expresses k in terms of the maximum radius R_0 along the path, where $R' = 0$. The integration gives a cycloid

$$R = \tfrac{1}{2}R_0(1 + \cos\omega),$$
$$t = \tfrac{1}{2}R_0(R_0/2M)^{\frac{1}{2}}(\omega + \sin\omega). \quad (72)$$

Either R or the cycloid parameter ω can be used in place of proper time t to identify points on this geodesic, and thus serve as a time coordinate in the comoving frame. Thus,

$$dt^2 = \frac{dR^2}{2M/R - 2M/R_0} = \frac{R_0}{2M}R^2\,d\omega^2. \quad (73)$$

After choosing a geodesic, the next step in constructing Fermi normal coordinates is to choose an orthonormal frame along the geodesic. The timelike base vector must be the tangent $\partial/\partial t$, and the symmetry of the present example determines the others. Thus,

$$\begin{aligned}
\mathbf{e}_0 &= \partial/\partial t|_G = T'\,\partial/\partial T + R'\,\partial/\partial R,\\
\mathbf{e}_1 &= \partial/\partial x|_G = X^{-1}R'\,\partial/\partial T + XT'\,\partial/\partial R,\\
\mathbf{e}_2 &= \partial/\partial y|_G = 1/R\,\partial/\partial\Theta,\\
\mathbf{e}_3 &= \partial/\partial z|_G = 1/R\sin\Theta\,\partial/\partial\Phi,
\end{aligned} \tag{74}$$

where x^μ or $xyzt$ are to be Fermi normal coordinates. It is also easily verified from the components $(\mathbf{e}_\alpha)^{\mu'}$ displayed here that these vectors satisfy the necessary parallel transport condition

$$\delta\mathbf{e}_\alpha/\delta t = 0 = (\mathbf{e}_\alpha)^{\mu'}{}_{;\nu'}(\mathbf{e}_0)^{\nu'}. \tag{75}$$

We must now compute the curvatures in the Fermi frame by the tensor transformation law

$$R_{\alpha\beta\gamma\delta} = R_{\mu'\nu'\sigma'\tau'}(\mathbf{e}_\alpha)^{\mu'}(\mathbf{e}_\beta)^{\nu'}(\mathbf{e}_\gamma)^{\sigma'}(\mathbf{e}_\delta)^{\tau'}, \tag{76}$$

which states that a tensor component is the contraction of the tensor with the base vectors indicated by the indices. The Fermi base vectors $\mathbf{e}_\alpha$ we have in Eq. (74), while the curvature components $R_{\mu'\nu'\sigma'\tau'}$ with respect to the Schwarzschild frame are well known as

$$\begin{aligned}
R_{1'0'1'0'} &= 2M/R^3,\\
R_{3'0'3'0'} &= -(MX/R)\sin^2\Theta,\\
R_{1'2'1'2'} &= M/RX,\\
R_{2'0'2'0'} &= -MX/R,\\
R_{2'3'2'3'} &= -2MR\sin^2\Theta,\\
R_{1'3'1'3'} &= (M/RX)\sin^2\Theta.
\end{aligned} \tag{77}$$

(Here and below, only the independent nonvanishing components are listed.) The computation then yields

$$\begin{aligned}
R_{1010} &= 2M/R^3,\\
R_{2020} &= R_{3030} = -M/R^3,\\
R_{1212} &= R_{1313} = M/R^3,\\
R_{2323} &= -2M/R^3.
\end{aligned} \tag{78}$$

Some of the simplicity of Eq. (78) as compared to Eq. (77) was, of course, to be expected, for the Fermi frame is orthonormal so that all components must at least have the same dimensions, and the equivalence of the Θ and Φ directions must become evident. However, a very surprising feature is that the gravitational field gradients in Eq. (78) depend on the observer's position R, but *not* upon his velocity R' (or energy k) with respect to the mass M. Thus the preferred rest frame indicated locally by the Killing vector field $\partial/\partial T$ cannot be recognized by an observer who measures all the gravitational field gradients (78) at one point. He can only discover the direction of the vector $\partial/\partial T$ by finding a velocity (i.e. direction in the $R - T$ plane) which makes the field gradients constant in time, i.e., by measuring $R_{\mu\nu\sigma\tau;\alpha}$.

The Fermi normal metric from Eq. (66) is

$$\begin{aligned}
ds^2 = &-\left[1 + \frac{M}{R^3}(y^2 + z^2 - 2x^2)\right]dt^2\\
&- \frac{2M}{3R^3}[xz\,dx\,dz + xy\,dx\,dy - 2yz\,dy\,dz]\\
&+ \left[1 + \frac{M}{3R^3}(y^2 + z^2)\right]dx^2\\
&+ \left[1 + \frac{M}{3R^3}(x^2 - 2z^2)\right]dy^2\\
&+ \left[1 + \frac{M}{3R^3}(x^2 - 2y^2)\right]dz^2.
\end{aligned} \tag{79}$$

The entire dependence of this metric on t is through the geodesic equation (72) which gives $R(t)$.

A more compact form for the Fermi metric (79) is obtained by introducing spherical coordinates r, θ, φ related to x, y, z by the standard formulas. Taking the x direction as the polar axis we get a diagonal metric,

$$\begin{aligned}
ds^2 = &-(1 - q\mu)\,dt^2 + dr^2 + (1 + \tfrac{1}{3}\mu)(r\,d\theta)^2\\
&+ (1 + \tfrac{1}{3}q\mu - \tfrac{1}{3}\mu)(r\sin\theta\,d\varphi)^2,
\end{aligned} \tag{80}$$

where

$$\mu = Mr^2/R^3, \tag{81a}$$

and

$$q = 3\cos^2\theta - 1. \tag{81b}$$

Again, R must be considered the function of t given in Eqs (72), or equivalently one may take R as the time coordinate and use Eq. (73) to eliminate dt^2 in favor of dR^2 in Eq. (80).

In the following paper[8] this metric provides boundary conditions for a computation of tidal deformations of a freely falling Schwarzschild singularity (wormhole mouth). It is also evidently well suited to a calculation of tides in an elastic test body whose center of mass would define the geodesic $x^i = 0$. We content ourselves here with

a mathematical example and investigate the shape of a sphere. Define a sphere Σ as the surface formed by all points a fixed *proper* distance r measured out orthogonally from some point on the central geodesic. For the coordinates of Eq. (80) this is the surface $t = \text{const}$, $r = \text{const}$, whose metric is, therefore,

$$(ds^2)_\Sigma = (1 + \tfrac{1}{3}\mu)(r\,d\theta)^2 + (1 + \tfrac{1}{3}q\mu - \tfrac{1}{3}\mu)(r \sin\theta\, d\varphi)^2. \quad (82a)$$

From this metric we find that the area of the sphere Σ is just $4\pi r^2$, independent of the small quantity $\mu = Mr^2/R^3$ in first order, but a change in intrinsic shape can be readily computed. The length of a great circle $\varphi = \text{const}$ over the poles of this sphere is

$$L_{\text{poles}} = r\int_0^{2\pi} (1 + \tfrac{1}{3}\mu)^{\frac{1}{2}}\, d\theta \approx 2\pi r(1 + \tfrac{1}{6}\mu). \quad (82b)$$

Similarly, the circumference of the equator, $\theta = \frac{1}{2}\pi$, is

$$L_{\text{equator}} \approx 2\pi r(1 - \tfrac{1}{3}\mu). \quad (82c)$$

As a measure of the distortion of the shape of this sphere, then, we may take

$$\eta = \frac{L_{\text{poles}} - L_{\text{equator}}}{L_{\text{poles}} + L_{\text{equator}}} \approx \frac{\mu}{4} = \frac{Mr^2}{4R^3}. \quad (83)$$

Thus, a sphere $r = \text{const}$ is a surface shaped like a football pointing toward the center of gravitation.

IX. RANGE OF VALIDITY OF THE FERMI EXPANSION

In this section we point out that in most situations where the Fermi metric expanded through quadratic terms is a useful description, the time dependence of the metric can be considered adiabatic, that is, time derivatives of the metric will be negligible in comparison to space derivatives. Order of magnitude-wise, the Fermi metric can be written

$$g \simeq 1 + r^2K(t) + r^3\left.\frac{\partial K}{\partial r}\right|_G + O(r^4), \quad (84)$$

where r is proper distance normal to the geodesic, t is proper time along the geodesic, and K represents a typical component of the curvature tensor. For this metric, the ratio of time to space derivatives (computed from the r^2 term only) is

$$\frac{\partial g/\partial t}{\partial g/\partial r} = \frac{r}{K}\frac{\partial K}{\partial r}. \quad (85)$$

But if we assume that the quadratic terms in (84) are an adequate approximation to the metric, then the cubic terms must be negligible in comparison to the quadratic ones,

$$\frac{r}{K}\frac{\partial K}{\partial r} \ll 1. \quad (86)$$

This small quantity is almost the one appearing in Eq. (85), except a space and time derivative are interchanged. But as a sort of causality condition, one expects that

$$\partial K/\partial t \lesssim \partial K/\partial r, \quad (87)$$

for in the contrary case, a disturbance would appear spontaneously at some point ($\partial K/\partial t$ large) without having arrived there as a wave propagating with velocity less than $c = 1$. Thus the quadratic Fermi approximation (86) together with causality in the sense of Eq. (87) imply in Eq. (85) the adiabatic condition

$$\partial g/\partial t \ll \partial g/\partial r. \quad (88)$$

These results can be specialized to the Schwarzschild case and give some surprises. For this metric we have from Eq. (78)

$$K = M/R^3\,. \quad (89)$$

Using Eqs. (74) for $\partial/\partial t$ and $\partial/\partial x$ we can test the causality conditions (87) and find that it reads

$$\frac{R'}{XT'} = \left[\left(\frac{2M}{R} - \frac{2M}{R_0}\right)\Big/\left(1 - \frac{2M}{R_0}\right)\right] \leq 1, \quad (90)$$

which is always *violated* for $R < 2M$. [Other situations which violate the causality condition of Eq. (87) are the expanding-universe cosmological models where one assumes $\partial K/\partial r = 0$.] A condition which will ensure the validity of the Fermi expansion is

$$Kr^2 = Mr^2/R^3 \ll 1, \quad (91)$$

and this can be satisfied by taking (r/R) small enough even if (M/R) is large. Thus the Fermi expansion is useful even inside the Schwarzschild "singularity." The adiabatic condition computed from Eqs. (85) and (74) reads

$$\frac{\partial g/\partial t}{\partial g/\partial r} = \frac{r}{R}R' \simeq \left(\frac{Mr^2}{R^3}\right)^{\frac{1}{2}} \ll 1, \quad (92)$$

and is satisfied as a consequence of Eq. (91) which is a stronger convergence requirement than Eq. (86).

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 4, NUMBER 6 JUNE 1963

Distortion in the Metric of a Small Center of Gravitational Attraction due to its Proximity to a Very Large Mass

F. K. MANASSE†
Palmer Physical Laboratory, Princeton University, Princeton, New Jersey
and
Bell Telephone Laboratories, Holmdel, New Jersey
(Received 31 May 1962)

The problem of two bodies in general relativity in an especially restricted mode is analyzed. One body is of small mass, and, under the influence of gravitational attraction, moves toward a much larger mass whose field produces "tidal" deformations in the geometry of the smaller one. To evaluate these deformations, we treat the Schwarzschild metric of the particle by a perturbation analysis similar to that performed by Regge and Wheeler. The boundary conditions for this analysis are obtained from the metric of the background field expressed in a novel set of comoving coordinates, here called Fermi normal coordinates. These coordinates have been described in the previous paper. In this paper we use the Schwarzschild metric given there in comoving coordinates as an asymptotic approximation for the full solution which is evaluated here.

A perturbation analysis using the "tidal deformation" (background curvature) as an expansion parameter is performed on the metric of a small Schwarzschild particle. The solution so obtained satisfies Einstein's equations for empty space for small deviations from a nonflat metric. This solution also reduces to the Fermi metric in the appropriate limit, namely, when the "distance" from the geodesic is large compared to the radius of the small mass but small compared to the separation of the two masses. Thus only the quadratic terms in the distance are important. This solution is then used to find the deformations in shape of the throat of the wormhole and to locate this region of symmetry (the throat) by finding a coordinate transformation which leaves the metric invariant.

I. INTRODUCTION

ONE of the most important developments in general relativity from Einstein's first papers to the present, was the discovery by Einstein and Grommer,[1] and by Einstein, Infeld, and Hoffman,[2] that the equations of motion of test particles can be determined from the field equations and do not have to be postulated independently. In more recent times one has tried to secure a better understanding of the character of the objects to which one applies such derivations.[3-8] Are they to be treated as singularities in the metric? As geons? As objects susceptible to break up? On one of these pictures—that of a singularity—new insight has been acquired from the work of Fronsdal[8] and Kruskal.[4] They find that the geometry at the center of attraction, for a single massive body, pursued systematically backwards closer to the center is *free of singularity*, but endowed with an unusual topology. Moreover, it remains free of singularity for a finite proper time. Their study gives a simple model to use for mass with a view to understanding more about the problem of motion. That is, mass arises from the energy of the gravitational or metric field itself strongly curved in the immediate vicinity of the bridge between the two nearly Euclidian spaces which form the two sheets of the space. This development brings to the fore the question of what happens to a small test mass of this character when it approaches a large center of attraction of the same character. Does it remain stable in view of the very large disruptive "tidal forces", or is it crushed out of existence?[9,10]

The question just asked about masses of purely geometrical origin can also be reformulated for the kind of masses which one sees in the physical world. An analogy of this kind would be a small mass of A_1 nucleons ($A_1 \ll A_{crit}$) which approaches a cold star (large mass) of the critical mass ($A_2 = A_{crit}$). Will the small mass be crushed out of existence?

† Based on a dissertation presented to the faculty of Princeton University in candidacy for the Ph.D. in Physics, September, 1961.

[1] A. Einstein and J. Grommer, Akad. Wiss. Preuss. Math. Phys. Kl. Sitz 2 (1927).

[2] A. Einstein, L. Infeld, and B. Hoffman, Ann. Math., 39, 65 (1938).

[3] J. A. Wheeler, Rev. Mod. Phys., 33, 63 (1961).

[4] M. D. Kruskal, Phys. Rev. 119, 1743 (1960).

[5] J. L. Synge and A. Schild, *Tensor Calculus* (Toronto University Press, Toronto, Canada, 1949).

[6] J. L. Synge, *Relativity, the General Theory* (North-Holland Publishing Co., Amsterdam, 1960).

[7] L. Infeld and J. Plebanski, *Motion and Relativity* (Pergamon Press Inc., New York, 1959).

[8] C. Fronsdal, Phys. Rev., 116, 778 (1959).

[9] V. A. Fock, *Theory of Space, Time and Gravitation* (Pergamon Press, Inc., New York, 1959).

[10] J. A. Wheeler and R. W. Lindquist, Rev. Mod. Phys. 29, (1957).

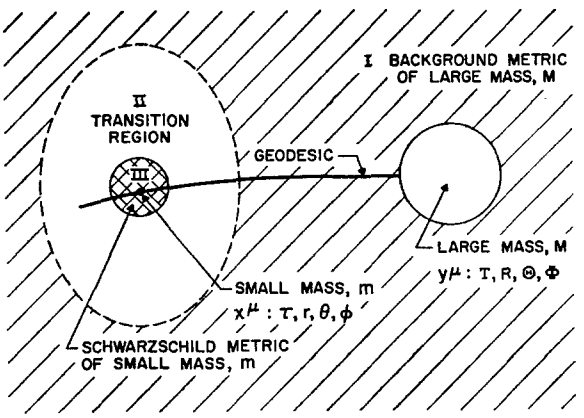

FIG. 1. Three regimes in restricted problem of two bodies. The three different regions considered in the problem are shown above. In Region I, the small mass has negligible effect and is treated as an ideal test particle moving on a geodesic of the background field. This field is that due to a center of attraction of large mass, M, and is a Schwarzschild field in the coordinates of the large mass y^μ. In Region III, immediately surrounding the small mass m, the field of the large mass is negligible in comparison with the self field of the particle and is neglected. Thus we have a Schwarzschild line element in the parameters and coordinates of the small mass. Region II is the transition region where both masses exert influence, and is the principal region of interest in this problem.

Will it amalgamate or stick to the large mass and lose its identity?[11]

These questions are very timely and their answers must be known if we are to learn more about primitive models for mass, and of their behavior in strong gravitational fields.

One simple technique suggested itself in the investigation of the basic issues of this problem: the dynamics of a small mass moving in the field of a large one. We consider the case of two bodies, one of very small but *not negligible* mass which is attracted toward a large mass idealized as stationary. We find a metric which accurately represents the gravitational field in which the small mass finds itself. The procedure is to analyze the perturbations of the metric around the small mass caused by the gravitational field of the large mass. The expansion parameter of the perturbations is proportional to the curvature of the space caused by the gravitational field of the large mass. We obtain the solution described above by analyzing the problem in three parts or regions (Fig. 1). Region I is far enough from the small mass that its effects can be neglected. The geometry is described by the Schwarzschild metric caused by the large mass. This metric is here expanded in powers of the deviation from the geodesic of an ideal test particle (Fermi normal coordinates). Region III is close enough to the small mass that the effects of the large mass are negligible by comparison. Here one has the Schwarzschild metric associated with the small mass. In Region II, both masses exert significant influence on the dynamics. It is here that the metric undergoes transition from I to III.

The first part of this paper deals with the solution in Region I. That is, we obtain a metric which is valid "far" from the small mass assuming it to be an ideal infinitesimal test particle. This is essentially a problem of translating the ordinary Schwarzschild metric associated with the large mass into a coordinate system comoving with the test particle. The previous paper describes the system of coordinates (Fermi normal coordinates) and gives the techniques necessary for the transformation of general metrics into these coordinates. That paper also exhibits the Schwarzschild metric as expressed in these comoving coordinates. The metric there described correctly represents at all times the curvature at the test particle produced by the large mass, but neglects curvature gradients. The second part of this paper deals with the transition region. In this we replace the ideal test particle by a small Schwarzschild-like mass and perform a perturbation analysis on its metric similar to that performed by Regge and Wheeler.[12] The boundary conditions on this problem are, however, not those of the above authors but are instead, exactly those of the solution to Region I previously obtained. Thus, nonlinear perturbations of the Schwarzschild metric at the small mass are not considered, nor are linear perturbations induced by curvature gradients of the field of the large mass. We also do not consider the reaction of the large mass to the small one in this paper. Thus, the problem that is here solved is essentially that of finding the effect on the metric of a small Schwarzschild mass embedded not in a flat space, but rather in a curved background space caused by the presence of the large mass.

The effects that we look for can be qualitatively discussed here. In Region II we have the largest effect on the motion and metric of the small mass due both to its own field and that produced by the presence of the large mass. The deviation in the pure Schwarzschild caused by the small mass is essentially given by terms which are proportional to the tidal force exerted between two bodies.[13] Let us obtain an order-of-magnitude estimate of this tidal force. Let M represent the mass of the

[11] J. A. Wheeler, Varenna Lectures, Rend. Scuola Intern. Fis. "Enrico Fermi," **11**, 67 (1960).

[12] T. Regge and J. A. Wheeler, Phys. Rev. **108**, 1063 (1957).

[13] Encyclopedia Britannica, (1911), 11th ed., Vol. XXVI.

large body and m that of the little one. Let their separation—suitably defined—be called R. Then, the departure from "flatness"—more precisely, the departure of the metric coefficients from their Lorentz values—is of the order of Mm^2/R^3 in Region II. This causes a departure from "sphericity" —i.e., a distortion in shape—for the small mass of the same magnitude. As R tends toward zero, the effect becomes larger regardless of the smallness of size of m so long as it is nonzero. Some rather more quantitative results are listed in Table 1. The entries give order-of-magnitude effects for the distortions due to the "tidal forces" described in this paper. It can be seen from this table that, in Region II, the deviation from flatness is of the same order of magnitude as that due to pure Schwarzschild fields in either Region I or III. However, here it is due to the *combined* effects of both masses. A typical component of the metric in Region II will be shown below to behave as

$$g \sim [1 - (2m/r)][1 + (Mr^2/R^3)]. \tag{1.1}$$

In the Region I the small mass is negligible and thus we have that the metric and consequently the curvature associated with it consists only of the terms in the second parentheses. Here we must remember that our metric is expressed in Fermi normal coordinates which accounts for its unusual form. Similarly in Region III the large mass exerts only a negligible effect and thus our metric and curvature depend only on the terms in the first parentheses. Here the ordinary expression for the Schwarzschild center of gravitation appears. In Region II, however, the metric and the curvature associated with it are very much dependent on both m and M. That is, let us call m/r^3 the curvature associated with the space about just the small mass, the *intrinsic* curvature, and M/R^3 the curvature associated with the space about just the large mass, the *extrinsic* curvature. Then, the curvature in Region II, for appropriately chosen values of the parameter M, R, m, r, can be varied over a rather large range since it is a direct combination of the intrinsic and extrinsic curvature. All this is shown in quantitative detail in the body of the paper.

The attack described above can be summarized in the following manner. We "break up" our space into three different regions (Fig. 1). Region I is that region about the large center of gravitation, M, where the small mass, m, has a negligible effect. There, the metric has the familiar Schwarzschild character, as if the large mass alone were present. We consider the small mass as an ideal test particle which thus moves on a geodesic toward the large mass. By using Fermi normal coordinates, we transform the metric into the coordinates relative to the test mass (comoving coordinates). We thus obtain a metric which describes an ideal test particle imbedded in a background space of local curvature $\sim M/R^3$. Region III, very near the small mass, is described by a metric which in first approximation neglects the effect of the large mass, M, and is thus an ordinary Schwarzschild metric in the coordinates of the small mass, m. The curvature at the distance r is of the order m/r^3. This curvature reaches a maximum value $\sim 1/m^2$ at the throat of the metric associated with the small mass. By performing a perturbation analysis on this metric we arrive at the metric of the intermediary Region II where both masses are taken into account. The boundary conditions for solving the perturbation equations are provided by the metric of Region I in our Fermi coordinates.

II. THE SCHWARZSCHILD PROBLEM FOR ONE BODY IN COMOVING COORDINATES

The previous paper has given a scheme of coordinates and an expression for the Schwarzschild metric in terms of these coordinates.[14] For the special two-body problem to be discussed below, this metric will serve as the boundary condition of our perturbed metric about a small Schwarzschild center of attraction. The metric as obtained from that paper is

$$d\tau^2 = \left[1 - \frac{Mr^2}{R^3}(3\cos^2\theta - 1)\right] \frac{dR^2}{\left(\frac{2M}{R} - \frac{2M}{R_0}\right)} - dr^2 - r^2\left[1 + \frac{Mr^2}{3R^3}\right] d\theta^2 - r^2 \sin^2\theta \times \left[1 + \frac{Mr^2}{3R^3}(3\cos^2\theta - 2)\right] d\varphi^2. \tag{2.1}$$

If we change the notation somewhat for compactness, we have

$$d\tau^2 = (1 - Q) \frac{dR^2}{\left(\frac{2M}{R} - \frac{2M}{R_0}\right)} - dr^2 - (1 + \tfrac{1}{3}\mu)r^2\, d\theta^2 - (1 + \tfrac{1}{3}(Q - \mu))r^2 \sin^2\theta\, d\varphi^2. \tag{2.2}$$

Here we have put

$$Q = +(3\cos^2\theta - 1)Mr^2/R^3 = q\mu, \qquad \mu = +Mr^2/R^3, \tag{2.3}$$

[14] F. K. Manasse and C. W. Misner, J. Math. Phys. 4 735 (1963) (previous paper).

where Q is a quadrupole term and μ a monopole term. To compute the curvature components referred to our new coordinates R, r, θ, φ, we must use the orthonormal tetrad

$$\begin{aligned} W^{\mu}_{(0)} &= (1, \quad 0, \quad 0, \quad 0), \\ W^{\mu}_{(1)} &= (0, \quad \cos\theta, \quad \sin\theta\ \cos\varphi, \sin\theta\sin\varphi), \\ W^{\mu}_{(2)} &= (0, -\sin\theta, \quad \cos\theta\ \cos\varphi, \cos\theta\sin\varphi), \\ W_{(3)} &= (0, \quad 0, -\sin\varphi, \cos\varphi), \end{aligned} \tag{2.4}$$

where our vectors, in notation similar to that of the previous paper, are defined by unit vectors along the t, r, θ, φ coordinate lines as

$$\begin{aligned} W_{(0)} &\equiv W_{(t)} \equiv \partial/\partial t|_{r=0}, \\ W_{(1)} &\equiv W_{(r)} \equiv \partial/\partial r|_{r=0}, \\ W_{(2)} &\equiv W_{(\theta)} \equiv \partial/r\ \partial\theta|_{r=0}, \\ W_{(3)} &\equiv W_{(\varphi)} \equiv (\sin\theta)^{-1}\ \partial/r\ \partial\varphi|_{r=0}. \end{aligned} \tag{2.5}$$

Here we again have put

$$dt^2 = \frac{dR^2}{2M(1/R - 1/R_0)}, \tag{2.6}$$

since from the previous paper, these are equivalent. The curvature components evaluated in this coordinate system are

$$\begin{aligned} R_{1010} &= +Q/r^2, \quad & R_{2020} &= -(Q-\mu)/r^2, \\ R_{3030} &= -\mu/r^2, \quad & R_{2323} &= -Q/r^2, \\ R_{1212} &= +\mu/r^2, \quad & R_{1313} &= +(Q-\mu)/r^2, \\ & & R_{1323} &= R_{1020} = -(3\mu\sin 2\theta)/2r^2. \end{aligned} \tag{2.7}$$

Here are exhibited all the independent nonvanishing components of the Riemann curvature tensor evaluated on the given radial geodesic described by

$$R, r = 0, \theta, \varphi. \tag{2.8}$$

That we do indeed have a reasonable metric can be easily checked, for on the geodesic line we must have

$$r = 0, \qquad R = R(t). \tag{2.9}$$

If we put this into expression (2.1), we obtain

$$d\tau^2|_G = \frac{dR^2}{(2M/R) - (2M/R_0)} \equiv dt^2. \tag{2.10}$$

Thus on the given line we do indeed obtain the correct result. In addition, if the center of attraction is removed, that is $M \to 0$, the metric reduces to

$$d\tau^2 = dt^2 - dr^2 - r^2\,d\theta^2 - r^2\sin^2\theta\,d\varphi^2. \tag{2.11}$$

This is just the Lorentz metric in spherical coordinates, and states that geodesics in flat space are straight lines as required.

The advantage of exhibiting the metric in the form 2.1 as against a similar form using Riemann normal coordinates will now be described. First, our result is true over a much wider range, namely, in a region surrounding a line rather than just near a point. This larger region of validity makes the metric much more useful to us. Second, our problem of "two bodies," as it is proposed in this paper, assumes the smaller one to move on, or very near, a godesic of the background space due to the large mass. Thus, our metric "accurately" represents the "asymptotic" metric of the small mass, or its metric in Region I (Fig. 1) of the space. Also, the axial symmetry of the metric about the geodesic line allows us to effectively and naturally eliminate the local time coordinate of the large mass from consideration. Its effect is lumped into our new "effective" local time coordinate R which determines essentially our radial distance from the large mass. Thus, if we specify only R, we already specify everything about the gross structure of the position of the two bodies. Third, our metric is valid everywhere in the region surrounding the geodesic line, to the second order in small departures from this line, regardless of the value of R, for a massless test particle. The metric expansion as expressed in (1) will serve as the boundary condition for the perturbed Schwarzschild problem considered below.

III. RESTRICTED TWO-BODY PROBLEM

A. Range of Validity of the Fermi Expansion in the One-Body Problem

The previous section has been concerned with the well-known Schwarzschild one-body problem. Utilizing the techniques of the previous paper, one has cast this solution into a form which is (1) no longer "static", since the metric components now depend on the time parameter explicitly, and (2) no longer exact since the metric holds only for values of r which are "small". The range of validity of the metric given by expression 2.1 needs to be discussed further before we can proceed to the restricted two-body problem. The previous paper has already discussed this to some extent but we treat it here again somewhat differently for ease of reference. Here also, we specifically concern ourselves with discussing this range of validity for a perturbation problem which follows. The solution 2.1 holds provided that the terms of order $(x)^3$ and all higher orders of x can be neglected in the metric expansion.

The criterion is essentially that

$$g_{\alpha\beta,ijk}|_G\, x^i x^j x^k \ll g_{\alpha\beta,ij}|_G\, x^i x^j. \tag{3.1}$$

Here we have left out constant factors. Now from our previous work we know that the $g_{\alpha\beta,ij}|_G$ are fixed by the components of the curvature tensor.[14] By a similar process we can easily show[15] that the $g_{\alpha\beta,ijk}$ are given by the gradients of the curvature components. That is, the terms which we must compute are terms like

$$R_{i\alpha\beta j,k}|_G\, x^i x^j x^k \ll R_{i\alpha\beta j}|_G\, x^i x^j. \tag{3.2}$$

Now the curvature components for the Schwarzschild case all go as

$$R_{i\alpha\beta j}|_G \sim (M/R^3) = K \sim (R_{\mu\gamma\alpha\beta}R^{\mu\gamma\alpha\beta})^{\frac{1}{2}}. \tag{3.3}$$

Here, we denote a typical nonzero curvature component by K, the Gaussian curvature. The metric will be well represented by the first two terms in the Fermi expansion if and only if the following condition is satisfied:

$$|r^3(\partial K/\partial r)| \ll r^2 K. \tag{3.4}$$

Here r is a typical spatial coordinate of our space. This expression can be cast into the form

$$|(r/K)(\partial K/\partial r)| \ll 1, \tag{3.5}$$

or

$$|(\partial/\partial r)(\ln K)| \ll 1/r. \tag{3.6}$$

The expression (3.5) is the relation which defines the field of applicability of the Fermi expansion. If we are to use the expansion as a basis for a perturbation expansion (as we do later), then an even stricter requirement is imposed. This condition is that the term Kr^2 in the metric expansion must itself be small compared to 1. That is, for a Fermi expansion to be valid we must have relation (3.5), and if it is to be used as a boundary condition for a perturbation problem, we must in addition have

$$Kr^2 \ll 1. \tag{3.7}$$

This condition (3.7) includes (3.5) and is in fact a more restrictive one. A more precise method of stating this result is outlined below. Let

$$Kr^2 = (M/R^3)r^2 = \epsilon_{10}. \tag{3.8}$$

Here ϵ_{10} means (see Table I) that Region II covers a 10:1 range of the coordinate r. Then we have that

$$r \lesssim (R^3\epsilon_{10}/M)^{\frac{1}{3}}. \tag{3.9}$$

Now consider the Fermi condition (3.5):

$$\frac{r}{K}\frac{\partial K}{\partial r} \sim \frac{r}{R} \sim \left(\frac{R\epsilon_{10}}{M}\right)^{\frac{1}{3}}. \tag{3.10}$$

However, we have from Table I, that

$$\epsilon_{10} \sim \epsilon' = (m/M)^{\frac{1}{3}}(M/R), \tag{3.11}$$

where we consider orders of magnitude only. Thus we have finally

$$\left|\frac{r}{K}\frac{\partial K}{\partial r}\right| \sim \left(\frac{m}{M}\right)^{\frac{1}{3}} \equiv (\alpha)^{\frac{1}{3}} \ll 1. \tag{3.12}$$

We thus see that if

$$Kr^2 = \epsilon_{10} \ll 1, \tag{3.13}$$

then we must have also condition (3.5) holding since α must be small for our problem as set up to be valid.

In the restricted two-body problem, where one mass is small compared to the other, we have just this condition holding. That is, we know that for distances of the order of M, from the large mass, the small mass has no effect on the metric of the large one. Similarly, in order to determine the metric *very near* the small mass m, we do not need to know the mass M. In the intermediary space however, where

$$r \sim [10^p]m \ll M, \qquad p > 0, \tag{3.14}$$

we wish to find the metric. Now the criterion becomes

$$\epsilon_{10} = Kr^2 = (M/R^3)m^2 10^{2p} \ll 1. \tag{3.15}$$

If we now concern ourselves again with a limiting process, regardless of how large we make p, we can find an m such that the ratio $\alpha = m/M$ is so small that ϵ_{10} remains much smaller than 1. Therefore, so long as we keep $(m/M) \ll 1$, we will satisfy the condition (3.7) and as long as we keep $R \sim M$ and $r \sim 10^p m$, we can rightly drop all terms in the metric expansion of order in r higher than two. See Table I for more details.

B. The Perturbed One-Body Problem

For the restricted two-body problem considered—a small mass moving toward a large stationary mass along a geodesic—we propose now a method of solution. We perturb the ordinary one-body Schwarzschild solution for the small mass m in the manner of Regge and Wheeler.[12] The boundary conditions, and the choice of independent perturbation amplitudes are chosen to match "asymptotically" with the one-body Schwarzschild of large mass in Fermi or comoving coordinates. So long as

[15] L. P. Eisenhart, *Riemannian Geometry* (Princeton University Press, Princeton, New Jersey, 1926).

TABLE 1. Regimes in analysis of metric in two-body problem (Due to J. A. Wheeler).

	Regime III	Regime II	Regime I
Property	m dominates; effects of M negligible	departure from flatness due to m *and* to M are *both* small (curvature times r^2 small compared to 1)	M dominates; effects of m negligible
Departure from flatness caused by M at distance r from test mass ($\sim Mr^2/R^3$)	$\ll\epsilon_{10}$	$<\epsilon_{10}$ | $\sim\epsilon_{10}$ | $>\epsilon_{10}$; at $r_{I-II} = (R^3\epsilon_{10}/M)^{\frac{1}{3}}$	
Departure from flatness caused by m at distance r from test mass ($\sim m/r$)	$>\epsilon_{10}$ | $\sim\epsilon_{10}$ | $<\epsilon_{10}$; at $r_{II-III} = m/\epsilon_{10}$		$\ll\epsilon_{10}$
Condition that $r_{I-II} = 10r_{II-III}$	$\epsilon_{10} = (10m/M)^{2/3}(M/R)$		
Values of interzone radii for this value of ϵ_{10}	$r_{II-III} = 10^{-2/3}(m/M)^{1/3}R$; $r_{I-II} = 10^{1/3}(m/M)^{1/3}R$		
Deviation from flatness caused by M	$10^{-4/3}(m/M)(M/R)$		$10^{2/3}(m/M)^{2/3}(M/R)$
Deviation from flatness caused by m	$10^{2/3}(m/M)^{2/3}(M/R)$		$10^{-1/3}(m/M)^{2/3}(M/R)$
Maximum value of qualitative expression for typical term in metric	Maximum of $[1 - (2m/r)][1 - (Mr^2/R^3)]$ $\doteq [1 - 2(m/M)^{2/3}(M/R)][1 - (m/M)^{2/3}(M/R)]$ $\doteq 1 - 3(m/M)^{2/3}(M/R) = 1 - 3\epsilon_{10}/10^{2/3}$ (deviation from flatness $3\epsilon_{10}/10^{2/3}$)		
Total deviation from flatness in units $(m/M)^{2/3}(M/R)$	$\sim 10^{-4/3} + 10^{2/3}$	$\sim 1 + 2 = 3$	$\sim 10^{2/3} + 10^{-1/3}$
Relevant type of approximation	Schwarzschild expression associated with small mass	Schwarzschild expression associated with small mass plus small corrections matched to metric in Region I	Schwarzschild expression associated with large mass expressed in Fermi coordinates referred to world line of ideal infinitesimal test particle.
Accuracy achieved in description of metric for 2-body problem	$\epsilon' \sim (m/M)^{2/3}(M/R)$ (for the given R and for all larger values of R).		
Value to which m is limited if 2-body problem is to be treated to error less than ϵ' from $R = \infty$ down to $R = R$	$m < \epsilon'^{3/2}(R/M)^{3/2}M$		
	$\epsilon_{10} = 10^{2/3}\epsilon'$		

we choose the ratio $\alpha = m/M$ sufficiently small, (see previous section) we will be able to match the solutions at the boundaries. This is certainly reasonable for analogy with any astrophysical problem of interest, say an earth satellite, or a comet approaching the sun, etc. Thus, we expect that our metric in the intermediary Region II (Fig. 1) will, in the limit r "large" go over into expression (2.1), and in the limit r "small", go into the ordinary Schwarzschild of mass m.

We begin with the expression for the metric of a point mass m, the familiar Schwarzschild line element. The expression is

$$y^\mu: t', r, \theta, \varphi$$

$$(ds')^2 = -[1 - (2m/r)](dt')^2 + [1 - (2m/r)]^{-1}\,dr^2 + r^2\,d\theta^2 + r^2\sin^2\theta\,d\varphi^2. \qquad (3.16)$$

(The notation throughout the paper, except where specifically stated otherwise, is: Greek indices run

from 0 to 3, Latin indices run from 1 to 3, commas indicate ordinary differentiation, semicolons indicate covariant differentiation, and the summation convention on repeated indices is utilized.) If we normalize the coordinates for ease of writing by dividing through by $4m^2$ and renaming the coordinates, we have

$$x^\mu: t, x, \theta, \varphi$$

$$ds^2 = -\left[1 - \frac{1}{x}\right] dt^2 + \frac{1}{(1 - 1/x)}\, dx^2 + x^2(d\theta^2 + \sin^2\theta\, d\varphi^2). \tag{3.17}$$

If we follow Regge and Wheeler,[12] and perform a perturbation of the metric, we will go from the above metric denoted by $g_{\mu\nu}$ to one which can be written as

$$g'_{\mu\nu} = g_{\mu\nu} + h_{\mu\nu}. \tag{3.18}$$

If we make the $h_{\mu\nu}$ small in comparison with the $g_{\mu\nu}$, then we can write the Einstein equations correct to first order as

$$\delta R_{\mu\nu} \equiv \delta\Gamma^\beta_{\mu\beta;\nu} - \delta\Gamma^\beta_{\mu\nu;\beta} = 0, \tag{3.19}$$

where we have put

$$\delta\Gamma^\alpha_{\beta\gamma} = \tfrac{1}{2}g^{\alpha\nu}(h_{\beta\nu;\gamma} + h_{\gamma\nu;\beta} - h_{\beta\gamma;\nu}). \tag{3.20}$$

We analyze these equations, separate them into polar coordinates, and inquire as to the most general perturbation which we can have. Now Eq. (3.19) is analogous to the general equation for a massless, spin-two particle in quantum physics—that for a graviton. Under rotations on the two-dimensional submanifold formed by holding $x^0 = t$ = constant, $x^1 = x$ = constant, the ten independent components of a symmetric tensor transform as three scalars (h_{00}, h_{11}, h_{01}), two vectors (h_{02}, h_{03}, h_{12}, h_{13}), and a second-order tensor. The eigenvalues of the expansion are those of the angular-momentum operator, and we obtain two independent odd-parity modes and four independent even-parity modes (this count is detailed below). This expansion is useful for our problem, and we then expand the metric components in terms of generalized spherical harmonics. We obtain a perturbation of odd-parity type and one of even parity. These most general expansions of the $h_{\mu\nu}$ are given with eigenvalues corresponding to fixed: L, angular momentum; μ, projection along z-axis of angular momentum; and P, parity multiplied by $(-1)^{L+1}$. However, different waves can represent the same physical phenomena but viewed in a different coordinate system. Thus, we can still make infinitesimal coordinate transformations. If we choose to eliminate the most complicated of the $h_{\mu\nu}$ which we can do with our coordinate transformations, specialize to the case where $\mu = 0$ since our base metric is independent of φ, and choose a perturbation of definite normalized frequency $k = k'/2m = (\omega/c)/2m$, our perturbation metric becomes finally,

$$h_{\mu\nu} = \begin{pmatrix} 0 & 0 & 0 & h_0(x) \\ 0 & 0 & 0 & h_1(x) \\ 0 & 0 & 0 & 0 \\ h_0(x) & h_1(x) & 0 & 0 \end{pmatrix} \times e^{-ikt} \sin\theta(\partial P_L/\partial\theta) \tag{3.21}$$

for the odd perturbation, where P_L denotes the Legendre polynomial. The even purturbation becomes

$$h_{\mu\nu} = \begin{pmatrix} -\left(1 - \frac{1}{x}\right)H_0(x) & H_1(x) & 0 & 0 \\ H_1(x) & \left(1 - \frac{1}{x}\right)^{-1} H_2(x) & 0 & 0 \\ 0 & 0 & x^2 H_3(x) & 0 \\ 0 & 0 & 0 & x^2 \sin^2\theta H_3^{(x)} \end{pmatrix} \times e^{-ikt} P_L. \tag{3.22}$$

Thus we have, as previously noted, two odd-parity independent components h_0, h_1, and four even-parity independent components H_0, H_1, H_2, H_3. These correspond to the *gross* number of degrees of freedom of the problem.

If we now substitute these most general perturbations into the Einstein equations, we obtain equations for the unknown radial functions which depend on two parameters k and L. Several of the radial equations of even parity as given by Regge and Wheeler[12] were improperly evaluated. Consequently, their resulting solutions for the static case ($k = 0$) for example, do not satisfy the Einstein equation as they should.

C. Justification of Method of Analysis

The metric which is asymptotically to be matched with the solution to the Regge–Wheeler problem is our metric of the ideal test particle moving under the influence of a large mass expressed in comoving coordinates. This metric is from (2.1), with a change of notation for compactness:

$$x^\mu\colon \rho,\ x,\ \theta,\ \varphi$$

$$ds^2 = -\left(1 - \frac{qx^2}{2\alpha\rho^3}\right)\frac{d\rho^2/\alpha}{(1/\rho - 1/\rho_0)} + dx^2 + x^2\left(1 + \frac{qx^2}{6\alpha\rho^3}\right)d\theta^2 + x^2\sin^2\theta\left(1 + \frac{(q-1)x^2}{6\alpha\rho^3}\right)d\varphi^2. \quad (3.23)$$

Here we have divided (2.1) by $4m^2$, and called $R/2m = \rho$, $r/2m = x$, $(d\tau)^2 = -(ds')^2/4m^2 = -ds^2$, $\alpha = m/M$, and $q = 3\cos^2\theta - 1$. If we desire to test Einstein equations for this metric, or rather for a metric which is asymptotically this one, we must first compute curvature components.

In computing these curvature components, which are necessary to obtain Einstein equations, we will need to compute both the space and the "time" derivatives of the $g_{\mu\nu}$ given by (3.23). These time derivatives will now be compared in magnitude with spatial derivatives in order to obtain a very interesting and important result. Neglecting angular dependence, the $g_{\mu\nu}$ with which we are concerned all go like $(x^2/\alpha\rho^3) \sim r^2K$. Now compute this "time" derivative:

$$\left|\frac{\partial}{\partial t}(r^2K)\right| = \left|r^2\frac{\partial K}{\partial t}\right| = \frac{3r^2M}{R^4}\frac{\partial R}{\partial t}. \quad (3.24)$$

Now recall from (2.6) that

$$dt^2 = \frac{dR^2}{(2M/R - 2M/R_0)}. \quad (3.25)$$

For simplicity suppose that $R_0 \to \infty$, which does not change the generality but merely states that the test particle starts out at infinity. Then we have, combining these results into (3.24) with the definition of K, that

$$|(\partial/\partial t)(r^2K)| = 3\sqrt{2}\,r^2K^{\frac{3}{2}}. \quad (3.26)$$

Now if we compute the radial derivative of the metric component we have:

$$\left|\frac{\partial}{\partial r}(r^2K)\right| = \frac{2}{r}\left[r^2K + \frac{1}{2}\left|r^3\frac{\partial K}{\partial r}\right|\right]. \quad (3.27)$$

From the previous discussions we have from (3.4), the result

$$|r^3(\partial K/\partial r)| \ll Kr^2. \quad (3.28)$$

Thus we have approximately, dividing (3.26) by (3.27) and using (3.28), the result that

$$\frac{|\partial(r^2K)/\partial t|}{|\partial(r^2K)/\partial r|} \sim \text{const}\ (r^2K)^{\frac{1}{2}}. \quad (3.29)$$

Assume that our second-order expansion for the Fermi metric is used under conditions where it is valid; that is, under conditions where

$$Kr^2 \ll 1. \quad (3.30)$$

Then we are able to conclude:

$$\frac{|\partial(r^2K)/\partial t|}{|\partial(r^2K)/\partial r|} \ll 1. \quad (3.31)$$

Thus, *derivatives of the metric components taken with respect to time are much smaller than those taken with respect to spatial coordinates.* A more physical way of obtaining the same result will now be given.

We inquire as to whether the system is adiabatic. That is, we ask whether the tidal force that acts on the internal structure of the smaller mass changes by a significant fraction of its own magnitude in the characteristic time interval associated with the inner structure of the smaller mass. If so, the tidal force can produce nonadiabatic changes in this mass as puffs of wind at properly timed intervals can impart great energy to a swing. On the other hand, if the rate of change of the tidal force is *slow* in this sense, then the structure of the smaller mass will respond as if to a static perturbation (adiabatic case).

We can attack the problem in two ways. In one case we consider the fractional change in the tidal force quasiclassically. That is, the tidal force is approximately given by:

$$F \text{ tidal} \sim (GMm/R^3), \quad \text{masses in grams.} \quad (3.32)$$

The magnitude of the fractional rate of change of this accelerative force is given by

$$\left|\frac{dF \text{ tidal}/d\tau}{F \text{ tidal}}\right| \sim \frac{3\,dR/d\tau}{R}, \quad \tau = \text{proper time.} \quad (3.33)$$

Now in consistent units, the time required for gravitons to cross the small mass is

$$t \sim m \qquad \text{(centimeters)}. \quad (3.34)$$

The product of the fractional force change and this time are to be compared to 1. We ask that this product is significantly smaller than 1 in order for the motion to be abiabatic. That is, in orders of magnitude, we ask whether

$$\frac{m\,dR/d\tau}{R} \ll 1. \tag{3.35}$$

If we use the metric expansion given by (2.1), and specialize to a fixed spatial point with coordinates $r \sim m$ = constant, $\theta = \frac{1}{2}\pi$ = constant, $\varphi = 0$ = constant, we have

$$d\tau^2 = \left(1 + \frac{Mr^2}{R^3}\right) \frac{dR^2}{(2M/R - 2M/R_0)}. \tag{3.36}$$

If we choose for simplicity $R_0 \to \infty$ as before, and recall the smallness of Mr^2/R^3 in comparison to 1, we have that (3.35) becomes

$$m/R(2M/R)^{\frac{1}{2}} \ll 1. \tag{3.37}$$

This can be rewritten as

$$m^2M/R^3 = m^2K \ll 1. \tag{3.38}$$

Since we choose $r \sim m$ we again obtain the by-now well-known result that $Kr^2 \ll 1$. Thus, if our expansion is to be valid at all, it must be that the motion is adiabatic.

An alternate derivation of this important result utilizes the concept of frequency change in a gravitational wave passing across the small mass. The change in frequency Δf is given by

$$\Delta f \sim \left|\frac{\partial(\Delta g_{00})/\partial\tau}{\Delta g_{00}}\right|, \qquad \tau = \text{proper time}. \tag{3.39}$$

We are concerned with a region of space about the small mass and therefore $r \sim m$. Also we choose again a fixed space point $\theta = \frac{1}{2}\pi$ = constant, $\varphi = 0$ = constant. Thus we have that expression (3.39) becomes

$$\Delta f \sim \left|\frac{\partial(Km^2)/\partial\tau}{Km^2}\right|. \tag{3.40}$$

Using (3.36) we obtain, after some manipulation,

$$\Delta f = \left|\frac{\partial(\Delta g_{00})/\partial\tau}{\Delta g_{00}}\right| \sim K^{\frac{1}{2}}. \tag{3.41}$$

Now if this is the fractional frequency change, the wavelength change is the inverse of this, or

$$\Delta(\text{wavelength})_{\text{system}} \sim K^{-\frac{1}{2}}. \tag{3.42}$$

We must compare this with the appropriate wavelength for the wave radiated by the small moving body. This wavelength is of order $r \sim m$. Thus we must have

$$\Delta(\text{wavelength})_{\text{system}} \gg (\text{wavelength})_{\text{small mass}}, \tag{3.43}$$

or mathematically,

$$K^{-\frac{1}{2}} \gg r. \tag{3.44}$$

This can easily be put into the by now well-known form

$$r^2K \ll 1. \tag{3.45}$$

Thus the motion is adiabatic and time derivatives can be neglected.

The preceding section can be summarized as follows. When we limit attention to regions so close to the ideal geodesic that it is legitimate to use only the quadratic terms in the Fermi metric, then the Fermi form of the one-body problem presupposes—indeed, if we keep only quadratic terms—forces the motion to be adiabatic. Thus all time derivatives can be neglected in comparison with spatial derivatives, and we treat essentially a quasi-static problem. This will greatly simplify the solution of the Einstein equations of the perturbed one-body problem.

D. Preliminaries to Obtaining the Einstein Equations for the Perturbed Schwarzschild Metric

In view of the nature of the asymptotic solution (3.23), and the foregoing result, we make some simplifying assumptions. First, the replacement of x by $-x$ in (3.23) leaves the metric unchanged. Thus we expect that our solution is of even parity and can thus concentrate on that general perturbation, namely (3.22). The time dependence of the metric is not to be considered in our approximation and thus we can put $k = 0$. Also, the angular dependence of our metric, since it has even parity, must be even, and since $L = 0$ represents only a change in magnitude for the small mass, $L = 2$ angular dependence is sufficient to cover our problem. This is even more obvious from the fact that

$$P_{L=2} = \tfrac{1}{2}(3\cos^2\theta - 1) \equiv \tfrac{1}{2}q, \tag{3.46}$$

which is exactly the angular dependence of our asymptotic metric. Already at this point we have established an important fact. Namely, that the motion of the small mass m is, still in this finite m approximation, geodesic. For it was the $L = 1$ perturbation that Regge–Wheeler[12] found which would correspond to a translation of the Schwarzschild solution. In the present case this would imply a translation of m away from the ($r = 0$) geodesic of the M field. However, since $L = 1$ is not necessary to satisfy the Einstein equation, the motion is geodesic to second order.

One further change is made for simplicity. Recall that the perturbation (3.22) was not the most general, but had been simplified by the use of infinitesimal coordinate transformations. Our as-

ymptotic metric is diagonal, and rather than transform it to a form such as (3.22), we choose to change the coordinate transformation used to obtain (3.22). That is, we eliminate $H_1(x)$, but complicate somewhat the angular perturbations h_{22}, h_{33}. Thus we propose to try the perturbation shown below. (Here we use γ's to bring out clearly that these do not correspond to the h's used in Regge–Wheeler[12]):

$$\gamma_{\mu\nu} = \begin{bmatrix} [\gamma_0 q] & 0 & 0 & 0 \\ 0 & [\gamma_1 q] & 0 & 0 \\ 0 & 0 & [\gamma_3 + (\gamma_2 - \gamma_3)q] & 0 \\ 0 & 0 & 0 & [-\gamma_3 + \gamma_2 q] \end{bmatrix}. \tag{3.47}$$

The metric is given by

$$x^\mu: t, x, \theta, \varphi \tag{3.48}$$

$$g_{\mu\nu} + h_{\mu\nu} = \begin{bmatrix} -\left[\left(1 - \frac{1}{x}\right)(1 - 2\gamma_0 q)\right] & 0 & 0 & 0 \\ 0 & \left[\left(1 - \frac{1}{x}\right)^{-1}(1 + 2\gamma_1 q)\right] & 0 & 0 \\ 0 & 0 & [x^2(1 + 2\gamma_3 + 2(\gamma_2 - \gamma_3)q)] & 0 \\ 0 & 0 & 0 & [(1 - 2\gamma_3 + 2\gamma_2 q)x^2 \sin^2\theta] \end{bmatrix}. \tag{3.49}$$

Thus, in the asymptotic limits, we recover the one-body Fermi metric by substitution of

$$\gamma_0 = 3\epsilon x^2, \quad \gamma_1 = 0, \quad \gamma_2 = \gamma_3 = \epsilon x^2. \qquad \epsilon = 2m^2M/3R^3, \quad \left(1 - \frac{1}{x}\right) \to 1. \tag{3.50}$$

Another reason for choosing a diagonal form of the metric is that it is somewhat easier to obtain the curvature components in this form. The use of some standard simple formulas given in Eisenhart,[15] gives an alternate derivation of the Einstein equations for a perturbed Schwarzschild which is useful as it can be used to check the Regge–Wheeler[12] expressions. The expressions are

$$ds^2 = \sum_\alpha e_\alpha (H_\alpha)^2 (dx^\alpha)^2 = g_{\alpha\alpha}(dx^\alpha)^2 \qquad \alpha = 0, 1, 2, 3, \tag{3.51}$$

$$e_\alpha = (-1, 1, 1, 1), \tag{3.52}$$

$$\beta_{\mu\nu} = \frac{1}{H_\mu} H_{\nu,\mu}, \tag{3.53}$$

$$R_{\mu\sigma\sigma\nu} = e_\sigma H_\sigma \{H_{\sigma,\mu\nu} - H_{\nu}\beta_{\nu\sigma}\beta_{\mu\nu} - H_\nu \beta_{\mu\sigma}\beta_{\nu\mu}\} \qquad \mu \neq \sigma \neq \nu, \tag{3.54}$$

$$R_{\mu\sigma\sigma\mu} = H_\mu H_\sigma \left\{ e_\sigma \beta_{\mu\sigma,\mu} + e_\mu \beta_{\sigma\mu,\sigma} + \sum_{\substack{\lambda\neq\mu \\ \lambda\neq\sigma}}^{\lambda\neq\mu} e_\lambda e_\sigma e_\mu \beta_{\lambda\sigma}\beta_{\lambda\mu} \right\} \qquad \mu \neq \sigma. \tag{3.55}$$

Here, the indices run from 0 to 3, and the summation convention is *not* employed. These are the only independent nonzero components of the curvature tensor. In view of our previous statements, the time dependence of the metric is a second-order effect insofar as it affects the computation of curvature tensor components. Thus for our calculations, the time dependence of the metric is neglected and is lumped into the "constant" ϵ. Also, we set all derivatives with respect to either "time" (t), or φ exactly zero. These approximations simplify the computations markedly. If we compute the nonzero curvature components and evaluate their asymptotic forms we obtain

$$R_{1332} = +9\epsilon x^3 \sin^2\theta \sin 2\theta, \quad R_{1002} = +9\epsilon x \sin 2\theta,$$
$$R_{1001} = -6\epsilon q, \quad R_{1221} = -6\epsilon x^2,$$
$$R_{1331} = -6\epsilon x^2 \sin^2\theta (q - 1), \quad R_{2002} = +6\epsilon x^2 (q - 1),$$
$$R_{3003} = +6\epsilon x^2 \sin^2\theta, \quad R_{2332} = +6\epsilon x^4 \sin^2\theta q, \tag{3.56}$$

which are the only independent nonzero components of the curvature tensor. Here we have neglected all terms in x lower in order than the highest. That is, for example, R_{1001} should be $-6\epsilon q[1 - (1/2x)]$, but to the same order of approximation that $[1 - (1/x)] \to 1$, we have $R_{1001} = -6\epsilon q$.

If we compute Einstein equations we must have

$$G_{\mu\nu} \equiv R_{\mu\nu} - \tfrac{1}{2} g_{\mu\nu} R = 0, \tag{3.57}$$

or since the space is empty of stress energy,

$$R = 0, \quad G_{\mu\nu} = R_{\mu\nu} \equiv 0, \quad R_{\mu\nu} = g^{\alpha\beta} R_{\alpha\mu\nu\beta}. \tag{3.58}$$

Now since $g_{\alpha\beta}$ is diagonal, $g^{\alpha\beta}$ is also, and thus we have

$$R_{\mu\nu} = +g^{\alpha\alpha}R_{\alpha\mu\nu\alpha}. \tag{3.59}$$

Since $R_{i;0i} \equiv 0$ for our case, we have

$$R_{0i} \equiv 0. \tag{3.60}$$

Now for our case,

$$R_{\alpha 13\alpha} = R_{\alpha 23\alpha} \equiv 0; \qquad \alpha = 0, 1, 2, 3, \tag{3.61}$$

and thus we have

$$R_{13} = R_{23} \equiv 0. \tag{3.62}$$

Therefore, the only nontrivial Einstein equations are

$$\begin{aligned} R_{\mu\mu} &= 0 \qquad \mu = 0, 1, 2, 3, \\ R_{12} &= 0. \end{aligned} \tag{3.63}$$

When the metric of (3.23) is developed in powers of $1/x$, the Einstein equations above are satisfied correct up to—but not through—order $(1/x)$. That is, Eqs. (3.63) are satisfied more nearly as $x \to \infty$. The deviations go as

$$R_{\alpha\beta} \sim \epsilon[(1/x) + O(1/x^2)]. \tag{3.64}$$

To improve this result to all orders of x is the purpose of the next section. The Regge–Wheeler results, computed from their expressions gave, in the same limits,

$$R_{\alpha\beta} \sim \epsilon[1 + O(1/x)]. \tag{3.65}$$

Thus it is clear that their result is inconsistent with the Einstein equations even if $x \to \infty$.

E. The Einstein Equations for our Problem and Their Solution

The equations to be obtained are just the equations corresponding to (3.63). These are five in number but they may be related and thus not independent. Using the metric given by (3.49) and the expressions for curvature, Einstein equations, etc., we arrive, after much computation, at the required expressions. In these, the angle-dependent parts and the angle-independent parts have been individually set equal to zero. Thus, potentially, from our five nontrivial Einstein equations, we could obtain ten equations for the four unknown γ_μ. However, with only some algebraic manipulations we can reduce the number of "independent" equations to five again. These are:

$$\begin{aligned} \tfrac{1}{2}G_{22} \equiv\ & x(x-1)\tfrac{1}{2}\gamma_3'' + x\gamma_3' \\ & - \tfrac{1}{2}\gamma_3' + (\gamma_1 - \gamma_0) = 0, \end{aligned} \tag{3.66a}$$

$$\begin{aligned} [x(x-1)/2q]G_{11} \equiv\ & x(\gamma_0' - \gamma_2' + \tfrac{1}{2}\gamma_3') \\ & + \tfrac{1}{2}(\gamma_2' - 2\gamma_0' - \tfrac{1}{2}\gamma_3') \\ & + (2\gamma_2 + 2\gamma_3 + \gamma_1 - 3\gamma_0) = 0, \end{aligned} \tag{3.66b}$$

$$\begin{aligned} [x^3/2q(x-1)]G_{00} \equiv\ & x(x-1)(\gamma_2'' - \tfrac{1}{2}\gamma_3'') \\ & + 3x(\gamma_2' - \tfrac{1}{2}\gamma_3' - \tfrac{1}{3}\gamma_1') \\ & - \frac{5}{2}(\gamma_2' - \tfrac{1}{2}\gamma_3' - \tfrac{2}{5}\gamma_1') \\ & - 2(\gamma_2 + \gamma_3 + 2\gamma_1) = 0, \end{aligned} \tag{3.66c}$$

$$\begin{aligned} [x(x-1)/3\sin^2\theta]G_{12} \equiv\ & x(x-1)(\gamma_0' - \gamma_2' - \tfrac{1}{2}\gamma_3') \\ & + x(\gamma_1 - \gamma_0) + \tfrac{1}{2}(3\gamma_0 - \gamma_1) = 0, \end{aligned} \tag{3.66d}$$

$$\begin{aligned} (1/q)G_{33} \equiv\ & x(x-1)(\gamma_0'' - \gamma_2'') + x(\gamma_0' + \gamma_1' - 2\gamma_2') \\ & + \tfrac{1}{2}(\gamma_0' - \gamma_1' + 2\gamma_2') + 2(\gamma_1 - \gamma_0) = 0. \end{aligned} \tag{3.66e}$$

Here the equations $G_{\mu\mu} = 0$ are chosen since they are somewhat more compact than the equations $R_{\mu\mu} = 0$. The Bianchi identities

$$G^{\mu\nu}_{;\nu} = 0 \tag{3.67}$$

give only two nontrivial expressions, since

$$G^{0\nu}_{;\nu} = G^{3\nu}_{;\nu} \equiv 0. \tag{3.68}$$

Thus it would appear that, out of four identities, there are two independent additional pieces of knowledge. However, because the $G^{\mu\nu}$ are themselves zero, the remaining Bianchi identities reduce to

$$\begin{aligned} G^{1\nu}_{;\nu} = G^{1i}_{;i} = G^{11}_{,1} + G^{12}_{,2} = 0, \\ G^{2\nu}_{;\nu} = G^{2i}_{;i} = G^{12}_{,1} + G^{22}_{,2} = 0. \end{aligned} \tag{3.69}$$

These expressions give essentially combinations of the previous equations. However, they lead to the elimination of one further expression from the group (3.66). It can be shown that

$$G_{12,1} + G_{22} \to G_{33}. \tag{3.70}$$

That is, in expression (3.66), the equation labeled (d) differentiated, plus the equation (a) gives the equation (e), where all multiplying factors have been omitted. Thus we have reduced our problems to four equations, for four unknowns; Eqs. (3.66 a–d).

The equations we have are linear differential equations, of which two are of second order, and two of first order. Attempts to simplify them further have failed and so seem to indicate that four independent functions γ_μ are necessary to solve them. We try a polynomial series with four independent degrees of freedom. That is,

$$\gamma_\mu = \sum_{n=-\infty}^{\infty} A(\mu)_n x^n. \tag{3.71}$$

In view of the known asymptotic behavior of the γ_μ, the series is cut off at $n = 2$, with values for the

$A(\mu)$ as given by (3.50). After a great deal of manipulation we arrive at the result in closed form:

$$\gamma_0/\epsilon = 3x^2 - \tfrac{5}{2}x - \tfrac{1}{8}, \qquad \gamma_1/\epsilon = -\tfrac{3}{8},$$
$$\gamma_2/\epsilon = x^2 + \tfrac{7}{4}x - \tfrac{7}{4}, \qquad \gamma_3/\epsilon = x^2 - \tfrac{1}{2}x. \tag{3.72}$$

These expressions, within the framework of the approximation—namely ϵ is small, and so terms of order higher than one in ϵ are neglected—exactly solve Einstein equations to *all* orders of x. That ϵ is small is known, since

$$\epsilon = \tfrac{2}{3}Mm^2/R^3 = \tfrac{2}{3}Km^2 \sim Kr^2 \ll 1. \tag{3.73}$$

In terms of the above γ_μ, the metric in the intermediate Region II (Fig. 1) is

$$x^\mu\colon t, x, \theta, \varphi$$

$$\begin{aligned} ds^2 = &-\left(1 - \frac{1}{x}\right)\left[1 - q\epsilon\left(6x^2 - 5x - \frac{1}{4}\right)\right]dt^2 \\ &+ \left(1 - \frac{1}{x}\right)^{-1}\left[1 - \frac{3\epsilon q}{4}\right]dx^2 \\ &+ x^2\left[1 + \epsilon\left\{x(2x-1) + \frac{q}{2}(9x-7)\right\}\right]d\theta^2 \\ &+ x^2\sin^2\theta\left[1 - \epsilon\left\{x(2x-1)\right.\right. \\ &\left.\left. - \frac{q}{2}(4x^2 + 7x - 7)\right\}\right]d\varphi^2, \end{aligned} \tag{3.74}$$

with the notation

$$\epsilon = 2m^2M/3R^3, \qquad q = 3\cos^2\theta - 1,$$

$$dt^2 = \frac{dR^2/2}{4m^2M[(1/R) - (1/R_0)]} = \frac{d\epsilon^2/2}{54\epsilon^3[1 - (\epsilon_0/\epsilon)^{\frac{1}{3}}]}. \tag{3.75}$$

This metric gives the transition metric for "going from a small Schwarzschild mass to a large mass." In the limit where x is small, that is, of order 1, the metric reduces to the Schwarzschild expansion in the parameters of the small mass. That is, to recover the metric of the small mass in the absence of the large mass, put $\epsilon \sim 0$ and let t be no longer related to ϵ. In the limit where x is large, say $x > 10$, we can neglect all powers of x smaller than the second, and we get our metric as viewed from a *massless* test particle moving toward a large mass. Both limiting cases and the intermediate case satisfy Einstein empty-space equations within their appropriate limits of convergence.

The curvature components may be computed for this metric and are exhibited below. Here time derivatives have been neglected because of their "smallness" in agreement with our previous results. The components are

$$\begin{aligned} -R_{1332} &= -9\epsilon x^3\sin 2\theta\sin^2\theta\left[1 + \frac{1}{2x} + \frac{1}{8x^3}\right], \\ -R_{1002} &= -9\epsilon x\sin 2\theta\left[1 - \frac{1}{2x} - \frac{3}{8x^2} - \frac{1}{8x^3}\right], \\ -R_{1001} &= \frac{1}{x^3} + 6\epsilon q\left[1 - \frac{1}{2x} + \frac{5}{24x^2} + \frac{1}{24x^3}\right], \\ -R_{1221} &= \frac{1}{2(x-1)} + \frac{6\epsilon x^3}{(x-1)}\left[\left(1 - \frac{5}{6x} - \frac{1}{12x^2}\right)\right. \\ &\qquad \left. + q\left(\frac{3}{4x} + \frac{1}{16x^2} - \frac{7}{24x^3}\right)\right], \\ -R_{1331} &= \frac{\sin^2\theta}{2(x-1)} - \frac{6\epsilon x^3\sin^2\theta}{(x-1)}\left[\left(1 - \frac{5}{6x} - \frac{1}{12x^2}\right)\right. \\ &\qquad \left. + q\left(-1 + \frac{1}{12x} + \frac{1}{48x^2} + \frac{7}{24x^3}\right)\right], \\ -R_{2002} &= -\frac{(x-1)}{2x^2} + 6\epsilon x(x-1)\left[\left(1 - \frac{7}{6x} + \frac{1}{12x^2}\right)\right. \\ &\qquad \left. - q\left(1 - \frac{3}{4x} + \frac{23}{48x^2} - \frac{5}{24x^3}\right)\right] \\ -R_{3003} &= -\frac{(x-1)\sin^2\theta}{2x^2} - 6\epsilon x(x-1)\sin^2\theta \\ &\quad \times\left[\left(1 - \frac{7}{6x} + \frac{1}{12x^2}\right) + q\left(\frac{5}{12x} + \frac{19}{48x^2} - \frac{5}{24x^3}\right)\right], \\ -R_{2332} &= -x\sin^2\theta - 6\epsilon x^4\sin^2\theta q \\ &\qquad \times\left[1 + \frac{5}{7x} + \frac{17}{24x^2} - \frac{25}{24x^3}\right]. \end{aligned} \tag{3.76}$$

It is evident from this that if ϵ is negligible in comparison with 1, that is if $M \to 0$, the curvatures are just those one would get for an ordinary Schwarzschild metric, namely:

$$\begin{aligned} &R_{1001} = \frac{-1}{x^3}, \qquad R_{1221} = \frac{-1}{2(x-1)}, \\ &R_{1331} = \frac{-1}{2(x-1)}\sin^2\theta, \quad R_{2002} = +\frac{(x-1)}{2x^2}, \\ &R_{3003} = +\frac{(x-1)}{2x^2}\sin^2\theta, \quad R_{2332} = +x\sin^2\theta, \\ &\qquad \text{all other } R_{\mu\nu\alpha\beta} \equiv 0. \end{aligned} \tag{3.77}$$

Here all curvatures are expressed in dimensionless units which can easily be put into the standard form. For example,

$$\bar{R}_{1001} = +(1/4m^2)R_{1001} = -(2m/r^3). \tag{3.78}$$

On the other hand, if $x \gg 10$, that is, in our terms $x \to \infty$ (the asymptotic case), we see that the expressions (3.76) go into (3.56) as they should. That is, we neglect all terms within the square brackets which go as $O(1/x)$ or higher. The terms which are not multiplied by ϵ are neglected, since they all go as $O(1/x)$ or higher. Thus, our metric gives the right limiting values for the metric *and* for the curvature components as well as satisfying the Einstein equations.

IV. USES OF SOLUTION

A. Aspects and Comments on the Solution Obtained

We first make some comments on the results just obtained. If we list here, for convenience, all the approximations and inequalities which we must have for our answer to be correct, we have

$$\text{(a)}\quad Kr^2 \sim \epsilon x^2 \ll 1, \quad \text{(b)}\quad |r^3(\partial K/\partial r)| \ll Kr^2, \tag{4.1}$$
$$\text{(c)}\quad \alpha = (m/M) \ll 1, \quad \text{(d)}\quad \epsilon = (2m^2M/3R^3) \ll 1.$$

The first (a) is necessary for our perturbation expansion to be valid. The second (b) is needed so that we are justified in dropping higher-order terms than the second in the Fermi expansion. The third (c) is needed in order to apply our results near the large mass ($R \sim M$), i.e. the small mass moving through the neck of the large Schwarzschild singularity. The fourth (d) is necessary in order that our *procedure* for perturbing a Schwarzschild metric be valid, since the $h_{\mu\nu}(\gamma_\mu)$ must be small compared to 1, and they all depend on the expansion parameter ϵ (small distortion in shape of neck or bridge in Schwarzschild metric). This condition also permits our interpretation of the Fermi metric as a boundary condition for our perturbation problem.

Now we can satisfy all these requirements and yet obtain an additional piece of information not already explicitly shown which is extremely important. Since the metric components go as (neglecting constants)

$$g_{\mu\nu} \sim (1 - \epsilon q\, O(x^2)), \tag{4.2}$$

and since ϵx^2 is very much less than 1, it would appear that in our metric, our expansion is not going to change the physics much. That is, we have almost flat space anyhow about our geodesic line. However, the place where our physics shows up the most is not in the metric but in its derivatives: that is, in our curvature components (coefficients of x^2 in tide-producing forces). These give the physics, and in them, the terms involving ϵ are not necessarily negligible. We show this below.

The term R_{1001} will be taken as a typical component, and x will be assumed large ($x > 10$) so that the asymptotic metric and curvatures hold. We have then

$$-R_{1001} = (1/x^3) + 6\epsilon q. \tag{4.3}$$

If we compare these terms we will have, for order-of-magnitude computations only, to compare ϵx^3 to 1. Now, we know that $\epsilon x^2 \ll 1$. However, suppose we put in some numbers. We have (limiting ourselves to $R \sim M$)

$$\epsilon = (2m^2M/3R^3) \sim (m^2/M^2) = \alpha^2. \tag{4.4}$$

Suppose that $\alpha \sim 10^{-3} \ll 1$, $x \sim 10^2 > 10$, as per our previously quoted values. Thus we have

$$\epsilon x^2 = 10^{-2} \ll 1, \qquad \epsilon x^3 = 1. \tag{4.5}$$

Thus, the curvature component is very importantly a combination of the effect due to the mass $m(1/x^3$ term) and the effect of the background field due to M (ϵ term). If, however, x is small ($x \sim 1$), then still keeping $\alpha \sim 10^{-3}$, we have

$$\epsilon x^2 = 10^{-6} \ll 1, \qquad \epsilon x^3 = 10^{-6} \ll 1. \tag{4.6}$$

Now, the ϵ terms in the curvature as well as in the metric vanish in importance compared to the terms due to the small mass m. Thus, by choosing the mass of the small particle in relation to the large one *properly*, we can probe the various regions of interest throughout all space. That is to say, since we can choose the mass of the small particle arbitrarily, we can make its effects either large or small at any given point in space, so long as we do not violate the condition $Kr^2 \ll 1$.

The evaluation of the deformation of a test sphere located at the position of our "test mass" has been performed using the Fermi expansion in the preceding paper.[14] We now proceed to perform the analysis for general x. That is, we use the metric just developed for the restricted two-body problem. From (3.74) we have, if we hold both t and x constant,

$$\begin{aligned} ds^2 = {} & x^2\{1 + \epsilon[x(2x - 1) + (\tfrac{1}{2}q)(9x - 7)]\}\, d\theta^2 \\ & + x^2 \sin^2\theta\{1 - \epsilon[x(2x - 1) \\ & - (\tfrac{1}{2}q)(4x^2 + 7x - 7)]\}\, d\varphi^2. \end{aligned} \tag{4.7}$$

Let η denote the relative difference between circumferences around the poles and equator. We evaluate the arc lengths, expanding the square roots as in the previous paper,[14] and find the deformation coefficient

$$\eta = (S_\theta - S_\varphi)/(S_\theta + S_\varphi) \approx \tfrac{3}{16}\epsilon(8x^2 + 5x - 7). \quad (4.8)$$

If x is large we can neglect all but the term in x^2. Putting in the definition of ϵ and x, we have

$$\eta \approx \frac{3}{16}\frac{2m^2M}{3R^3}\cdot 8\frac{r^2}{4m^2} = \frac{Mr^2}{4R^3}, \quad (4.9)$$

where we have that $x > 10$.

This is the same result as the expression obtained from the one-body Schwarzschild expressed in comoving coordinates gave, as it should be. The expression (4.8) gives us a measure of the deformation over the whole region of interest (Region I, II, III) and is thus an extremely useful equation. Most interesting is the value of this distortion coefficient at the throat (see below) where $x = 1$. For this we have that

$$\eta = (Mm^2/R^3)(\tfrac{3}{4}), \quad (4.10)$$

which shows that the wormhole neck is more rigid than empty space by a factor of $\frac{3}{4}$.

B. Coordinate Transformation on Metric to Obtain Symmetries

Suppose we perform a coordinate transformation on our metric for "two bodies" in order to observe more clearly certain relationships which it is customary to obtain in the case of the one-body problem. Let us use the same transformation that takes an ordinary Schwarzschild metric into an isotropic Schwarzschild metric. That is, the coordinate transformation[16-19]

$$t \to t, \quad x = (r/2m) \to y = (2r/m), \quad \theta \to \theta, \quad \varphi \to \varphi, \quad x = \{\tfrac{1}{2}[1 + (1/y)]\}^2 y. \quad (4.11)$$

Performing the necessary operations for transforming a second-order covariant tensor (e.g. the metric tensor $g_{\mu\nu}$), we obtain, after some manipulations,

$$ds^2 = -\left(\frac{\omega}{\psi}\right)^2\left[1 - 6\epsilon q y^2\psi^4 \times \left(1 - \frac{5}{6y\psi^2} - \frac{1}{24y^2\psi^4}\right)\right] dt^2 + \left(1 - \frac{3\epsilon q}{4}\right)\psi^4\, dy^2 + y^2\psi^4\left[1 + 2\epsilon y^2\psi^4 \times \left(1 - \frac{1}{2y\psi^2} + q\left\{\frac{9}{4y\psi^2} - \frac{7}{4y^2\psi^4}\right\}\right)\right] d\theta^2 + y^2\sin^2\theta\psi^4\left[1 - 2\epsilon y^2\psi^4\left(1 - \frac{1}{2y\psi^2} - q\left\{1 + \frac{7}{4y\psi^2} - \frac{7}{4y^2\psi^4}\right\}\right)\right] d\varphi^2. \quad (4.12)$$

Here we have put for convenience

$$\psi = \tfrac{1}{2}(1 + 1/y), \qquad \omega = \tfrac{1}{2}(1 - 1/y). \quad (4.13)$$

If we let $x \to \infty$ in our expression, we have $y \to \infty$ also, so that both ω and $\psi \to \frac{1}{2}$. Putting this into (4.12), we obtain

$$ds^2 \approx -[1 - \tfrac{3}{8}\epsilon q y^2]\, dt^2 + \tfrac{1}{16}[1 - \tfrac{3}{4}\epsilon q]\, dy^2 + \frac{y^2}{16}[1 + \tfrac{1}{8}\epsilon y^2]\, d\theta^2 + \tfrac{1}{16}y^2 \times \sin^2\theta[1 - \tfrac{1}{8}\epsilon y^2(1 - q)]\, d\varphi^2. \quad (4.14)$$

This is just the expression for the one-body Schwarzschild in comoving coordinates if we take into account the fact that for $y \to \infty$, $x \approx \frac{1}{4}y$.

If we now perform an additional coordinate transformation given by

$$t \to t, \quad y = (2r/m) \to z = (m/2\rho), \quad \theta \to \theta, \quad \varphi \to \varphi, \quad y = 1/z, \quad (4.15)$$

and carry out the manipulations just as in the above, we observe that the metric expressed in the z coordinate system is of the *same form* as is that expressed in the y coordinate system. Thus, the metric for $z \to \infty$ is the same as that for $y \to \infty$ [Eq. (4.14)] expressed in terms of z. But $y = 1/z$, so that if we have covered the region $1 \leq y \leq \infty$, we have also covered the region $1 \geq y > 0$ since that corresponds to $1 \leq z \leq \infty$. Thus the metric is found to be symmetric with respect to reflection in the world tube $y = 1$ which thus defines the location of the throat. This behavior is exactly the same as that observed for the Schwarzschild one-body problem in isotropic coordinates. Thus, either the two-sheeted or wormhole picture is suitable to describe our new problem.

C. Evaluation of the Minimal Imbedded 2 Surface (Throat)

If we use the concept of a wormhole[3,20] or multiply connected topology, we can ask for the shape and

[16] L. Landau and E. Lifshitz, *Classical Theory of Fields* (Addison-Wesley Publishing Company, Reading, Massachusetts, 1951).
[17] P. G. Bergmann, *Introduction to Theory of Relativity*, (Prentice-Hall, Inc., Englewood Cliffs, New Jersey, 1942).
[18] G. D. Birkhoff, *Relativity and Modern Physics* (Harvard University Press, Cambridge, Massachusetts, 1923).
[19] L. Infeld and A. Schild, Rev. Mod. Phys. 21, 408 (1949).
[20] J. A. Wheeler, Lecture notes in Course on General Relativity, Princeton University, Princeton, New Jersey, 1960.

area of a 2 surface imbedded in the space; that is, what is the geometry of a cross section of the wormhole. We proceed in a manner analogous to that used for the study of this surface in the case of the one-body problem,[21] and in the problem of two equal masses (symmetric wormhole).[20,22]

The element of area for a 2 surface imbedded in a 4 volume is given as[15]:

$$dA = a^{\frac{1}{2}}\, d\eta^2\, d\eta^3. \tag{4.16}$$

Here the coordinates η^2 and η^3 describe the 2 surface and a is the determinant of the metric tensor on the surface. That is, if the line element is given in the 4-space as

$$ds^2 = g_{\alpha\beta}\, d\xi^\alpha\, d\xi^\beta, \qquad \alpha = 0, 1, 2, 3, \tag{4.17}$$

it is given on the 2 surface by

$$ds^2 = a_{cd}\, d\eta^c\, d\eta^d, \qquad c, d = 2, 3. \tag{4.18}$$

For the 2 surface of interest we hold t = const. (That is, R = const $\langle = \rangle t'$ = const). We also look for a two-dimensional surface where we hold x = const (that is r = const and therefore y = const). We have from elementary tensor analysis that[23]

$$a_{cd} = g_{\alpha\beta} \frac{\partial \xi^\alpha}{\partial \eta^c} \frac{\partial \xi^\beta}{\partial \eta^d}. \tag{4.19}$$

Because of our metric and the independence of the ξ^α, we have, after some computations,

$$a_{22} = g_{22}, \quad a_{33} = g_{33}, \quad a_{32} = a_{23} = 0, \quad a^{\frac{1}{2}} = (g_{22}g_{33})^{\frac{1}{2}}. \tag{4.20}$$

Thus, we finally have

$$dA = (g_{22}g_{33})^{\frac{1}{2}}\, d\theta\, d\varphi, \tag{4.21}$$

having taken θ and φ as our coordinates on the surface.

If we evaluate this expression using the values of $g_{\mu\nu}$ from (4.12) we will have

$$dA = y^2\psi^4 \sin\theta \left[1 + 2\epsilon y^2\psi^4 q \times \left(1 + \frac{4}{y\psi^2} - \frac{7}{2y^2\psi^4}\right)\right] d\theta\, d\varphi. \tag{4.22}$$

Here we have neglected terms of order ϵ^2 as in all our previous work. Integrating this expression from 0 to 2π for both variables θ and φ, noting that ψ depends only on y which is constant, and that q is the second Legendre polynomial and depends only on θ we obtain *exactly*, from integrating (4.22),

$$A = 4\pi y^2 \psi^4. \tag{4.23}$$

This simple expression is *independent* of ϵ [to the present order of approximation in the parameter $\epsilon \sim (Mm^2/R^3)$] since the terms containing ϵ depend on $q \sin\theta$ whose average value (i.e. $1/2\pi \int_0^{2\pi} q \sin\theta\, d\theta$) vanishes. Thus we obtain the same result as for the one body-problem[21] correct up to the order ϵ^2.

Now, from our symmetry investigations we have previously found that the world tube $y = 1$ represents the throat of the wormhole. That is, our coordinate transformation about this tube leaves the metric invariant and this then locates an extremum. We can now discover more by differentiating expression (4.23) in order to obtain a minimum. That is

$$(\partial A/\partial y) = 0 = 4\pi y\psi^3[1 - (1/y)] \Rightarrow y = 1. \tag{4.24}$$

Thus both our symmetry arguments and our extremization give the same result, namely, the throat *occurs* at $y = 1$ and is the *minimal* surface of the space. Putting $y = 1$ into (4.23) we obtain

$$A = 4\pi. \tag{4.25}$$

To put this in more customary or unnormalized units we must multiply by $4m^2$. Thus we finally obtain

$$\text{Area of throat} = A'_{\min} = 16\pi m^2. \tag{4.26}$$

This result is *identical* with the one-body Schwarzschild problem.

We have shown above that the throat is the minimal surface among those surfaces which have y = constant (r = constant). However, we can remove even this restriction. Eisenhart,[15] proves that the minimal surface is always found where the extrinsic curvature invariant (second fundamental form) vanishes. That is, in somewhat more standard notation, where

$$a^{cd}K_{cd} \equiv 0, \qquad c, d = 2, 3. \tag{4.27}$$

For a diagonal metric, K_{cd}, the second fundamental form, is proportional to the derivative of the metric. From our expression for the metric, we see that a_{22} and a_{33} depend on the function

$$f = y\psi^2. \tag{4.28}$$

We have that

$$K_{cd} \sim \frac{\partial a_{cd}}{\partial y} \equiv \frac{\partial a_{cd}}{\partial f} \frac{\partial f}{\partial y}. \tag{4.29}$$

[21] F. K. Manasse and R. Lindquist, "Solutions of Schwarzschild Metric" (unpublished).
[22] F. K. Manasse, "Two Wormhole Throat Investigations" (unpublished).
[23] J. A. Schouten, *Ricci Calculus* (Julius Springer-Verlag Berlin, 1954).

Evaluating $\partial f/\partial y$ we have:

$$\partial f/\partial y = \tfrac{1}{4}[1 - (1/y^2)]. \tag{4.30}$$

If $y = 1$, the world tube of the throat, this quantity is evidently zero and consequently the second fundamental form K_{cd} is identically zero for this value of y. Thus the throat is the minimal surface among *all* submanifolds of the Riemannian 3 space t = constant.

V. CONCLUSION

We have evaluated the metric and inquired into some of the dynamics of a restricted two-body problem. The problem was extremely specialized to that of one large massive body gravitationally attracting a small mass toward it on a "radial" path. The curvature in the neighborhood of the small particle was viewed as due to two causes. (1) the self mass of the small body producing a Schwarzschild-like gravitational field (mass parameter m); and (2) the "tide-producing" forces represented by curvature of space near the small mass because of the presence of the large mass (M). This second force induced a change in the Schwarzschild metric near the small mass so that it would satisfy Einstein equations throughout its motion. This change was analyzed by a perturbation approach similar to that of Regge–Wheeler[12] with the "normalized curvature" caused by the large mass as the expansion parameter [e.g. $\epsilon \sim (m^2M/R^3)$].

The Einstein equations for the perturbation were solved by noting that the "asymptotic" behavior necessary for the two-body problem must be identical with the solution of the one-body problem cast into comoving coordinates (Fermi normal coordinates).

Among the interesting results obtained by using this perturbed metric is the deformation in shape of a small sphere. It is shown that the tidal effect of the large mass is significant. However, it is not too large to prevent the particle from crossing the so-called Schwarzschild singularity at $R = 2M$ for a sufficiently small test mass. However, the tidal forces grow larger and larger as one approaches the essential singularity at $R = 0$, or the center of the large mass. Indeed, the analysis indicates that we can get no closer than $\sim(Mm^2)^{\frac{1}{3}}$ before disruption occurs. What occurs when this happens is a new problem which is not analyzed here. This paper treats only that regime where the tidal forces can be treated as a small perturbation on the intrinsic forces which govern the dynamics of evolution of a small mass.

VI. ACKNOWLEDGMENTS

I should like to thank especially Professor J. A. Wheeler, under whose supervision this work was carried out. It is a pleasure to acknowledge his invaluable guidance and support of this investigation. Appreciative thanks are also due to Professor C. W. Misner for his generous help and stimulating discussions. Appreciation is expressed to Bell Telephone Laboratories for the award of a Communications Development Training Program Fellowship under which all of this work was carried out.

Parameter Differentiation of Quantum-Mechanical Linear Operators

KÊITSIRO AIZU
Hitachi Central Research Laboratory, Kokubunzi, Tokyo, Japan
(Received 26 December 1962)

Quantum-mechanical linear operators have some real-number parameters in them. The differential coefficients of functions, including eigenvalues and eigenvectors, of such an operator with respect to a parameter are calculable from the differential coefficient of the operator with respect to the parameter. The formulas for this purpose are presented together with their proofs.

I. INTRODUCTION

A LINEAR operator in quantum mechanics usually includes some real parameters, which are, for example, masses, spacings, and field strengths, and which may also be artificial parameters. If one of these parameters varies, the linear operator and its functions may also vary; these functions can be numbers, vectors, or linear operators. Therefore, we can think of the differentiation of the functions with respect to the parameter. In this article, the connection of the differential coefficients of the functions with the differential coefficient of the linear operator is investigated in a systematic way. Some number of formulas and their proofs are presented according to the order of reasoning. The same subject has been studied, if only partially, by some people. For instance, Salem proposed a formula[1]

$$\sum_{j(\neq i)} \frac{\langle K_i|\,\partial K/\partial\lambda\,|K_j\rangle\langle K_j|\,A\,|K_i\rangle + \langle K_i|\,A\,|K_j\rangle\langle K_j|\,\partial K/\partial\lambda\,|K_i\rangle}{K_i - K_j} = (\partial/\partial\lambda)\langle K_i|\,A\,|K_i\rangle - \langle K_i|\,\partial A/\partial\lambda\,|K_i\rangle.$$

(This notation is not the same as Salem's. The meaning of each symbol will soon become clear.) Several already proposed formulas such as this are included in the present system of theory as necessary parts.

This theory may serve for calculating the differential coefficients of functions of a linear operator from the differential coefficient of the linear operator. It may also be useful in quantum-mechanical perturbation theory; in next section, this "parameter-differentiation" method is employed.

In the subsequent sections, most symbols are always used in a definite meaning. They are listed below.

λ	a real parameter;
μ	another real parameter;
K	a Hermitian linear operator the eigenvectors of which form a complete set;
A	a linear operator;
K_i	the ith eigenvalue of K;
$\|K_i\rangle$	the *normalized* eigenvector of K belonging to the eigenvalue K_i;
$\langle K_i\|$	the vector conjugate with $\|K_i\rangle$;
$f(z)$	a function of z which is a complex number when z is a complex number and which is a linear operator when z is a linear operator;
$f'(z)$	the derivative of $f(z)$.

K and A are dependent on a single parameter λ in Sec. II, and dependent on two parameters λ and μ in Sec. III. It is assumed throughout this article that the eigenvalues of K are discrete and nondegenerate (although this assumption is not absolutely necessary for all theorems in this article).

II. A SINGLE PARAMETER

[1]

$$\partial K_i/\partial\lambda = \langle K_i\,|\partial K/\partial\lambda|\,K_i\rangle.$$

Proof. Each side of the identical equation

$$f(K_i) = \langle K_i\,|f(K)|\,K_i\rangle$$

is differentiated with respect to λ:

$$\begin{aligned} f'(K_i)\,\partial K_i/\partial\lambda &= (\partial\langle K_i|/\partial\lambda)f(K)\,|K_i\rangle + \langle K_i|\,\partial f(K)/\partial\lambda\,|K_i\rangle \\ &\quad + \langle K_i|\,f(K)(\partial\,|K_i\rangle/\partial\lambda), \\ &= f(K_i)(\partial\langle K_i|/\partial\lambda)\,|K_i\rangle + \langle K_i|\,\partial f(K)/\partial\lambda\,|K_i\rangle \\ &\quad + f(K_i)\,\langle K_i|\,(\partial\,|K_i\rangle/\partial\lambda), \\ &= \langle K_i|\,\partial f(K)/\partial\lambda\,|K_i\rangle + f(K_i)\,\partial\langle K_i|K_i\rangle/\partial\lambda, \\ &= \langle K_i|\,\partial f(K)/\partial\lambda\,|K_i\rangle, \qquad (\because \langle K_i|K_i\rangle \equiv 1). \end{aligned}$$

[1] L. Salem, Phys. Rev. 125, 1788 (1962).

When especially $f(K) = K$, the above calculation is for deriving the relevant formula.

Remark. This formula is sometimes called Feynman's theorem. If we want to express $\langle K_i| \partial K/\partial\lambda |K_j\rangle$ in terms of the wavefunctions ψ_i and ψ_j corresponding to the vectors $|K_i\rangle$ and $|K_j\rangle$, respectively, it becomes

$$\int \bar{\psi}_i \frac{\partial K}{\partial \lambda} \psi_j \, d\tau,$$

where $d\tau$ is the volume element of the configuration space, and $\bar{\psi}_i$ is complex-conjugate with ψ_i.

[2]

$$\langle K_i| \frac{\partial f(K)}{\partial \lambda} |K_j\rangle = \frac{f(K_i) - f(K_j)}{K_i - K_j} \langle K_i| \frac{\partial K}{\partial \lambda} |K_j\rangle,$$

where i and j may as well be equal, in this case $\{f(K_i) - f(K_j)\}/(K_i - K_j)$ meaning $f'(K_i)$.

Proof. The proof in $i = j$ is obtained immediately from [1]; that is,

$$\langle K_i| \partial f(K)/\partial\lambda |K_i\rangle = f'(K_i) \partial K_i/\partial\lambda = f'(K_i) \langle K_i| \partial K/\partial\lambda |K_i\rangle.$$

The proof in $i \neq j$ is made as follows. We note that the commutation relation

$$f(K)K = Kf(K)$$

holds. This equation is differentiated with respect to λ:

$$(\partial f(K)/\partial\lambda)K + f(K)(\partial K/\partial\lambda) = (\partial K/\partial\lambda)f(K) + K(\partial f(K)/\partial\lambda),$$

and rearranged:

$$K(\partial f(K)/\partial\lambda) - (\partial f(K)/\partial\lambda)K = f(K)(\partial K/\partial\lambda) - (\partial K/\partial\lambda)f(K).$$

The last equation shows that the commutator between K and $\partial f(K)/\partial\lambda$ is equal to the commutator between $f(K)$ and $\partial K/\partial\lambda$. We multiply each side of this equation by the vectors $\langle K_i|$ and $|K_j\rangle$ from the left and right, respectively:

$$\langle K_i| K(\partial f(K)/\partial\lambda) |K_j\rangle - \langle K_i| (\partial f(K)/\partial\lambda)K |K_j\rangle = \langle K_i| f(K)(\partial K/\partial\lambda) |K_j\rangle - \langle K_i| (\partial K/\partial\lambda)f(K) |K_j\rangle.$$

This becomes

$$(K_i - K_j) \langle K_i| \partial f(K)/\partial\lambda |K_j\rangle = \{f(K_i) - f(K_j)\} \langle K_i| \partial K/\partial\lambda |K_j\rangle.$$

[3]

$$\left.\begin{aligned} \frac{\partial \langle K_i|}{\partial\lambda} |K_j\rangle &= \frac{\langle K_i| \partial K/\partial\lambda |K_j\rangle}{K_i - K_j} \\ \langle K_j| \frac{\partial |K_i\rangle}{\partial\lambda} &= \frac{\langle K_j| \partial K/\partial\lambda |K_i\rangle}{K_i - K_j} \end{aligned}\right\} \quad (i \neq j).$$

Proof. When $i \neq j$, the equation

$$\langle K_i| K |K_j\rangle = 0$$

holds. This is differentiated with respect to λ to

$$(\partial \langle K_i|/\partial\lambda)K |K_j\rangle + \langle K_i| \partial K/\partial\lambda |K_j\rangle + \langle K_i| K(\partial |K_j\rangle/\partial\lambda) = 0,$$

which becomes

$$K_i(\partial \langle K_i|/\partial\lambda) |K_j\rangle + \langle K_i| \partial K/\partial\lambda |K_j\rangle + K_i \langle K_i| (\partial |K_j\rangle/\partial\lambda) = 0. \quad \text{(a)}$$

Also, the identical equation

$$\langle K_i|K_j\rangle = 0$$

is differentiated with respect to λ to

$$(\partial \langle K_i|/\partial\lambda) |K_j\rangle + \langle K_i| (\partial |K_j\rangle/\partial\lambda) = 0. \quad \text{(b)}$$

From (a) and (b), the equation

$$\frac{\partial \langle K_i|}{\partial\lambda} |K_j\rangle = \frac{\langle K_i| \partial K/\partial\lambda |K_j\rangle}{K_i - K_j}$$

follows, of which the conjugate form is

$$\langle K_j| \frac{\partial |K_i\rangle}{\partial\lambda} = \frac{\langle K_j| \partial K/\partial\lambda |K_i\rangle}{K_i - K_j}.$$

Remark. If we want the left-hand sides of the formulas [3] to be expressed in terms of the wavefunctions ψ_i, ψ_j corresponding to $|K_i\rangle$, $|K_j\rangle$, they become

$$\int \frac{\partial \bar{\psi}_i}{\partial\lambda} \psi_j \, d\tau, \qquad \int \bar{\psi}_i \frac{\partial \psi_i}{\partial\lambda} \, d\tau.$$

[4] It is possible to choose a phase factor of $|K_i\rangle$ such that

$$\langle K_i| \frac{\partial |K_i\rangle}{\partial\lambda} \equiv 0.$$

Proof.

$$g_i(\lambda) \equiv \langle K_i| (\partial |K_i\rangle/\partial\lambda)$$

is assumed to be not identically zero; this is the general case. Since

$$g_i(\lambda) + \bar{g}_i(\lambda) = \langle K_i| (\partial |K_i\rangle/\partial\lambda) + (\partial \langle K_i|/\partial\lambda) |K_i\rangle = \partial\langle K_i|K_i\rangle/\partial\lambda = 0,$$

$g_i(\lambda)$ is purely imaginary. We consider another vector

$$|K_i\rangle^* = \{\exp i\theta_i(\lambda)\} |K_i\rangle \qquad (\theta_i(\lambda): \text{a real function of } \lambda),$$

which differs from $|K_i\rangle$ only in phase factor. (Let the imaginary unit i be distinguished from the subscript i.) The conjugate form of this is

$${}^*\langle K_i| = \{\exp(-i)\theta_i(\lambda)\} \langle K_i|.$$

After a calculation, the equation

$$*\langle K_i|\ (\partial\ |K_i\rangle^*/\partial\lambda) = g_i(\lambda) + i\, d\theta_i(\lambda)/d\lambda$$

is obtained. Therefore, we see that if

$$\theta_i(\lambda) = i\int g_i(\lambda)\, d\lambda,$$

then

$$*\langle K_i|\ (\partial\ |K_i\rangle^*/\partial\lambda) \equiv 0.$$

This $\theta_i(\lambda)$ can certainly be real, since $g_i(\lambda)$ is purely imaginary.

[5] If the phase factor of $|K_i\rangle$ is chosen appropriately, then

$$\frac{\partial\ |K_i\rangle}{\partial\lambda} = \sum_{j(\neq i)} |K_j\rangle \frac{\langle K_j|\ \partial K/\partial\lambda\ |K_i\rangle}{K_i - K_j},$$

$$\frac{\partial\ \langle K_i|}{\partial\lambda} = \sum_{j(\neq i)} \frac{\langle K_i|\ \partial K/\partial\lambda\ |K_j\rangle}{K_i - K_j} \langle K_j|.$$

Proof. If the phase factor of $|K_i\rangle$ is chosen such that

$$\langle K_i|\ (\partial\ |K_i\rangle/\partial\lambda) = 0,$$

then

$$\partial\ |K_i\rangle/\partial\lambda = \sum_{j(\neq i)} |K_j\rangle\ \langle K_j|\ (\partial\ |K_i\rangle/\partial\lambda),$$

where $\sum_{j(\neq i)}$ means the summation over all j but i. Substituting [3] into the right-hand side, we obtain the upper of the relevant two formulas. Taking the conjugate form of this, we obtain the lower.

Remark. The norm of $|K_i\rangle$, i.e., $\langle K_i|K_i\rangle$, is invariant under any alteration in the phase factor, but the norm of $\partial\ |K_i\rangle/\partial\lambda$ is not, in general. Let us consider the vector

$$|K_i\rangle^* = e^{i\theta}\ |K_i\rangle \qquad (\theta \text{ is a real function of } \lambda),$$

which differs from $|K_i\rangle$ only in phase factor. If we put

$$\langle K_i|\ (\partial\ |K_i\rangle/\partial\lambda) = g,$$

which, we know, is a purely imaginary function of λ, we have, after a calculation,

$$\frac{\partial^*\ \langle K_i|}{\partial\lambda}\frac{\partial\ |K_i\rangle^*}{\partial\lambda} - \frac{\partial\ \langle K_i|}{\partial\lambda}\frac{\partial\ |K_i\rangle}{\partial\lambda} = \frac{d\theta}{d\lambda}\left(\frac{d\theta}{d\lambda} - 2ig\right).$$

It is obvious that the right-hand side is not zero, in general. Thus, the norm of $\partial|K_i\rangle/\partial\lambda$ is affected by the way of choosing the phase factor of $|K_i\rangle$. The norm of $|K_i\rangle$ is always taken as unity, but the norm of $\partial\ |K_i\rangle/\partial\lambda$ may no longer be unity. It follows from [5] that

$$\frac{\partial\ \langle K_i|}{\partial\lambda}\frac{\partial\ |K_i\rangle}{\partial\lambda} = \sum_{j(\neq i)} \left|\frac{\langle K_i|\ \partial K/\partial\lambda\ |K_j\rangle}{K_i - K_j}\right|^2,$$

if the phase factor of $|K_i\rangle$ is chosen appropriately. It is noted that the right-hand side itself is invariant under any alteration in phase factors.

[6]

$$\frac{\partial^2 K_i}{\partial\lambda^2} = 2\sum_{j(\neq i)} \frac{|\langle K_i|\ \partial K/\partial\lambda\ |K_j\rangle|^2}{K_i - K_j} + \langle K_i|\ \frac{\partial^2 K}{\partial\lambda^2}\ |K_i\rangle.$$

(The proof of this is given in the "remark" below.)

[7]

$$\frac{\partial^3 K_i}{\partial\lambda^3} = 6\sum_{j(\neq i)}\sum_{k(\neq i)} \frac{(\partial K/\partial\lambda)_{ij}(\partial K/\partial\lambda)_{jk}(\partial K/\partial\lambda)_{ki}}{(K_i - K_j)(K_i - K_k)}$$

$$- 6(\partial K/\partial\lambda)_{ii}\sum_{j(\neq i)} \left|\frac{(\partial K/\partial\lambda)_{ij}}{K_i - K_j}\right|^2$$

$$+3\sum_{j(\neq i)} \frac{(\partial K/\partial\lambda)_{ij}(\partial^2 K/\partial\lambda^2)_{ji} + (\partial^2 K/\partial\lambda^2)_{ij}(\partial K/\partial\lambda)_{ji}}{K_i - K_j}$$

$$+ (\partial^3 K/\partial\lambda^3)_{ii},$$

where $(\partial K/\partial\lambda)_{ij}$, etc., are $\langle K_i|\ \partial K/\partial\lambda\ |K_j\rangle$, etc., abbreviated.

Remark. Below is given a general method for obtaining the formula which enables us to calculate $\partial^n K_i/\partial\lambda^n$ (n: an integer) from eigenvalues of K and matrix elements of $\partial^m K/\partial\lambda^m$ (m: an integer) based on the eigenvectors of K. This method is (i) to differentiate each side of the formula for $\partial^{n-1}K_i/\partial\lambda^{n-1}$ with respect to λ, and then (ii) to use [1] and [5] for the forms $\partial\langle K_i|/\partial\lambda$, $\partial\ |K_i\rangle/\partial\lambda$, and $\partial K_i/\partial\lambda$ that necessarily appear as a result of (i). For instance, in order to obtain the formula for $\partial^2 K_i/\partial\lambda^2$, we differentiate each side of [1] with respect to λ:

$$\frac{\partial^2 K_i}{\partial\lambda^2} = \frac{\partial\ \langle K_i|}{\partial\lambda}\frac{\partial K}{\partial\lambda}\ |K_i\rangle + \langle K_i|\ \frac{\partial^2 K}{\partial\lambda^2}\ |K_i\rangle + \langle K_i|\ \frac{\partial K}{\partial\lambda}\frac{\partial\ |K_i\rangle}{\partial\lambda},$$

and then substitute [5] into the right-hand first and third terms. We, thus, obtain the formula [6], which was first proposed by Epstein.[2] In order to obtain the formula for $\partial^3 K_i/\partial\lambda^3$, we differentiate each side of [6] with respect to λ:

$$\frac{\partial^3 K_i}{\partial\lambda^3} = 2\sum_{j(\neq i)} \frac{\partial}{\partial\lambda}\frac{\langle K_i|\ \partial K/\partial\lambda\ |K_j\rangle\ \langle K_j|\ \partial K/\partial\lambda\ |K_i\rangle}{K_i - K_j} + \frac{\partial}{\partial\lambda}\langle K_i|\ \frac{\partial^2 K}{\partial\lambda^2}\ |K_i\rangle.$$

The right-hand second term becomes

[2] S. T. Epstein, Am. J. Phys. **22**, 613 (1954).

$$\frac{\partial}{\partial\lambda}\langle K_i|\frac{\partial^2 K}{\partial\lambda^2}|K_i\rangle = \frac{\partial\langle K_i|}{\partial\lambda}\frac{\partial^2 K}{\partial\lambda^2}|K_i\rangle + \langle K_i|\frac{\partial^3 K}{\partial\lambda^3}|K_i\rangle + \langle K_i|\frac{\partial^2 K}{\partial\lambda^2}\frac{\partial|K_i\rangle}{\partial\lambda},$$

and the first term becomes

$$2\sum_{j(\neq i)}\frac{\partial}{\partial\lambda}\frac{\langle K_i|\,\partial K/\partial\lambda\,|K_j\rangle\langle K_j|\,\partial K/\partial\lambda\,|K_i\rangle}{K_i - K_j}$$

$$= 2\sum_{j(\neq i)}\Bigg[\frac{1}{K_i - K_j}\Bigg\{\frac{\partial\langle K_i|}{\partial\lambda}\frac{\partial K}{\partial\lambda}|K_j\rangle\langle K_j|\frac{\partial K}{\partial\lambda}|K_i\rangle + \langle K_i|\frac{\partial^2 K}{\partial\lambda^2}|K_j\rangle\langle K_j|\frac{\partial K}{\partial\lambda}|K_i\rangle + \langle K_i|\frac{\partial K}{\partial\lambda}\frac{\partial|K_j\rangle}{\partial\lambda}\langle K_j|\frac{\partial K}{\partial\lambda}|K_i\rangle + \langle K_i|\frac{\partial K}{\partial\lambda}|K_j\rangle\frac{\partial\langle K_j|}{\partial\lambda}\frac{\partial K}{\partial\lambda}|K_i\rangle + \langle K_i|\frac{\partial K}{\partial\lambda}|K_j\rangle\langle K_j|\frac{\partial^2 K}{\partial\lambda^2}|K_i\rangle + \langle K_i|\frac{\partial K}{\partial\lambda}|K_j\rangle\langle K_j|\frac{\partial K}{\partial\lambda}\frac{\partial|K_i\rangle}{\partial\lambda}\Bigg\} + \frac{\langle K_i|\,\partial K/\partial\lambda\,|K_j\rangle\langle K_j|\,\partial K/\partial\lambda\,|K_i\rangle}{(K_i - K_j)^2}\times\left(\frac{\partial K_j}{\partial\lambda} - \frac{\partial K_i}{\partial\lambda}\right)\Bigg].$$

Subsequently, [1] and [5] are used for the forms $\partial\langle K_j|/\partial\lambda$, $\partial|K_j\rangle/\partial\lambda$, and $\partial K_j/\partial\lambda$. The final result is given in [7]. We can, in this way, obtain the formula for $\partial^n K_i/\partial\lambda^n$, with as high an n as we want, only by the repetition of differentiation with respect to λ and replacement by [1] and [5]. Now a commentary must be made on the question of phase factors; for, the equations in [5] are valid only for certain particular phase factors. We temporarily assume, for the use of these equations, that the phase factors of all $|K_i\rangle$ are chosen such that $\langle K_i|\,(\partial|K_i\rangle/\partial\lambda) \equiv 0$. The formula for $\partial^n K_i/\partial\lambda^n$ obtained in this way is, however, invariant under any alteration in phase factors, since any vector is seen in a pair with its conjugate vector. Therefore, the above assumption is unnecessary for the validity of the formula; or the formula is valid for any phase factors.

[8] When E and V are Hermitian linear operators and $H = E + V$, then

$$f(H_i) = f(E_i) + f'(E_i)V_{ii} + f'(E_i)\times\sum_{j(\neq i)}\frac{V_{ij}V_{ji}}{E_i - E_j} + \tfrac{1}{2}f''(E_i)V_{ii}^2 + f'(E_i)\Bigg\{\sum_{j(\neq i)}\sum_{k(\neq i)}\frac{V_{ij}V_{jk}V_{ki}}{(E_i - E_j)(E_i - E_k)} - V_{ii}\sum_{j(\neq i)}\frac{V_{ij}V_{ji}}{(E_i - E_j)^2}\Bigg\} + f''(E_i)V_{ii}\sum_{j(\neq i)}\frac{V_{ij}V_{ji}}{E_i - E_j} + \tfrac{1}{6}f'''(E_i)V_{ii}^3 + \cdots,$$

where V_{ij}, etc., are $\langle E_i|V|E_j\rangle$, etc., abbreviated.

Proof. We put

$$K(\lambda) = E + \lambda V;$$

then

$$\partial K/\partial\lambda = V, \qquad \partial^2 K/\partial\lambda^2 = 0.$$

We regard $f(K_i)$ as a function of λ and apply to it Taylor's expansion:

$$f[K_i(\lambda + \Delta\lambda)] = f[K_i(\lambda)] + \Delta\lambda\,\partial f(K_i)/\partial\lambda + \tfrac{1}{2}\Delta\lambda^2\,\partial^2 f(K_i)/\partial\lambda^2 + \tfrac{1}{6}\Delta\lambda^3\,\partial^3 f(K_i)/\partial\lambda^3 + \cdots.$$

In order to evaluate the coefficients $\partial f(K_i)/\partial\lambda$, $\partial^2 f(K_i)/\partial\lambda^2$, etc., we modify these as

$$\partial f(K_i)/\partial\lambda = f'(K_i)\,\partial K_i/\partial\lambda,$$

$$\partial^2 f(K_i)/\partial\lambda^2 = f'(K_i)\,\partial^2 K_i/\partial\lambda^2 + f''(K_i)(\partial K_i/\partial\lambda)^2,$$

$$\partial^3 f(K_i)/\partial\lambda^3 = f'(K_i)\,\partial^3 K_i/\partial\lambda^3 + 3f''(K_i)(\partial K_i/\partial\lambda)(\partial^2 K_i/\partial\lambda^2) + f'''(K_i)(\partial K_i/\partial\lambda)^3,$$

and then replace $\partial K_i/\partial\lambda$, $\partial^2 K_i/\partial\lambda^2$, etc. in the right-hand sides by

$$\partial K_i/\partial\lambda = V_{ii} \quad \text{(from [1])},$$

$$\partial^2 K_i/\partial\lambda^2 = 2\sum_{j(\neq i)}V_{ij}V_{ji}/(K_i - K_j) \quad \text{(from [6])},$$

$$\frac{\partial^3 K_i}{\partial\lambda^3} = 6\sum_{j(\neq i)}\sum_{k(\neq i)}\frac{V_{ij}V_{jk}V_{ki}}{(K_i - K_j)(K_i - K_k)} - 6V_{ii}\sum_{j(\neq i)}\frac{V_{ij}V_{ji}}{(K_i - K_j)^2} \quad \text{(from [7])}.$$

If we especially put $\lambda = 0$ and $\Delta\lambda = 1$, we obtain the relevant formula.

[9] If the phase factor of $|K_i\rangle$ is chosen appropriately, then

$$\frac{\partial^2 |K_i\rangle}{\partial\lambda^2} = \sum_{j(\neq i)} |K_j\rangle \times \left\{ 2 \sum_{k(\neq i)} \frac{\langle K_j| \partial K/\partial\lambda |K_k\rangle \langle K_k| \partial K/\partial\lambda |K_i\rangle}{(K_i - K_j)(K_i - K_k)} - 2\frac{\langle K_j| \partial K/\partial\lambda |K_i\rangle \langle K_i| \partial K/\partial\lambda |K_i\rangle}{(K_i - K_j)^2} + \frac{\langle K_j| \partial^2 K/\partial\lambda^2 |K_i\rangle}{K_i - K_j} \right\} - |K_i\rangle \sum_{k(\neq i)} \left| \frac{\langle K_i| \partial K/\partial\lambda |K_k\rangle}{K_i - K_k} \right|^2 .$$

(The proof of this is given in the "remark" below.)

[10] If the phase factor of $|K_i\rangle$ is chosen appropriately, then

$$\langle K_j| \frac{\partial^3 |K_i\rangle}{\partial\lambda^3} = 6 \sum_{k(\neq i)} \sum_{l(\neq i)} \frac{(\partial K/\partial\lambda)_{jk}(\partial K/\partial\lambda)_{kl}(\partial K/\partial\lambda)_{li}}{(K_i - K_j)(K_i - K_k)(K_i - K_l)} - 6 \sum_{k(\neq i)} \left\{ \frac{1}{(K_i - K_j)^2(K_i - K_k)} + \frac{1}{(K_i - K_j)(K_i - K_k)^2} \right\} \left(\frac{\partial K}{\partial\lambda}\right)_{jk} \left(\frac{\partial K}{\partial\lambda}\right)_{ki} \left(\frac{\partial K}{\partial\lambda}\right)_{ii} - 6 \frac{(\partial K/\partial\lambda)_{ji}}{(K_i - K_j)^2} \sum_{k(\neq i)} \frac{|(\partial K/\partial\lambda)_{ik}|^2}{K_i - K_k} - 3 \frac{(\partial K/\partial\lambda)_{ji}}{K_i - K_j} \sum_{k(\neq i)} \left| \frac{(\partial K/\partial\lambda)_{ik}}{K_i - K_k} \right|^2 + 6 \frac{(\partial K/\partial\lambda)_{ji}(\partial K/\partial\lambda)_{ii}^2}{(K_i - K_j)^3} + 3 \sum_{k(\neq i)} \frac{(\partial K/\partial\lambda)_{jk}(\partial^2 K/\partial\lambda^2)_{ki} + (\partial^2 K/\partial\lambda^2)_{jk}(\partial K/\partial\lambda)_{ki}}{(K_i - K_j)(K_i - K_k)} - 3 \frac{(\partial K/\partial\lambda)_{ji}(\partial^2 K/\partial\lambda^2)_{ii} + (\partial^2 K/\partial\lambda^2)_{ji}(\partial K/\partial\lambda)_{ii}}{(K_i - K_j)^2} + \frac{(\partial^3 K/\partial\lambda^3)_{ji}}{K_i - K_j} \qquad (i \neq j),$$

and

$$\langle K_i| \frac{\partial^3 |K_i\rangle}{\partial\lambda^3} = -2 \sum_{k(\neq i)} \sum_{l(\neq i)} \left\{ \frac{2}{(K_i - K_k)^2(K_i - K_l)} + \frac{1}{(K_i - K_k)(K_i - K_l)^2} \right\} \times \left(\frac{\partial K}{\partial\lambda}\right)_{ik} \left(\frac{\partial K}{\partial\lambda}\right)_{kl} \left(\frac{\partial K}{\partial\lambda}\right)_{li} + 6 \left(\frac{\partial K}{\partial\lambda}\right)_{ii} \sum_{k(\neq i)} \frac{|(\partial K/\partial\lambda)_{ik}|^2}{(K_i - K_k)^3} - \sum_{k(\neq i)} \frac{2(\partial K/\partial\lambda)_{ik}(\partial^2 K/\partial\lambda^2)_{ki} + (\partial^2 K/\partial\lambda^2)_{ik}(\partial K/\partial\lambda)_{ki}}{(K_i - K_k)^2},$$

where $(\partial K/\partial\lambda)_{ik}$, etc., are $\langle K_i| \partial K/\partial\lambda |K_k\rangle$, etc., abbreviated.

Remark. We can obtain the formula which enables us to calculate $\partial^n |K_i\rangle/\partial\lambda^n$ from the eigenvectors and eigenvalues of K and the matrix elements of $\partial^m K/\partial\lambda^m$ based by the eigenvectors of K. The method is similar to that for $\partial^n K_i/\partial\lambda^n$; we have only to perform the differentiation of the formula for $\partial^{n-1} |K_i\rangle/\partial\lambda^{n-1}$ with respect to λ and the replacement by [1] and [5]. For the use of [5], we temporarily assume that the phase factors of all $|K_i\rangle$ are chosen such that $\langle K_i| (\partial |K_i\rangle/\partial\lambda) \equiv 0$. But it is found that the final formula for $\partial^n |K_i\rangle/\partial\lambda^n$ obtained in this way is invariant under any alteration in the phase factors of all $|K_i\rangle$ except $|K_i\rangle$. (For, any vector $|K_j\rangle$ other than $|K_i\rangle$ is always seen in a pair with its conjugate vector $\langle K_j|$.) Hence, for the validity of our formula, only the phase factor of $|K_i\rangle$ must satisfy the above requirement. [9] gives the formula for $\partial^2 |K_i\rangle/\partial\lambda^2$, which was first proposed by Epstein.[2] [10] gives the formula for $\partial^3 |K_i\rangle/\partial\lambda^3$.

[11] Let E and V be Hermitian linear operators, and let $H = E + V$. If the phase factor of $|H_i\rangle$ or $|E_i\rangle$ is chosen appropriately, then

$$|H_i\rangle = |E_i\rangle + \sum_{j(\neq i)} |E_j\rangle \frac{V_{ji}}{E_i - E_j} + \text{the second-order term} + \text{the third-order term} + \cdots ;$$

the second-order term =

$$\sum_{j(\neq i)} |E_j\rangle \left\{ \sum_{k(\neq i)} \frac{V_{jk}V_{ki}}{(E_i - E_j)(E_i - E_k)} - \frac{V_{ji}V_{ii}}{(E_i - E_j)^2} \right\} - \tfrac{1}{2} |E_i\rangle \sum_{k(\neq i)} \frac{V_{ik}V_{ki}}{(E_i - E_k)^2},$$

the third-order term =

$$\sum_{j(\neq i)} |E_j\rangle \left[\sum_{k(\neq i)} \sum_{l(\neq i)} \frac{V_{jk}V_{kl}V_{li}}{(E_i - E_j)(E_i - E_k)(E_i - E_l)} - \sum_{k(\neq i)} \left\{ \frac{1}{(E_i - E_j)^2(E_i - E_k)} + \frac{1}{(E_i - E_j)(E_i - E_k)^2} \right\} V_{jk}V_{ki}V_{ii} - \frac{V_{ji}}{(E_i - E_j)^2} \sum_{k(\neq i)} \frac{V_{ik}V_{ki}}{E_i - E_k} - \frac{1}{2} \frac{V_{ji}}{E_i - E_j} \sum_{k(\neq i)} \frac{V_{ik}V_{ki}}{(E_i - E_k)^2}$$

$$+\frac{V_{ji}V_{ii}V_{ii}}{(E_i-E_j)^3}\Big]$$
$$+|E_i\rangle\Big[-\tfrac{1}{3}\sum_{k(\neq i)}\sum_{l(\neq i)}\Big\{\frac{2}{(E_i-E_k)^2(E_i-E_l)}$$
$$+\frac{1}{(E_i-E_k)(E_i-E_l)^2}\Big\}V_{ik}V_{kl}V_{li}+V_{ii}$$
$$\times\sum_{k(\neq i)}\frac{V_{ik}V_{ki}}{(E_i-E_k)^3}\Big].$$

Proof. We put

$$K(\lambda)=E+\lambda V.$$

When $\langle a|$ is an arbitrary vector not dependent on λ, $\langle a|K_i(\lambda)\rangle$ is a complex-number function of λ. We expand $\langle a|K_i(\lambda+\Delta\lambda)\rangle$ into a Taylor's series:

$$\langle a|K_i(\lambda+\Delta\lambda)\rangle=\langle a|K_i(\lambda)\rangle+\Delta\lambda(\partial\langle a|K_i\rangle/\partial\lambda)$$
$$+\tfrac{1}{2}\Delta\lambda^2(\partial^2\langle a|K_i\rangle/\partial\lambda^2)+\tfrac{1}{6}\Delta\lambda^3(\partial^3\langle a|K_i\rangle/\partial\lambda^3)+\cdots.$$

The right-hand side is equal to

$$\langle a|\{|K_i(\lambda)\rangle+\Delta\lambda(\partial|K_i\rangle/\partial\lambda)+\tfrac{1}{2}\Delta\lambda^2(\partial^2|K_i\rangle/\partial\lambda^2)$$
$$+\tfrac{1}{6}\Delta\lambda^3(\partial^3|K_i\rangle/\partial\lambda^3)+\cdots\}.$$

We remove $\langle a|$ from the above equation:

$$|K_i(\lambda+\Delta\lambda)\rangle=|K_i(\lambda)\rangle+\Delta\lambda(\partial|K_i\rangle/\partial\lambda)$$
$$+\tfrac{1}{2}\Delta\lambda^2(\partial^2|K_i\rangle/\partial\lambda^2)+\tfrac{1}{6}\Delta\lambda^3(\partial^3|K_i\rangle/\partial\lambda^3)+\cdots.$$

If we here use [5], [9], and [10], and put $\lambda=0$, $\Delta\lambda=1$, then we obtain the relevant formula. It may be obvious that, for the validity of this formula, there must be a certain relation between the phase factors of $|E_i\rangle$ and $|H_i\rangle$.

Remark. This formula is well known in the conventional perturbation theory. But the methods of derivation are not the same. In the conventional standard method, the expansion of $|H_i\rangle$ is advanced in conjugation with the expansion of H_i. Therefore, if one of the two expansions is suspended at a certain order of approximation, it is impossible to put the other expansion as forward as we want. In the present method, which is the same as Epstein proposed,[2] the two expansions are advanced, quite separately and only by the repetition of differentiation with respect to λ and replacement by [1] and [5].

[12]

$$\langle K_i|\frac{\partial^2 f(K)}{\partial\lambda^2}|K_j\rangle=\sum_k\frac{2}{K_i-K_j}$$
$$\times\Big\{\frac{f(K_i)-f(K_k)}{K_i-K_k}-\frac{f(K_k)-f(K_j)}{K_k-K_j}\Big\}$$
$$\times\langle K_i|\frac{\partial K}{\partial\lambda}|K_k\rangle\langle K_k|\frac{\partial K}{\partial\lambda}|K_j\rangle$$
$$+\frac{f(K_i)-f(K_j)}{K_i-K_j}\langle K_i|\frac{\partial^2 K}{\partial\lambda^2}|K_j\rangle,$$

where i and j may as well be equal. (The proof of this is given in the "remark" below.)

[13]

$$\langle K_i|\frac{\partial^3 f(K)}{\partial\lambda^3}|K_j\rangle=6\sum_k\sum_l\Big[\frac{1}{K_i-K_l}$$
$$\times\Big\{\frac{f(K_i)-f(K_k)}{K_i-K_k}-\frac{f(K_k)-f(K_l)}{K_k-K_l}\Big\}$$
$$-\frac{1}{K_k-K_j}\Big\{\frac{f(K_k)-f(K_l)}{K_k-K_l}-\frac{f(K_l)-f(K_j)}{K_l-K_j}\Big\}\Big]$$
$$\times\frac{(\partial K/\partial\lambda)_{ik}(\partial K/\partial\lambda)_{kl}(\partial K/\partial\lambda)_{lj}}{K_i-K_j}$$
$$+3\sum_k\frac{1}{K_i-K_j}\Big\{\frac{f(K_i)-f(K_k)}{K_i-K_k}$$
$$-\frac{f(K_k)-f(K_j)}{K_k-K_j}\Big\}\Big\{\Big(\frac{\partial K}{\partial\lambda}\Big)_{ik}\Big(\frac{\partial^2 K}{\partial\lambda^2}\Big)_{kj}$$
$$+\Big(\frac{\partial^2 K}{\partial\lambda^2}\Big)_{ik}\Big(\frac{\partial K}{\partial\lambda}\Big)_{kj}\Big\}+\frac{f(K_i)-f(K_j)}{K_i-K_j}\Big(\frac{\partial^3 K}{\partial\lambda^3}\Big)_{ij},$$

where i and j may as well be equal.

Remark. Differentiating the equation

$$f(K)K=Kf(K)$$

with respect to λ successively, we have

$$\frac{\partial f(K)}{\partial\lambda}K+f(K)\frac{\partial K}{\partial\lambda}=\frac{\partial K}{\partial\lambda}f(K)+K\frac{\partial f(K)}{\partial\lambda},$$

$$\frac{\partial^2 f(K)}{\partial\lambda^2}K+2\frac{\partial f(K)}{\partial\lambda}\frac{\partial K}{\partial\lambda}+f(K)\frac{\partial^2 K}{\partial\lambda^2}$$
$$=\frac{\partial^2 K}{\partial\lambda^2}f(K)+2\frac{\partial K}{\partial\lambda}\frac{\partial f(K)}{\partial\lambda}+K\frac{\partial^2 f(K)}{\partial\lambda^2},$$

$$\frac{\partial^3 f(K)}{\partial\lambda^3}K+3\frac{\partial^2 f(K)}{\partial\lambda^2}\frac{\partial K}{\partial\lambda}+3\frac{\partial f(K)}{\partial\lambda}\frac{\partial^2 K}{\partial\lambda^2}+f(K)\frac{\partial^3 K}{\partial\lambda^3}$$
$$=\frac{\partial^3 K}{\partial\lambda^3}f(K)+3\frac{\partial^2 K}{\partial\lambda^2}\frac{\partial f(K)}{\partial\lambda}+3\frac{\partial K}{\partial\lambda}\frac{\partial^2 f(K)}{\partial\lambda^2}+K\frac{\partial^3 f(K)}{\partial\lambda^3}.$$

Multiplying each side by $\langle K_i|$ and $|K_j\rangle$ from the left and right, respectively, we get

$$K_j\langle K_i|\frac{\partial f(K)}{\partial\lambda}|K_j\rangle+f(K_i)\langle K_i|\frac{\partial K}{\partial\lambda}|K_j\rangle$$
$$=f(K_j)\langle K_i|\frac{\partial K}{\partial\lambda}|K_j\rangle+K_i\langle K_i|\frac{\partial f(K)}{\partial\lambda}|K_j\rangle,$$

$$K_j\langle K_i|\frac{\partial^2 f(K)}{\partial\lambda^2}|K_j\rangle+2\langle K_i|\frac{\partial f(K)}{\partial\lambda}\frac{\partial K}{\partial\lambda}|K_j\rangle$$
$$+f(K_i)\langle K_i|\frac{\partial^2 K}{\partial\lambda^2}|K_j\rangle=f(K_j)\langle K_i|\frac{\partial^2 K}{\partial\lambda^2}|K_j\rangle$$
$$+2\langle K_i|\frac{\partial K}{\partial\lambda}\frac{\partial f(K)}{\partial\lambda}|K_j\rangle+K_i\langle K_i|\frac{\partial^2 f(K)}{\partial\lambda^2}|K_j\rangle,$$

$$K_i \langle K_i| \frac{\partial^3 f(K)}{\partial\lambda^3} |K_i\rangle + 3 \langle K_i| \frac{\partial^2 f(K)}{\partial\lambda^2} \frac{\partial K}{\partial\lambda} |K_i\rangle$$
$$+ 3 \langle K_i| \frac{\partial f(K)}{\partial\lambda} \frac{\partial^2 K}{\partial\lambda^2} |K_i\rangle + f(K_i) \langle K_i| \frac{\partial^3 K}{\partial\lambda^3} |K_i\rangle$$
$$= f(K_j) \langle K_i| \frac{\partial^3 K}{\partial\lambda^3} |K_i\rangle + 3 \langle K_i| \frac{\partial^2 K}{\partial\lambda^2} \frac{\partial f(K)}{\partial\lambda} |K_i\rangle$$
$$+ 3 \langle K_i| \frac{\partial K}{\partial\lambda} \frac{\partial^2 f(K)}{\partial\lambda^2} |K_i\rangle + K_i \langle K_i| \frac{\partial^3 f(K)}{\partial\lambda^3} |K_i\rangle.$$

The first equation leads to [2]. In the second equation, we note

$$\langle K_i| \frac{\partial f(K)}{\partial\lambda} \frac{\partial K}{\partial\lambda} |K_i\rangle = \sum_k \langle K_i| \frac{\partial f(K)}{\partial\lambda} |K_k\rangle \langle K_k| \frac{\partial K}{\partial\lambda} |K_i\rangle,$$

$$\langle K_i| \frac{\partial K}{\partial\lambda} \frac{\partial f(K)}{\partial\lambda} |K_i\rangle = \sum_k \langle K_i| \frac{\partial K}{\partial\lambda} |K_k\rangle \langle K_k| \frac{\partial f(K)}{\partial\lambda} |K_i\rangle,$$

and use [2] for the right-hand sides; the final result is given in [12]. In the third equation, we note

$$\langle K_i| \frac{\partial^2 K}{\partial\lambda^2} \frac{\partial f(K)}{\partial\lambda} |K_i\rangle = \sum_k \langle K_i| \frac{\partial^2 K}{\partial\lambda^2} |K_k\rangle \langle K_k| \frac{\partial f(K)}{\partial\lambda} |K_i\rangle,$$

$$\langle K_i| \frac{\partial K}{\partial\lambda} \frac{\partial^2 f(K)}{\partial\lambda^2} |K_i\rangle = \sum_k \langle K_i| \frac{\partial K}{\partial\lambda} |K_k\rangle \langle K_k| \frac{\partial^2 f(K)}{\partial\lambda^2} |K_i\rangle,$$

and substitute [2] into the upper right-hand side and [12] into the lower right-hand side; the final result is given in [13]. [12] and [13] are valid even in case of $i = j$, as [2] is. In this case, if we use the following equations, in elementary analytics:

$$\lim_{x\to a} \frac{f(a) - f(x)}{a - x} = f'(a),$$

$$\lim_{x\to a} \frac{1}{a - x} \left\{ \frac{f(a) - f(b)}{a - b} - \frac{f(b) - f(x)}{b - x} \right\} = \frac{1}{a - b} \left\{ f'(a) - \frac{f(a) - f(b)}{a - b} \right\},$$

$$\lim_{x\to a} \frac{1}{a - x} \left[\frac{1}{a - c} \left\{ \frac{f(a) - f(b)}{a - b} - \frac{f(b) - f(c)}{b - c} \right\} - \frac{1}{b - x} \left\{ \frac{f(b) - f(c)}{b - c} - \frac{f(c) - f(x)}{c - x} \right\} \right]$$
$$= \frac{1}{b - c} \left[\frac{1}{a - b} \left\{ f'(a) - \frac{f(a) - f(b)}{a - b} \right\} - \frac{1}{a - c} \left\{ f'(a) - \frac{f(a) - f(c)}{a - c} \right\} \right],$$

where a, b, c, and x are real numbers, then we can write [12] and [13] as

$$\langle K_i| \frac{\partial^2 f(K)}{\partial\lambda^2} |K_i\rangle = 2 \sum_k \left\{ f'(K_i) - \frac{f(K_i) - f(K_k)}{K_i - K_k} \right\} \times \frac{|\langle K_i| \partial K/\partial\lambda |K_k\rangle|^2}{K_i - K_k} + f'(K_i) \langle K_i| \partial^2 K/\partial\lambda^2 |K_i\rangle,$$

$$\langle K_i| \frac{\partial^3 f(K)}{\partial\lambda^3} |K_i\rangle = 6 \sum_k \sum_l \left[\frac{1}{K_i - K_k} \left\{ f'(K_i) - \frac{f(K_i) - f(K_k)}{K_i - K_k} \right\} - \frac{1}{K_i - K_l} \left\{ f'(K_i) - \frac{f(K_i) - f(K_l)}{K_i - K_l} \right\} \right] \times \frac{(\partial K/\partial\lambda)_{ik} (\partial K/\partial\lambda)_{kl} (\partial K/\partial\lambda)_{li}}{K_k - K_l}$$
$$+ 3 \sum_k \frac{1}{K_i - K_k} \left\{ f'(K_i) - \frac{f(K_i) - f(K_k)}{K_i - K_k} \right\} \times \left\{ \left(\frac{\partial K}{\partial\lambda}\right)_{ik} \left(\frac{\partial^2 K}{\partial\lambda^2}\right)_{ki} + \left(\frac{\partial^2 K}{\partial\lambda^2}\right)_{ik} \left(\frac{\partial K}{\partial\lambda}\right)_{ki} \right\}$$
$$+ f'(K_i) \left(\frac{\partial^3 K}{\partial\lambda^3}\right)_{ii},$$

respectively. The expression

$$\frac{1}{K_i - K_k} \left\{ f'(K_i) - \frac{f(K_i) - f(K_k)}{K_i - K_k} \right\}$$

within the right-hand sides means $\frac{1}{2} f''(K_i)$ when $k = i$; the expression

$$\frac{1}{K_k - K_l} \left[\frac{1}{K_i - K_k} \left\{ f'(K_i) - \frac{f(K_i) - f(K_k)}{K_i - K_k} \right\} - \frac{1}{K_i - K_l} \left\{ f'(K_i) - \frac{f(K_i) - f(K_l)}{K_i - K_l} \right\} \right]$$

means

$$\frac{2}{(K_i - K_k)^2} \left\{ \frac{f'(K_i) + f'(K_k)}{2} - \frac{f(K_i) - f(K_k)}{K_i - K_k} \right\}$$

when $l = k$, and, further, means $\frac{1}{6} f'''(K_i)$ when $l = k = i$.

The formula for $\langle K_i| \partial^n f(K)/\partial\lambda^n |K_i\rangle$ can also be derived by differentiating the formula for $\langle K_i| \partial^{n-1} f(K)/\partial\lambda^{n-1} |K_i\rangle$ with respect to λ and then using [1] and [5], as in the case of $\partial^n K_i/\partial\lambda^n$ and $\partial^n |K_i\rangle/\partial\lambda^n$. The form, however, of the formula

as directly obtained by this method is considerably different from the previous form so that the new form does not readily remind us of the previous one. It may be more convenient to employ the previous method, since a more elegant form of the formula is obtained with less effort.

[14] When E and V are Hermitian linear operators and $H = E + V$, then

$$\langle E_i|\, f(H)\, |E_i\rangle = f(E_i)\,\langle E_i|E_i\rangle + \frac{f(E_i) - f(E_j)}{E_i - E_j}\, V_{ij}$$
$$+ \sum_k \left\{\frac{f(E_i) - f(E_k)}{E_i - E_k} - \frac{f(E_k) - f(E_j)}{E_k - E_j}\right\}\frac{V_{ik}V_{kj}}{E_i - E_j}$$
$$+ \sum_k \sum_l \left[\frac{1}{E_i - E_l}\left\{\frac{f(E_i) - f(E_k)}{E_i - E_k} - \frac{f(E_k) - f(E_l)}{E_k - E_l}\right\}\right.$$
$$- \frac{1}{E_k - E_j}\left\{\frac{f(E_k) - f(E_l)}{E_k - E_l}\right.$$
$$\left.\left. - \frac{f(E_l) - f(E_j)}{E_l - E_j}\right\}\right]\frac{V_{ik}V_{kl}V_{lj}}{E_i - E_j} + \cdots .$$

Proof. We put

$$K(\lambda) = E + \lambda V.$$

When $\langle a|$ and $|b\rangle$ are arbitrary vectors not dependent on λ, $\langle a|\, f[K(\lambda)]\, |b\rangle$ is a complex-number function of λ. We expand $\langle a|\, f[K(\lambda + \Delta\lambda)]\, |b\rangle$ into a Taylor's series:

$$\langle a|\, f[K(\lambda + \Delta\lambda)]\, |b\rangle = \langle a|\, f[K(\lambda)]\, |b\rangle$$
$$+ \Delta\lambda(\partial/\partial\lambda)\,\langle a|\, f(K)\, |b\rangle$$
$$+ \tfrac{1}{2}\Delta\lambda^2(\partial^2/\partial\lambda^2)\,\langle a|\, f(K)\, |b\rangle$$
$$+ \tfrac{1}{6}\Delta\lambda^3(\partial^3/\partial\lambda^3)\,\langle a|\, f(K)\, |b\rangle + \cdots .$$

Since

$$(\partial^n/\partial\lambda^n)\,\langle a|\, f(K)\, |b\rangle = \langle a|\, \partial^n f(K)/\partial\lambda^n\, |b\rangle,$$

the above equation becomes

$$\langle a|\, f[K(\lambda + \Delta\lambda)]\, |b\rangle = \langle a|\, f[K(\lambda)]\, |b\rangle$$
$$+ \Delta\lambda\,\langle a|\, \partial f(K)/\partial\lambda\, |b\rangle$$
$$+ \tfrac{1}{2}\Delta\lambda^2\,\langle a|\, \partial^2 f(K)/\partial\lambda^2\, |b\rangle$$
$$+ \tfrac{1}{6}\Delta\lambda^3\,\langle a|\, \partial^3 f(K)/\partial\lambda^3\, |b\rangle + \cdots .$$

This equation holds, even when $\langle a| = \langle K_i(\lambda)|$ and $|b\rangle = |K_j(\lambda)\rangle$. Hence,

$$\langle K_i(\lambda)|\, f[K(\lambda + \Delta\lambda)]\, |K_j(\lambda)\rangle$$
$$= f[K_i(\lambda)]\,\langle K_i(\lambda)|K_j(\lambda)\rangle + \Delta\lambda\,\langle K_i|\,\frac{\partial f(K)}{\partial\lambda}\,|K_j\rangle$$
$$+ \tfrac{1}{2}\Delta\lambda^2\,\langle K_i|\,\frac{\partial^2 f(K)}{\partial\lambda^2}\,|K_j\rangle$$
$$+ \tfrac{1}{6}\Delta\lambda^3\,\langle K_i|\,\frac{\partial^3 f(K)}{\partial\lambda^3}\,|K_j\rangle + \cdots .$$

If we here use [2], [12], and [13], and put $\lambda = 0$, $\Delta\lambda = 1$, then we get the relevant formula.

Remark. The above formula is valid even when $i = j$. In this case, it can be written as

$$\langle E_i|\, f(H)\, |E_i\rangle = f(E_i) + f'(E_i)V_{ii}$$
$$+ \sum_k \left\{f'(E_i) - \frac{f(E_i) - f(E_k)}{E_i - E_k}\right\}\frac{V_{ik}V_{ki}}{E_i - E_k}$$
$$+ \sum_k \sum_l \left[\frac{1}{E_i - E_k}\left\{f'(E_i) - \frac{f(E_i) - f(E_k)}{E_i - E_k}\right\}\right.$$
$$\left. - \frac{1}{E_i - E_l}\left\{f'(E_i) - \frac{f(E_i) - f(E_l)}{E_i - E_l}\right\}\right]\frac{V_{ik}V_{kl}V_{li}}{E_k - E_l}$$
$$+ \cdots .$$

[15]

$$\frac{\partial}{\partial\lambda}\,\langle K_i|\, A\, |K_i\rangle = \sum_{j(\neq i)} \frac{\langle K_i|\, \partial K/\partial\lambda\, |K_j\rangle\langle K_j|\, A\, |K_i\rangle + \langle K_i|\, A\, |K_j\rangle\langle K_j|\, \partial K/\partial\lambda\, |K_i\rangle}{K_i - K_j} + \langle K_i|\,\frac{\partial A}{\partial\lambda}\,|K_i\rangle.$$

(The proof of this is given in the "remark" below.)

[16]

$$\frac{\partial^2 A_{ii}}{\partial\lambda^2} = 2\sum_{j(\neq i)}\sum_{k(\neq i)}\frac{1}{(K_i - K_j)(K_i - K_k)}$$
$$\times\left\{\left(\frac{\partial K}{\partial\lambda}\right)_{ij}\left(\frac{\partial K}{\partial\lambda}\right)_{jk}A_{ki} + \left(\frac{\partial K}{\partial\lambda}\right)_{ij}A_{jk}\left(\frac{\partial K}{\partial\lambda}\right)_{ki}\right.$$
$$\left. + A_{ij}\left(\frac{\partial K}{\partial\lambda}\right)_{jk}\left(\frac{\partial K}{\partial\lambda}\right)_{ki}\right\}$$
$$- 2\sum_{j(\neq i)}\frac{1}{(K_i - K_j)^2}\left\{\left(\frac{\partial K}{\partial\lambda}\right)_{ii}\left(\frac{\partial K}{\partial\lambda}\right)_{ij}A_{ij}\right.$$
$$\left. + \left(\frac{\partial K}{\partial\lambda}\right)_{ij}A_{ij}\left(\frac{\partial K}{\partial\lambda}\right)_{jj} + A_{ij}\left(\frac{\partial K}{\partial\lambda}\right)_{ij}\left(\frac{\partial K}{\partial\lambda}\right)_{jj}\right\}$$
$$+ \sum_{j(\neq i)}\frac{(\partial^2 K/\partial\lambda^2)_{ij}A_{ji} + A_{ij}(\partial^2 K/\partial\lambda^2)_{ji}}{K_i - K_j}$$
$$+ 2\sum_{j(\neq i)}\frac{(\partial K/\partial\lambda)_{ij}(\partial A/\partial\lambda)_{ji} + (\partial A/\partial\lambda)_{ij}(\partial K/\partial\lambda)_{ji}}{K_i - K_j}$$
$$+ \left(\frac{\partial^2 A}{\partial\lambda^2}\right)_{ii},$$

where A_{ij}, $(\partial K/\partial\lambda)_{ij}$, etc., are $\langle K_i|\, A\, |K_j\rangle$, $\langle K_i|\, \partial K/\partial\lambda\, |K_j\rangle$, etc., abbreviated.

Remark. If $\langle K_i|\ A\ |K_i\rangle$ is differentiated with respect to λ simply, then

$$(\partial/\partial\lambda)\ \langle K_i|\ A\ |K_i\rangle = (\partial\ \langle K_i|/\partial\lambda)A\ |K_i\rangle + \langle K_i|\ A(\partial\ |K_i\rangle/\partial\lambda) + \langle K_i|\ \partial A/\partial\lambda\ |K_i\rangle,$$

$$\frac{\partial^2}{\partial\lambda^2}\langle K_i|\ A\ |K_i\rangle = \frac{\partial^2\langle K_i|}{\partial\lambda^2}A\ |K_i\rangle + \langle K_i|\ A\frac{\partial^2|K_i\rangle}{\partial\lambda^2} + 2\frac{\partial\langle K_i|}{\partial\lambda}A\frac{\partial|K_i\rangle}{\partial\lambda} + 2\left\{\frac{\partial\langle K_i|}{\partial\lambda}\frac{\partial A}{\partial\lambda}|K_i\rangle + \langle K_i|\frac{\partial A}{\partial\lambda}\frac{\partial|K_i\rangle}{\partial\lambda}\right\} + \langle K_i|\frac{\partial^2 A}{\partial\lambda^2}|K_i\rangle.$$

If [5] and [9] are substituted into the right-hand sides, [15] and [16] are obtained. These formulas are valid for any phase factors. [15] was first proposed by Salem.[1] His intention was different from ours. That was to reduce the sum over all $|K_i\rangle$, except $|K_i\rangle$, to a simpler expression which depends on the properties of $|K_i\rangle$ alone. Therefore, he described this formula in the form as shown in Sec. I.

[17]

$$\mathrm{Tr}\ \varphi(K)\frac{\partial f(K)}{\partial\lambda} = \mathrm{Tr}\ \varphi(K)f'(K)\frac{\partial K}{\partial\lambda},$$

where φ and f are arbitrary functions.

Proof.

$$\mathrm{Tr}\ \varphi(K)\ \partial f(K)/\partial\lambda = \sum_i \langle K_i|\ \varphi(K)\ \partial f(K)/\partial\lambda\ |K_i\rangle$$
$$= \sum_i \varphi(K_i)\ \langle K_i|\ \partial f(K)/\partial\lambda\ |K_i\rangle$$
$$= \sum_i \varphi(K_i)f'(K_i)\ \langle K_i|\ \partial K/\partial\lambda\ |K_i\rangle \quad \text{(by [2])}$$
$$= \sum_i \langle K_i|\ \varphi(K)f'(K)\ \partial K/\partial\lambda\ |K_i\rangle$$
$$= \mathrm{Tr}\ \varphi(K)f'(K)\ \partial K/\partial\lambda.$$

[18]

$$\mathrm{Tr}\ \varphi(K)\frac{\partial^2 f(K)}{\partial\lambda^2} = \mathrm{Tr}\ \varphi(K)\frac{\partial f'(K)}{\partial\lambda}\frac{\partial K}{\partial\lambda} + \mathrm{Tr}\ \varphi(K)f'(K)\frac{\partial^2 K}{\partial\lambda^2} + \mathrm{Tr}\frac{\partial\varphi(K)}{\partial\lambda}f'(K)\frac{\partial K}{\partial\lambda} - \mathrm{Tr}\frac{\partial\varphi(K)}{\partial\lambda}\frac{\partial f(K)}{\partial\lambda},$$

where φ and f are arbitrary functions.

Proof. Differentiating the left-hand side of Eq. [17] with respect to λ, we have

$$(\partial/\partial\lambda)\ \mathrm{Tr}\ \varphi(K)\ \partial f(K)/\partial\lambda = \mathrm{Tr}\ (\partial/\partial\lambda)\varphi(K)\ \partial f(K)/\partial\lambda$$
$$= \mathrm{Tr}\ \varphi(K)\ \partial^2 f(K)/\partial\lambda^2 + \mathrm{Tr}\ (\partial\varphi(K)/\partial\lambda)(\partial f(K)/\partial\lambda). \quad \text{(a)}$$

Differentiating the right-hand side of [17], we have

$$(\partial/\partial\lambda)\ \mathrm{Tr}\ \varphi(K)f'(K)\ \partial K/\partial\lambda$$
$$= \mathrm{Tr}\ (\partial/\partial\lambda)\varphi(K)f'(K)\ \partial K/\partial\lambda$$
$$= \mathrm{Tr}\ \varphi(K)(\partial f'(K)/\partial\lambda)(\partial K/\partial\lambda) + \mathrm{Tr}\ \varphi(K)f'(K)\ \partial^2 K/\partial\lambda^2 + \mathrm{Tr}\ (\partial\varphi(K)/\partial\lambda)f'(K)(\partial K/\partial\lambda). \quad \text{(b)}$$

Equating (a) and (b), we obtain the relevant formula.

Remark. We cannot expect the general validity of the equation

$$\mathrm{Tr}\ \varphi(K)\frac{\partial^2 f(K)}{\partial\lambda^2} = \mathrm{Tr}\ \varphi(K)f''(K)\left(\frac{\partial K}{\partial\lambda}\right)^2 + \mathrm{Tr}\ \varphi(K)f'(K)\frac{\partial^2 K}{\partial\lambda^2},$$

but only of Eq. [18]. In case of $\varphi(K) = 1$, [17] becomes

$$\mathrm{Tr}\ \partial f(K)/\partial\lambda = \mathrm{Tr}\ f'(K)\ \partial K/\partial\lambda, \quad \text{(c)}$$

and [18] becomes

$$\mathrm{Tr}\ \partial^2 f(K)/\partial\lambda^2 = \mathrm{Tr}\ (\partial f'(K)/\partial\lambda)(\partial K/\partial\lambda) + \mathrm{Tr}\ f'(K)\ \partial^2 K/\partial\lambda^2. \quad \text{(d)}$$

Even in this case, we cannot expect the general validity of the equation

$$\mathrm{Tr}\ \partial^2 f(K)/\partial\lambda^2 = \mathrm{Tr}\ f''(K)(\partial K/\partial\lambda)^2 + \mathrm{Tr}\ f'(K)\ \partial^2 K/\partial\lambda^2. \quad \text{(e)}$$

In order that (e) is valid, the right-hand first term of (e) must be equal to the right-hand first term of (d):

$$\mathrm{Tr}\ f''(K)(\partial K/\partial\lambda)^2 = \mathrm{Tr}\ (\partial f'(K)/\partial\lambda)(\partial K/\partial\lambda).$$

But this is, in general, not the case, although the equation

$$\mathrm{Tr}\ f''(K)\ \partial K/\partial\lambda = \mathrm{Tr}\ \partial f'(K)/\partial\lambda$$

is valid by (c).

Remark. If $K = E + \lambda V$, where E and V are Hermitian linear operators not dependent on λ, then [17] becomes, at $\lambda = 0$,

$$\mathrm{Tr}\ \varphi(E)[\partial f(E + \lambda V)/\partial\lambda]_{\lambda=0} = \mathrm{Tr}\ \varphi(E)f'(E)V. \quad \text{(f)}$$

$\mathrm{Tr}\ \varphi(E)f(E + \lambda V)$, which is a complex-number function of λ, may approximately be developed as

$$\mathrm{Tr}\ \varphi(E)f(E + \lambda V) \approx \mathrm{Tr}\ \varphi(E)f(E) + \lambda[(\partial/\partial\lambda)\ \mathrm{Tr}\ \varphi(E)f(E + \lambda V)]_{\lambda=0}. \quad \text{(g)}$$

If in (g) we put $\lambda = 1$ and use (f) for the right-hand second term ($= \mathrm{Tr}\, \varphi(E)[\partial f(E + \lambda V)/\partial\lambda]_{\lambda=0}$), then (g) becomes

$$\mathrm{Tr}\, \varphi(E)f(E + V) \approx \mathrm{Tr}\, \varphi(E)f(E) + \mathrm{Tr}\, \varphi(E)f'(E)V.$$

This approximation formula can also be obtained by use of [14].

[19]

$$\mathrm{Tr}\, \varphi(K) \frac{\partial f(K)}{\partial\lambda} \frac{\partial g(K)}{\partial\lambda} = \mathrm{Tr}\, \varphi(K) \frac{\partial g(K)}{\partial\lambda} \frac{\partial f(K)}{\partial\lambda},$$

where φ, f, and g are arbitrary functions.

Proof.

$$\begin{aligned}
&\mathrm{Tr}\, \varphi(K)(\partial f(K)/\partial\lambda)(\partial g(K)/\partial\lambda) \\
&= \sum_i \sum_j \varphi(K_i) \langle K_i| \partial f(K)/\partial\lambda |K_j\rangle \langle K_j| \partial g(K)/\partial\lambda |K_i\rangle \\
&= \sum_i \sum_j \varphi(K_i)\{f(K_i) - f(K_j)\} \\
&\quad \times \{g(K_i) - g(K_j)\}(K_i - K_j)^{-2} \\
&\quad \times \langle K_i| \partial K/\partial\lambda |K_j\rangle \langle K_j| \partial K/\partial\lambda |K_i\rangle \quad \text{(by [2])}.
\end{aligned}$$

The last expression is, obviously, unaffected by the interchange between f and g. Hence, the relevant formula holds.

The present theorem can also be proved as follows. We differentiate the identical equation

$$f(K)g(K) = g(K)f(K)$$

with respect to λ two times, and then multiply each side of the resultant equation by $\varphi(K)$ from the left, and finally take the trace of each side. The result is

$$\begin{aligned}
&\mathrm{Tr}\, \varphi(K) \frac{\partial^2 f(K)}{\partial\lambda^2} g(K) + 2\, \mathrm{Tr}\, \varphi(K) \frac{\partial f(K)}{\partial\lambda} \frac{\partial g(K)}{\partial\lambda} \\
&\quad + \mathrm{Tr}\, \varphi(K)f(K) \frac{\partial^2 g(K)}{\partial\lambda^2} = \mathrm{Tr}\, \varphi(K) \frac{\partial^2 g(K)}{\partial\lambda^2} f(K) \\
&\quad + 2\, \mathrm{Tr}\, \varphi(K) \frac{\partial g(K)}{\partial\lambda} \frac{\partial f(K)}{\partial\lambda} + \mathrm{Tr}\, \varphi(K)g(K) \frac{\partial^2 f(K)}{\partial\lambda^2}.
\end{aligned}$$

The left-hand first term is equal to the right-hand third term; for,

$$\mathrm{Tr}\, \varphi(K)(\partial^2 f(K)/\partial\lambda^2)g(K) = \mathrm{Tr}\, g(K)\varphi(K)(\partial^2 f(K)/\partial\lambda^2)$$

(by the commutability inside the trace)

$$= \mathrm{Tr}\, \varphi(K)g(K)(\partial^2 f(K)/\partial\lambda^2)$$

(by the commutability between $\varphi(K)$ and $g(K)$). Similarly, the left-hand third term is equal to the right-hand first term. Consequently, the second terms on both sides are equal.

III. TWO PARAMETERS

[20]

$$\frac{\partial^2 K_i}{\partial\lambda\, \partial\mu} = \sum_{j(\neq i)} \frac{\langle K_i| \partial K/\partial\lambda |K_j\rangle\langle K_j| \partial K/\partial\mu |K_i\rangle + \langle K_i| \partial K/\partial\mu |K_j\rangle\langle K_j| \partial K/\partial\lambda |K_i\rangle}{K_i - K_j} + \langle K_i| \frac{\partial^2 K}{\partial\lambda\, \partial\mu} |K_i\rangle.$$

Proof. The equation (from [1])

$$\partial K_i/\partial\mu = \langle K_i| \partial K/\partial\mu |K_i\rangle$$

is differentiated with respect to λ to

$$\frac{\partial^2 K_i}{\partial\lambda\, \partial\mu} = \frac{\partial \langle K_i|}{\partial\lambda} \frac{\partial K}{\partial\mu} |K_i\rangle + \langle K_i| \frac{\partial^2 K}{\partial\lambda\, \partial\mu} |K_i\rangle + \langle K_i| \frac{\partial K}{\partial\mu} \frac{\partial |K_i\rangle}{\partial\lambda}.$$

Substituting [5] into the right-hand first and third terms, we obtain the relevant formula, which is valid for any phase factors.

Remark. This formula was first proposed by Brown.[3] Generally, $\partial^{m+n}K_i/\partial\lambda^m\, \partial\mu^n$ (m, n: integers) can also be evaluated in a similar way.

[21]

$$\begin{aligned}
\langle K_j| \frac{\partial^2 |K_i\rangle}{\partial\lambda\, \partial\mu} &= \sum_{k(\neq i)} \frac{\langle K_j| \partial K/\partial\lambda |K_k\rangle\langle K_k| \partial K/\partial\mu |K_i\rangle + \langle K_j| \partial K/\partial\mu |K_k\rangle\langle K_k| \partial K/\partial\lambda |K_i\rangle}{(K_i - K_j)(K_i - K_k)} \\
&\quad - \frac{\langle K_j| \partial K/\partial\lambda |K_i\rangle\langle K_i| \partial K/\partial\mu |K_i\rangle + \langle K_j| \partial K/\partial\mu |K_i\rangle\langle K_i| \partial K/\partial\lambda |K_i\rangle}{(K_i - K_j)^2} + \frac{\langle K_j| \partial^2 K/\partial\lambda\, \partial\mu |K_i\rangle}{K_i - K_j} \\
&\quad + \frac{\langle K_j| \partial K/\partial\lambda |K_i\rangle}{K_i - K_j} \langle K_i| \frac{\partial |K_i\rangle}{\partial\mu} + \frac{\langle K_j| \partial K/\partial\mu |K_i\rangle}{K_i - K_j} \langle K_i| \frac{\partial |K_i\rangle}{\partial\lambda} \qquad (i \neq j).
\end{aligned}$$

[3] W. B. Brown, Proc. Cambridge Phil. Soc. 54, 251 (1958).

Proof. The equation (from [3])

$$\langle K_j| \frac{\partial |K_i\rangle}{\partial \mu} = \frac{\langle K_j| \partial K/\partial \mu |K_i\rangle}{K_i - K_j}$$

is differentiated with respect to λ to

$$\langle K_j| \frac{\partial^2 |K_i\rangle}{\partial\lambda\,\partial\mu} + \frac{\partial \langle K_j|}{\partial\lambda}\frac{\partial |K_i\rangle}{\partial\mu} = \frac{\partial \langle K_j|}{\partial\lambda}\frac{\partial K/\partial\mu |K_i\rangle}{K_i - K_j} + \frac{\langle K_j| \partial K/\partial\mu}{K_i - K_j}\frac{\partial |K_i\rangle}{\partial\lambda} + \frac{\langle K_j| \partial^2 K/\partial\lambda\,\partial\mu |K_i\rangle}{K_i - K_j} + \frac{\langle K_j| \partial K/\partial\mu |K_i\rangle}{(K_i - K_j)^2}\left(\frac{\partial K_j}{\partial\lambda} - \frac{\partial K_i}{\partial\lambda}\right).$$

The right-hand last term is subjected to the replacement [1]. The left-hand second term and the right-hand first and second terms are, respectively, modified to

$$\frac{\partial \langle K_j|}{\partial\lambda}\frac{\partial |K_i\rangle}{\partial\mu} = \sum_{k(\neq i,j)} \frac{\partial \langle K_j|}{\partial\lambda}|K_k\rangle\langle K_k|\frac{\partial |K_i\rangle}{\partial\mu} + \frac{\partial \langle K_j|}{\partial\lambda}|K_i\rangle\langle K_i|\frac{\partial |K_i\rangle}{\partial\mu} + \frac{\partial \langle K_j|}{\partial\lambda}|K_j\rangle\langle K_j|\frac{\partial |K_i\rangle}{\partial\mu},$$

$$\frac{\partial \langle K_j|}{\partial\lambda}\frac{\partial K/\partial\mu |K_i\rangle}{K_i - K_j} = \sum_{k(\neq j)} \frac{\partial \langle K_j|}{\partial\lambda}|K_k\rangle \times \frac{\langle K_k| \partial K/\partial\mu |K_i\rangle}{K_i - K_j} + \frac{\partial \langle K_j|}{\partial\lambda}|K_j\rangle\frac{\langle K_j| \partial K/\partial\mu |K_i\rangle}{K_i - K_j},$$

$$\frac{\langle K_j| \partial K/\partial\mu}{K_i - K_j}\frac{\partial |K_i\rangle}{\partial\lambda} = \sum_{k(\neq i)} \frac{\langle K_j| \partial K/\partial\mu |K_k\rangle}{K_i - K_j}\langle K_k| \times \frac{\partial |K_i\rangle}{\partial\lambda} + \frac{\langle K_j| \partial K/\partial\mu |K_i\rangle}{K_i - K_j}\langle K_i|\frac{\partial |K_i\rangle}{\partial\lambda}.$$

Subsequently, these right-hand sides are subjected to the replacement [3].

Remark. In this proof, [3], not [5], has been used; the use of [5] is hazardous in the present instance, since [5] is connected with a phase-factor requirement. The formula [21] is valid for any phase factors. In the case of a single parameter, we saw the presence of a phase-factor function satisfying

$$\langle K_i| (\partial |K_i\rangle/\partial\lambda) \equiv 0.$$

If, in the case of two parameters also, we could choose a phase-factor function of $|K_i\rangle$ such that simultaneously

$$\langle K_i| (\partial |K_i\rangle/\partial\lambda) \equiv 0, \quad \langle K_i| (\partial |K_i\rangle/\partial\mu) \equiv 0, \qquad \text{(a)}$$

then in the formula [21] we could always make absent the right-hand last two terms.

Let us examine this possibility. We set up another vector

$$|K_i\rangle^* = (\exp i\theta_i)\,|K_i\rangle \quad (\theta_i\text{: a real function of } \lambda,\ \mu),$$

which differs from $|K_i\rangle$ only in phase factor, and we put

$$\langle K_i| (\partial |K_i\rangle/\partial\lambda) = g_i, \qquad \langle K_i| (\partial |K_i\rangle/\partial\mu) = h_i,$$

$$^*\langle K_i| (\partial |K_i\rangle^*/\partial\lambda) = g_i^*, \quad ^*\langle K_i| (\partial |K_i\rangle^*/\partial\mu) = h_i^*,$$

which are, we know, purely imaginary unless zero. Between g_i and g_i^*, and between h_i and h_i^*, there are, obviously present, the relations

$$g_i^* = g_i + i\,\partial\theta_i/\partial\lambda, \qquad h_i^* = h_i + i\,\partial\theta_i/\partial\mu.$$

Hence, the equation

$$\partial h_i^*/\partial\lambda - \partial g_i^*/\partial\mu = \partial h_i/\partial\lambda - \partial g_i/\partial\mu$$

must hold. This shows that the function of λ, μ

$$\frac{\partial h_i}{\partial\lambda} - \frac{\partial g_i}{\partial\mu} \equiv \frac{\partial \langle K_i|}{\partial\lambda}\frac{\partial |K_i\rangle}{\partial\mu} - \frac{\partial \langle K_i|}{\partial\mu}\frac{\partial |K_i\rangle}{\partial\lambda}$$

is independent of a particular phase-factor function chosen. Now, let us suppose the possibility of realizing (a) by choosing an appropriate phase-factor function of $|K_i\rangle$, or the presence of a phase-factor function θ_i satisfying

$$g_i + i\,\partial\theta_i/\partial\lambda \equiv 0, \qquad h_i + i\,\partial\theta_i/\partial\mu \equiv 0, \qquad \text{(b)}$$

for given g_i and h_i. Then we see that there is a necessary condition,

$$\partial h_i/\partial\lambda - \partial g_i/\partial\mu \equiv 0, \qquad \text{(c)}$$

or

$$\frac{\partial \langle K_i|}{\partial\lambda}\frac{\partial |K_i\rangle}{\partial\mu} - \frac{\partial \langle K_i|}{\partial\mu}\frac{\partial |K_i\rangle}{\partial\lambda} \equiv 0. \qquad \text{(d)}$$

This condition is, obviously, also sufficient; if it is satisfied, the two equations of (b) are simultaneously satisfied, for example, by

$$\theta_i(\lambda, \mu) = i\int_0^\lambda g_i(\lambda, \mu)\,d\lambda + i\int_0^\mu h_i(0, \mu)\,d\mu.$$

Condition (d) can also be replaced by

$$\frac{\partial \langle K_i|}{\partial\lambda}\frac{\partial |K_i\rangle}{\partial\mu} \equiv \text{real}. \qquad \text{(e)}$$

If the wavefunction ψ_i is used in place of $|K_i\rangle$, (d) and (e) become, respectively,

$$\int\left(\frac{\partial\bar{\psi}_i}{\partial\lambda}\frac{\partial\psi_i}{\partial\mu} - \frac{\partial\bar{\psi}_i}{\partial\mu}\frac{\partial\psi_i}{\partial\lambda}\right)d\tau \equiv 0, \qquad \text{(d')}$$

$$\int\frac{\partial\bar{\psi}_i}{\partial\lambda}\frac{\partial\psi_i}{\partial\mu}\,d\tau \equiv \text{real}. \qquad \text{(e')}$$

It depends on cases whether the above condition is satisfied or not. It may readily be found that if ψ_i is real, (d') or (e') holds.

Let us consider a one-dimensional problem. K is expressed in terms of the canonical coordinate x and momentum p as

$$K = ap^2 + bp + U(x),$$

where a, b are real numbers and include λ, μ, but do not depend on x; $U(x)$ is a real function of x and also may depend on λ, μ. If $\psi(x)$ is the eigenfunction belonging to an eigenvalue K' of K, then the differential equation

$$\left\{\frac{d^2}{dx^2} + \frac{ib}{\hbar a}\frac{d}{dx} - \frac{1}{\hbar^2 a}(U - K')\right\}\psi = 0$$

holds, where p has been replaced by $-i\hbar(d/dx)$. This equation has solutions of the form

$$\psi(x) = \varphi(x)e^{-i\beta x}, \qquad \beta = b/2\hbar a,$$

where $\varphi(x)$ is a *real* function of x satisfying the differential equation

$$\left\{\frac{d^2}{dx^2} - \frac{1}{\hbar^2 a}\left(U - K' - \frac{b^2}{4a}\right)\right\}\varphi = 0,$$

and is normalized:

$$\int_{-\infty}^{\infty} \varphi^2\, dx = 1.$$

$\varphi(x)$ may depend on λ and μ. For this $\psi(x)$, the left-hand side of (d′) becomes

$$i\frac{\partial\beta}{\partial\lambda}\frac{\partial}{\partial\mu}\int_{-\infty}^{\infty} x\varphi^2\, dx - i\frac{\partial\beta}{\partial\mu}\frac{\partial}{\partial\lambda}\int_{-\infty}^{\infty} x\varphi^2\, dx.$$

We here assume that $U(x)$ is an even function in some domain of λ, μ. Then, $\varphi(x)$ is obviously an even or odd function in this domain. Hence, the above integral is identically zero and condition (d′) is satisfied. Thus, we see that if $U(x)$ is an even function in some domain of λ, μ, then (d′) is satisfied. We next consider the case that $U(x)$ is not a complete even function but an asymmetric one. Let us also take $\mu = \beta$. Then, $\varphi(x)$ may be neither even nor odd, and the left-hand side of (d′) which becomes

$$-i\frac{\partial}{\partial\lambda}\int_{-\infty}^{\infty} x\varphi^2\, dx$$

may be a function of λ, β not identically equal to zero. Thus, in this case, condition (d′) may not be satisfied.

[22] The function of λ, μ

$$\frac{\partial\langle K_i|}{\partial\lambda}\frac{\partial|K_i\rangle}{\partial\mu} - \frac{\partial\langle K_i|}{\partial\mu}\frac{\partial|K_i\rangle}{\partial\lambda}$$

is independent of a particular phase factor of $|K_i\rangle$ chosen. The identical equality of this to zero is the necessary and sufficient condition for the presence of some vector $|K_i\rangle^*$ that differs from $|K_i\rangle$ only in phase factor and satisfies

$${}^*\langle K_i|\frac{\partial|K_i\rangle^*}{\partial\lambda} \equiv 0, \qquad {}^*\langle K_i|\frac{\partial|K_i\rangle^*}{\partial\mu} \equiv 0$$

simultaneously. (This theorem is a result of the consideration in the "remark" above.)

[23] If the phase factor of $|K_i\rangle$ is chosen appropriately, then

$$\langle K_i|\frac{\partial^2|K_i\rangle}{\partial\lambda\,\partial\mu} = -\sum_{j(\neq i)}\frac{\langle K_i|\,\partial K/\partial\lambda\,|K_j\rangle\langle K_j|\,\partial K/\partial\mu\,|K_i\rangle}{(K_i - K_j)^2},$$

or

$$\langle K_i|\frac{\partial^2|K_i\rangle}{\partial\lambda\,\partial\mu} = -\sum_{j(\neq i)}\frac{\langle K_i|\,\partial K/\partial\mu\,|K_j\rangle\langle K_j|\,\partial K/\partial\lambda\,|K_i\rangle}{(K_i - K_j)^2}.$$

(Which equation holds is dependent on phase factors chosen.)

Proof. Let the phase factor of $|K_i\rangle$ be chosen such that

$$\langle K_i|\,(\partial\,|K_i\rangle/\partial\mu) \equiv 0.$$

This equation is differentiated with respect to λ and modified to

$$\langle K_i|\frac{\partial^2|K_i\rangle}{\partial\lambda\,\partial\mu} = -\sum_{j(\neq i)}\frac{\partial\langle K_i|}{\partial\lambda}|K_j\rangle\langle K_j|\frac{\partial|K_i\rangle}{\partial\mu}.$$

Subsequently, the right-hand side is subjected to the replacement [3]. Thus results the former of the relevant two formulas. If the phase factor is chosen such that

$$\langle K_i|\,(\partial\,|K_i\rangle/\partial\lambda) \equiv 0,$$

the latter formula is derived.

Remark. It is noted that the two right-hand expressions in [23] are complex-conjugate.

[24] That

$$\frac{\partial\langle K_i|}{\partial\lambda}\frac{\partial|K_i\rangle}{\partial\mu}$$

is real is necessary and sufficient for that

$$\sum_{j(\neq i)}\frac{\langle K_i|\,\partial K/\partial\lambda\,|K_j\rangle\langle K_j|\,\partial K/\partial\mu\,|K_i\rangle}{(K_i - K_j)^2}$$

is real. (Whether $(\partial\,\langle K_i|/\partial\lambda)(\partial\,|K_i\rangle/\partial\mu)$ is real or

not is independent of a particular phase factor of $|K_i\rangle$ chosen.)

Proof.

$$\frac{\partial \langle K_i|}{\partial\lambda}\frac{\partial |K_i\rangle}{\partial\mu} - \frac{\partial \langle K_i|}{\partial\mu}\frac{\partial |K_i\rangle}{\partial\lambda}$$

$$= \sum_{j(\neq i)} \frac{\partial \langle K_i|}{\partial\lambda}|K_j\rangle\langle K_j|\frac{\partial |K_i\rangle}{\partial\mu} - \sum_{j(\neq i)} \frac{\partial \langle K_i|}{\partial\mu}|K_j\rangle\langle K_j|\frac{\partial |K_i\rangle}{\partial\lambda}$$

$$+ \left(\frac{\partial \langle K_i|}{\partial\lambda}|K_i\rangle + \langle K_i|\frac{\partial |K_i\rangle}{\partial\lambda}\right)\langle K_i|\frac{\partial |K_i\rangle}{\partial\mu} - \langle K_i|\frac{\partial |K_i\rangle}{\partial\lambda}\left(\frac{\partial \langle K_i|}{\partial\mu}|K_i\rangle + \langle K_i|\frac{\partial |K_i\rangle}{\partial\mu}\right)$$

$$= \sum_{j(\neq i)} \frac{\partial \langle K_i|}{\partial\lambda}|K_j\rangle\langle K_j|\frac{\partial |K_i\rangle}{\partial\mu} - \sum_{j(\neq i)} \frac{\partial \langle K_i|}{\partial\mu}|K_j\rangle\langle K_j|\frac{\partial |K_i\rangle}{\partial\lambda}$$

$$= \sum_{j(\neq i)} \frac{\langle K_i|\,\partial K/\partial\lambda\,|K_j\rangle\langle K_j|\,\partial K/\partial\mu\,|K_i\rangle}{(K_i-K_j)^2} - \sum_{j(\neq i)} \frac{\langle K_i|\,\partial K/\partial\mu\,|K_j\rangle\langle K_j|\,\partial K/\partial\lambda\,|K_i\rangle}{(K_i-K_j)^2} \quad \text{(by [3])}.$$

[25]

$$\langle K_i|\frac{\partial^2 f(K)}{\partial\lambda\,\partial\mu}|K_i\rangle = \sum_k \frac{1}{K_i-K_j} \times \left\{\frac{f(K_i)-f(K_k)}{K_i-K_k} - \frac{f(K_k)-f(K_j)}{K_k-K_j}\right\} \times \left(\langle K_i|\frac{\partial K}{\partial\lambda}|K_k\rangle\langle K_k|\frac{\partial K}{\partial\mu}|K_j\rangle + \langle K_i|\frac{\partial K}{\partial\mu}|K_k\rangle\langle K_k|\frac{\partial K}{\partial\lambda}|K_j\rangle\right) + \frac{f(K_i)-f(K_j)}{K_i-K_j}\langle K_i|\frac{\partial^2 K}{\partial\lambda\,\partial\mu}|K_j\rangle.$$

Proof. The equation

$$f(K)K = Kf(K)$$

is differentiated with respect to λ and μ:

$$\frac{\partial^2 f(K)}{\partial\lambda\,\partial\mu}K + \frac{\partial f(K)}{\partial\lambda}\frac{\partial K}{\partial\mu} + \frac{\partial f(K)}{\partial\mu}\frac{\partial K}{\partial\lambda} + f(K)\frac{\partial^2 K}{\partial\lambda\,\partial\mu} = \frac{\partial^2 K}{\partial\lambda\,\partial\mu}f(K) + \frac{\partial K}{\partial\lambda}\frac{\partial f(K)}{\partial\mu} + \frac{\partial K}{\partial\mu}\frac{\partial f(K)}{\partial\lambda} + K\frac{\partial^2 f(K)}{\partial\lambda\,\partial\mu}.$$

Each side is multiplied by $\langle K_i|$ and $|K_i\rangle$ from the left and right, respectively; the subsequent proceeding is similar to that for [12].

Remark. The above formula is valid even when $i = j$. In this case, it can be written

$$\langle K_i|\frac{\partial^2 f(K)}{\partial\lambda\,\partial\mu}|K_i\rangle = \sum_k \frac{1}{K_i-K_k} \times \left\{f'(K_i) - \frac{f(K_i)-f(K_k)}{K_i-K_k}\right\} \times \left(\langle K_i|\frac{\partial K}{\partial\lambda}|K_k\rangle\langle K_k|\frac{\partial K}{\partial\mu}|K_i\rangle + \langle K_i|\frac{\partial K}{\partial\mu}|K_k\rangle\langle K_k|\frac{\partial K}{\partial\lambda}|K_i\rangle\right) + f'(K_i)\langle K_i|\frac{\partial^2 K}{\partial\lambda\,\partial\mu}|K_i\rangle.$$

[26]

$$\frac{\partial^2 A_{ii}}{\partial\lambda\,\partial\mu} = \sum_{j(\neq i)}\sum_{k(\neq i)} \frac{1}{(K_i-K_j)(K_i-K_k)} \times \left\{\left(\frac{\partial K}{\partial\lambda}\right)_{ij}\left(\frac{\partial K}{\partial\mu}\right)_{jk}A_{ki} + \left(\frac{\partial K}{\partial\mu}\right)_{ij}\left(\frac{\partial K}{\partial\lambda}\right)_{jk}A_{ki} + \left(\frac{\partial K}{\partial\mu}\right)_{ij}A_{jk}\left(\frac{\partial K}{\partial\lambda}\right)_{ki} + \left(\frac{\partial K}{\partial\lambda}\right)_{ij}A_{jk}\left(\frac{\partial K}{\partial\mu}\right)_{ki} + A_{ij}\left(\frac{\partial K}{\partial\lambda}\right)_{jk}\left(\frac{\partial K}{\partial\mu}\right)_{ki} + A_{ij}\left(\frac{\partial K}{\partial\mu}\right)_{jk}\left(\frac{\partial K}{\partial\lambda}\right)_{ki}\right\}$$

$$- \sum_{j(\neq i)} \frac{1}{(K_i-K_j)^2}\left\{\left(\frac{\partial K}{\partial\lambda}\right)_{ii}\left(\frac{\partial K}{\partial\mu}\right)_{ij}A_{ji} + \left(\frac{\partial K}{\partial\mu}\right)_{ii}\left(\frac{\partial K}{\partial\lambda}\right)_{ij}A_{ji} + \left(\frac{\partial K}{\partial\mu}\right)_{ii}A_{ij}\left(\frac{\partial K}{\partial\lambda}\right)_{ji} + \left(\frac{\partial K}{\partial\lambda}\right)_{ii}A_{ij}\left(\frac{\partial K}{\partial\mu}\right)_{ji} + A_{ij}\left(\frac{\partial K}{\partial\lambda}\right)_{ij}\left(\frac{\partial K}{\partial\mu}\right)_{ji} + A_{ii}\left(\frac{\partial K}{\partial\mu}\right)_{ij}\left(\frac{\partial K}{\partial\lambda}\right)_{ji}\right\}$$

$$+ \sum_{j(\neq i)} \frac{1}{K_i-K_j}\left\{\left(\frac{\partial^2 K}{\partial\lambda\,\partial\mu}\right)_{ij}A_{ji} + A_{ij}\left(\frac{\partial^2 K}{\partial\lambda\,\partial\mu}\right)_{ji}\right\}$$

$$+ \sum_{j(\neq i)} \frac{1}{K_i-K_j}\left\{\left(\frac{\partial K}{\partial\lambda}\right)_{ij}\left(\frac{\partial A}{\partial\mu}\right)_{ji} + \left(\frac{\partial K}{\partial\mu}\right)_{ij}\left(\frac{\partial A}{\partial\lambda}\right)_{ji} + \left(\frac{\partial A}{\partial\lambda}\right)_{ij}\left(\frac{\partial K}{\partial\mu}\right)_{ji} + \left(\frac{\partial A}{\partial\mu}\right)_{ij}\left(\frac{\partial K}{\partial\lambda}\right)_{ji}\right\} + \left(\frac{\partial^2 A}{\partial\lambda\,\partial\mu}\right)_{ii},$$

where A_{ij}, $(\partial K/\partial\lambda)_{ij}$, etc., are $\langle K_i|\,A\,|K_j\rangle$, $\langle K_i|\,\partial K/\partial\lambda\,|K_j\rangle$, etc., abbreviated.

Proof. The formula [15] is differentiated with respect to μ, and then [1] and [5] are used for the forms $\partial K_i/\partial\mu$, $\partial\,|K_i\rangle/\partial\mu$, and $\partial\,\langle K_i\,|/\partial\mu$. The relevant formula thus obtained is valid for any phase factors.

[27]

$$\mathrm{Tr}\,\varphi(K)\,\frac{\partial f(K)}{\partial\lambda}\,\frac{\partial g(K)}{\partial\mu} = \mathrm{Tr}\,\varphi(K)\,\frac{\partial g(K)}{\partial\lambda}\,\frac{\partial f(K)}{\partial\mu},$$

where φ, f, and g are arbitrary functions.

Proof.

$$\begin{aligned}
&\mathrm{Tr}\,\varphi(K)(\partial f(K)/\partial\lambda)(\partial g(K)/\partial\mu)\\
&\quad=\sum_i\sum_j \varphi(K_i)\,\langle K_i|\,\partial f(K)/\partial\lambda\,|K_j\rangle\\
&\qquad\times\langle K_j|\,\partial g(K)/\partial\mu\,|K_i\rangle\\
&\quad=\sum_i\sum_j \varphi(K_i)\{f(K_i)-f(K_j)\}\\
&\qquad\times\{g(K_i)-g(K_j)\}(K_i-K_j)^{-2}\\
&\qquad\times\langle K_i|\,\partial K/\partial\lambda\,|K_j\rangle\langle K_j|\,\partial K/\partial\mu\,|K_i\rangle\quad\text{(by [2]).}
\end{aligned}$$

The last expression is unaffected by the interchange between f and g. Hence, the relevant formula holds.

This way of proof is analogous to the first way of proof in [19]. In the present case, the method analogous to the second one in [19] is not useful.

Remark. According to [17], we have

$$\mathrm{Tr}\,\varphi(K)\,\partial f(K)/\partial\lambda = \mathrm{Tr}\,\varphi(K)f'(K)\,\partial K/\partial\lambda,$$

and

$$\mathrm{Tr}\,\varphi(K)\,\partial f(K)/\partial\mu = \mathrm{Tr}\,\varphi(K)f'(K)\,\partial K/\partial\mu.$$

Differentiating the former equation with respect to μ and the latter with respect to λ, we get

$$\begin{aligned}
&\mathrm{Tr}\,\varphi(K)\,\frac{\partial^2 f(K)}{\partial\lambda\,\partial\mu}\\
&\quad=\mathrm{Tr}\,\varphi(K)\,\frac{\partial f'(K)}{\partial\mu}\,\frac{\partial K}{\partial\lambda}+\mathrm{Tr}\,\varphi(K)f'(K)\,\frac{\partial^2 K}{\partial\lambda\,\partial\mu}\\
&\qquad+\mathrm{Tr}\,\frac{\partial\varphi(K)}{\partial\mu}\,f'(K)\,\frac{\partial K}{\partial\lambda}-\mathrm{Tr}\,\frac{\partial\varphi(K)}{\partial\mu}\,\frac{\partial f(K)}{\partial\lambda},
\end{aligned}$$

and

$$\begin{aligned}
&\mathrm{Tr}\,\varphi(K)\,\frac{\partial^2 f(K)}{\partial\lambda\,\partial\mu}\\
&\quad=\mathrm{Tr}\,\varphi(K)\,\frac{\partial f'(K)}{\partial\lambda}\,\frac{\partial K}{\partial\mu}+\mathrm{Tr}\,\varphi(K)f'(K)\,\frac{\partial^2 K}{\partial\lambda\,\partial\mu}\\
&\qquad+\mathrm{Tr}\,\frac{\partial\varphi(K)}{\partial\lambda}\,f'(K)\,\frac{\partial K}{\partial\mu}-\mathrm{Tr}\,\frac{\partial\varphi(K)}{\partial\lambda}\,\frac{\partial f(K)}{\partial\mu},
\end{aligned}$$

respectively. We notice that the right-hand sides of both these equations have different forms. They must, of course, be equal. The fourth terms are readily found to be equal to one another from [27]. Therefore, the equation

$$\begin{aligned}
&\mathrm{Tr}\,\varphi(K)(\partial f'(K)/\partial\mu)(\partial K/\partial\lambda)\\
&\qquad+\mathrm{Tr}\,(\partial\varphi(K)/\partial\mu)f'(K)(\partial K/\partial\lambda)\\
&\quad=\mathrm{Tr}\,\varphi(K)(\partial f'(K)/\partial\lambda)(\partial K/\partial\mu)\\
&\qquad+\mathrm{Tr}\,(\partial\varphi(K)/\partial\lambda)f'(K)(\partial K/\partial\mu)
\end{aligned}$$

must hold. This validity are also found from [27], since the left- and right-hand sides are equal to

$$\mathrm{Tr}\,(\partial\varphi(K)f'(K)/\partial\mu)(\partial K/\partial\lambda),$$

and

$$\mathrm{Tr}\,(\partial\varphi(K)f'(K)/\partial\lambda)(\partial K/\partial\mu),$$

respectively.

[28] If A and B are Hermitian linear operators dependent on λ alone and μ alone, respectively, and if $A = B\ (= C)$ at $\lambda = \lambda_1$, $\mu = \mu_1$, then

$$\mathrm{Tr}\,\varphi(C)\,\frac{\partial f(A)}{\partial\lambda}\,\frac{\partial g(B)}{\partial\mu} = \mathrm{Tr}\,\varphi(C)\,\frac{\partial g(A)}{\partial\lambda}\,\frac{\partial f(B)}{\partial\mu}$$

at $\lambda = \lambda_1$, $\mu = \mu_1$, where φ, f, and g are arbitrary functions.

Proof. We put

$$K = A + B - C.$$

According to [27], we have

$$\begin{aligned}
&\lim_{\lambda\to\lambda_1}\lim_{\mu\to\mu_1}\mathrm{Tr}\,\varphi(K)(\partial f(K)/\partial\lambda)(\partial g(K)/\partial\mu)\\
&\quad=\lim_{\lambda\to\lambda_1}\lim_{\mu\to\mu_1}\mathrm{Tr}\,\varphi(K)(\partial g(K)/\partial\lambda)(\partial f(K)/\partial\mu).\qquad\text{(a)}
\end{aligned}$$

Since

$$\begin{aligned}
&\lim_{\lambda\to\lambda_1}\lim_{\mu\to\mu_1}\varphi(K)=\varphi(C),\\
&\lim_{\mu\to\mu_1}\partial f(K)/\partial\lambda=\partial f(A)/\partial\lambda,\\
&\lim_{\mu\to\mu_1}\partial g(K)/\partial\lambda=\partial g(A)/\partial\lambda,\\
&\lim_{\lambda\to\lambda_1}\partial f(K)/\partial\mu=\partial f(B)/\partial\mu,
\end{aligned}$$

and

$$\lim_{\lambda\to\lambda_1}\partial g(K)/\partial\mu=\partial g(B)/\partial\mu,$$

the left- and right-hand sides of (a) become

$$\lim_{\lambda\to\lambda_1}\lim_{\mu\to\mu_1}\mathrm{Tr}\,\varphi(C)(\partial f(A)/\partial\lambda)(\partial g(B)/\partial\mu),$$

and

$$\lim_{\lambda\to\lambda_1}\lim_{\mu\to\mu_1}\mathrm{Tr}\,\varphi(C)(\partial g(A)/\partial\lambda)(\partial f(B)/\partial\mu),$$

respectively.

ACKNOWLEDGMENTS

The author thanks his colleagues in the Laboratory for their reading and criticizing the manuscript.

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 4, NUMBER 6 JUNE 1963

Galilei Group and Nonrelativistic Quantum Mechanics

JEAN-MARC LEVY-LEBLOND
Laboratoire de Physique Théorique et Hautes Energies, Orsay, France
(Received 16 January 1963)

This paper is devoted to the study of the Galilei group and its representations. The Galilei group presents a certain number of essential differences with respect to the Poincaré group. As Bargmann showed, its physical representations, here explicitly constructed, are not true representations but only up-to-a-factor ones. Consequently, in nonrelativistic quantum mechanics, the mass has a very special role, and in fact, gives rise to a superselection rule which prevents the existence of unstable particles. The internal energy of a nonrelativistic system is known to be an arbitrary parameter; this is shown to come also from Galilean invariance, because of a nontrivial concept of equivalence between physical representations. On the contrary, the behavior of an elementary system with respect to rotations, is very similar to the relativistic case. We show here, in particular, how the number of polarization states reduces to two for the zero-mass case (though in fact there are no physical zero-mass systems in nonrelativistic mechanics). Finally, we study the two-particle system, where the orbital angular momenta quite naturally introduce themselves through the decomposition of the tensor product of two physical representations.

INTRODUCTION

SINCE the work by Wigner[1] came out, the Poincaré group (inhomogeneous Lorentz group) and its unitary representations have become well known. In particular, each relativistic wavefunction corresponds to some unitary representation of the Poincaré group, and in a certain sense, one usually says that an elementary particle is associated to a unitary irreducible representation of the group. Within such a definition, an elementary particle is characterized by its mass and spin.

It was much later that such a work was undertaken for the Galilei group, the invariance group of nonrelativistic mechanics. The Galilei group has, in fact, a rather more intricate structure than has the Poincaré group and this has important repercussions in the study of the group representations. Indeed, in quantum mechanics, we deal with the unitary projective (i.e. up-to-a-factor) representations of the group concerned. But in most of the physically interesting cases, as Bargmann showed,[2] the study of unitary projective representations of the group can be reduced to the study of true unitary representations of its universal covering group. Such is the case of the rotation, Lorentz, and Poincaré groups.

On the other hand, as Bargmann also showed,[2] the Galilei group owns a (one-dimensional) infinity of projective representations classes, nonequivalent to true representations, and, what is more troublesome, the physically meaningful representations are precisely these nontrivial projective representations. Inönu and Wigner[3] have indeed shown that under no condition can the basis functions of the Galilei-group true representations be interpreted as wavefunctions of physical particles. With these functions, one can construct neither localized states, nor even states with definite velocity. Conversely, Hamermesh[4] studying the infinitesimal group operations, has shown that one can construct a position operator only in the case of nontrivial projective representations. One easily sees that the solutions of the Schrödinger equation for a free particle transform precisely according to such representations.

In the following pages, these unitary, irreducible, nontrivial projective representations of the Galilei group will be called, for short, "physical representations."

In the first section, we recall some generalities about the Galilei group and its structure. In the second one, we explicitly construct the physical representations of the Galilei group, with the help of the "little group" technique.[5] The third section is devoted to a physical discussion of these representations where we exhibit their connection with the free-particle Schrödinger equation, and obtain a group-theoretical characterization of a nonrelativistic elementary system by its spin, mass and internal energy—this last parameter in fact revealing itself to be arbitrary. We next study the zero-mass case. In this fourth section, we rediscover some of the true representations already studied by Inönu

[1] E. P. Wigner, Ann. Math. 40, 149 (1939).
[2] V. Bargmann, Ann. Math. 59, 1 (1954).
[3] E. Inönu and E. P. Wigner, Nuovo Cimento 9, 705 (1952).
[4] M. Hamermesh, Ann. Phys. 9, 518 (1960).
[5] See, for instance, M. Hamermesh, *Group Theory* (Addison-Wesley Publishing Company, Inc., Reading, Massachusetts, 1962), Sec. 12-7.

and Wigner.[3] One can then give to these representations some vague physical meaning, and we observe, as in the zero-mass case of the Poincaré group, the uncoupling of the different helicity states. The Lie algebra of the Galilei group and its representations are investigated in the fifth section where we emphasize the difference between the parts played by the mass in relativistic and nonrelativistic quantum mechanics (Bargmann's superselection rule). In the sixth and last section, we decompose the tensor product of two physical representations of the Galilei group into a direct sum (integral) of such physical (irreducible) representations.

I. THE GALILEI GROUP

The proper Galilei group to which we restrict ourselves (we exclude the inversions) contains the translations in space and time, the rotations and the pure Galilei transformations, i.e., transitions to a uniformly moving coordinate system. Let us note the general element of the group by

$$G = (b, \mathbf{a}, \mathbf{v}, R), \tag{I.1}$$

where b is a time translation, $\mathbf{a}$ a space translation, $\mathbf{v}$ a pure Galilei transformation, and R a rotation. The group acts on the coordinates $(\mathbf{x}, t)$ of an event in space–time according to

$$\mathbf{x}' = R\mathbf{x} + \mathbf{v}t + \mathbf{a}, \qquad t' = t + b. \tag{I.2}$$

We thus get the multiplication law for the group:

$$G'G = (b', \mathbf{a}', \mathbf{v}', R')(b, \mathbf{a}, \mathbf{v}, R)$$

$$= (b' + b, \mathbf{a}' + R'\mathbf{a} + b\mathbf{v}', \mathbf{v}' + R'\mathbf{v}, R'R). \tag{I.3}$$

The identity for the group is

$$\mathbf{1} = (0, 0, 0, 1), \tag{I.4}$$

and the inverse element of $G = (b, \mathbf{a}, \mathbf{v}, R)$ is given by

$$G^{-1} = (-b, -R^{-1}(\mathbf{a} - b\mathbf{v}), -R^{-1}\mathbf{v}, R^{-1}). \tag{I.5}$$

One notices at once the complexity of the Galilei group structure.

The Poincaré group $\mathcal{P}$ admits a maximal abelian invariant subgroup $\mathcal{C}$ (the space and time translations) and the factor group $\mathcal{P}/\mathcal{C}$ is a simple group $\mathcal{L}$, the Lorentz group.

Here the situation is more complicated. The maximal abelian invariant subgroup of the Galilei group $\mathcal{G}$, is a six-parameter group $\mathcal{U}$ (space translations plus pure Galilei transformations). The factor group $\mathcal{G}/\mathcal{U}$ itself admits a one-parameter invariant subgroup $\mathcal{D}$, the time translations, and it is only the factor group $(\mathcal{G}/\mathcal{U})/\mathcal{D}$ which is a simple group $\mathcal{R}$, the rotation group.

In other words, the Poincaré group can be written

$$\mathcal{P} = \mathcal{L} \times \mathcal{C}, \tag{I.6}$$

that is, the semidirect product of a simple group $\mathcal{L}$ by an abelian group $\mathcal{C}$. But the Galilei group is

$$\mathcal{G} = (\mathcal{R} \times \mathcal{D}) \times \mathcal{U}, \tag{I.7}$$

the semidirect product by an abelian group $\mathcal{U}$ of the semidirect product by an abelian group $\mathcal{D}$ of a simple group $\mathcal{R}$.

We will now use the following result, due to Bargmann,[2] which we reproduce without proof:

"The physical representations of the Galilei group are obtained from the projective unitary representations of its universal covering group characterized by the system of factors:

$$\omega(G', G) = \exp\,[i(\tfrac{1}{2}m)(\mathbf{a}'\cdot R'\mathbf{v} - \mathbf{v}'\cdot R\mathbf{a} + b\mathbf{v}'\cdot R'\mathbf{v})], \tag{I.8}$$

where $G = (b, \mathbf{a}, \mathbf{v}, R)$, $G' = (b', \mathbf{a}', \mathbf{v}', R')$ and m is any real number.

This means that to each element G of the universal covering group (which one obtains merely by replacing the rotations R by the elements of the unitary unimodular group), corresponds a unitary operator $U(G)$ such that the multiplication law

$$U(G')U(G) = \omega(G', G)U(G'G) \tag{I.9}$$

holds, where G', G, $\omega(G', G)$ have been defined above.

The ensuing transition from the covering group to the original Galilei group only adds a possible sign ambiguity.

In a more elaborate language, we are looking for true unitary representations of some nontrivial central extension of the Galilei group universal covering group by a one-dimensional abelian group.[6]

Let us note

$$\tilde{G} = (\theta, G), \qquad \theta \text{ real}, \tag{I.10}$$

which are the elements of this extension. We have then the multiplication law

$$\tilde{G}'\tilde{G} = (\theta' + \theta + \xi(G', G), G'G), \tag{I.11}$$

where $\xi(G', G)$, given by

$$\omega(G', G) = \exp\,[i\xi(G', G)], \tag{I.12}$$

is an exponent of the group.[2]

The eleven-parameter group $\tilde{\mathcal{G}}$, with which we

[6] F. Lurcat and L. Michel (unpublished). L. Michel, Lectures at the Istanbul Summer School (1962) (to be published).

deal now, has a structure rather different from the one of the original Galilei group. In fact, the maximal abelian invariant subgroup of $\tilde{\mathcal{G}}$ is $\tilde{\mathcal{C}}$, the space–time translations plus the one-parameter central subgroup. The factor group $\tilde{\mathcal{G}}/\tilde{\mathcal{C}}$ in turn admits a maximal abelian invariant subgroup $\mathcal{U}$, made up of the pure Galilei transformations (3 parameters). Finally, the factor group $(\tilde{\mathcal{G}}/\tilde{\mathcal{C}})/\mathcal{U}$ is a simple group $\mathcal{R}$, the rotation group. We can write

$$\tilde{\mathcal{G}} = (\mathcal{R} \times \mathcal{U}) \times \tilde{\mathcal{C}}, \tag{I.13}$$

where the products are semidirect products. This is the structure we shall be concerned with. Let us finally notice that setting $m = 0$ in (I.8) brings us back to the study of true representations of the Galilei group:

$$m = 0 \Rightarrow \omega(G', G) = 1 \Rightarrow U(G')U(G) = U(G'G). \tag{I.14}$$

The central extension $\tilde{\mathcal{G}}$ becomes a trivial one (direct product).

Unless otherwise specified, we will deal exclusively from now on with the case $m \neq 0$.

II. PHYSICAL REPRESENTIONS OF THE GALILEI GROUP

We now proceed to construct the physical (i.e., irreducible, unitary, nontrivial projective) representations of the Galilei group, making use of its structure as studied above and following the "little group" technique.[5]

Let us suppose we have found some physical representations of the Galilei group. If we restrict ourselves to the abelian subgroup $\mathcal{C}$ of space–time translations, this representation will decompose into a direct integral of unitary irreducible representations of the subgroup $\mathcal{C}$. These representations are well-known, they are designated by a real vector $\mathbf{p}$ and a real number E. We can then choose as a set of basis functions, square-integrable functions $\psi(\mathbf{p}, E, \zeta)$, where ζ is an additional set of variables which may be needed to distinguish the basis functions belonging to the same irreducible representation of the translation group $\mathcal{C}$. We now know the representation of this subgroup:

$$U(b, \mathbf{a}, 0, 1)\psi(\mathbf{p}, E, \zeta) = \exp(-ibE + i\mathbf{a}\cdot\mathbf{p})\psi(\mathbf{p}, E, \zeta). \tag{II.1}$$

Using the multiplication law of the group representation (I.9), a factor system, i.e., a real number m (I.8) having been chosen, we look for the representation of the factor group $\mathcal{G}/\mathcal{C}$. Let us note first the following equalities, inferred from (I.3), (I.8), and (I.9):

$$U(b, \mathbf{a}, \mathbf{v}, R) = \exp(-i(\tfrac{1}{2}m)\mathbf{a}\cdot\mathbf{v}) \times U(b, \mathbf{a}, 0, 1)U(0, 0, \mathbf{v}, R), \tag{II.2}$$

$$U(b, \mathbf{a}, \mathbf{v}, R) = \exp[i(\tfrac{1}{2}m)(\mathbf{a}\cdot\mathbf{v} - b\mathbf{v}\cdot\mathbf{v})] \times U(0, 0, \mathbf{v}, R)U(b, R^{-1}(\mathbf{a} - b\mathbf{v}), 0, 1), \tag{II.2'}$$

whence

$$U(b, \mathbf{a}, 0, 1)U(0, 0, \mathbf{v}, R) = \exp[im(\mathbf{a}\cdot\mathbf{v} - \tfrac{1}{2}b\mathbf{v}\cdot\mathbf{v})] \times U(0, 0, \mathbf{v}, R)U(b, R^{-1}(\mathbf{a} - b\mathbf{v}), 0, 1). \tag{II.3}$$

Letting each member of this equality between operators act upon some basis function $\psi(\mathbf{p}, E, \zeta)$, and taking (II.1) into account, we get

$$U(b, \mathbf{a}, 0, 1)U(0, 0, \mathbf{v}, R)\psi(\mathbf{p}, E, \zeta = \exp[im(\mathbf{a}\cdot\mathbf{v} - \tfrac{1}{2}b\mathbf{v}\cdot\mathbf{v})] \times \exp[-ibE + i(R^{-1}\mathbf{a}\cdot bR^{-1}\mathbf{v})\cdot\mathbf{p}] \times U(0, 0, \mathbf{v}, R)\psi(\mathbf{p}, E, \zeta), \tag{II.4}$$

$$U(b, \mathbf{a}, 0, 1)U(0, 0, \mathbf{v}, R)\psi(\mathbf{p}, E, \zeta) = \exp[-ib(E + \mathbf{v}\cdot R\mathbf{p} + \tfrac{1}{2}m\mathbf{v}^2) + i\mathbf{a}\cdot(R\mathbf{p} + m\mathbf{v})] \times U(0, 0, \mathbf{v}, R)\psi(\mathbf{p}, E, \zeta). \tag{II.5}$$

That is to say, the function $U(0, 0, \mathbf{v}, R)\psi(\mathbf{p}, E, \zeta)$ transforms according to the representation $(\mathbf{p}', E')$ of the space–time translations group, where

$$\begin{aligned} \mathbf{p}' &= R\mathbf{p} + m\mathbf{v}, \\ E' &= E + \mathbf{v}\cdot R\mathbf{p} + \tfrac{1}{2}m\mathbf{v}^2. \end{aligned} \tag{II.6}$$

Thus, if $(\mathbf{p}', E')$ and $(\mathbf{p}, E)$ are connected by the relation

$$E' - (\mathbf{p}'^2/2m) = E - (\mathbf{p}^2/2m), \tag{II.7}$$

it is always possible to find an element $(0, 0, \mathbf{v}, R)$ of the Galilei group (more precisely of the factor group $\mathcal{G}/\mathcal{C}$) such that (II.6) holds.

In other words, if a physical representation of the Galilei group contains an irreducible representation $(\mathbf{p}, E)$ of the translation group, it contains all the representations $(\mathbf{p}', E')$ given by (II.7). Therefore, there is a one-to-one correspondence between the points of the paraboloid:

$$E - (\mathbf{p}^2/2m) = \mathcal{U} = C^{te}, \tag{II.7'}$$

and the irreducible representations of the translation group contained in a physical representation of the Galilei group.

This enables us to construct the Hilbert space $\mathcal{H}$

of the representation as a direct integral of the Hilbert spaces $\mathfrak{H}_{\mathbf{p},E}$ where acts the irreducible representation $(\mathbf{p}, E)$ of $\mathcal{C}$:

$$\mathfrak{H} = \int d\mu(\mathbf{p}, E)\, \mathfrak{H}_{\mathbf{p},E}, \tag{II.8}$$

where $\mu(\mathbf{p}, E)$ is an invariant measure on the paraboloid (II.8). As a matter of fact,

$$d\mu(\mathbf{p}, E) = d\mathbf{p}\, dE\, \delta[E - (\mathbf{p}^2/2m) - \mathcal{V}]. \tag{II.9}$$

The Little Group

We now search for the "little group",[6] i.e., the subgroup $\mathfrak{h}(\mathbf{p}, E)$ of the Galilei group constituted by those elements $(0, 0, \mathbf{v}, R)$ of the factor group $\mathcal{G}/\mathcal{C}$ such that the function $U(0, 0, \mathbf{v}, R)\psi(\mathbf{p}, E, \zeta)$ still belongs to the irreducible representation $(\mathbf{p}, E)$ of $\mathcal{C}$. $\mathfrak{h}(\mathbf{p}, E)$ is what mathematicians call the "stabilisator" of $(\mathbf{p}, E)$. After (II.6), we get for $(0, 0, \mathbf{v}, R)$ the conditions

$$\begin{aligned} \mathbf{p} &= R\mathbf{p} + m\mathbf{v}, \\ E &= E + \mathbf{v}\cdot R\mathbf{p} + \tfrac{1}{2}m\mathbf{v}^2, \end{aligned} \tag{II.10}$$

which can also be written as

$$\begin{aligned} \mathbf{p} &= R\mathbf{p} + m\mathbf{v}, \\ \mathbf{p}^2 &= (R\mathbf{p} + m\mathbf{v})^2. \end{aligned} \tag{II.10'}$$

These two conditions then are not independent, the second being implied by the first one.

Thus an element of the little group $\mathfrak{h}(\mathbf{p}, E)$ is uniquely defined by the choice of a rotation R, since in that case, the condition (II.10') determines $\mathbf{v}$.

This correspondence between the little group $\mathfrak{h}(\mathbf{p}, E)$ and the rotation group $\mathfrak{R}$ is an isomorphism. Indeed,

$$\mathbf{p} = R\mathbf{p} + m\mathbf{v}, \qquad \mathbf{p} = R'\mathbf{p} + m\mathbf{v}'$$

implies

$$\mathbf{p} = R'(R\mathbf{p} + m\mathbf{v}) + m\mathbf{v}' = R'R\mathbf{p} + m(R'\mathbf{v} + \mathbf{v}').$$

The product of $(0, 0, \mathbf{v}', R')$ and $(0, 0, \mathbf{v}, R)$ corresponds to the product of R and R'.

The little group representations are then well known: they are the rotation group representations. The irreducible ones are labeled by an integer of half-integer number s. We denote them by D^s. They are $(2s + 1)$-dimensional.

We show now that choosing a paraboloid (II.7') (which fixes the possible representations of the translation group) and a representation D^s of the little group completely determines a representation of the whole Galilei group.

First we set up a point $(\mathbf{p}_0, E_0)$ on the paraboloid. Then for each point $(\mathbf{p}, E)$ of that same paraboloid, one can select one element

$$V_{pE} = (0, 0, v_{pE}, R_{pE}) \tag{II.11}$$

of the Galilei group such that V_{pE} acting on $(\mathbf{p}_0, E_0)$ according to (II.6) transforms it into $(\mathbf{p}, E)$, which we denote by

$$V_{pE}(\mathbf{p}_0, E_0) = (\mathbf{p}, E). \tag{II.11'}$$

Now let V be any element of the factor group:

$$V = (0, 0, \mathbf{v}, R), \tag{II.12}$$

and let $(\mathbf{p}', E')$ be the result of the action of V $(\mathbf{p}, E)$ according to (II.6):

$$V(\mathbf{p}, E) = (\mathbf{p}', E'). \tag{II.13}$$

This can also be written as

$$VV_{pE}(\mathbf{p}_0, E_0) = V_{p'E'}(\mathbf{p}_0, E_0),$$

or

$$V_{p'E'}^{-1} V V_{pE}(\mathbf{p}_0, E_0) = (\mathbf{p}_0, E_0).$$

Thus,

$$V_{p',E'}^{-1} V V_{pE} = V_0^{(p,E),V} \tag{II.14}$$

is an element of the little group $\mathfrak{h}(\mathbf{p}_0, E_0)$, which, of course, depends on $(\mathbf{p}, E)$ and V.

Conversely, every element V of the factor group $\mathcal{G}/\mathcal{C}$ can be written in the form

$$V = V_{p'E'} V_0^{(p,E),V} V_{pE}^{-1}, \tag{II.15}$$

with the definitions (II.11'), (II.12), and (II.13).

Now, after (II.5), (II.6), and (II.11'), $U(V_{p,E})\cdot\psi(\mathbf{p}_0, E_0)$ is proportional to $\psi(\mathbf{p}, E)$. The simplest choice is then to define $U(V_{p,E})$ by

$$U(V_{pE})\psi(\mathbf{p}_0, E_0, \zeta) = \psi(\mathbf{p}, E, \zeta). \tag{II.16}$$

Finally, we choose a representation D^s of the little group. The variable ζ is merely an index, running from $-s$ to $+s$, on which act the $(2s + 1)$-dimensional matrices of the representation D^s:

$$\begin{aligned} V_0 \in \mathfrak{h}(\mathbf{p}_0, E_0) \Rightarrow U(V_0)\psi(\mathbf{p}_0, E_0, \zeta) \\ = \sum_{\xi=-s}^{+s} \psi(\mathbf{p}_0, E_0, \xi)[D^s(V_0)]_{\xi\zeta}. \end{aligned} \tag{II.17}$$

Therefore, for any $V = (0, 0, \mathbf{v}, R)$ and any $\psi(p, E, \zeta)$, we have, after (II.15), (II.16), and (II.17),

$$\begin{aligned} &U(V)\psi(\mathbf{p}, E, \zeta) \\ &= U(V_{p'E'})U(V_0^{(pE),V})U^{-1}(V_{pE})\psi(\mathbf{p}, E, \zeta) \\ &= U(V_{p'E'})U(V_0^{(pE),V})\psi(\mathbf{p}_0, E_0, \zeta) \\ &= \sum_{\xi} U(V_{p'E'})\psi(\mathbf{p}_0, E_0, \xi)[D^s(V_0^{(pE),V})]_{\xi\zeta} \end{aligned}$$

$$U(V)\psi(\mathbf{p}, E, \zeta) = \sum_{\xi} \psi(\mathbf{p}', E', \xi)[D^s(V_0^{(\mathbf{p},E),V})]_{\xi\zeta}. \quad \text{(II.18)}$$

Since we also know the operators which represent the translation, the decomposition (II.2) of any element of the whole group, gives us at once the complete solution to our problem:

$$U(b, \mathbf{a}, \mathbf{v}, R)\psi(\mathbf{p}, E, \zeta) = \exp[-i(\tfrac{1}{2}m)\mathbf{a}\cdot\mathbf{v} + i\mathbf{a}\cdot\mathbf{p}' - ibE'] \times \sum_{\xi} \psi(\mathbf{p}', E', \xi)[D^s(V_0^{(\mathbf{p},E)(\mathbf{v},R)})]_{\xi\zeta}, \quad \text{(II.19)}$$

where

$$\mathbf{p}' = R\mathbf{p} + m\mathbf{v},$$
$$E' = E + \mathbf{v}\cdot R\mathbf{p} + \tfrac{1}{2}m\mathbf{v}^2,$$

and

$$V_0^{(\mathbf{p}E),(\mathbf{v},R)} = V_{\mathbf{p}'E'}^{-1}(0, 0, \mathbf{v}, R)V_{\mathbf{p}E}.$$

These representations are clearly irreducible. They are unitary with the scalar product:

$$(\phi, \psi) = \int d\mathbf{p}\, dE\, \delta\left[E - \frac{\mathbf{p}^2}{2m} - \mathcal{U}\right] \times \sum_{\alpha} \bar{\Phi}(\mathbf{p}, E, \alpha)\psi(\mathbf{p}, E, \alpha). \quad \text{(II.20)}$$

Finally, one can verify directly that they are indeed projective representations of the universal covering group of the Galilei group with the system of factors (I.8) which we started from.

The transition to the Galilei group itself merely introduces the usual sign ambiguity in the case of half-integers. We have thus obtained the following result:

> The physical representations of the Galilei group are characterized by two real numbers m and $\mathcal{U}$ and an integer or half-integer number s. We designate them by $[m \mid \mathcal{U}, s]$ and they are given explicitly by (II.19).

III. PHYSICAL DISCUSSION

The interpretation of the preceding results is straightforward. Eq. (II.20) defines the functional space of the representation as the space of square-integrable functions on the paraboloid:

$$E - (\mathbf{p}^2/2m) = \mathcal{U}. \quad \text{(II.7')}$$

This, and the rotation properties of these basis functions, impel us to establish a one-to-one correspondence between a (free) particle of mass m, internal energy $\mathcal{U}$, spin s, and the physical representation $[m \mid \mathcal{U}, s]$ of the Galilei group. We will now study this correspondence.

Schrödinger Equation and Galilei Group

For the time being, we shall disregard the spin and consider the Schrödinger equation for a spinless particle:

$$i(\partial\psi/\partial t) + (1/2m)\Delta\psi = \mathcal{U}\psi. \quad \text{(III.1)}$$

We are only concerned with the free-particle case, so that in (III.1), $\mathcal{U}$ is a constant. We wish to study the invariance properties of the Schrödinger equation with respect to the Galilei group transformation. (Let us note that most of quantum mechanics textbooks thoroughly investigate the Lorentz invariance of the Dirac equation but completely overlook the Galilean invariance of the Schrödinger equation.)

We follow the *passive* point of view. That is, we look at the same state described by (III.1), in a transformed frame of reference defined by

$$\mathbf{x}' = R\mathbf{x} + \mathbf{v}t + \mathbf{a}, \qquad t' = t + b. \quad \text{(I.2)}$$

$G = (b, \mathbf{a}, \mathbf{v}, R)$ is the Galilean transformation we consider. In the new frame of reference, the state must be described by some wavefunction ψ'. The physical predictions we get from the two descriptions will be identical if and only if the transformed wavefunction at any point differs from the original wavefunction at the transformed point by at most a phase factor (the density of particles being a scalar):

$$\psi'(\mathbf{x}, t) = e^{-if(\mathbf{x}',t')}\,\psi(\mathbf{x}', t'), \quad \text{(III.2)}$$

where $(\mathbf{x}', t')$ depend on $(\mathbf{x}, t)$ according to (I.2). Obviously, the new wavefunction has to satisfy the Schrödinger equation

$$i(\partial\psi'/\partial t) + (1/2m)\Delta\psi' = \mathcal{U}\psi'. \quad \text{(III.3)}$$

We may now determine the unknown function f. Using (I.2), we find

$$\partial/\partial t = \partial/\partial t' + \mathbf{v}\cdot\nabla', \qquad \nabla = R\nabla'. \quad \text{(III.4)}$$

The Schrödinger equation (III.3) for the new wavefunction can be rewritten as an equation in f and ψ:

$$[i(\partial/\partial t') + i\mathbf{v}\cdot\nabla' + (1/2m)\Delta' - \mathcal{U}] \times e^{-if(x',t')}\psi(x', t') = 0. \quad \text{(III.5)}$$

Dropping the primes and expanding,

$$\left(\frac{\partial f}{\partial t} + \mathbf{v}\cdot\nabla f - \frac{1}{2m}\Delta f\right)e^{-if}\psi - i\left(\mathbf{v} - \frac{1}{m}\nabla f\right)e^{-if}$$
$$\cdot\nabla\psi + e^{-if}\left(i\frac{\partial}{\partial t} + \frac{1}{2m}\Delta - \mathcal{V}\right)\psi = 0. \qquad \text{(III.6)}$$

$\psi(\mathbf{x}, t)$ satisfies the Schrödinger equation (III.1), so that we are left with the two conditions

$$\begin{aligned} \nabla f - m\mathbf{v} &= 0, \\ (\partial f/\partial t) - (1/2m)\Delta f + m\mathbf{v}^2 &= 0. \end{aligned} \qquad \text{(III.7)}$$

We easily integrate these equations, and get

$$f(\mathbf{x}, t) = m\mathbf{v}\cdot\mathbf{x} - \tfrac{1}{2}m\mathbf{v}^2 t + C, \qquad \text{(III.8)}$$

where C is a constant.

We see that, unlike the relativistic case, the phase factor cannot be eliminated. The transformation properties of the Schrödinger wavefunctions are then

$$\begin{aligned} \psi'(\mathbf{x}, t) &= \exp\,[-im\mathbf{v}\cdot\mathbf{x}' \\ &\quad + \tfrac{1}{2}im\mathbf{v}^2 t' - iC]\psi(\mathbf{x}', t'), \\ \mathbf{x}' &= R\mathbf{x} + \mathbf{v}t + \mathbf{a}, \\ t' &= t + b. \end{aligned} \qquad \text{(III.9)}$$

In the momentum space, we deal with wavefunctions:

$$\phi(\mathbf{p}, E) = \int e^{-i\mathbf{p}\cdot\mathbf{x} + iEt}\psi(\mathbf{x}, t)\, d\mathbf{x}\, dt. \qquad \text{(III.10)}$$

The Schrödinger equation reads

$$[E - (\mathbf{p}^2/2m)]\phi(\mathbf{p}, E) = \mathcal{V}\phi(\mathbf{p}, E), \qquad \text{(III.11)}$$

so that the support of ϕ is

$$E - (\mathbf{p}^2/2m) = \mathcal{V}, \qquad \text{(III.11')}$$

and the wavefunctions have to be square-integrable:

$$\int \left|\phi\left(\mathbf{p}, \frac{\mathbf{p}^2}{2m} + \mathcal{V}\right)\right|^2 dp = \int |\phi(\mathbf{p}, E)|^2$$
$$\times\, \delta\left(E - \frac{\mathbf{p}^2}{2m} - \mathcal{V}\right) d\mathbf{p}\, dE < \infty. \qquad \text{(III.12)}$$

Let us study the Galilean transformation in the momentum space:

$$\phi'(\mathbf{p}, E) = \int e^{-i\mathbf{p}\cdot\mathbf{x} + iEt}\psi'(\mathbf{x}, t)\, d\mathbf{x}\, dt$$
$$= \int \exp\,[-i\mathbf{p}\cdot R^{-1}(\mathbf{x}' - \mathbf{v}t' + \mathbf{v}b - \mathbf{a})$$
$$+ iE(t' - b)] \times \exp\Big[-im\mathbf{v}\cdot\mathbf{x}'$$
$$+ \tfrac{1}{2}im\mathbf{v}^2 t' - iC\Big]\psi(\mathbf{x}', t')\, d\mathbf{x}'\, dt'. \qquad \text{(III.13)}$$

We drop the primes and re-order the sum in the exponential:

$$\phi'(\mathbf{p}, E) = \exp\,[iR\mathbf{p}\cdot\mathbf{a} - ibR\mathbf{p}\cdot\mathbf{v} - iEb - iC]$$
$$\times \int \exp\,[-i(R\mathbf{p} + m\mathbf{v})\cdot\mathbf{x} + i(E + \mathbf{v}\cdot R\mathbf{p}$$
$$+ \tfrac{1}{2}m\mathbf{v}^2)t]\psi(\mathbf{x}, t)\, d\mathbf{x}\, dt.$$

We write

$$\begin{aligned} \mathbf{p}' &= R\mathbf{p} + m\mathbf{v}, \\ E' &= E + \mathbf{v}\cdot R\mathbf{p} + \tfrac{1}{2}m\mathbf{v}^2. \end{aligned} \qquad \text{(III.14)}$$

Expressing $(\mathbf{p}, E)$ in terms of $(\mathbf{p}', E')$, we obtain

$$\phi'(\mathbf{p}, E) = \exp\,[-im\mathbf{v}\cdot\mathbf{a} + \tfrac{1}{2}imb\mathbf{v}^2 - iC$$
$$+ i\mathbf{a}\cdot\mathbf{p}' - ibE']\phi(\mathbf{p}', E'). \qquad \text{(III.15)}$$

We now choose

$$C = -\tfrac{1}{2}m\mathbf{a}\cdot\mathbf{v} + \tfrac{1}{2}mb\mathbf{v}^2. \qquad \text{(III.16)}$$

Then

$$\phi'(\mathbf{p}, E) = \exp\,[-i\tfrac{1}{2}m\mathbf{v}\cdot\mathbf{a} + i\mathbf{a}\cdot\mathbf{p}'$$
$$- ibE']\phi(\mathbf{p}', E'). \qquad \text{(III.17)}$$

One sees at once that (III.17), (III.14), and (III.12) are identical with (II.19) and (II.20); (III.11') with (II.7'). We conclude that the solution of the Schrödinger equation for a free spinless particle of mass m and internal energy $\mathcal{V}$ belongs to the physical representation $[m \mid \mathcal{V}, 0]$ of the Galilei group.

Clearly, any choice other than (III.16) for the constant C, would have led us to an irreducible, projective representation of the Galilei group different from but equivalent to the one we consider here.[7]

The Spin

We now want to interpret s as the spin of the particle described by the physical representation $[m \mid \mathcal{V}, s]$. This, however, is not a Galilei invariant concept. Indeed, in order to understand the meaning

[7] The concept of equivalence in the case of projective representations is somewhat distinct from the case of true representations. In fact if $\{U_r\}$ and $\{U_r'\}$ are two projective representations of the group G, they are said to be equivalent if $\mathbf{U}_r' = \mathbf{V}\mathbf{U}_r\mathbf{V}^{-1}$ holds between operator rays[2] ($\mathbf{U}_r$ is the operator ray generated by U_r, i.e. the set of all operators τU_r, $\tau \in \mathbf{C}$, $|\tau| = 1$). For the operators themselves, we have $U_r' = \phi(r)VU_rV^{-1}$, where $\phi(r)$ is some complex function of modulus 1 on the group. We see that U_r' is indeed a projective representation of the group with a factor system $\omega'(r, s) = [\phi(r)\phi(s)/\phi(rs)]\omega(r, s)$ equivalent to the factor system $\omega(r, s)$ of $\{U_r\}$.

of s, we have to define the little group $\mathfrak{h}(\mathbf{p}_0, E_0)$, that is to pick up some particular point $(\mathbf{p}_0, E_0)$ of the paraboloid. Let us recall however that all the representations $[m \mid \mathcal{V}, s]$ obtained from different points $(\mathbf{p}_0, E_0)$ (of the same paraboloid!) are equivalent.[5]

We now choose $(\mathbf{p}_0, E_0) = (0, \mathcal{V})$, i.e. the paraboloid top. We then have a most natural choice for $V_{\mathbf{p},E}$:

$$V_{\mathbf{p},E} = (0, 0, m^{-1}\mathbf{p}, 1). \qquad \text{(III.18)}$$

Starting from (I.3), we now easily obtain

$$\begin{aligned} V_0^{(\mathbf{p},E)(\mathbf{v},R)} &= V_{\mathbf{p}',E'}^{-1} V V_{\mathbf{p},E} \\ &= (0, 0, 0, R), \end{aligned} \qquad \text{(III.19)}$$

with any $V = (0, 0, \mathbf{v}, R)$ and $(\mathbf{p}', E') = V(\mathbf{p}, E)$. Consequently, we rewrite (II.19) as

$$\begin{aligned} &U(b, \mathbf{a}, \mathbf{v}, R)\psi(\mathbf{p}, E, \zeta) \\ &\quad = \exp\left[-i\tfrac{1}{2}m\mathbf{a}\cdot\mathbf{v} + i\mathbf{a}\cdot\mathbf{p}' - ibE'\right] \\ &\qquad \times \sum_\xi \psi(\mathbf{p}', E', \xi)[D^s(R)]_{\xi\zeta}\,, \end{aligned} \qquad \text{(III.20)}$$

where

$$\begin{aligned} \mathbf{p}' &= R\mathbf{p} + m\mathbf{v}, \\ E' &= E + \mathbf{v}\cdot R\mathbf{p} + \tfrac{1}{2}m\mathbf{v}^2. \end{aligned}$$

s characterizes now exclusively the behavior of our particle with respect to rotations; it is really its intrinsic angular momentum.

Let us notice that the choice (III.18) which led us to this result amounts to bringing back the particle at rest by accelerating the initial coordinate system, without rotating it around the direction of the movement (pure Galilei transformation).

Internal Energy

Physically, we are used to saying that, in nonrelativistic mechanics, we can freely choose the origin from which we count the energies. In the particular case of one free particle, this amounts to saying that the internal energy is completely arbitrary. We would like to rediscover this feature from Galilean invariance. This is done quite easily.

Let $\psi(\mathbf{p}, E, \zeta)$ be some basis function of the $[m \mid \mathcal{V}, s]$ physical representation of the Galilei group. Let us now define an operator $\mathfrak{U}$ by:

$$\begin{aligned} \tilde{\psi}(\mathbf{p}, E, \zeta) &= (\mathfrak{U}\psi)(\mathbf{p}, E, \zeta) \\ &= \psi(\mathbf{p}, E + \mathcal{V}, \zeta), \end{aligned} \qquad \text{(III.21)}$$

We call $\mathfrak{K}_\mathcal{V}$ the Hilbert spaces of functions $\psi(\mathbf{p}, E, \zeta)$ with the scalar product:

$$(\phi, \psi) = \int \sum_\zeta \bar{\Phi}(\mathbf{p}, E, \zeta)\psi(\mathbf{p}, E, \zeta) \times \delta\left(E - \frac{\mathbf{p}^2}{2m} - \mathcal{V}\right) d\mathbf{p}\, dE. \qquad \text{(II.20)}$$

It is clear that V realizes a mapping, in fact an homeomorphism of $\mathfrak{K}_\mathcal{V}$ on $\mathfrak{K}_0$. Since

$$\begin{aligned} (\tilde{\phi}, \tilde{\psi}) &= \int \sum_\zeta \bar{\tilde{\phi}}(\mathbf{p}, E, \zeta)\tilde{\psi}(\mathbf{p}, E, \zeta) \times \delta\left(E - \frac{\mathbf{p}^2}{2m}\right) d\mathbf{p}\, dE \\ &= \int \sum_\zeta \bar{\phi}(\mathbf{p}, E + \mathcal{V}, \zeta)\psi(\mathbf{p}, E + \mathcal{V}, \zeta) \times \delta\left(E - \frac{\mathbf{p}^2}{2m}\right) d\mathbf{p}\, dE \\ &= \int \sum_\zeta \bar{\phi}(\mathbf{p}, E', \zeta)\psi(\mathbf{p}, E', \zeta) \times \delta\left(E' - \frac{\mathbf{p}^2}{2m} - \mathcal{V}\right) d\mathbf{p}\, dE', \end{aligned}$$

we have

$$(\tilde{\phi}, \tilde{\psi}) = (\mathfrak{U}\phi, \mathfrak{U}\psi) = (\phi, \psi), \qquad \text{(III.22)}$$

that is to say, $\mathfrak{U}$ is an isometric operator and $\mathfrak{K}_\mathcal{V}$ and $\mathfrak{K}_0$ are isomorphic Hilbert spaces.

If $U(G)$ is the operator corresponding to $G = (b, \mathbf{a}, \mathbf{v}, R)$ in the $[m \mid \mathcal{V}, s]$ representation, we define

$$\tilde{U}(G) = \mathfrak{U}U(G)\mathfrak{U}^{-1}. \qquad \text{(III.23)}$$

Letting now $\tilde{U}(G)$ act upon some function $\psi(\mathbf{p}, E)$ of $\mathfrak{K}_0$, using (III.21) and (II.19), we get

$$\tilde{U}(G) = e^{ib\mathcal{V}}U_0(G), \qquad \text{(III.24)}$$

where $U_0(G)$ is the operator corresponding to G, in the representation $[m \mid 0, s]$ according to the definition (II.19), or else

$$U_\mathcal{V}(G) = e^{ib\mathcal{V}}\mathfrak{U}^{-1}U_0(G)\mathfrak{U}, \qquad \text{(III.25)}$$

where we add a subscript $\mathcal{V}$ to $U(G)$ in order to emphasize the fact that it belongs to the representation $[m \mid \mathcal{V}, s]$. Obviously then, the representations $[m \mid \mathcal{V}, s]$ and $[m \mid 0, s]$ are equivalent, in the sense of projective representations equivalence.[7] In other words, for an isolated particle, the internal energy $\mathcal{V}$ has no physical significance.

Antiparticles

Until now, when looking for a physical interpretation of our results, we implicitly assumed that m, which we interpreted as the mass, was positive. Actually, the construction of the first section is

valid for any nonzero value of m. What can be said of the negative m case?

Casting a glance at Eq. (II.19), we see at once that if $\psi(\mathbf{p}, E, \zeta)$ transforms according to the $[m \mid \mathcal{U}, s]$ representation, then

$$\psi^c(\mathbf{p}, E, \zeta) = \psi(-\mathbf{p}, -E, \zeta) \qquad \text{(III.26)}$$

transforms according to $[-m \mid -\mathcal{U}, \bar{s}]$, where by $\bar{s}$ we mean that ψ^c transforms under rotations by $\bar{D}^s(R)$, the complex conjugate representation of $D^s(R)$. But this does not alter the physical interpretation of s as the spin of the particle.

We can immediately verify that the operation $\psi \to \psi^c$ is an antiunitary one and then, that the representations $[m \mid \mathcal{U}, s]$ and $[-m \mid -\mathcal{U}, \bar{s}]$ are antiunitarily equivalent.

If we consider now charged particles, by the replacement

$$\mathbf{p} \to \mathbf{p} - Q\mathbf{A}, \qquad E \to E - Q\varphi, \qquad \text{(III.27)}$$

we see that the above-mentioned antiunitary transformation takes a particle of mass m, internal energy $\mathcal{U}$, and charge Q into another particle characterized respectively by $(-m, -\mathcal{U}, -Q)$.

Let us remark that a Fourier transformation, or else direct dealing with the Schrödinger equation in space–time, leads us to the same result with the transformation

$$\psi^c(\mathbf{x}, t) = \bar{\psi}(\mathbf{x}, t). \qquad \text{(III.28)}$$

In other words, if the representation $[m \mid \mathcal{U}, s]$ and the charge Q describe some particle, we may describe its antiparticle either by the same representation $[m \mid \mathcal{U}, s]$ and the charge $-Q$, or else by the representation $[-m \mid -\mathcal{U}, \bar{s}]$ and the same charge Q.

This gives some meaning to the negative m case.

We notice, however, that, as it is well-known, if the same Dirac equation describes particle and antiparticle, their nonrelativistic description needs two different Schrödinger equations.

IV. THE ZERO-MASS CASE

We now plainly make $m = 0$ in the realization (III.20) we obtained for the physical representations $[m \mid \mathcal{U}, s]$ of the Galilei group. We thus get some *true* representation of the group:

$$U(0, \mathbf{a}, \mathbf{v}, R)\psi(\mathbf{p}, E, \zeta) = \exp(i\mathbf{a}\cdot\mathbf{p}' - ibE') \times \sum_\xi \psi(\mathbf{p}', E', \xi)[D^s(R)]_{\xi\zeta}, \qquad \text{(IV.1)}$$

where

$$\mathbf{p}' = R\mathbf{p},$$

$$E' = E + \mathbf{v}\cdot R\mathbf{p}.$$

The basis functions of the representation are now defined onto the "cylinder":

$$\mathbf{p}^2 = P^2 = C^{te} \quad (E \text{ arbitrary}). \qquad \text{(IV.2)}$$

But these representations are no longer irreducible. In fact, let $\mathbf{p}_0$ (lying on the cylinder IV.2) be the vector around which the rotations, in the D^s representation chosen, are diagonal matrices. We call $\mathfrak{h}(\mathbf{p}_0)$ the group of rotations around $\mathbf{p}_0$. R_0 being such a rotation, with an angle φ_0, one has

$$[D^s(R_0)]_{\rho\tau} = \delta_{\rho\tau}\, e^{i\rho\varphi_0}. \qquad \text{(IV.3)}$$

We once more apply the little group technique.[8] Starting from any vector $\mathbf{p}$ of the cylinder (IV.2), we can choose one rotation $r_\mathbf{p}$ such that

$$r_\mathbf{p}\mathbf{p}_0 = \mathbf{p}. \qquad \text{(IV.4)}$$

Now, R being some rotation which takes $\mathbf{p}$ into $\mathbf{p}'$:

$$\mathbf{p}' = R\mathbf{p}, \qquad \text{(IV.5)}$$

we get

$$p_0 = r_{\mathbf{p}'}^{-1} R r_\mathbf{p} \mathbf{p}_0, \quad \text{i.e.,} \quad r_{\mathbf{p}'}^{-1} R r_\mathbf{p} = R_0^{\mathbf{p}', R} \in \mathfrak{h}(\mathbf{p}_0).$$

That is, any rotation can be written in the form

$$R = r_{\mathbf{p}'} R_0^{\mathbf{p}', R} r_\mathbf{p}^{-1}. \qquad \text{(IV.6)}$$

We next define new functions $\phi(\mathbf{p}, E, \rho)$ on our cylinder:

$$\phi(\mathbf{p}, E, \rho) = \sum_\zeta \psi(\mathbf{p}, E, \zeta)[D^s(r_\mathbf{p})]_{\zeta\rho}. \qquad \text{(IV.7)}$$

They transform according to

$$U(b, \mathbf{a}, \mathbf{v}, R)\phi(\mathbf{p}, E, \rho) = \sum_\zeta \exp(i\mathbf{a}\mathbf{p}' - ibE') \times \sum_\xi \psi(\mathbf{p}', E', \xi)[D^s(R)]_{\xi\zeta}[D^s(r_\mathbf{p})]_{\zeta\rho}.$$

Inverting the summations, and using (IV.7),

$$U(b, \mathbf{a}, \mathbf{v}, R)\phi(\mathbf{p}, E, \rho) = \exp(i\mathbf{a}\mathbf{p}' - ibE') \times \sum_\xi \sum_\tau \psi(\mathbf{p}', E', \xi)[D^s(r_{\mathbf{p}'})]_{\xi\tau}[D^s(R_0^{\mathbf{p}', R})]_{\tau\rho}.$$

Lastly,

$$U(b, \mathbf{a}, \mathbf{v}, R)\phi(\mathbf{p}, E, \rho) = \exp[i\mathbf{a}\mathbf{p}' - ibE' + i\rho\varphi_0(R, \mathbf{p})]\phi(\mathbf{p}', E'\rho). \qquad \text{(IV.8)}$$

We calculate explicitly the function $\varphi_0(R, \mathbf{p})$ in the Appendix. The subspaces of the functions $\phi(\mathbf{p}, E, \rho)$, with a given ρ, are thus invariant. We have reduced the primitive representation (IV.1) into $(2s + 1)$ representations which are now irreducible. These are the true representations named "class II" by

Inönu and Wigner.[3] They are characterized by an integer ρ and a positive number P^2. We denote them by $\{0 \mid P, \rho\}$. The very construction of these representations and as we will see later, their Lie algebra and her invariants, designate ρ as the component of the angular momentum along the direction of the linear momentum, i.e., helicity.

Added to the fact that we deal here with null-mass states, this suggests a close analogy between the representations just found and the irreducible representations of the Poincaré group in the zero-mass case. In fact, at least in the case of vanishing P, we may interpret the representations $\{0 \mid P, \rho\}$ $(P \to 0)$ as describing zero-mass and infinite-speed particles, which are indeed the nonrelativistic limit of the zero-mass particles. Naturally, there is no completely consistent interpretation of the representations just found. Nevertheless, we have been able to give some vague meaning to them. And mainly, they display certain features we usually think to be characteristic of the relativistic case (uncoupling of different helicity states) and which, in fact, are latent in the nonrelativistic case.[8,9]

V. THE LIE ALGEBRA OF THE GALILEI GROUP AND THE ROLE OF THE MASS IN NONRELATIVISTIC QUANTUM MECHANICS

Taking the infinitesimal elements of the one-parameter subgroups of the Galilei group (considered as a Lie group) and using the group law (I.3), we calculate their commutators and thus obtain the Lie brackets for the Lie algebra of the group.

We make the most natural choice for the basis elements of the algebra:

τ for the time translations,
$k_i (i = 1, 2, 3)$ for the space translations,
$u_i (i = 1, 2, 3)$ for the pure Galilei transformations,
$M_i (i = 1, 2, 3)$ for the rotations.

We then have

$$\begin{aligned} [M_i, M_j] &= \epsilon_{ijk} M_k, & [u_i, u_j] &= [k_i, k_j] = 0, \\ [M_i, u_j] &= \epsilon_{ijk} u_k, & [k_i, \tau] &= 0, \\ [M_i, k_j] &= \epsilon_{ijk} k_k, & [u_i, k_j] &= 0, \\ [M_i, \tau] &= 0, & [u_i, \tau] &= k_i. \end{aligned} \tag{V.1}$$

Let us now compute explicitly the infinitesimal elements of the physical representation (III.20):

$$\begin{aligned} \mathbf{M} &= -\mathbf{p} \times (\partial/\partial \mathbf{p}) - i\mathbf{S}, \\ \mathbf{u} &= m(\partial/\partial \mathbf{p}) + \mathbf{p}(\partial/\partial E), \\ \mathbf{k} &= \mathbf{p}, \qquad \tau = E, \end{aligned} \tag{V.2}$$

with obvious notations. Let us notice, as usual, the splitting of the total angular momentum in orbital and intrinsic (spin) parts.

It is now easy to see that the realization (V.2) fulfills all equations (V.1), except that translations and pure Galilei transformations no longer commute. Instead,

$$[u_i, k_j] = m\delta_{ij}. \tag{V.3}$$

This is quite natural. We know that we deal in fact with a projective representation. This means that in (V.2) we obtain a representation of the Lie algebra of a central extension of the Galilei group, and no longer of the Galilei group itself. The Lie algebra element of the one-parameter subgroup by which the extension is made can be called μ. Here, it is represented by $\mu = m$. The extension is central, so that μ commutes with all other Lie algebra elements. But it is nontrivial, so that μ appears in some Lie bracket [see (V.3)].

The enveloping algebra admits the following invariants:[4]

$$\begin{aligned} 2\mu\tau - \mathbf{k}^2 &= 2mE - \mathbf{p}^2 = 2m\mathcal{V}, \\ -(\mu M + \mathbf{k} \times \mathbf{u})^2 &= m^2 \mathbf{S}^2 = m^2 s(s+1), \\ \mu &= m. \end{aligned} \tag{V.4}$$

We recover, of course, the characterization of physical representations by $[m \mid \mathcal{V}, s]$. There is however a rather subtle point we have yet to make clear. We have seen that physical knowledge as well as mathematical considerations on the Galilei group physical representations allow us to conclude that, in fact, the internal energy of an isolated particle is an arbitrary parameter. Precisely we showed that all representations $[m \mid \mathcal{V}, s]$ and $[m \mid \mathcal{V}', s]$ are physically equivalent. But we now find $\mathcal{V}$ as an element of the center of the group algebra. How can any equivalence transformation modify this center? The answer is to be found in the fact that we deal here with an extension of the Galilei group, and such an extension as we consider here has not a uniquely defined Lie algebra. There is a whole class of algebras, in one-to-one correspondence with the unlike but equivalent systems of factors of the projective representation associated with the extension. Here, going from some algebra to another equivalent one, we modify precisely the center element $\mathcal{V}$, and that one only.

[8] It was Wigner[9] who emphasized (in the relativistic case but the same remark is valid here) that a zero-mass system possesses two polarization states but only if we consider the space reflections, since otherwise they would not be connected to each other. On the other hand, for the nonzero mass systems, proper rotational invariance is sufficient for deducing the $(2s + 1)$ polarization states from any one among them.

[9] E. P. Wigner, Rev. Mod. Phys. **29**, 255 (1957).

Now the existence within the center of the enveloping algebra of the group, of some basis element of this algebra (here μ, the mass operator), involves a most important physical consequence: it leads to a superselection rule.[10]

Let us for instance consider a state vector which results from the superposition of two state vectors having different masses:

$$\psi = \psi_1 + \psi_2, \tag{V.5}$$

where ψ_1 and ψ_2 respectively belong to the physical representations $[m_1 \mid \mathcal{U}, s]$ and $[m_2 \mid \mathcal{U}, s]$ of the Galilei group. We now consider the behavior of this composed state under the following series of transformations: a translation $\mathbf{a}$; a pure Galilei transformation $\mathbf{v}$; the inverse translation, and the inverse Galilei transformation. Within the group we know that these all commute, whence,

$$(0, 0, -\mathbf{v}, 1)(0, -\mathbf{a}, 0, 1)(0, 0, \mathbf{v}, 1)(0, \mathbf{a}, 0, 1) = (0, 0, 0, 1), \tag{V.6}$$

i.e., the identical transformation.

With respect to some physical representation, that series is obviously represented by some phase factor at most. In fact, using (III.20), we find at once

$$U(0, 0, -\mathbf{v}, 1)U(0, -\mathbf{a}, 0, 1)U(0, 0, \mathbf{v}, 1) \times U(0, \mathbf{a}, 0, 1) = e^{-im\mathbf{a}\cdot\mathbf{v}}. \tag{V.7}$$

Thus, our compound state becomes

$$\psi = \psi_1 + \psi_2 \rightarrow \psi = e^{-im_1\mathbf{a}\cdot\mathbf{v}}\psi_1 + e^{-im_2\mathbf{a}\cdot\mathbf{v}}\psi_2. \tag{V.8}$$

The superposition principle cannot have any meaning for ψ_1 and ψ_2 if $m_1 \neq m_2$, since that would mean that an identical transformation could affect the norm of any of their compound states. The relative phase of two states having different masses is completely arbitrary. This is known as the "Bragmann superselection rule."[11] It prevents the existence, in nonrelativistic quantum mechanics, of states with a mass spectrum, and therefore of unstable particles.

We see here how the mass plays different parts in relativistic and nonrelativistic quantum theories.

The Lie Algebra in the Zero-Mass Case

We will deal now really with the Lie algebra of the Galilei group itself: the central extension becomes trivially a direct product and has no more physical consequences.

Starting from the true representation (IV.9) of the Galilei group, and using the explicit form $\varphi_0(R, \mathbf{p})$ derived in the Appendix, we obtain the following representation:

$$\begin{aligned} M_1 &= -p_2\frac{\partial}{\partial p_3} + p_3\frac{\partial}{\partial p_2} + i\rho\frac{p_2}{p + p_3} \\ M_2 &= -p_3\frac{\partial}{\partial p_1} + p_1\frac{\partial}{\partial p_3} + i\rho\frac{p_2}{p + p_3} \\ M_3 &= -p_1\frac{\partial}{\partial p_2} + p_2\frac{\partial}{\partial p_1} + i\rho \\ \mathbf{u} &= \mathbf{p}(\partial/\partial E), \qquad \mathbf{k} = \mathbf{p}, \qquad \tau = E. \end{aligned} \tag{V.9}$$

We may now verify that this representation leads us to the rules (V.1). We also notice the close analogy between these expressions and those obtained for the Poincaré-group Lie algebra in the zero-mass case.[12]

The enveloping algebra invariants and their values for the $\{0 \mid P, \rho\}$ representation are

$$\begin{aligned} \mathbf{k}^2 &= \mathbf{p}^2 = P^2, \\ \mathbf{M}\cdot\mathbf{k} &= i\rho p = i\rho P. \end{aligned} \tag{V.10}$$

This confirms our interpretation of ρ as the helicity.

VI. DECOMPOSITION OF THE TENSOR PRODUCT OF TWO PHYSICAL REPRESENTATIONS

The tensor product of two physical representations of the Galilei group,

$$[m_1 \mid \mathcal{U}_1, s_1] \otimes [m_2 \mid \mathcal{U}_2, s_2], \tag{VI.1}$$

is still a (projective) representation whose operators act onto the square-integrable basis functions

$$\psi(\mathbf{p}_1, \mathbf{p}_2, E_1, E_2, \zeta_1, \zeta_2)$$

according to

$$\begin{aligned} U(b, \mathbf{a}, \mathbf{v}, R)\psi(\mathbf{p}_1, \mathbf{p}_2, E_1, E_2, \zeta_1, \zeta_2) &= \exp\left[-\tfrac{1}{2}i(m_1 + m_2)\mathbf{a}\cdot\mathbf{v}\right. \\ &\quad \left. + i\mathbf{a}(\mathbf{p}_1' + \mathbf{p}_2') - ib(E_1' + E_2')\right] \\ &\quad \times \sum_{\xi_1\xi_2} \psi(\mathbf{p}_1', \mathbf{p}_2', E_1', E_2', \xi_1, \xi_2) \\ &\quad \times [D^{s_1}(R)]_{\xi_1\zeta_1}[D^{s_2}(R)]_{\xi_2\zeta_2}, \end{aligned} \tag{VI.2}$$

where

$$\begin{aligned} \mathbf{p}_i' &= R\mathbf{p}_i + m_i\mathbf{v}, \\ E_i' &= E_i + \mathbf{v}\cdot R\mathbf{p}_i + \tfrac{1}{2}m_i\mathbf{v}^2. \end{aligned} \qquad (i = 1, 2)$$

[10] G. C. Wick, A. S. Wightman, and E. P. Wigner, Phys. Rev. **88**, 101 (1952).

[11] A. S. Wightman, "Lectures on Relativistic Invariance," in *Les Houches 1960 Summer School Proceedings* (Hermann et Cie., Paris, 1960), pp. 159–226.

[12] J. S. Lomont and H. E. Moses, J. Math. Phys. **3**, 405 (1962).

These representations are unitary, the scalar product being defined as

$$(\phi, \psi) = \int d\mathbf{p}_1\, dE_1\, d\mathbf{p}_2\, dE_2$$
$$\times\, \delta\Big(E_1 - \frac{\mathbf{p}_1^2}{2m_1} - \mathcal{V}_1\Big)\, \delta\Big(E_2 - \frac{\mathbf{p}_2^2}{2m_2} - \mathcal{V}_2\Big)$$
$$\times \sum_{\zeta_1\zeta_2} \bar{\phi}(\mathbf{p}_1, \mathbf{p}_2, E_1, E_2, \zeta_1, \zeta_2)$$
$$\times\, \psi(\mathbf{p}_1, \mathbf{p}_2, E_1, E_2, \zeta_1, \zeta_2). \qquad \text{(VI.3)}$$

We are going to reduce the Hilbert space, which we just determined, into a direct sum (possibly a direct integral) of invariant Hilbert spaces. We proceed quite similarly to Wightman.[11]

We first define new variables:

$$\mathbf{P} = \mathbf{p}_1 + \mathbf{p}_2\,, \quad E = E_1 + E_2\,,$$
$$\mathbf{q} = \frac{m_2\mathbf{p}_1 - m_1\mathbf{p}_2}{m_1 + m_2}, \quad \epsilon = \frac{m_2E_1 - m_1E_2}{m_1 + m_2}\,, \qquad \text{(VI.4)}$$

and call

$$M = m_1 + m_2\,, \quad \mu = \frac{m_1m_2}{m_1 + m_2}\,,$$

where we have assumed $m_1 + m_2 \neq 0$.

These variables are precisely those corresponding to the usual separation of the center-of-mass and relative motions for our two-particle system.

We also introduce the internal energy of the compound system, i.e., the difference between its total energy and the center-of-mass kinetic energy:

$$\mathcal{V} = E_1 + E_2 - (\mathbf{P}^2/2M). \qquad \text{(VI.5)}$$

We have now the following expression for the scalar product:

$$(\phi, \psi) = \int_{\mathcal{V}_1+\mathcal{V}_2}^{\infty} d\mathcal{V} \int d\mathbf{P}\, dE\, \delta\Big(E - \frac{P^2}{2M} - \mathcal{V}\Big)$$
$$\times \int dq\, d\epsilon\, \delta\Big(\mathcal{V} - \frac{q^2}{2} - \mathcal{V}_1 - \mathcal{V}_2\Big)$$
$$\times\, \delta\Big(\epsilon - \frac{\mathbf{P}\cdot\mathbf{q}}{m_1 + m_2}\, q^2 - \frac{m_2\mathcal{V}_1 - m_1\mathcal{V}_2}{m_1 + m_2}\Big)$$
$$\times \sum_{\zeta_1\zeta_2} \bar{\phi}(\mathbf{P}, E, \mathbf{q}, \epsilon, \zeta_1, \zeta_2)\psi(\mathbf{P}, E, \mathbf{q}, \epsilon, \zeta_1, \zeta_2), \qquad \text{(VI.6)}$$

which has been directly derived from (VI.3).

The Hilbert space $\mathfrak{C}$ of the representation is thus reduced to a direct integral of Hilbert spaces $\mathfrak{C}_{\mathcal{V}}$, each associated to a paraboloid:

$$E = (\mathbf{P}^2/2M) = \mathcal{V}. \qquad \text{(VI.7)}$$

Since $(\mathbf{P}, E)$ precisely characterizes the representation of the translation subgroup in the tensor product (see VI.1), the condition (VI.7) is then a necessary one for the Hilbert spaces $\mathfrak{C}_{\mathcal{V}}$ to be invariant. In fact, we can reduce them no further with respect to the translation subgroup representation, and it suffices now to look in each $\mathfrak{C}_{\mathcal{V}}$ for the subspaces invariant with respect to the little group, which may be chosen simply as the rotation group.

Equation (VI.6) shows now that, if some point $(\mathbf{P}, E)$ is fixed, the basis functions of our representation only depend on some vector $\mathbf{q}$ whose length is fixed. Since these are uniform functions onto the sphere $\mathbf{q}^2 = C^{te}$, we expand them in spherical harmonics, i.e., basis functions of irreducible representations of the rotation group. We then obtain, for the little group (rotation group), the representation

$$D^{s_1} \otimes D^{s_2} \otimes \Big(\bigoplus_{l=0}^{\infty} D^l\Big). \qquad \text{(VI.8)}$$

The decomposition of this tensor product into irreducible representations is immediate. Since we have seen at the beginning that a paraboloid $\mathcal{V}$ and an irreducible representation of the little group uniquely define a physical representation of the Galilei group, we finally obtain the complete solution to our problem, which we symbolically write as:

$$[m_1 \mid \mathcal{V}_1, s_1] \otimes [m_2 \mid \mathcal{V}_2, s_2] = \int_{\oplus\mathcal{V}_1+\mathcal{V}_2}^{\infty} d\mathcal{V} \bigoplus_{l=0}^{\infty} \bigoplus_{s=|s_1-s_2|}^{s_1+s_2} \bigoplus_{j=|l-s|}^{l+s} [m_1 + m_2 \mid \mathcal{V}, j].$$

We see, exactly as in the case of the Poincaré group,[11] the most natural appearance of the orbital angular momenta.

This provides also the profound reason why, when studying the Schrödinger equation, one keeps only integer (and not half-integer) relative angular momenta. But in contradistinction with the relativistic case, the mass is now conserved (and even superconserved as we have seen). The kinetic energy of the relative motion of the components is to be found now in the internal energy of the compound system. The weakened concept of equivalence which we introduced earlier and which led us to the arbitrariness of the internal energy of an isolated particle, has now also, in the two-particle case, a quite interesting application. It enables us to change

$[m_1 \mid \upsilon_1, s_1]$ and $[m_2 \mid \upsilon_2, s_2]$ into $[m_1 \mid 0, s_1]$ and $[m_2 \mid 0, s_2]$, but not to change simultaneously all the $[m_1 + m_2 \mid \upsilon, j]$ of their tensor product decomposition into $[m_1 + m_2 \mid 0, j]$. All we can do is to "renormalize" the internal energy of the possible compound states by an amount $\upsilon_1 + \upsilon_2$. Of course this agrees entirely with our previous physical knowledge: once we have fixed up the internal energy of two isolated particles, their compound state has an internal energy which is no longer arbitrary.

In the case where $m_1 + m_2 = 0$, with the help of techniques quite similar to those just used, we obtain the following result, which we quote here for completeness and without proof:

$$[m \mid \upsilon_1, s_1] \otimes [-m \mid \upsilon_2, s_2] = \int_{\oplus 0}^{\infty} dP \bigoplus_{l=-\infty}^{\infty} \bigoplus_{\rho_1=-s_1}^{+s_1} \bigoplus_{\rho_2=-s_2}^{+s_2} \{0 \mid P, \rho_1 + \rho_2 + l\}, \tag{VI.10}$$

or else

$$[m \mid \upsilon_1, s_1] \otimes [-m \mid \upsilon_2, s_2] = \int_{\oplus 0}^{\infty} dP \bigoplus_{\rho=-\infty}^{+\infty} \{0 \mid P, \rho\}^{\otimes (2s_1+1)(2s_2+1)}. \tag{VI.10'}$$

Such a result is valid also if we replace one (or two) of the representations $[m \mid \upsilon, s]$ by $[m \mid \upsilon, \bar{s}]$ (see Sec. III); particularly,

$$[m \mid \upsilon, s] \otimes [-m \mid \upsilon', \bar{s}] = \int_{\oplus 0}^{\infty} dP \bigoplus_{\rho=-\infty}^{+\infty} \{0 \mid P, \rho\}^{\otimes (2s+1)^2}, \tag{VI.11}$$

and this justifies our choosing the representations $[m \mid \upsilon, s]$ and $[-m \mid \upsilon', \bar{s}]$ in order to represent, respectively, a particle and its antiparticle.

All the results derived here are of course well known. It is however stimulating to obtain them from Galilean invariance only, and this provides an agreeable and unifying piont of view.

ACKNOWLEDGMENTS

The author is very grateful to Professor Louis Michel and Professor François Lurçat for the interest they have taken in this work, as well as for their many suggestions and critical remarks.

APPENDIX

We here derive an explicit formula for the angle φ_0 of the rotation:

$$R_0(\mathbf{p}, R) = r_{\mathbf{p}'}^{-1} R r_{\mathbf{p}}, \tag{A.1}$$

where $\mathbf{p}' = R\mathbf{p}$, and the $r_{\mathbf{p}}$'s are well defined rotations which take some fixed $\mathbf{p}_0$ into $\mathbf{p}$:

$$r_{\mathbf{p}}\mathbf{p}_0 = \mathbf{p}.$$

We use the spinorial representation of the rotation group. A rotation by angle around an axis $\mathbf{n}(|\mathbf{n}| = 1)$ may be written as

$$R(\mathbf{n}, \varphi) = \cos(\varphi/2) - i\hat{n} \sin(\varphi/2), \tag{A.2}$$

where

$$\hat{n} = \boldsymbol{\tau}\cdot\mathbf{n}, \tag{A.3}$$

and the τ_i's are the usual Pauli matrices.

We choose as $r_{\mathbf{p}}$, the rotation in the $(\mathbf{p}_0, \mathbf{p})$ plane which brings $\mathbf{p}_0$ into $\mathbf{p}$. Writing it as the product of two plane symmetries, we have

$$r_{\mathbf{k}} = \frac{(\hat{k} + \hat{k}_0)\hat{k}_0}{(2 + 2\mathbf{k}\cdot\mathbf{k}_0)^{\frac{1}{2}}} = \frac{\hat{k}(\hat{k} + \hat{k}_0)}{(2 + 2\mathbf{k}\cdot\mathbf{k}_0)^{\frac{1}{2}}}, \tag{A.4}$$

where $\mathbf{k}$ and $\mathbf{k}_0$ are the unit vectors lying on $\mathbf{p}$ and $\mathbf{p}_0(\mathbf{k} = \mathbf{p}/|\mathbf{p}|$, etc. $\cdots)$, and with the same notations as in (A.3).

Now, from the definition of $\mathbf{k}'$ and (A.2), we have

$$\hat{k}' = \left(\cos\frac{\varphi}{2} - i\hat{n}\sin\frac{\varphi}{2}\right)\hat{k}\left(\cos\frac{\varphi}{2} + i\hat{n}\sin\frac{\varphi}{2}\right). \tag{A.5}$$

Bringing (A.5), (A.4), and (A.2) into (A.1), using also, repeatedly, the well-known identity

$$\hat{a}\hat{b} = \mathbf{a}\cdot\mathbf{b} + i\boldsymbol{\tau}\cdot(\mathbf{a} \times \mathbf{b}), \tag{A.6}$$

and

$$\hat{a}\hat{b} = \mathbf{a}^2,$$

we get

$$R_0 = \frac{\hat{k}_0(\hat{k}_0 + \hat{k}')}{(2 + 2\mathbf{k}_0\cdot\mathbf{k}')^{\frac{1}{2}}}\left(\cos\frac{\varphi}{2} - i\hat{n}\sin\frac{\varphi}{2}\right)\frac{\hat{k}(\hat{k} + \hat{k}_0)}{(2 + 2\mathbf{k}_0\cdot\mathbf{k})^{\frac{1}{2}}},$$

$$R_0 = \frac{(2 + \hat{k}\hat{k}_0 + \hat{k}_0\hat{k})\cos\frac{\varphi}{2} - i(\hat{n} + \hat{k}_0\hat{n}\hat{k}_0 + \hat{n}\hat{k}\hat{k}_0 + \hat{k}_0\hat{n}\hat{k}_0)\sin\frac{\varphi}{2}}{2(1 + \mathbf{k}_0\cdot\mathbf{k}')^{\frac{1}{2}}(1 + \mathbf{k}_0\cdot\mathbf{k})^{\frac{1}{2}}}.$$

Finally,

$$R_0 = \frac{(1 + \mathbf{k}\cdot\mathbf{k}_0)\cos(\varphi/2) + (\mathbf{k}_0, \mathbf{n}, \mathbf{k})\sin(\varphi/2) - i\hat{k}_0(\mathbf{n}\cdot\mathbf{k}_0 + \mathbf{n}\cdot\mathbf{k})\sin(\varphi/2)}{(1 + \mathbf{k}\cdot\mathbf{k}_0)^{\frac{1}{2}}[1 + \cos\varphi\mathbf{k}_0\cdot\mathbf{k} + (1 - \cos\varphi)(\mathbf{n}\cdot\mathbf{k}_0)(\mathbf{n}\cdot\mathbf{k}) + \sin\varphi(\mathbf{k}_0, \mathbf{n}, \mathbf{k})]^{\frac{1}{2}}}. \tag{A.7}$$

The expression (A.7) for R_0 shows clearly the axis $\mathbf{k}_0$ of the rotation R_0, whose angle φ_0 is immediately obtained by identifying (A.7), and

$$R_0 = \cos(\varphi_0/2) - i\hat{k}_0 \sin(\varphi_0/2). \quad (A.7')$$

In particular, for an infinitesimal rotation around $\mathbf{n}$, with an angle of $\varphi \ll 1$,

$$\varphi_0 \simeq (\mathbf{n}\cdot\mathbf{k}_0 + \mathbf{n}\cdot\mathbf{k}/1 + \mathbf{k}\cdot\mathbf{k}_0)\varphi. \quad (A.8$$

Choosing $\mathbf{k}_0$ as our z axis, and taking successively infinitesimal rotations around the x, y, and z axes, we obtain

$$\begin{aligned}(\varphi_0)_x &= [k_x/(1 + k_z]\varphi = [p_x/(p + p_z)]\varphi,\\ (\varphi_0)_y &= [k_y/(1 + k_z)]\varphi = [p_y/(p + p_z)]\varphi, \quad (A.9)\\ (\varphi_0)_z &= [1 + k_z/(1 + k_z)]\varphi = \varphi.\end{aligned}$$

We can then immediately derive the expressions (V.9) for the Lie algebra in the zero-mass case.

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 4, NUMBER 6 JUNE 1963

Principle of General Q Covariance

D. FINKELSTEIN,*
Yeshiva University, New York, New York
J. M. JAUCH,
University of Geneva and *CERN, Geneva, Switzerland*
S. SCHIMINOVICH,†
Yeshiva University, New York, New York
AND
D. SPEISER,‡
University of Geneva, Geneva, Switzerland
(Received 10 December 1962)

In this paper the physical implications of quaternion quantum mechanics are further explored. In a quanternionic Hilbert space $\mathfrak{H}_Q$, the lattice of subspaces has a symmetry group which is isomorphic to the group of all co-unitary transformations in $\mathfrak{H}_Q$. In contrast to the complex space $\mathfrak{H}_C$ (ordinary Hilbert space), this group is connected, while for $\mathfrak{H}_C$ it consists of two disconnected pieces.

The subgroup of transformations in $\mathfrak{H}_Q$ which associates with every quaternion q of magnitude 1, the correspondence $\psi \to q\psi q^{-1}$ for all $\psi \in \mathfrak{H}_Q$ (called Q conjugations), is isomorphic to the three-dimensional rotation group. We postulate the principle of Q covariance: The physical laws are invariant under Q conjugations. The full significance of this postulate is brought to light in localizable systems where it can be generalized to the principle of general Q covariance: Physical laws are invariant under general Q conjugations. Under the latter we understand conjugation transformations which vary continuously from point to point.

The implementation of this principle forces us to construct a theory of parallel transport of quaternions. The notions of Q-covariant derivative and Q curvature are natural consequences thereof.

There is a further new structure built into the quaternionic frame through the equations of motion. These equations single out a purely imaginary quaternion $\eta(x)$ which may be a continuous function of the space–time coordinates. It corresponds to the i in the Schrödinger equation of ordinary quantum mechanics. We consider $\eta(x)$ as a fundamental field, much like the tensor $g_{\mu\nu}$ in the general theory of relativity. We give here a classical theory of this field by assuming the simplest invariant Lagrangian which can be constructed out of η and the covariant Q connection. It is shown that this theory describes three vector fields, two of them with mass and charge, and one massless and neutral. The latter is identifiable with the classical electromagnetic field.

1. INTRODUCTION

IN the development from Galilean to special to general relativity, it was shown by Einstein that the concepts of Euclidean geometry have only an approximate validity and that the true laws of geometry are subject to disturbances from place to place. Still more fundamental than the laws of geometry are those of classical logic as expressed in the propositional calculus. In the development from classical to quantum physics it was shown by Bohr that the concepts of classical logic have

* Supported by the National Science Foundation.
† Now at the University of Buenos Aires.
‡ Supported by the Swiss Commision for Atomic Research.

The expression (A.7) for R_0 shows clearly the axis $\mathbf{k}_0$ of the rotation R_0, whose angle φ_0 is immediately obtained by identifying (A.7), and

$$R_0 = \cos(\varphi_0/2) - i\hat{k}_0 \sin(\varphi_0/2). \quad (A.7')$$

In particular, for an infinitesimal rotation around $\mathbf{n}$, with an angle of $\varphi \ll 1$,

$$\varphi_0 \simeq (\mathbf{n}\cdot\mathbf{k}_0 + \mathbf{n}\cdot\mathbf{k}/1 + \mathbf{k}\cdot\mathbf{k}_0)\varphi. \quad (A.8$$

Choosing $\mathbf{k}_0$ as our z axis, and taking successively infinitesimal rotations around the x, y, and z axes, we obtain

$$\begin{aligned}(\varphi_0)_x &= [k_x/(1 + k_z]\varphi = [p_x/(p + p_z)]\varphi,\\ (\varphi_0)_y &= [k_y/(1 + k_z)]\varphi = [p_y/(p + p_z)]\varphi, \quad (A.9)\\ (\varphi_0)_z &= [1 + k_z/(1 + k_z)]\varphi = \varphi.\end{aligned}$$

We can then immediately derive the expressions (V.9) for the Lie algebra in the zero-mass case.

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 4, NUMBER 6 JUNE 1963

Principle of General Q Covariance

D. FINKELSTEIN,*
Yeshiva University, New York, New York
J. M. JAUCH,
University of Geneva and *CERN, Geneva, Switzerland*
S. SCHIMINOVICH,†
Yeshiva University, New York, New York
AND
D. SPEISER,‡
University of Geneva, Geneva, Switzerland
(Received 10 December 1962)

In this paper the physical implications of quaternion quantum mechanics are further explored. In a quanternionic Hilbert space $\mathfrak{H}_Q$, the lattice of subspaces has a symmetry group which is isomorphic to the group of all co-unitary transformations in $\mathfrak{H}_Q$. In contrast to the complex space $\mathfrak{H}_C$ (ordinary Hilbert space), this group is connected, while for $\mathfrak{H}_C$ it consists of two disconnected pieces.

The subgroup of transformations in $\mathfrak{H}_Q$ which associates with every quaternion q of magnitude 1, the correspondence $\psi \to q\psi q^{-1}$ for all $\psi \in \mathfrak{H}_Q$ (called Q conjugations), is isomorphic to the three-dimensional rotation group. We postulate the principle of Q covariance: The physical laws are invariant under Q conjugations. The full significance of this postulate is brought to light in localizable systems where it can be generalized to the principle of general Q covariance: Physical laws are invariant under general Q conjugations. Under the latter we understand conjugation transformations which vary continuously from point to point.

The implementation of this principle forces us to construct a theory of parallel transport of quaternions. The notions of Q-covariant derivative and Q curvature are natural consequences thereof.

There is a further new structure built into the quaternionic frame through the equations of motion. These equations single out a purely imaginary quaternion $\eta(x)$ which may be a continuous function of the space–time coordinates. It corresponds to the i in the Schrödinger equation of ordinary quantum mechanics. We consider $\eta(x)$ as a fundamental field, much like the tensor $g_{\mu\nu}$ in the general theory of relativity. We give here a classical theory of this field by assuming the simplest invariant Lagrangian which can be constructed out of η and the covariant Q connection. It is shown that this theory describes three vector fields, two of them with mass and charge, and one massless and neutral. The latter is identifiable with the classical electromagnetic field.

1. INTRODUCTION

IN the development from Galilean to special to general relativity, it was shown by Einstein that the concepts of Euclidean geometry have only an approximate validity and that the true laws of geometry are subject to disturbances from place to place. Still more fundamental than the laws of geometry are those of classical logic as expressed in the propositional calculus. In the development from classical to quantum physics it was shown by Bohr that the concepts of classical logic have

* Supported by the National Science Foundation.
† Now at the University of Buenos Aires.
‡ Supported by the Swiss Commision for Atomic Research.

only an approximate validity due to complementarity. Indeed, as has been shown by Birkhoff and von Neumann[1] in a fundamental paper, the class calculus of classical mechanics is a Boolean algebra, while that of quantum mechanics is a nondistributive lattice.

Here we consider the next step: the true laws of logic may be subject to disturbances from place to place. A preliminary examination of such a local variation of the logical structure shows that there is no nontrivial possibility for such a generalization of ordinary quantum mechanics. It is too simple a structure.

A richer structure is obtained with quaternion quantum mechanics.

In order to illustrate the new features introduced by quaternion quantum mechanics, we consider side by side the three simplest possibilities of a propositional calculus for a quantum system. In such a calculus, the propositions are represented as subspaces (or projection operators) in an infinite-dimensional linear vector space $\mathfrak{H}_F$ with coefficients from a field F. If F is continuous then there exist only three possibilities for F; the reals (R), the complex numbers (C), and the quaternions (Q). If $F = C$, we obtain ordinary quantum mechanics. It can be shown[2] that $F = R$ is in contradiction with nature unless it is supplemented by a superselection rule which makes it essentially equivalent with $F = C$. There remains finally $F = Q$ or quaternion quantum mechanics (Q quantum mechanics).

The difference in the logical structure for the three cases is best shown if we consider the groups of symmetry transformations of the lattice of subspaces. Let E be a projection operator in $\mathfrak{H}_F$ and $M = E\mathfrak{H}_F$ the subspace consisting of all elements of the form $E\psi$ with $\psi \in \mathfrak{H}_F$. We say the projection E_1 is contained in E_2, and we write $E_1 < E_2$, if $M_1 = E_1\mathfrak{H}_F$ is contained in $M_2 = E_2\mathfrak{H}_F$. An alternate equivalent way of stating this is

$$E_1E_2 = E_1.$$

Consider now a permutation of the subspaces of $\mathfrak{H}_F$, that is a one-to-one correspondence $\phi(E)$ of the projections, such that

$$E_1 < E_2 \quad \text{implies} \quad \phi(E_1) < \phi(E_2).$$

Such a permutation is called a *symmetry transformation* of the lattice. The transformations of the Hilbert space that induce the symmetry transformations form the symmetry group g_F of the space. These groups are known. For $F = R$, it is the group O of all orthogonal transformations in $\mathfrak{H}_R$. For $F = C$, it is the direct product of the group U_C of all unitary transformations in $\mathfrak{H}_C$, with a cyclic group Z_2 of order two. Finally, for $F = Q$, it is the direct product of the group U_Q of all unitary transformations in $\mathfrak{H}_Q$, with the group R_3 of all rotations in three dimensions;

$$g_R = O, \qquad g_C = U_C \times Z_2, \qquad g_Q = U_Q \times R_3.$$

There is an important difference between the groups g_C and g_Q, which will be decisive for the content of this paper: g_C consists of two disconnected pieces while g_Q consists of one piece only. This difference is directly connected with the difference in the group of automorphisms for the number field F. [A one-to-one correspondence $\alpha \to \alpha'(\alpha, \alpha' \in F)$ is an automorphism of F if it is continuous and

$$(\alpha + \beta)' = \alpha' + \beta',$$

$$(\alpha\beta)' = \alpha'\beta',$$

for all $\alpha, \beta \in F$.] The following facts are well known: the only automorphism for R is the identity, for C the group of automorphisms is the cylic group of order two (the generator being in this case complex conjugation), while for Q it is the rotation group in three dimensions and it is implemented by

$$q \to q' = pqp^{-1} \quad \text{for all} \quad q, p \in Q \qquad (|p| = 1).$$

The advantage of a larger group of automorphisms of the quaternions is offset by the loss of commutativity. Quaternions do not commute. This leads to a major difficulty in Q quantum mechanics: Because of the noncommutativity of Q, a unique tensor product of $\mathfrak{H}_Q$ with itself does not exist. In ordinary quantum mechanics, this tensor product is necessary in order to formulate the quantum mechanics of composite systems.

This deficiency can be turned into an advantage leading to a new basic principle, if we introduce the postulate that the laws of physics should be invariant with respect to this nonuniqueness in its mathematical description. We have called this the *principle of general Q covariance.*

We shall show that this principle leads in a natural way to a new basic field which is a generalization of the Maxwell field and contains it as a special case. This field describes in addition to the neutral photons of mass zero, a pair of charged vector mesons with finite mass and a magnetic moment. These fields have a basic character in the same

[1] G. Birkhoff and J. von Neumann, Ann. Math. **37**, 823 (1936).
[2] E. C. G. Stueckelberg, Helv. Phys. Acta **33**, 727 (1960).

sense as the $g_{\mu\nu}$-fields in general relativity. They are directly connected with the intrinsic structure of the propositional calculus in its relation to the space–time continuum. We leave open, for the time being, how the fields of stable or unstable physical particles can be fitted into this scheme. We feel that the point of view presented here could be the clue to a deeper understanding of these questions.

2. GENERAL Q COVARIANCE AND THE ELECTROMAGNETIC AXIS

In a previous publication[3,4] we have developed a quantum mechanics with noncommutative c numbers (probability amplitudes). We have shown that the only reasonable choice for these numbers is (if not the real numbers R, or the complex numbers C) the quaternion field Q. More general hypercomplex number systems may be excluded.

The most characteristic property of the quaternions is their "roundness," which expresses itself mathematically in the existence of a group of automorphisms which is identical with the rotation group R_3 in three dimensions. Switching to a number field with more automorphisms seems a natural way of obtaining multiplets with more members. This was the original motivation for the study of quaternion quantum mechanics. Therefore we postulate that the laws of the new quantum mechanics should be invariant under automorphisms of the quaternion number field. This invariance principle we call Q *covariance*, and the invariance group R_3 we refer to as the Q group.

In the attempt to apply quaternion quantum mechanics to localizable systems we are led to introduce an even stronger hypothesis which we call *general Q covariance*. We understand by it invariance of the physical laws under automorphisms which vary continuously from point to point of space–time.

In order to formalize this principle, we recall some basic notions of quaternion quantum mechanics.[3,4] The pure states are represented as normalized state vectors from a quaternion Hilbert space $\mathfrak{H}_Q$. In such a space we have postulated two kinds of multiplications with scalars, one from the left and one from the right. This means if $\psi \in \mathfrak{H}_Q$ and $q \in Q$, there is defined an element ψq and an element $q\psi$. However while $\psi \to q\psi$ is a *linear* operator, $\psi \to \psi p$ is only *colinear*, defined below. These two operators commute:

$$(q\psi)p = q(\psi p).$$

In general, the elements $q\psi$ and ψq are different from one another. If they are equal for all q we call ψ *real*. One can show that the real elements form a real linear space $\mathfrak{H}_R$ in $\mathfrak{H}_Q$ which is complete in the sense that its orthogonal complement in $\mathfrak{H}_Q$ is zero: $\mathfrak{H}_R^{\perp} = 0$.

For every element $\psi \in \mathfrak{H}_Q$ and $q \in Q$, we can define the transformation

$$K(q)\colon \psi \to \psi^q \equiv q\psi q^{-1},$$

which we shall call *conjugation*. It leaves all the real elements in $\mathfrak{H}_Q$ invariant. This transformation induces a transformation of all operators defined on $\mathfrak{H}_Q$. If O is such an operator, the transformed operator O^q is defined by

$$O^q = qOq^{-1}.$$

We have previously introduced the notion of a *colinear operator*. An operator O is colinear if O is additive, and if for any $\psi \in \mathfrak{H}_Q$, $\alpha \in Q$, there exists a scalar α' such that

$$O(\psi\alpha) = (O\psi)\alpha'.$$

It can then be shown there exists a $q \in Q$ determined by O such that

$$\alpha' = \alpha^q = q\alpha q^{-1}.$$

The transformation $K(q)$ is colinear in this sense, and moreover, every colinear operator O has a unique representation as a product of a linear operator O_L with a $K(q)$:

$$O = K(q)O_L.$$

An operator O is said to be real if $O^q = O$ for all q. It has then the property of leaving the set of real elements invariant.

The observables are self-adjoint operators A. They are not necessarily real. If they represent yes–no experiments (questions), they are projections. Special Q covariance requires that all the physical laws are invariant under the conjugations $\psi \to \psi^q$. At this point we might compare the situation with the much simpler one in ordinary quantum mechanics. The transformation which corresponds to $\psi \to \psi^q$ is a complex conjugation in $\mathfrak{H}_C$. It assigns to every $\psi \in \mathfrak{H}_C$ a vector $\psi^c = \mathcal{J}\psi$. This transformation is antilinear:

$$\mathcal{J}\lambda\psi = \lambda^*\mathcal{J}\psi,$$

$$(\mathcal{J}\psi, \mathcal{J}\varphi) = (\varphi, \psi),$$

and involutory:

$$\mathcal{J}^2 = I.$$

[3] D. Finkelstein, J. M. Jauch, and D. Speiser, "Notes on Quaternion Quantum Mechanics," CERN Repts. I, II, and III, (1959).

[4] D. Finkelstein, J. M. Jauch, S. Schiminovich, and D. Speiser, J. Math. Phys. 3, 207 (1962).

The corresponding transformation of operators is given by

$$A^c = gAg^{-1}.$$

An operator is *real* if $A^c = A$.

The physical laws are invariant under this transformation in the sense that all the observable consequences of the theory are expressible in real numbers. This is true even though these transformations are not canonical. The commutation rules are, for instance, not invariant.

We may paraphrase the meaning of this invariance property: There is no physical distinction between complex numbers having the same intrinsic algebraic properties. Likewise we would require for quaternion quantum mechanics: There is no physical distinction between quaternions having the same intrinsic algebraic properties. In this case, this means all those algebraic properties which are invariant under the transformation $\alpha \to \alpha^q = q\alpha q^{-1}$.

The analogy to the ordinary quantum mechanics applies only for *special* Q covariance. We go now beyond this and formulate *general* Q covariance. This notion is related to that of localizability of a system. In order to avoid too long a discussion which would obscure our main point, we shall not insist here on a precise and rigorous definition of this notion, but we shall use instead a heuristic language guided by physical interpretations and analogies.

The physical properties of a system at a given point of space–time are described by a system of operators $O(x)$ representing the observables associated with this point. The observables at different space–time points commute if the points are situated spacelike (microcausality). We can consider the algebra $\mathfrak{A}(x)$ of bounded operators generated by the $O(x)$. If we denote by $\mathfrak{A}'(x)$ the commutator algebra of $\mathfrak{A}(x)$, it follows from what has been said so far that

$$\mathfrak{A}(y) \leq \mathfrak{A}'(x) \quad \text{if} \quad x - y \quad \text{is spacelike.}$$

The simplest way to realize such commuting sets of operators associated with spacelike-situated points is to assume that we are dealing with a family of Hilbert spaces $\mathfrak{H}(x)$, one for each x, and that the operators of $\mathfrak{A}(x)$ operate on $\mathfrak{H}(x)$ only and are unit operators in all $\mathfrak{H}(y)$ with $y \neq x$. In the manipulations of quantum field theory, one unites the family of Hilbert spaces $\mathfrak{H}(x)$ to a large Hilbert space $\mathfrak{H}$ which has the formal properties of a direct product of the spaces $\mathfrak{H}(x)$.

If one attempts a similar thing for a quaternion field theory, one runs straight into the ambiguity of the direct product of quaternionic Hilbert spaces. This state of affairs suggests that the principle of Q covariance should be extended to a principle of *general* Q covariance which requires that *it is physically meaningless to compare quaternions at different space–time points except in their intrinsic algebraic properties.* An arbitrary choice must be made which relates quaternions at neighboring space–time points. But the physical content of the theory is independent of that choice.

The formal expression of this principle would be the following. Let $O(x)$ be the system of local observables (the shaky foundation on which this notion rests does not deter us), and consider the transformation

$$O(x) \to O^q(x) \equiv q(x)O(x)q^{-1}(x), \qquad q(x) \in Q. \quad (1)$$

Simultaneously, every $\psi(x) \in \mathfrak{H}_Q(x)$ undergoes the transformation

$$\psi(x) \to \psi^q(x) \equiv q(x)\psi(x)q^{-1}(x). \quad (2)$$

The family of transformations of this kind make up the invariance group of general Q covariance; briefly the general Q group.

There are some simple consequences of such invariance requirements which appear already for special Q covariance.

Let us look for instance at the Schrödinger equation, which in Q quantum mechanics takes on the form

$$H\psi = \eta\psi, \quad (3)$$

where H, the Hamiltonian, is a nonnegative self-adjoint operator, and η is an imaginary unit quaternion operator which commutes with H;

$$\eta^2 = -1, \qquad [\eta, H] = 0.$$

If we carry out the transformations (1), (2) of the Q group, Eq. (3) transforms into

$$H^q\psi^q = \eta^q\psi^q.$$

Since in general $\eta^q \neq \eta$, we see that the Schrödinger equation is not invariant under the Q group unless we transform η as if it were a dynamical variable. The only feature that resembles this in ordinary quantum mechanics is the behavior of the Schrödinger equation under the time-reversal transformation. There the i in the Schrödinger equation changes its sign because time reversal is an anti-unitary transformation.

A local field theory which satisfies the principle of general Q covariance will exhibit even more

explicitly the dynamical character of η. The reason for this is that we must keep the formalism sufficiently general to permit a variation of η from point to point. Already in ordinary field theory it is possible to write the equations of motion in the form of Tomonaga. In this form, the state vector of the system is considered as a functional of a spacelike surface σ and the equation of motion which expresses a local variation of σ has the form

$$H(x)\psi[\sigma] = i\{\delta\psi[\sigma]/\delta\sigma(x)\},$$

where $H(x)$ is the Hamiltonian density. In Q quantum mechanics, the corresponding equation would be

$$H(x)\psi[\sigma] = \eta(x)\{\delta\psi[\sigma]/\delta\sigma(x)\},$$

where $\eta(x)^2 = -1$ and $[H(x), \eta(x)] = 0$.

Under general Q transformations, the operator η transforms just like any other dynamical variables according to

$$\eta(x) \rightarrow \eta(x)^q = q(x)\eta(x)q^{-1}(x), \qquad q(x) \in Q.$$

Evidently the Schrödinger equation is not invariant under general conjugations unless $q(x)$ commutes with $\eta(x)$.

What can we make of this apparent asymmetry of Q quantum mechanics under general conjugations? A clue might be obtained if we consider those conjugations which leave the $\eta(x)$ invariant. These conjugations form a certain subgroup of the group of all conjugations. They are generated by all those $q(x)$ which commute with $\eta(x)$. These transformations resemble gauge transformations of ordinary quantum field theory. In this analogy $\eta(x)$ takes the role of the isotopic spin axis.

It is well known that the requirement of the invariance of a field theory under local phase transformations leads in a natural way to a theory of the interaction of the electromagnetic field with a charged field. The above remarks indicate that the principle of general Q covariance contains the germ of a more general type of electromagnetic field theory, in which $\eta(x)$ appears as a fundamental dynamical variable.

The decision to raise the $\eta(x)$ to the status of a field variable is a step which resembles in many ways the treatment of the metric tensor $g_{\mu\nu}(x)$ in the theory of general relativity. There it is geometry which becomes a part of the dynamical structure, here it is the "logic" of propositions which is incorporated into the fundamental dynamical laws.

However there is a fundamental difference between the present attempt and general relativity which must be kept in mind: General relativity is a generalization of *classical* mechanics, Q quantum mechanics is a generalization of *quantum mechanics*.

The geometrical implications of the former are such that they lead to a *curved space* described by a Riemannian differential geometry. The logical implications of the latter are such that they lead to a logic of propositions which changes from point to point so that we may speak of a "*warped logic.*" In calling this the theory of warped logic, we have in mind that the lattices of subspaces of $\mathfrak{H}_Q(x)$ make up the class calculus of quantum mechanics because of the correspondence between subspaces and properties or classes.

3. Q CONNECTION AND Q CURVATURE

We now take the next step that is necessary for the construction of physical theories with quaternionic field operators subject to the principle of general Q covariance. The problem is the formation of derivatives of quaternionic field operators with respect to space–time coordinates. If $A(x)$ is such a field operator, we can introduce the space–time derivatives $\partial_\mu A(x)$. The definition of these quantities requires the comparison of the field $A(x)$ with a field at a neighboring point according to the usual interpretation

$$A(x + dx) = A(x) + dx^\mu\, \partial_\mu A(x) + \cdots .$$

In a general Q covariant theory, the derivative $\partial_\mu A(x)$ is not a suitable formation since it is by itself not invariant under general Q transformations. Indeed we find

$$\partial_\mu(q(x)A(x)q^{-1}(x)) = q(x)\,\partial_\mu A(x)q^{-1}(x) + [\partial_\mu q(x)q^{-1}(x),\ q(x)A(x)q^{-1}(x)].$$

What we are faced with is the problem of composite systems and the tensor product all over again. We must subtract an operator at x from an operator at $x + dx$. This makes sense only as the difference of two operators on the tensor product of the Q Hilbert spaces $\mathfrak{H}_Q(x)$ and $\mathfrak{H}_Q(x + dx)$. To define this product uniquely we have found that we must give an isomorphism between the quaternions at x and the quaternions at $x + dx$. If general Q covariance is not to be a trivial hypothesis, this isomorphism itself must be taken as one of the dynamical variables of the theory, since it is not invariant under the general Q group.

The most general quaternionic isomorphism that goes over smoothly into the identity when the two points x and $x + dx$ merge, is given by the first-order change in dx,

$$q(x + dx) = q(x) - \tfrac{1}{2}[C_\mu(x)\, dx^\mu,\ q(x)],$$

where $C_\mu(x)$ is a covariant space–time vector with pure imaginary Q components. In the analogy to general relativity, they correspond to the affine connection $\Gamma^\lambda_{k\mu}$. We call the $C_\mu(x)$ the Q *connection.* They may be taken as the potentials of a sort of generalized electromagnetic field. However, to guarantee general Q covariance, we shall mention the C_μ as seldom as possible, working always with the covariant combination D_μ defined by

$$D_\mu \cdot A(x) \equiv \partial_\mu A(x) + \tfrac{1}{2}[C_\mu(x), A(x)]. \tag{4}$$

The most general Q connection satisfies the following identities, valid for any pair of fields $A(x)$, $B(x)$:

$$\begin{aligned} D_\mu \cdot AB &= (D_\mu \cdot A)B + A(D_\mu \cdot B), \\ D_\mu \cdot (A + B) &= D_\mu \cdot A + D_\mu \cdot B, \\ (D_\mu \cdot A)^Q &= D_\mu \cdot A^Q, \end{aligned} \tag{5}$$

where the superscript Q is the operation of quaternion conjugation. Furthermore,

$$D_\mu x^\nu = \delta^\nu_\mu.$$

A point of notation: It is useful to be able to drop the operand $A(x)$ in the definition of D_μ given by (4). If A is any operator, we shall write ΔA for the operator on operators defined by

$$\Delta A \cdot B = [A, B].$$

The properties of this Δ operation have been studied elsewhere.[5] Here we observe that Jacobi's identity takes the form

$$\Delta[A, B] = [\Delta A, \Delta B].$$

We can now rewrite (4) as $D_\mu = \partial_\mu + \frac{1}{2}\Delta C_\mu$.

The invariant Q connection provides us with a transport of the quaternion number system from one point to a neighboring point. This transport can be extended to any point along a given curve. If $q = q(0)$ is a quaternion at a point $x = 0$, the quaternion at a point x transported along a curve Γ: $x = x(s)$ which corresponds to q would be given by integrating the differential equation

$$Dq = 0$$

along Γ, where $D = (dx^\mu/ds)D_\mu$. In general this transport of the quaternion from one point to another will depend on the curve Γ. If the transport is independent of Γ for given end points, we say that it is *integrable.* The necessary and sufficient condition for the integrability of the transport is the vanishing of the commutator $[D_\mu, D_\nu]$, which we shall call the Q *curvature.*

[5] D. Finkelstein, Comm. Pure Appl. Math. 8, 245 (1955).

A direct evaluation of this expression shows that it has the form

$$[D_\mu, D_\nu] = \tfrac{1}{2}\Delta K_{\mu\nu}, \qquad K^Q_{\mu\nu} = -K_{\mu\nu},$$

with

$$K_{\mu\nu} = \partial_\mu C_\nu - \partial_\nu C_\mu + \tfrac{1}{2}[C_\mu, C_\nu]. \tag{6}$$

The existence of the Q curvature expresses the fact that the transport of a quaternion from one point to another is only unique up to an automorphism, belonging to the same automorphism class as the initial amplitude. The particular member of the class depends on the path of transport and the physical conditions.

This possibility does not exist in C quantum mechanics, where the automorphisms of the number form a disconnected set consisting of just two elements. Thus an amplitude cannot vary continuously within its class.

Under the general Q group, the Q connection transforms according to

$$D_\mu \to D^q_\mu, \quad \text{where} \quad D^q_\mu A^q = (D_\mu A)^q.$$

Explicit evaluation for D^q_μ shows that it is of the form

$$D^q_\mu = \partial_\mu + \tfrac{1}{2}\Delta C'_\mu,$$

with

$$C'_\mu = qC_\mu q^{-1} - 2(\partial_\mu q)q^{-1}.$$

The second term shows that C_μ does not transform like a quaternion under the general Q group. This is in complete analogy with the affine connection $\Gamma^\lambda_{\mu\nu}$ which does not transform like a tensor in spite of its appearance.

Because of this transformation law, it is possible to transform the C_μ locally to zero by a suitable transformation. It suffices indeed to choose local values for q and $\partial_\mu q$ such that they satisfy

$$0 = qC_\mu - 2\partial_\mu q.$$

This condition can be extended to a finite region surrounding a given point if the partial differential equation

$$\partial_\mu q = \tfrac{1}{2}qC_\mu$$

is integrable. The necessary and sufficient condition for this is

$$\partial_\nu(qC_\mu) - \partial_\mu(qC_\nu) = 0,$$

which can be written in the equivalent form

$$K_{\nu\mu} \equiv \partial_\nu C_\mu - \partial_\mu C_\nu + \tfrac{1}{2}[C_\nu, C_\mu] = 0.$$

Thus we see the vanishing of the Q curvature is the necessary and sufficient condition for the existence of a coordinate system such that $D_\mu = \partial_\mu$.

4. FIELD EQUATIONS FOR THE FUNDAMENTAL FIELDS

In this section we discuss the dynamical properties of the fundamental fields η and D_μ. We treat these here as classical fields, leaving the quantization to a later treatment. Just as the quantization of affine connections in general relativity, the quantization of these fields raises fundamental difficulties which have not yet been overcome. Rather than dwell on these difficulties we shall extract as much as possible from the classical part of the theory in order to gain an insight into the possible interpretation of the formalism. Thus we assume that the field $\eta(x)$ is a pure, imaginary, quaternionic scalar in each of the Hilbert spaces $\mathfrak{H}(x)$.

In order to obtain a dynamical structure, that is, a set of field equations, we start with a Lagrangian which we assume in the simplest possible form. For the action density of the η field we assume $\frac{1}{2}(D^\mu \cdot \eta)(D_\mu \cdot \eta)$ and for that of the D field we assume $\frac{1}{4}K^{\mu\nu}K_{\mu\nu}$. We do not consider terms like η^2 since $\eta^2 = -1$.

Accordingly, the simplest action density for the full theory is a linear combination of the two:

$$L = -\frac{1}{4\alpha} K^{\mu\nu}K_{\mu\nu} - \frac{1}{2\beta}(D^\mu \cdot \eta)(D_\mu \cdot \eta) \qquad \alpha, \beta > 0. \tag{7}$$

We have chosen the signs in such a way that the "kinetic" terms in the energy (those quadratic in the time derivatives) make a positive-definite quadratic form.

The field equations which follow from this action density are

$$\frac{1}{\beta} D_\mu D^\mu \eta + \lambda\eta = 0, \tag{8}$$

and

$$D_\mu K^{\mu\nu} = -(\alpha/\beta)\eta D^\nu \eta. \tag{9}$$

Here $\lambda = \lambda(x)$ is the Lagrange multiplier associated with the subsidiary condition $\eta^2 = -1$. From the definition of $K_{\mu\nu}$ follows further the identity

$$D_\lambda K_{\mu\nu} + D_\mu K_{\nu\lambda} + D_\nu K_{\lambda\mu} = 0. \tag{10}$$

The steps which lead to these field equations are similar to those leading to the equations of general relativity. Here, as there, the field equations are obtained from the simplest invariants which can be formed out of the fundamental fields. But there is an important break with the analogy to general relativity. In Einstein's theory, an action for the combined system of the metric tensor and the covariant differential operator consists of just one term, the doubly-contracted curvature tensor R, and it involves just one physical constant. We need two terms and two constants α, β. The analogy broke because the metrical curvature describes the transport of one space–time direction in another space–time direction; but the Q curvature describes the transport of a quaternion in a space–time direction, and it makes no sense to contract one with the other. There is no analogue to R for us. From the action R follows by Palatini's identity that the covariant derivative of the metric tensor is zero. The corresponding equation for the η field would be $D_\mu \cdot \eta = 0$ and this equation does *not* follow from our action except in a limiting case to be discussed later.

5. ELECTROMAGNETIC FIELD AND MAXWELL'S EQUATIONS

The physical content of the Eqs. (8), (9), and (10) can best be extracted if we separate from the Q-covariant differential operator D_μ that part which does not change the η field. Any D_μ can be decomposed into

$$D_\mu = D^0_\mu + D^1_\mu, \tag{11}$$

where the first term is a derivative having the property

$$\text{and} \qquad D^0_\mu \cdot \eta(x) = 0, \qquad D^1_\mu = D_\mu - D^0_\mu. \tag{12}$$

In order to make this decomposition unique, we impose the further condition that D^1_μ, which can be shown to be a purely algebraic operator of the form

$$D^1_\mu = \tfrac{1}{2}\Delta B_\mu, \tag{13}$$

is associated with a B_μ that anticommutes with η:

$$\{B_\mu, \eta\} = 0. \tag{14}$$

Under this restriction the decomposition (11) is unique and Q invariant. Anticipating the physical interpretation, we call D^0_μ the "neutral" and D^1_μ the "charged" part of the Q connection D_μ.

In order to give an explicit expression for this decomposition, we lapse into noncovariant language and write

$$D_\mu = \partial_\mu + \tfrac{1}{2}\Delta C_\mu.$$

We observe the anticommuting properties of η, $\eta_\mu \equiv \partial_\mu\eta$, and $\eta\eta_\mu$ that derive from the relation $\eta^2 = -1$:

$$0 = \partial_\mu \eta^2 = (\partial_\mu \eta)\eta + \eta\, \partial_\mu \eta.$$

Let us define C_μ^0 by writing

$$D_\mu^0 = \partial_\mu + \tfrac{1}{2}\Delta C_\mu^0,$$

so that

$$C_\mu^0 + B_\mu = C_\mu.$$

The condition (12) together with (14) gives immediately the result

$$C_\mu^0 = A_\mu\eta - \eta\eta_\mu, \tag{15}$$

where

$$A_\mu \equiv -\tfrac{1}{2}\{C_\mu, \eta\}. \tag{16}$$

Along with the decomposition of D into two parts goes a decomposition of its self-commutator and thus of the Q curvature $K_{\mu\nu}$ as well. Direct calculation shows that

$$K_{\mu\nu} = F_{\mu\nu}\eta + B_{\mu\nu}, \tag{17}$$

where $F_{\mu\nu}$ is the real skew tensor defined by

$$[D_\mu^0, D_\nu^0] = \tfrac{1}{2}F_{\mu\nu}\Delta\eta, \tag{18}$$

and given explicitly by

$$F_{\mu\nu} = \partial_\mu A_\nu - \partial_\nu A_\mu + \tfrac{1}{2}[\eta_\mu, \eta_\nu]\eta. \tag{19}$$

The pure imaginary skew tensor $B_{\mu\nu}$ is defined by

$$B_{\mu\nu} = D_\mu^0 B_\nu - D_\nu^0 B_\mu + \tfrac{1}{2}[B_\mu, B_\nu]. \tag{20}$$

The tensor $F_{\mu\nu}$ identically satisfies the relations

$$\partial_\lambda F_{\mu\nu} + \partial_\mu F_{\nu\lambda} + \partial_\nu F_{\lambda\mu} = 0. \tag{21}$$

We obtain a further equation for the $F_{\mu\nu}$ field by extracting from Eq. (9), that part which commutes with $\eta(x)$. This part is obtained by taking the anticommutator of (9) with $\eta(x)$. The result is the following:

$$\partial^\mu F_{\mu\nu} = j_\nu, \tag{22}$$

with

$$j_\nu = \tfrac{1}{2}(B_\mu\eta B^\mu{}_\nu - B_{\mu\nu}\eta B^\mu - \partial^\mu\{\eta, B_{\mu\nu}\}). \tag{23}$$

Equations (21) and (22) are Maxwell's equations for an electromagnetic field with a source field given by (23).

This result justifies identifying the $F_{\lambda\nu}$ with the electromagnetic field. The gauge transformations of the electromagnetic potentials are a subgroup of the general Q transformations consisting of all those which leave the field η invariant. Such transformations can be written

$$A(x) \to A^q(x) = q(x)A(x)q^{-1}(x),$$

where

$$q(x) = e^{\frac{1}{2}\eta(x)\varphi(x)},$$

with $\varphi(x)$ a real but otherwise arbitrary space–time function. The transformed Q connection becomes

$$C_\mu^q(x) = q(x)C_\mu(x)q^{-1}(x) - 2q_\mu(x)q^{-1}(x),$$

where

$$q_\mu(x) \equiv \partial_\mu q(x) = \tfrac{1}{2}\eta_\mu(x)\varphi(x)q(x) + \tfrac{1}{2}\eta(x)\varphi_\mu(x)q(x).$$

From this, with the help of Eq. (16), we obtain the transformed potentials $A_\mu^q(x)$ in the form

$$A_\mu^q(x) = -\tfrac{1}{2}\{C_\mu^q(x), \eta(x)\}.$$

Because $q(x)$ commutes with $\eta(x)$, this becomes

$$A_\mu^q(x) = A_\mu(x) + \tfrac{1}{2}\{(\eta_\mu(x)\varphi(x) - \eta(x)\varphi_\mu(x)), \eta(x)\},$$

and because $\eta_\mu(x)$ anticommutes with $\eta(x)$, the last term simplifies to $\varphi_\mu(x)$. There results the transformation

$$A_\mu(x) \to A_\mu^q(x) = A_\mu(x) + \partial_\mu\varphi(x),$$

which we recognize as the usual gauge transformation of electromagnetic theory. We have thus recovered the entire formalism of the classical Maxwell field, but with a special current given by the expression (23).

We now turn our attention to the "charged" part of the Q-covariant differential operator.

6. HEAVY PHOTONS

The vector field B_μ satisfies an equation of motion which is obtained from (9) by taking the quaternionic component of (9) that is "orthogonal" to (i.e., anticommutes with) the electromagnetic axis η. We therefore take the commutator of (9) with η, for this is the quaternionic transcription of the usual vector product in three dimensions; the result of the calculation is the equation

$$D_\mu^0 D^{0\mu}B^\nu - 2\eta B_\mu F^{\mu\nu} + (\alpha/\beta)B^\nu + \tfrac{1}{4}[B_\mu, [B^\mu, B^\nu]] = 0. \tag{24}$$

In addition one obtains the equation

$$D_\mu^0 \cdot B^\mu = 0. \tag{25}$$

These equations describe a field with a rest mass

$$M_B = (\alpha/\beta)^{\frac{1}{2}}.$$

The field is electrically charged. This is shown by the wave equation (24) which contains the electromagnetic coupling in D_μ^0 in addition to an anomalous magnetic moment coupling [the second term in (24)]. The charge of the field is

$$e = \alpha^{1/2}.$$

Thus we see that the same Q connection contains the makings of both neutral massless particles and charged massive particles in a general Q-covariant way. Moreover, the mass is a necessary consequence of the dynamics of the electromagnetic axis η. It can be interpreted as due to the reaction back on the charged part B of the Q connection, resulting from the action of B on the electromagnetic axis. Both effects arise from the same term in the action density.

The analysis of the resulting field equations shows that they describe a pair of charged vector bosons with a finite rest mass together with the interacting electromagnetic field. The equations for the Maxwell field are exact and can be derived without approximations as a special case of the general field equations. But the equations for the vector bosons are nonlinear and their usual interpretation is only valid in a linear approximation.

7. THE ELECTROMAGNETIC LIMIT

Just as in general relativity where the limit of vanishing curvature is important in understanding the success of special relativity, in Q quantum mechanics the limit of vanishing Q curvature $K_{\mu\nu} = 0$ is important. We call this the case *Q flat*. In the Q-flat limit, the neutral and charged photons vanish.

There are two interesting intermediate cases between the general case and the flat limit. The first is characterized by $D_\mu \cdot \eta = 0$, the second by $\Delta K_{\mu\nu} \cdot \eta = 0$. The first of these two conditions implies the second, since from $D_\mu \cdot \eta = 0$ follows

$$0 = [D_\mu, D_\nu] \cdot \eta = \Delta K_{\mu\nu} \cdot \eta.$$

But the second condition is weaker. Indeed from the first we obtain immediately

$$0 = D_\mu \cdot \eta = D^0_\mu \cdot \eta + \tfrac{1}{2}[B_\mu, \eta] = B_\mu \eta.$$

Since $\eta \neq 0$, it follows that $B_\mu = 0$, and consequently also $K_{\mu\nu} = F_{\mu\nu}\eta$. In this case the charged field alone vanishes. We could therefore call the case $D_\mu \cdot \eta = 0$ the *electromagnetic limit* of the theory.

[In case $\Delta K_{\mu\nu} \cdot \eta = 0$, we can only conclude

$$D^0_\mu B_\nu - D^0_\nu B_\mu = 0,$$

from which follows, with the help of (24) and (25),

$$(\alpha/\beta)B^\nu + \tfrac{1}{4}[B_\mu, [B^\mu, B^\nu]] - F^{\mu\nu}\eta B_\mu = 0,$$

as an algebraic relation. The physical meaning of this case is obscure.]

The electromagnetic limit characterized by $D_\mu \cdot \eta = 0$ can be interpreted more easily if we choose a Q frame in which $\eta(x)$ becomes a constant quaternion, say $\eta(x) = i_3$. Such a choice is always possible, for instance, by the explicit transformation

$$\eta \to \eta^q \equiv q\eta q^{-1},$$

with

$$q = \exp\{\tfrac{1}{2}(\cos^{-1}\eta_3)[\eta, i_3]/2(\eta_1^2 + \eta_2^2)^{\frac{1}{2}}\}$$
$$= \cos(\tfrac{1}{2}\cos^{-1}\eta_3) + \{[\eta, i_3]/2(\eta_4^2 + \eta_2^2)^{\frac{1}{2}}\}\sin(\tfrac{1}{2}\cos^{-1}\eta_3).$$

In this frame

$$D_\mu = \partial_\mu + \tfrac{1}{2}A_\mu\Delta\eta = D^0_\mu,$$
$$\eta_\mu = 0,$$
$$F_{\mu\nu} = \partial_\mu A_\nu - \partial_\nu A_\mu.$$

Since the B field is identically zero, the quaternion components in the directions i_1 and i_2 never appear. We can therefore set $\eta \equiv i_3 = i$ and identify it with the ordinary $(-1)^{\frac{1}{2}}$. The resulting field equations are identical with the classical equations of the source-free electromagnetic field.

ACKNOWLEDGMENTS

We are indebted to F. Gürsey for stimulating discussions. C. N. Yang contributed much constructive criticism, as well as other kinds. So also have Y. Aharanov, R. Behrends, L. Landowitz, and B. Zumino. The financial support of the U.S. National Science Foundation and the Swiss Commission for Atomic Science is gratefully acknowledged.

Reduction of S-Matrix Elements*

A. SYED
Department of Theoretical Physics, Imperial College, London, England

We have obtained expressions for the commutators of in- and out-fields with the generalized retarded operators of Burgoyne and Ruelle. These are then used to show that the matrix element

$$\langle p_1, \cdots, p_m \text{ out}/-p_{m+1}, \cdots, -p_n \text{ in}\rangle,$$

where $p \equiv (p_1, p_2, \cdots, p_n)$ is arbitrary, can be reduced in terms of retarded functions without using the assumption of stability of one-particle states.

1. INTRODUCTION

THE LSZ formulation of field theory[1,2] starts with the Heisenberg field $Q(x)$ (for neutral scalar field) and studies the consequences of its general properties, such as relativistic invariance locality, asymptotic conditions etc. It is well known that using asymptotic conditions one can derive certain reduction formulas for the time-ordered, retarded, and the advanced products of the field operator $Q(x)$. They are given by

$$T(x_1, x_2, \cdots, x_n)Q^{a*}_{\text{in}} - Q^{a*}_{\text{out}}T(x_1, x_2, \cdots, x_n) = i\int dz f_a(z)K_z T(x_1, x_2, \cdots, x_n, z), \quad (1)$$

$$[R(x; x_1, x_2, \cdots, x_n), Q^{a*}_{\text{in}}] = -i\int dz f_a(z)K_z R(x; x_1, \cdots, x_n, z), \quad (2)$$

$$[A(x; x_1, x_2, \cdots, x_n), Q^{a*}_{\text{out}}] = -i\int dz f_a(z)K_z A(x; x_1, \cdots, x_n, z), \quad (3)$$

where $\{f_a(z)\}$ is a complete set of c-number solution of Klein–Gordon equation, K_z is the Klein–Gordon operator corresponding to the variable z,

$$K_z = \Box_z + m^2 \equiv \frac{\partial^2}{\partial z_0^2} - \frac{\partial^2}{\partial z_1^2} - \frac{\partial^2}{\partial z_2^2} - \frac{\partial^2}{\partial z_3^2} + m^2,$$

and Q^a_{in}, Q^a_{out} are defined by

$$Q^{a*}_{\text{in,out}} = -i\int d^3x Q_{\text{in,out}}(x)\overleftrightarrow{\partial}_{x^0} f_a(x). \quad (4)$$

While the first of these formulas can be used to reduce the matrix element

$$\langle p_1, p_2, \cdots, p_m \text{ out} \mid k_1, k_2 \cdots, k_n \text{ in}\rangle \quad (5)$$

to what is essentially the vacuum expectation value of a time-ordered product,[3] it has been pointed out by LSZ[4] and Polkinghorne[5] that Eq. (2) [Eq. (3)] is not sufficient to reduce (5) to the vacuum expectation value of a retarded product (advanced product) if $m > 2(n > 2)$. However, Ruelle[6] and Burgoyne[7] have shown that the equation

$$\tilde{T}(p_1, p_2, \cdots, p_n) = \tilde{R}_\alpha(p_1, p_2, \cdots, p_n),$$

where R_α is a generalized retarded operator and the tilde denotes the Fourier transform, holds for certain region (depending on α) of the momentum space. We may, therefore, expect to be able to reduce (5) to retarded functions if we consider the generalized retarded functions also in addition to the ordinary ones. This reduction is desirable as the retarded functions are more useful than the time-ordered functions due to the fact that they can be continued analytically in the momentum space.

The difficulty which arises when one tries to reduce (5) to retarded functions is that one requires an expression for the commutator

$$[R(x; x_1, x_2, \cdots, x_n), Q^{a*}_{\text{out}}]$$

similar to the expression (2) for

$$[R(x; x_1, x_2, \cdots, x_n), Q^{a*}_{\text{in}}].$$

As we see later, this turns out to be one of the generalized retarded operators, so if we want to proceed one step further in the process of reduction of (5), we need the commutator of this generalized operator with Q^{a*}_{in} and Q^{a*}_{out}. Further reductions will require commutators of other generalized operators

* The research reported in this document has been sponsored in part by Air Force Office of Scientific Research, OAR, through the European Office, Aerospace Research, United States Air Force.

[1] H. Lehmann, K. Symanzik, and W. Zimmermann, Nuovo Cimento **1**, 425 (1955).

[2] H. Lehmann, K. Symanzik, and W. Zimmermann, Nuovo Cimento **6**, 319 (1957).

[3] See reference 1.

[4] See reference 2.

[5] J. C. Polkinghorne, Proc. Roy. Soc. (London) **A247**, 557 (1958).

[6] D. Ruelle, Nuovo Cimento **19**, 356 (1961).

[7] N. Burgoyne, Ph.D. thesis, Princeton University, 1961 (unpublished).

with the in- and out-fields and therefore, in the following, we study the commutator of an arbitrary generalized operator with $Q^{a^*}_{\text{in,out}}$.

First of all, let us give a precise definition of these operators.[8] If we consider the $(n-1)$-dimensional space Q^0 of $q^0 \equiv (q_1^0, q_2^0, \cdots, q_n^0)$ with

$$\sum_{i=1}^{n} q_i^0 = 0,$$

the $(2^{n-1}-1)$ distinct hyperplanes

$$q_J^0 = 0,$$

(with J any subset of $\{1, 2, \cdots, n\}$ and $q_J = \sum_{i\in J} q_i$) will divide it into a set of convex polyhedral cones, say C_α^0, each radiating from origin. It is clear that each cone C_α^0 determines uniquely whether $q_J^0 > 0$ or $-q_J^0 > 0$ for $q^0 \in C_\alpha^0$ and that there is at least one J, say J', such that $q_{J'}^0$ has different signs in C_α^0 and C_β^0 for $\alpha \neq \beta$. Thus there is a one-to-one correspondence between the cones C_α^0 and a consistent choice of sign of q_J^0 for all J.

Consider next the space Q of the 4-vectors

$$q \equiv (q_1, q_2, \cdots, q_n), \quad \text{with} \quad \sum_{i=1}^{n} q_i = 0.$$

We now define the convex cones C_α in this space by the condition that $q \in C_\alpha$ if

$$\begin{aligned} &\text{(i)}\ q_J \in V_+ \text{ for all } J \text{ such that} \\ &\quad q_J^0 > 0 \text{ when } q^0 \in C_\alpha^0; \\ &\text{(ii)}\ q_J \in V_- \text{ for all } J \text{ such that} \\ &\quad q_J^0 < 0 \text{ when } q^0 \in C_\alpha^0. \end{aligned} \tag{6}$$

In other words, $q \in C_\alpha$ if $q_J \in V_\pm$ for all J, and the subscript of V, for each J, is such that $q^0 \in C_\alpha^0$.

Corresponding to each of these cones C_α, Burgoyne now defines a generalized retarded operator by

$$R_\alpha(x_1, x_2, \cdots, x_n) = \sum_{\pi} \pm \theta_\pm(i_1, i_2) \cdots \times \theta_\pm(i_{n-1}, i_n) W(i_1, i_2, \cdots, i_n), \tag{7}$$

where

(i) π is the permutation $k \to i_k$ and

$$W(i_1, i_2, \cdots, i_n) = Q(x_{i_1})Q(x_{i_2}) \cdots Q(x_{i_n});$$

(ii) the sum is over all the permutations of $(1, 2, 3, \cdots, n)$;

(iii) the choice of sign in $\theta_\pm(i_k, i_{k+1}) \equiv \theta[\pm(x^0_{i_k} - x^0_{i_{k+1}})]$ is plus if $q_K \in V_+$ when $q \in C_\alpha$ and minus if $q_K \in V_-$ when $q \in C_\alpha$, K being the set $\{i_1, i_2, \cdots, i_k\}$;

(iv) the over-all sign of each term in the summation is plus or minus according to whether the term contains an even or an odd number of θ-minus functions.

[8] This is the definition given by Burgoyne in reference 7.

These rules define each R_α uniquely. Burgoyne[7] and Araki[9] have shown that its vacuum expectation value

$$r_\alpha(x_1, x_2, \cdots, x_n) = \langle 0 \,|R_\alpha(x_1, x_2, \cdots, x_n)|\, 0\rangle$$

is a Lorentz-invariant distribution vanishing outside the cone $\tilde{C}_\alpha^0$ (in x^0 space) which is conjugate of the cone C_α^0 (in q^0 space), i.e., it is the set of all x such that

$$x_1^0 q_1^0 + x_2^0 q_2^0 + \cdots + x_n^0 q_n^0 > 0$$

for all $q \in C_\alpha$. Hence, the Fourier transform

$$\tilde{r}_\alpha(p_1, p_2, \cdots, p_n)$$

will be the boundary value (as $q \to 0$) of a function

$$\tilde{r}_\alpha(s_1, s_2, \cdots, s_n) \qquad s \equiv p + iq$$

analytic for arbitrary p and q such that

$$q \equiv (q^0, 0), \qquad q^0 \in C_\alpha^0.$$

It can be easily seen that the ordinary retarded and advanced operators correspond to the cones

$$R(x_1; x_2, x_3 \cdots, x_n) \to C_1^-,$$

$$A(x_1; x_2, \cdots, x_n) \to C_1^+,$$

where C_1^- and C_1^+ are determined by

$$C_1^-\colon q_2, q_3, \cdots, q_n \in V_-,$$

$$C_1^+\colon q_2, q_3, \cdots, q_n \in V_+.$$

2. REDUCTION FORMULAS

Consider now the commutator $[R_\alpha, Q^{a^*}_{\text{in}}]$. We have

$$[R_\alpha(x_1, x_2, \cdots, x_n), Q^{a^*}_{\text{in}}] = \lim_{z^0 \to -\infty} -i \int d^3z [R_\alpha(x_1, x_2, \cdots, x_n), Q(z)] \overleftrightarrow{\partial_{z^0}} f_a(z),$$

in the sense of weak convergence. Here we have used the definition (4) for $Q^{a^*}_{\text{in}}$ and the asymptotic condition. Let us now assume that there exists a generalized retarded operator

$$R_{\alpha^-}(x_1, x_2, \cdots, x_n, z)$$

such that

$$\lim_{z^0 \to -\infty} [R_\alpha(x_1, x_2, \cdots, x_n), Q(z)] = \lim_{z^0 \to -\infty} R_{\alpha^-}(x_1, x_2, \cdots, x_n, z). \tag{8}$$

In Sec. 3, we will give a method which explicitly determines R_{α^-} when R_α is given. Using (8) we

[9] H. Araki, J. Math. Phys. 2, 163 (1961).

will get

$$[R_\alpha(x_1, x_2, \cdots, x_n), Q_{\text{in}}^{\alpha*}],$$

$$= \lim_{z^0 \to -\infty} -i \int d^3z R_\alpha\text{-}(x_1, x_2, \cdots, x_n, z) \overleftrightarrow{\partial_{z^0}} f_\alpha(z)$$

$$= i \int dz\, \partial_{z^0}[R_\alpha\text{-}(x_1, x_2, \cdots, x_n, z) \overleftrightarrow{\partial_{z^0}} f_\alpha(z)],$$

as it will be shown later that

$$\lim_{z^0 \to +\infty} R_\alpha\text{-}(x_1, x_2, \cdots, x_n, z) = 0. \tag{9}$$

Using now the fact that $f_\alpha(z)$ satisfies the K–G equation and performing integration by parts over spatial variables we will get the required formula for the commutator as

$$[R_\alpha(x_1, x_2, \cdots, x_n), Q_{\text{in}}^{\alpha*}]$$

$$= -i \int dz f_\alpha(z) K_z R_\alpha\text{-}(x_1, x_2, \cdots, x_n, z), \tag{10}$$

where $R_\alpha\text{-}$ is obtained from R_α according to (8).

By a similar method, we can show that

$$[R_\alpha(x_1, x_2, \cdots, x_n), Q_{\text{out}}^{\alpha*}]$$

$$= -i \int dz f_\alpha(z) R_{\alpha+}(x_1, x_2, \cdots, x_n, z), \tag{11}$$

where $R_{\alpha+}$ is the generalized retarded operator given by

$$\lim_{z^0 \to +\infty} [R_\alpha(x_1, x_2, \cdots, x_n), Q(z)]$$

$$= -\lim_{z^0 \to +\infty} R_{\alpha+}(x_1, x_2, \cdots, x_n, z). \tag{12}$$

3. DETERMINATIONS OF R_α^*

Let us suppose that

$$R_\alpha(x_1, x_2, \cdots, x_n)$$

is a given arbitrary generalized retarded operator, i.e., we know whether $q_J \in V_+$ or V_- when $q \in C_\alpha$, for all subsets J of the set

$$\{1, 2, 3, \cdots, n\}.$$

We now define the cone $C_\alpha\text{-}$ by the conditions

(i) $q_J \in V_+$ in $C_\alpha\text{-}$ for all J such that $q_J \in V_+$ in C_α;

(ii) $q_J \in V_-$ in $C_\alpha\text{-}$ for all J such that $q_J \in V_-$ in C_α;

(iii) $q_{n+1} \in V_-$ and hence $q_1 + q_2 + \cdots + q_n \in V_+$ in $C_\alpha\text{-}$. (13)

If J is such that $q_J \in V_-$ in C_α ,then according to (ii), $q_J \in V_-$ in $C_\alpha\text{-}$ also, so that $q_{n+1} + q_J \in V_-$ in $C_\alpha\text{-}$. If, on the other hand, it is such that $q_J \in V_+$ in C_α, then $q_{\bar{J}} \in V_-$ in C_α and hence, by (ii) again, $q_{\bar{J}} \in V_-$ in $C_\alpha\text{-}$ also, so that $q_J + q_{n+1} \in V_+$ in $C_\alpha\text{-}$. Thus the above conditions determine whether $q_J + q_{n+1}$ is in V_+ or in V_-, i.e., (12) defines the cone $C_\alpha\text{-}$ completely and uniquely.

We now show that the operator

$$R_\alpha\text{-}(x_1, x_2, \cdots, x_n, x_{n+1})$$

corresponding to the cone $C_\alpha\text{-}$ satisfies (8). From the definition of generalized retarded operators, we have

$$\lim_{x^0_{n+1} \to -\infty} R_\alpha\text{-}(x_1, x_2, \cdots, x_n, x_{n+1})$$

$$= \lim_{x^0_{n+1} \to -\infty} \sum_{\pi_{n+1}} \pm \theta_*(i_1 i_2) \cdots$$

$$\times \theta_*(i_n i_{n+1}) W(i_1, i_2, \cdots, i_n, i_{n+1})$$

$$= \lim_{x^0_{n+1} \to -\infty} \sum_{l=0}^{n} \sum_{\pi_n} \pm \theta_*(i_1, i_2) \cdots$$

$$\times \theta_*(i_l, n+1) \theta_*(n+1, i_{l+1}) \cdots \theta_*(i_{n-1}, i_n)$$

$$\times W(i_1, \cdots, i_l, n+1, i_{l+1}, \cdots, i_n). \tag{14}$$

Consider now the product

$$\theta_*(i_1, i_2) \cdots [\theta_*(i_l, n+1) \theta_*(n+1, i_{s+1})] \cdots$$

$$\times \theta_*(i_{n-1}, i_n). \tag{15}$$

The subscript of each of the θ functions is given by the rule (iii) in the definition of the retarded operator (7). As $x^0_{n+1} \to -\infty$, the product of the two θ functions in the square brackets, and hence the product (15) itself, will be zero unless the subscript of the first θ function is $(+)$ and that of the second is $(-)$. This will be true only when

$$q_L \in V_+,\ q_L + q_{n+1} \in V_- \text{ in } C_\alpha\text{-}$$

$$\text{where } L \equiv \{i_1, i_2, \cdots, i_l\}.$$

On the other hand the conditions (13) defining $C_\alpha\text{-}$ show that this is not true for $1 \leq l \leq n-1$. Hence, in the summation over l in (14), the only terms different from zero will be the ones corresponding to $l = 0$ and $l = n$. In the $l = 0$ case, (15) will become

$$\theta_-(n+1, i_1) \theta_*(i_1, i_2) \cdots \theta_*(i_{n-1}, i_n)$$

$$= -\theta_*(i_1, i_2) \cdots \theta_*(i_{n-1}, i_n),$$

as $q_{n+1} \in V_-$ in $C_\alpha\text{-}$ and $\theta_-(n+1, i_1) = 1$ in the limit $x^0_{n+1} \to -\infty$. The minus sign appears because the left-hand side has one θ-minus function more than the right-hand side. Thus, the term in (14)

corresponding to $l = 0$ will be

$$\lim_{x^0_{n+1}\to-\infty} -\sum_{\pi_n} \pm\theta_-(i_1, i_2)\cdots \times \theta_-(i_{n-1}, i_n)Q(x_{n+1})W(i_1, i_2, \cdots, i_n) = -\lim_{x^0_{n+1}\to-\infty} Q(x_{n+1})R_\alpha(x_1, x_2, \cdots, x_n), \quad (16a)$$

as for any J, q_J is in V_+ or in V_- in C_α- according as it is in V_+ or in V_- in C_α.

Similarly, when $l = n$, (15) will become

$$\theta_-(i_1, i_2)\cdots\theta_-(i_{n-1}, i_n)\theta_+(i_n, n+1) = \theta_-(i_1, i_2)\cdots\theta_-(i_{n-1}, i_n),$$

as $q_1+q_2+\cdots+q_n\in V_+$ in C_α- and $\theta_+(i_n, n+1)=1$ in the limit $x^0_{n+1}\to-\infty$. This time no minus sign appears because the left-hand side and the right-hand side have the same number of θ-minus functions. The term in (14) resulting from $l = n$ is, therefore,

$$\lim_{x^0_{n+1}\to-\infty} \sum_{\pi_n} \pm\theta_-(i_1, i_2)\cdots \times \theta_-(i_{n-1}, i_n)W(i_1, i_2, \cdots, i_n)Q(x_{n+1}) = \lim_{x^0_{n+1}\to-\infty} R_\alpha(x_1, x_2, \cdots, x_n)Q(x_{n+1}). \quad (16b)$$

Combining (16a) and (16b) we get

$$\lim_{x^0_{n+1}\to-\infty} R_\alpha\text{-}(x_1, x_2, \cdots, x_n, x_{n+1}) = \lim_{x^0_{n+1}\to-\infty} -Q(x_{n+1})R_\alpha(x_1, \cdots, x_n) + R_\alpha(x_1, \cdots, x_n)Q(x_{n+1}),$$

which shows that R_α-, defined by (13), satisfies (8).

In an exactly similar manner, we can show that the operator $R_{\alpha+}$, satisfying (12), is the generalized retarded operator corresponding to the cone $C_{\alpha+}$ defined by

(i) $q_J \in V_+$ in $C_{\alpha+}$ for all J such that $q_J \in V_+$ in C_α;
(ii) $q_J \in V_-$ in $C_{\alpha+}$ for all J such that $q_J \in V_-$ in C_α;
(iii) $q_{n+1} \in V_+$ and hence $q_1 + \cdots + q_n \in V_-$ in $C_{\alpha+}$. (17)

We now verify Eq. (9). We have

$$\lim_{x^0_{n+1}\to\infty} R_\alpha\text{-}(x_1, x_2, \cdots, x_n, x_{n+1}) = \lim_{x^0_{n+1}\to\infty} \sum_{l=0}^{n} \sum_{\pi_n} \pm\theta_+(i_1, i_2)\cdots \times \theta_+(i_l, n+1)\theta_+(n+1, i_{l+1})\cdots\theta_+(i_{n-1}, i_n) \times W(i_1, \cdots, i_l, n+1, i_{l+1}, \cdots, i_n). \quad (18)$$

Consider the product (15). As $x^0_{n+1} \to +\infty$, the product of the two θ functions in the square brackets will be different from zero only when

$$q_L \in V_- \quad \text{and} \quad q_L + q_{n+1} \in V_+, \quad \text{where } L \equiv \{i_1, i_2, \cdots, i_l\}.$$

This is impossible for $1 \le l \le n-1$, as q_L and $q_L + q_{n+1}$ are either both in V_+ or both in V_- in C_α-. When $l = 0$ and $l = n$, (15) will be nonzero only if $q_{n+1} \in V_+$ and $q_1 + \cdots + q_n \in V_-$, respectively, and this is again not the case, as $q_{n+1} \in V_-$ in C_α-. Hence, all the terms in the summation over in (18) vanish, so that (9) will be automatically satisfied.

To illustrate the results already obtained, let us evaluate the commutators

$$[R_0(x_0; x_1, x_2, \cdots, x_n), Q^{a*}_{\text{in}}],$$

and

$$[R_0(x_0; x_1, x_2, \cdots, x_n), Q^{a*}_{\text{out}}],$$

where R_0 is the ordinary retarded operator. The cone C_0 is defined by

$$C_0: \quad q_J \in V_-, \quad J \neq 0, \qquad q_0 + q_J \in V_+, \quad J \neq \{1, 2, \cdots, n\},$$

where J is any subset of $\{1, 2, \cdots, n\}$. Hence, if

$$[R_0(x_0; x_1, x_2, \cdots, x_n), Q^{a*}_{\text{in}}] = -iM^a_{n+1}R_0\text{-}(x_0; x_1, x_2, \cdots, x_n, x_{n+1}), \quad (2a)$$

where M is defined as the operator

$$M^a_i = \int dx_i f_a(x_i)K_{x_i}, \quad (19)$$

the cone C_0- is defined, according to (13) by

$$q_{n+1}, \quad q_J \in V_- \quad J \neq 0, \qquad q_0 + q_J, \quad q_0 + q_J + q_{n+1} \in V_+,$$

i.e., by

$$C_{0^-}: \quad q_I \in V_- \quad I \neq 0, \qquad q_0 + q_I \in V_+ \quad I \neq \{1, 2, \cdots, n, n+1\},$$

where I is any subset of $\{1, 2, \cdots, n, n+1\}$. This shows that R_0- is the ordinary retarded operator so that (2a) is identical with (2) as it should.

Similarly, if

$$[R_0(x_0; x_1, x_2, \cdots, x_n), Q^{a*}_{\text{out}}] = -iM^a_{n+1}R_{0+}(x_0; x_1, x_2, \cdots, x_n, x_{n+1}),$$

the cone C_{0+} will be defined by

$$q_J,\ q_J + q_{n+1} \in V_- \qquad J \neq 0,$$

$$q_{n+1},\ q_J + q_0 \in V_+ \qquad J \neq \{1, 2, \cdots, n\},$$

i.e., by

$$q_I \in V_-, \qquad I \neq 0$$

$$q_{n+1},\ q_0 + q_I \in V_+, \qquad I \neq \{1, 2, \cdots, n\},$$

$$\{1, 2, \cdots, n, n+1\}.$$

This shows that for $n \geq 2$, C_{0+} is the cone[10] $C^-_{0\,n+1}$ which is defined by

$$C^-_{0\,n+1}\colon -q_{n+1},\ q_1,\ q_2,\ \cdots,\ q_n \in V_-,$$

and to which corresponds the generalized retarded operator

$$R_{0\,n+1}(x_0; x_1, \cdots, x_n, x_{n+1}).$$

For $n = 1$, C_{0+} will be identical to the cone

$$C^+_1\colon q_0,\ q_2 \in V_+,$$

to which corresponds the advanced function

$$A_1(x_0, x_1, x_2).$$

Thus we will have

$$[R_0(x_0; x_1), Q^{a^*}_{\text{out}}] = -iM^a_2 A_1(x_0, x_1, x_2),$$

$$[R_0(x_0; x_1, \cdots x_n), Q^{a^*}_{\text{out}}] = -iM^a_{n+1} R_{0\,n+1}(x_0, x_1, \cdots x_n, x_{n+1}) \quad \text{for } n \geq 2. \tag{20}$$

4. CONNECTION WITH STEINMANN RETARDED OPERATOR

In this section, we study some of the properties of the retarded operators of the type $R_{\alpha^\pm}$ defined by (8) and (12). Let us start by introducing some new notations. If C_{α_n} is an arbitrary n-point cone, we can obtain from it, two $(n+1)$-point cones $C_{\alpha_{n+1}}$ and $C_{\alpha'_{n+1}}$ which are defined by

$C_{\alpha_{n+1}}\colon q_J \in V_\pm$ in $C_{\alpha_{n+1}}$

according as $q_J \in V_\pm$ in C_{α_n}, $q_{n+1} \in V_+$,

$C_{\alpha'_{n+1}}\colon q_J \in V_\pm$ in $C_{\alpha'_{n+1}}$

according as $q_J \in V_\pm$ in C_{α_n}, $q_{n+1} \in V_-$,

$J \equiv$ any proper subset of $\{1, 2, \cdots, n\}$.

We denote these cones by

$$C\{\alpha, +(n+1)\} \quad \text{and} \quad C\{\alpha, -(n+1)\}.$$

[10] For particular cones and corresponding retarded functions, we use the notation of Burgoyne given in H. Araki and N. Burgoyne, Nuovo Cimento, 8, 342 (1960).

If C_{α_n} is itself obtained from an $(n-1)$-point cone, i.e., if we have

$$C_{\alpha_n} = C\{\alpha_{n-1}, \pm n\},$$

where the prefix $\pm$ means either $(+)$ or $(-)$, and $C_{\alpha_{n-1}}$ an $(n-1)$-point cone, we will write

$$C_{\alpha_{n+1}} = C\{\alpha_{n-1}, \pm n, (+n+1)\}.$$

If we continue this process of expressing C_{α_n} in terms of C_{α_i}, then

either (i) it will terminate at a stage $i > 1$, i.e., we will have

$$C_{\alpha_{n+1}} = C\{\alpha_i, \pm(i+1), \cdots, \pm n, +(n+1)\},$$

where each of the prefixes is either $(+)$ or $(-)$ and C_{α_i} is an i-point cone such that it cannot be obtained from any of the $(i-1)$-point cones. As an example, C_{α_i} may be the 5-point cone C'^-_{ij} of reference 10 which cannot be obtained from any of the 4-point cones;

or (ii) it will go down to $i = 1$. In this case we have

$$C_{\alpha_{n+1}} = C\{\pm 1, \pm 2, \cdots, \pm n, +(n+1)\};$$

clearly the prefix of the first and the second index must be different. We also note that all the cones up to 4-point case can be expressed in this form.

Thus, we see that an arbitrary n-point cone C_{α_n} will either have the form

$$C_{\alpha_n} = C\{\pm i_1, \pm i_2, \cdots, \pm i_n\}, \tag{21a}$$

where $(i_1, i_2, \cdots, i_n)$ is a permutation of $(1, 2, \cdots, n)$, or the form

$$C_{\alpha_n} = C\{\alpha_j, \pm i_{j+1}, \cdots, \pm i_n\}, \tag{21b}$$

where the j-point cone C_{α_j} can not be obtained from any of the $(j-1)$-point cones.

We write

$$R_{\alpha_n}(x_1, x_2, \cdots, x_n) = R\{\alpha_j, \pm i_{j+1}, \cdots, \pm i_n\},$$

$$r_{\alpha_n}(x_1, x_2, \cdots, x_n) = r\{\alpha_j, \pm i_{j+1}, \cdots, \pm i_n\},$$

$$\tilde{r}_{\alpha_n}(p_1, p_2, \cdots, p_n) = \tilde{r}\{\alpha_j, \pm i_{j+1}, \cdots, \pm i_n\},$$

where C_{α_n} is the cone given by (21b).

Clearly the cones $C_{\alpha^\pm}$ defined by Eqs. (13) and (17), are given by

$$C_{\alpha^-} = C\{\alpha, -(n+1)\} C_{\alpha^+} = C\{\alpha, +(n+1)\},$$

so that the reduction formulas (10) and (11) can be written as

$$[R_\alpha(x_1, \cdots, x_n), Q^{a*}_{\text{in}}] = -iM^a_{n+1}R\{\alpha, -(n+1)\}, \quad (22a)$$

$$[R_\alpha(x_1, \cdots, x_n), Q^{a*}_{\text{out}}] = -iM^a_{n+1}R\{\alpha, +(n+1)\}. \quad (22b)$$

The complex (not the Hermitian) conjugate of these formulas will be

$$[Q^a_{\text{in}}, R_\alpha(x_1, \cdots, x_n)] = -iM^{a*}_{n+1}R\{\alpha, -(n+1)\}, \quad (22c)$$

$$[Q^a_{\text{out}}, R_\alpha(x_1, \cdots, x_n)] = -iM^{a*}_{n+1}R\{\alpha, +(n+1)\}, \quad (22d)$$

where $M^{a*}_i = \int dx_i f^*_a(x_i)K_{x_i}$.

Here we have used the fact that as $Q(x)$ is Hermitian the complex conjugates and the Hermitian conjugates of $Q^a_{\text{in,out}}$ will be the same, as can be seen from their definition (4). In the next section we will use (22a) and (22d) for the reduction of matrix elements.

In general, an arbitrary n-point cone C_α has a basis of more than $(n-1)$ elements, i.e., it can be completely defined only by more than $(n-1)$ conditions of the form

$$q_I \in V_\pm \quad \text{for} \quad q \in C_\alpha,$$
$$I \equiv \text{a subset of } \{1, 2, \cdots, n\}.$$

Thus, for example, the 5-point cone[10] C^-_{ij} referred to above, has a basis of 6 elements, namely

$$\begin{matrix} q_i + q_l, & q_i + q_m, & q_i + q_n \\ q_j + q_l, & q_j + q_m, & q_j + q_n \end{matrix} \in V_+, \quad (23)$$

(i, j, l, m, n) = a permutation of $(1, 2, 3, 4, 5)$.

However, we can show that the cones of the type (21a) have a basis of just $(n-1)$ elements. For let C_α be an n-point cone with a basis of $(n-1)$ elements and suppose that it is defined by

$$q_{I_1}, q_{I_2}, \cdots, q_{I_k} \in V_+, \quad q_{I_{k+1}}, \cdots, q_{I_{n-1}} \in V_-, \quad (24)$$
$$I_1, I_2, \cdots, I_{n-1} \equiv \text{subsets of } \{1, 2, \cdots, n\}.$$

Then it is easily seen that the cone

$$C\{\alpha, +(n+1)\}$$

is completely defined by the conditions

$$q_{n+1}, q_{I_1}, q_{I_2}, \cdots, q_{I_k} \in V_+,$$
$$q_{n+1} + q_{I_{k+1}}, q_{n+1} + q_{I_{k+2}}, \cdots, q_{n+1} + q_{I_{n-1}} \in V_-,$$

while the cone

$$C\{\alpha, -(n+1)\}$$

is defined by

$$q_{n+1} + q_{I_1}, q_{n+1} + q_{I_2}, \cdots, q_{n+1} + q_{I_k} \in V_+,$$
$$q_{n+1}, q_{I_{k+1}}, q_{I_{k+2}}, \cdots, q_{I_{n-1}} \in V_-.$$

Thus, if C_α is defined by $(n-1)$ conditions,

$$C\{\alpha, \pm(n+1)\}$$

is defined by n conditions. Hence, as

$$C\{+i_1, -i_2\} \quad \text{and} \quad C\{-i_1, +i_2\}$$

are defined by just one condition each, namely

$$q_{i_1} \in V_+ \quad \text{and} \quad q_{i_1} \in V_-,$$

respectively, it follows by induction that all the n-point cones of the type (21a) will have a basis of just $(n-1)$ elements.

Now let the cone

$$C_\alpha \equiv C\{\pm i_1, \pm i_2, \cdots, \pm i_n\}$$

be defined by (24). Then, as we can write

$$q_1x_1 + q_2x_2 + \cdots + q_nx_n = q_{I_1}X_1 + \cdots + q_{I_k}X_k - q_{I_{k+1}}X_{k+1} - \cdots - q_{I_{n-1}}X_{n-1},$$

where X_i are linear combination of terms of the form

$$x_u - x_v, \qquad u, v = 1, 2, \cdots, n,$$

and the conjugate cone $\tilde{C}_\alpha$ is defined by

$$X_1 \in V_+, \qquad X_2 \in V_+, \cdots, X_{n-1} \in V_+.$$

Thus, r_α, whose support is $\tilde{C}^0_0$, will vanish if

$$X^0_1 < 0 \quad \text{or} \quad X^0_2 < 0 \quad \text{or} \quad \cdots \quad \text{or} \quad X^0_{n-1} < 0,$$

and, therefore, due to Lorentz invariance, also if

$$X_1 \notin V_+ \quad \text{or} \quad X_2 \notin V_+ \quad \text{or} \quad \cdots \quad \text{or} \quad X_{n-1} \notin V_+,$$

i.e., if $x \notin \tilde{C}_\alpha$. For if there is an $X_i \notin V_+$ and $X^0_i > 0$, we can find a Lorentz transformation which will take X_i to a point with negative time component and hence r_α will vanish for such X_i. If we now consider the Fourier transform

$$r_\alpha(s_1, s_2, \cdots, s_n), \qquad s \equiv p + iq,$$

we see that it will be analytic in the tube

$$T_\alpha : p \text{ arbitrary}, \; q \in C_\alpha$$

and, hence, also in the extended tube

$$T'_\alpha = \bigcup_{\Lambda \in L(C)} \Lambda T_\alpha, \quad L(C) \equiv \text{complex Lorentz group},$$

by the theorem of Hall and Wightman.

It may be noted that although the analyticity in the tube T_α, and hence in the extended tube T'_α, cannot, in general, be proved when the cone C_α has a basis of more than $(n-1)$ elements, this can be explicitly done in the case of the 5-point function r'_{ij} corresponding to the cone C^-_{ij} defined by (23). For, using Steinmann relations,[11] it can be shown that

$$r'_{12} = 0$$

if $(x_i - x_l)$ and any one of $(x_i - x_m)$, $(x_i - x_n)$, $(x_j - x_l)$ is not in V_+.
Here,

$$(i, j) = \text{a permutation of } (1, 2),$$
$$(l, m, n) = \text{a permutation of } (3, 4, 5).$$

This support properly ensures that

$$x_1q_1 + x_2q_2 + x_3q_3 + x_4q_4 + x_5q_5$$

will be positive when $q \in C^-_{12}$ and x is such that $r'_{12} \neq 0$, so that the Fourier transform

$$r'_{12}(s_1, s_2, s_3, s_4, s_5)$$

will be analytic in the tube

$$T : p \text{ arbitrary}, q \in C^-_{12}.$$

The analyticity in the extended tube now follows from the theorem of Hall and Wightman which can be applied in this case also, even though the tube T differs from the tube $\mathcal{R}_4$ of the theorem, due to the fact that the cone C^-_{12} has a basis of 6 elements. This can be verified by looking at the proof of the theorem which remains unchanged if the tube $\mathcal{R}_{n-1}$ is replaced by a tube of the form

$$p \text{ arbitrary}, q \in C_\alpha,$$

and C_α has a basis of more than $(n-1)$ elements.

We now explicitly determine the support of

$$r\{\pm i_1, \pm i_2, \cdots, \pm i_n\}.$$

In fact, we will show that this function vanishes if any of the following $(n-1)$ conditions [corresponding to the $(n-1)$ values 2, 3, $\cdots$ n of l] is satisfied:

$$x^0_{i_l} - x^0_{i_m} < 0 \quad \text{if} \quad q_{i_l} \in V_+ \qquad (25a)$$
$$\text{for all } m < l \text{ and such that } q_{i_m} \in V_-,$$

and

$$x^0_{i_l} - x^0_{i_m} > 0 \quad \text{if} \quad q_{i_l} \in V_- \qquad (25b)$$
$$\text{for all } m < l \text{ and such that } q_{i_m} \in V_+.$$

[11] O. Steinmann, Helv. Phys. Acta **33**, 341 (1960).

Obviously, it is sufficient to prove the condition corresponding to the case $l = n$. Consider first the function

$$r\{\pm i_1, \pm i_2, \cdots, \pm i_{n-1}, +i_n\}; \qquad (26)$$

a general term in its expansion will be

$$\pm\theta_\pm(k_1, k_2) \cdots \theta_\pm(k_j, i_n)$$
$$\times \theta_\pm(i_n, k_{j+1}) \cdots \theta_\pm(k_{n+2}, k_{n-1}),$$
$$\times \langle 0 | W(k_1, k_2, \cdots, k_j, i_n, k_{j+1}, \cdots k_{n-1}) | 0 \rangle, \qquad (27)$$

where $(k_1, \cdots k_{n-1}) \equiv$ a permutation of $(i_1, \cdots, i_{n-1})$. Now if

$$q_{k_1} + q_{k_2} + \cdots + q_{k_j} + q_{i_n} \in V_+,$$

there will exist a smallest integer $t > j$ such that

$$q_{k_1} + q_{k_2} + \cdots + q_{k_j} + q_{i_n} + q_{k_{j+1}} + \cdots + q_{k_t} \in V .$$

Clearly, this will imply $q_{k_t} \in V_-$. The term (27) will contain the factor

$$\theta_+(i_n, k_{j+1})\theta_+(k_{j+1}, k_{j+2}) \cdots \theta_+(k_{t-1}, k_t),$$

and hence will be zero, if

$$x_{i_n} - x^0_{k_t} < 0.$$

On the other hand, if

$$q_{k_1} + \cdots + q_{k_j} + q_{n_i} \in V_-,$$

there will exist a largest integer $t < j$, such that

$$q_{k_1} + q_{k_2} + \cdots + q_{k_{t-1}} \in V_+,$$

and hence $q_{k_t} \in V_-$. The term (27) will now contain the factor

$$\theta_-(k_t, k_{t+1})\theta_-(k_{t+1}, k_{t+2}) \cdots \theta_-(k_{j-1}, k_j)\theta_-(k_j, i_n)$$

and will, therefore, be zero if

$$x^0_{i_n} - x^0_{k_t} < 0.$$

Thus, every term in the expansion of the function (26), and hence the function itself, will vanish if (25a), with $l = n$, is satisfied. To see that all $m < l$ are to be included in (25a), we note that the term

$$\pm\theta_-(k_1, i_n)\theta_-(i_n, k_2)\theta_\pm(k_2, k_3) \cdots \theta_\pm(k_{n-2}, k_{n-1}),$$
$$\langle 0 | W(k_1, i_n, k_2, k_3, \cdots, k_{n-1}) | 0 \rangle,$$

for an arbitrary k_1 such that $q_{k_1} \in V_-$ vanishes only when

$$x^0_{i_n} - x^0_{k_1} < 0.$$

Similarly, we can show that

$$r\{\pm i_1, \pm i_2, \cdots, \pm i_{n-1}, -i_n\}$$

vanishes if

$$x^0_{i_n} - x^0_{i_m} > 0,$$

for all $m < n$ and such that $q_{i_m} \in V_+$, as required according to (25b).

Lastly, we show that the set of retarded operators

$$R\{\pm i_1, \pm i_2, \cdots, \pm i_n\}$$

is identical with the set of Steinmann's retarded operators[11]

$$Q(x_{i_1} \updownarrow x_{i_2} \updownarrow \cdots \updownarrow x_{i_n}).$$

Here $\updownarrow$ means either $\uparrow$ or $\downarrow$, and these later are defined as follows:

Let $H(x_1, x_2, \cdots, x_n,)$ be an operator of the form

$$H(x_1, x_2, \cdots, x_n) = \sum_\pi f_\pi(x_1, \cdots, x_n)Q(x_{i_1}) \cdots Q(x_{i_n}),$$

where the summation extends over the set of all the permutations $\pi : j \to i_j$ and f_π are arbitrary C-number functions. Steinmann defines

$$\begin{aligned} &H(x_1, x_2, \cdots, x_n, \uparrow x_{n+1}) \\ &\quad \equiv H(x_1, x_2, \cdots, x_n) \uparrow Q(x_{n+1}) \\ &\quad = \sum_{j=1}^{n} \theta(x^0_j - x^0_{n+1})H(x_1, x_2, \cdots, \overline{x_j x_{n+1}}, \cdots, x_n), \\ &\qquad H(x_1, x_2, \cdots, x_n \downarrow x_{n+1}\lambda) \\ &\quad \equiv H(x_1, x_2, \cdots, x_n) \downarrow Q(x_{n+1}) \\ &\quad = \sum_{j=1}^{n} \theta(x^0_{n+1} - x^0_j)H(x_1, x_2, \cdots, \overline{x_j x_{n+1}}, \cdots, x_n), \end{aligned}$$

where $\overline{x_j x_{n+1}}$, means that in the definition of H, x_j is replaced by the commutator

$$[Q(x_j), Q(x_{n+1})].$$

It is easily seen that the identity of the two sets of functions will be established if we can show that

$$R_{\alpha_n}(x_1, \cdots, x_n) \uparrow Q(x_{n+1}) = R\{\alpha_n, -(n+1)\}, \tag{28a}$$

$$R_{\alpha_n}(x_1, \cdots, x_n) \downarrow Q(x_{n+1}) = -R\{\alpha_n, +(n+1)\}, \tag{28b}$$

where C_{α_n} is an arbitrary n-point cone. Now we have

$$\begin{aligned} &R_{\alpha_n}(x_1, x_2, \cdots, x_n) \uparrow Q(x_{n+1}) \\ &\quad = \sum_{j=1}^{n} \theta(j, n+1)R_{\alpha_n}(x_1, \cdots, \overline{x_j x_{n+1}}, \cdots, x_n) \\ &\quad = \sum_{j=1}^{n} \theta(i_j, n+1) \sum_\pi (-1)^\sigma \theta_{\alpha_n}(i_1, i_2) \cdots \\ &\qquad \times \theta_{\alpha_n}(i_{n-1}, i_n)Q(x_{i_1}) \cdots \\ &\qquad \times Q(x_{i_{j-1}})[Q(x_{i_j}), Q(x_{n+1})]Q(x_{i_{j+1}}) \cdots Q(x_{i_n}) \\ &\quad = \sum_\pi (-1)^\sigma \theta(i_1, n+1)\theta_{\alpha_n}(i_1, i_2) \cdots \\ &\qquad \times \theta_{\alpha_n}(i_{n-1}, i_n)W(x_{n+1}, \cdots, i_n) \\ &\qquad + \sum_\pi \sum_{j=1}^{n} (-1)^\sigma \theta_{\alpha_n}(i_1, i_2) \cdots \theta_{\alpha_n}(i_{j-1}, i_j) \\ &\qquad \times [\theta_{\alpha_n}(i_j, i_{j+1})\theta(i_j, n+1) - \theta_{\alpha_n}(i_j, i_{j+1}) \\ &\qquad \times \theta(i_{j+1}, n+1)]\theta_{\alpha_n}(i_{j+1}, i_{j+2}) \cdots \theta_{\alpha_n}(i_{n-1}, i_n) \\ &\qquad \times W(i_1, \cdots, i_j\, n+1, i_{j+1}, \cdots, i_n), \end{aligned}$$

where σ is the number of θ-minus functions. Using the identities

$$\begin{aligned} &\theta_+(i_j, i_{j+1})[\theta(i_j, n+1) - \theta(i_{j+1}, n+1)] \\ &\qquad = \theta_+(i_j, n+1)\theta_+(n+1, i_{j+1}), \\ &\theta_-(i_j, i_{j+1})[\theta(i_j, n+1) - \theta(i_{j+1}, n+1)] \\ &\qquad = -\theta_-(i_j, n+1)\theta_-(n+1, i_{j+1}), \end{aligned}$$

we get

$$\begin{aligned} &R_{\alpha_n}(x_1, x_2, \cdots, x_n) \uparrow Q(x_{n+1}) \\ &\quad = \sum_\pi (-1)^{\sigma'} \theta_-(n+1, i_1)\theta_{\alpha_n}(i_1, i_2) \cdots \\ &\qquad \times \theta_{\alpha_n}(i_{n-1}, i_n)W(n+1, i, \cdots, i_n) \\ &\qquad + \sum_{j=1}^{n} \sum_\pi (-1)^{\sigma'} \theta_{\alpha_n}(i_1, i_2) \cdots \\ &\qquad \times \theta_{\alpha_n}(i_{j-1}, i_j)[\theta_{\alpha_n}(i_j, n+1)\theta_{\alpha_n}(n+1, i_{j+1})], \\ &\qquad \times \theta_{\alpha_n}(i_{j+1}, i_{j+2}) \cdots \theta_{\alpha_n}(i_{n-1}, i_n) \\ &\qquad \times W(i_1, i_2, \cdots, i_j, n+1, i_{j+1} \cdots, i_n) \\ &\quad = \sum_{j=0}^{n} \sum_\pi (-1)^{\sigma'} \theta_{\alpha_n}(i_1, i_2) \cdots \\ &\qquad \times \theta_{\alpha_n}(i_{j-1}, i_j)[\theta_{\alpha_n}(i_j, n+1)\theta_{\alpha_n}(n+1, i_{j+1})] \\ &\qquad \times \theta_{\alpha_n}(i_{j+1}, i_{j+2}) \cdots \theta_{\alpha_n}(i_{n-1}, i_n) \\ &\qquad \times W(i_1, i_2, \cdots, i_j, n+1, i_{j+1}, \cdots, i_n), \end{aligned}$$

where σ' is again the number of θ-minus functions and the suffixes of the two θ functions in the squared brackets are either both plus or both minus. As this last expression is just equal to the operator

$$R\{\alpha_n, -(n+1)\},$$

we get Eq. (28a). The second part, Eq. (28b), can be proved in an exactly similar manner.

5. REDUCTION OF S-MATRIX ELEMENTS

Burgoyne has shown that if $\tilde{T}(p_1, p_2, \cdots, p_n)$ and $\tilde{R}_\alpha(p_1, p_2, \cdots, p_n)$ are the Fourier transform of the time-ordered product and the generalized retarded operator corresponding to the cone C_α, then if $p \in C_\alpha$, the two will be equal. We also know that by using (1), we can reduce the matrix element

$$\langle p_1, p_2, \cdots, p_m \text{ out} \mid -p_{m+1}, -p_{m+2}, \cdots, -p_n \text{ in}\rangle$$

to (29)

$$(-i)^n \int dx_1\, dx_2, \cdots dx_n\, e^{-i\sum p_i x_i} K_{x_1} K_{x_2} \cdots K_{x_n} \times \langle 0 \mid T(x_1, \cdots, x_n) \mid 0\rangle.$$

Hence, if we are able to carry out the reduction of (29) in terms of generalized retarded functions, we must expect that if $p \in C_\alpha$,

$$\langle p_1, p_2, \cdots, p_m \text{ out} \mid -p_{m+1}, -p_{m+2}, \cdots, -p_n \text{ in}\rangle = (-i)^n \int dx_1\, dx_2 \cdots dx_n \times e^{-i\sum p_i x_i} K_{x_1} \cdots K_{x_n} r_\alpha(x_1, \cdots, x_n). \quad (30)$$

We see in the following that this is, indeed, the case.

First of all, we give the following lemma.

Lemma:

$$Q^{p_1}_{\text{out}} Q^{p_2}_{\text{out}} \cdots Q^{p_n}_{\text{out}} \mid k_1, k_2, \cdots, k_m \text{ in}\rangle = 0,$$
$$Q^{p_1}_{\text{in}} Q^{p_2}_{\text{in}} \cdots Q^{p_n}_{\text{in}} \mid k_1, k_2, \cdots, k_m \text{ out}\rangle = 0,$$

if

$$p_1 + p_2 + \cdots + p_n - k_1 - k_2 \cdots - k_m \notin V_- \quad \text{and} \quad \neq 0.$$

These are the generalizations of

$$Q^p_{\text{in}} \mid k \text{ out}\rangle = 0 \quad \text{if} \quad p - k \notin V_- \quad \text{and} \quad \neq 0,$$

which one should expect to be true, as there are are no states which have their energy-momentum 4-vector outside the future light cone.

The proof runs as follows:

If P is the energy-momentum operator, we have

$$e^{iPx} Q^{p_1}_{\text{out}} \cdots Q^{p_n}_{\text{out}} \mid k_1 \cdots k_m \text{ in}\rangle = (-i)^n \lim_{y_i^0 \to +\infty} \int dy_1 \cdots dy_n \times e^{iPx} Q(y_1) \cdots Q(y_n) \overleftrightarrow{\partial}_{y_1^0} \cdots \times \overleftrightarrow{\partial}_{y_n^0} \exp\left(i \sum_1^n p_i y_i\right) \mid k_1, \cdots k_m \text{ in}\rangle$$

$$= (-i)^n \lim_{y_i^0 \to +\infty} \int dy_1 \cdots dy_n \times Q(y_1 + x) \cdots Q(y_n + x) \overleftrightarrow{\partial}_{y_1^0} \times \cdots \overleftrightarrow{\partial}_{y_n^0} \exp\left(i \sum_1^n p_i y_i\right) \times \exp\left(i \sum_1^m k_i x\right) \mid k_1 \cdots k_m \text{ in}\rangle.$$

Changing the variable of integration from y_i to $y_i - x$, we see that this becomes

$$\exp\left[-ix\left(\sum_1^n p_i - \sum_1^m k_i\right)\right] Q^{p_1}_{\text{out}} \cdots Q^{p_n}_{\text{out}} \times \mid k_1 \cdots k_m \text{ in}\rangle.$$

This shows that

$$Q^{p_1}_{\text{out}} \cdots Q^{p_n}_{\text{out}} \mid k_1 \cdots k_m \text{ in}\rangle$$

is an eigenstate of P corresponding to the e-value

$$-\left(\sum_1^n p_i - \sum_1^m k_i\right).$$

As the energy-momentum 4-vector of every eigenstate of P is to lie in the forward light cone, we have

$$Q^{p_1}_{\text{out}} \cdots Q^{p_n}_{\text{out}} \mid k_1 \cdots k_m \text{ in}\rangle = 0$$

if

$$\left(\sum_1^n p_i - \sum_1^m k_i\right) \notin V_- \quad \text{and} \quad \neq 0.$$

The second part of the lemma can be proved in exactly similar manner.

Next we prove

Theorem: The matrix element

$$\langle p_1, p_2, \cdots, p_m \text{ out} \mid -p_{m+1}, -p_{m+2}, \cdots, -p_n \text{ in}\rangle \quad (29)$$

with $p \in C_\alpha$, an arbitrary n-point cone, can be completely reduced to a linear combination of retarded functions of the form

$$r\{\pm i_1, \pm i_2, \cdots, \pm i_n\}.$$

Proof. Let us assume that the first $(k - j)$ reductions in (29) can be carried out, i.e., (29) can be expressed as a linear combination of terms of the form

$$X = \prod M\langle i_1, \cdots, i_j \mid R\{\pm i_{j+1}, \cdots, \pm i_k\} \times \mid i_{k+1}, \cdots, i_n\rangle, \quad (31)$$

where the product is over suitable M operators

defined by (19), and where we have used the notation

$$\langle p_{i_1}, p_{i_2}, \cdots, p_{i_j} \text{ out} \mid R\{\pm i_{j+1}, \cdots, \pm i_k\} \times |-p_{i_{k+1}}, \cdots, -p_{i_n} \text{ in}\rangle = \langle i_1, i_2, \cdots, i_j| \times R\{\pm i_{j+1}, \cdots, \pm i_k\} \; |i_{k+1} \cdots, i_n\rangle. \tag{32}$$

Further, let

$$-(p^0_{i_{k+1}} + p^0_{i_{k+2}} + \cdots + p^0_{i_n}) > p^0_{i_1} + \cdots + p^0_{i_j}. \tag{33}$$

Then there will exist a smallest integer l such that the sum of l of the energies

$$(-p^0_{i_{k+1}}, -p^0_{i_{k+2}}, \cdots, -p^0_{i_n})$$

will be greater than the rhs (right-hand side) of (33). Without any loss of generality, we may take them to be

$$(-p^0_{k+1}, -p^0_{k+2}, \cdots, -p^0_{i_{k+l}}). \tag{34}$$

consider first the case $l = 1$. We will have (writing $Q^{-p_j}_{\text{in}}{}^* \equiv Q^{j*}_{\text{in}}$)

$$X = \prod M\langle i_1, \cdots, i_j| \, R\{\pm i_{j+1}, \cdots, \pm i_k\} \times Q^{i_{k+1}*}_{\text{in}} \; |i_{k+2}, \cdots, i_n\rangle$$

$$= \prod M\langle i_1, \cdots, i_j| \, R\{\pm i_{j+1}, \cdots, \pm i_k, -i_{k+1}\} \times |i_{k+2}, \cdots, i_n\rangle + \prod M\langle i_1, \cdots, i_j| \, Q^{i_{k+1}*}_{\text{in}} \times R\{\pm i_{j+1}, \cdots, \pm i_k\} \; |i_{k+2}, \cdots, i_n\rangle$$

[using (22a)].

The second term vanishes according to the lemma as

$$-p^0_{i_{k+1}} > p^0_{i_1} + \cdots + p^0_{i_n},$$

so that X is equal to matrix element in which $k - j + 1$ reductions have been carried out.

For the case $l = 2$, we have

$$X = \prod M\langle i_1, \cdots, i_j| \, R\{\pm i_{j+1}, \cdots, \pm i_k\} \times Q^{i_{k+1}*}_{\text{in}} \; |i_{k+2}, \cdots i_n\rangle = \prod M\langle i_1, \cdots, i_j| \times R\{\pm i_{j+1}, \cdots, \pm i_{k_1} - i_{k+1}\} \; |i_{k+2}, i_{k+3}, \cdots, i_n\rangle + \prod M\langle i_1, \cdots, i_j| \, Q^{i_{k+1}*}_{\text{in}} R\{\pm i_{j+1}, \cdots, \pm i_k\} \times |i_{k+2}, \cdots, i_n\rangle = \prod M\langle i_1, \cdots, i_j| \times R\{\pm i_{j+1}, \cdots, \pm i_k, -i_{k+1}\} \; |i_{k+2}, i_{k+3}, \cdots, i_n\rangle + \prod M\langle i_1, \cdots, i_j| \, Q^{i_{k+1}*}_{\text{in}} R \times \{\pm i_{j+1}, \cdots, \pm i_k, -i_{k+2}\} \; |i_{k+3}, \cdots, i_n\rangle + \prod M\langle i_1, \cdots, i_j| \, Q^{i_{k+1}*}_{\text{in}} Q^{i_{k+2}*}_{\text{in}} \times R\{\pm i_{j+1}, \cdots, \pm i_k\} \; |i_{k+3}, \cdots, i_n\rangle.$$

The last term again vanishes according to the lemma as now we have

$$-p^0_{i_{k+1}} - p_{i_{k+2}} > +p^0_{i_1} + \cdots + p^0_{i_j}.$$

Hence, commuting $Q^{i_{k+1}*}_{\text{in}}$ back in the second term, we will get

$$X = \prod M\langle i_1, \cdots, i_j| \, R\{\pm i_j, \cdots, \pm i_k, -i_{k+1}\} \times |i_{k+2}, i_{k+3}, \cdots, i_n\rangle + \prod M\langle i_1, \cdots, i_j| \times R\{\pm i_j, \cdots, \pm i_k, -i_{k+2}\} \; |i_{k+1}, i_{k+3}, \cdots i_n\rangle - \prod M\langle i_1, \cdots, i_j| \, R\{\pm i_j, \cdots, + i_k, -i_{k+1}, -i_{k+2}\} \; |i_{k+3} \cdots, i_n\rangle,$$

i.e., X is equal to a sum of terms, in each of which either $(k - j + 1)$ or more reductions have been carried out.

Similar results can be obtained for the cases $l = 3, 4, \cdots, n - k$, which do not involve any further difficulties except that the calculations become a bit more lengthy. The case in which the inequality (33) is reversed can be similarly treated by commuting the out-states [using 22d)] instead of in-states. Thus, we can always express X as a linear combination of terms in each of which either $(k - j + 1)$ or more reductions have been carried out, and hence, the same will also be true for the matrix element (29). This shows that if $(k - j)$ reductions can be carried out in (29), the $(k - j + 1)$ th reduction can also be carried out.

We can perform the first reduction since (29) can always be put in the form

$$-iM_1\langle p_2, p_3, \cdots, p_m \text{ out}| \, Q(x_1) \times |-p_{m+1}, \cdots, -p_n \text{ in}\rangle$$

by using

$$Q^{p_1}_{\text{out}} = Q^{p_1}_{\text{in}} - iM_1Q(x_1) \equiv Q^{p_1}_{\text{in}} - i\int dx_1 f_{p_1}(x_1)K_{x_1}Q(x_1).$$

Hence, the proof of the theorem follows by induction.

This theorem shows that (29) can always be completely reduced, so that the only thing that remains to be discussed is whether or not the final result so obtained agrees with (30). Consider first the case when $p \in C_\alpha$ and C_α is an n-point cone of the type (21a). It is easily seen that if we reduce (29) in the order $i_1, i_2, i_3, \cdots$, all the n reductions can be carried out and the result will agree with (30). As an example, let us consider

$$S_5 = \langle p_1, p_2 \text{ out} \mid -p_3, -p_4, -p_5 \text{ in}\rangle,$$

where[10] $p \in C^-_{3,12} \equiv C\{+1, -3, +2, -4, -5\}$,

so that

$$p_1, p_2, p_1 + p_4, p_2 + p_4, p_1 + p_5, p_2 + p_5 \in V_+,$$
$$p_3, p_4, p_5, p_1 + p_3, p_2 + p_3 \in V_-.$$

We have

$$\begin{aligned} S_5 &= \langle p_2 \text{ out}| Q^{p_1}_{\text{out}} |-p_3, -p_4, -p_5 \text{ in}\rangle \\ &= -iM_1\langle p_2 \text{ out}| Q(x_1) |-p_3, -p_4, -p_5 \text{ in}\rangle, \\ &= -iM_1\langle p_2 \text{ out}| Q(x_1)Q^{-p_3*}_{\text{in}} |-p_4, -p_5 \text{ in}\rangle, \\ &= i^2M_1M_3\langle p_2 \text{ out}| R\{+1, -3\} \\ &\quad \times |-p_4, p_5, \text{ in}\rangle - iM_1 \langle p_2 \text{ out}| \\ &\quad \times Q^{-p_5*}_{\text{in}}Q(x_1) |-p_4, -p_5 \text{ in}\rangle, \end{aligned}$$

where $M_i \equiv M_i^{-p_i} = M_i^{-p_i*} = \int dx_i\ e^{-ip_ix_i}K_{x_i}$, and we have used (22a). The second term vanishes according to the lemma, as

$$-p_3 - p_2 \in V_+.$$

Hence,

$$\begin{aligned} S_5 &= i^{-2}M_1M_3\langle 0| Q^{p_2}_{\text{out}}R\{+1, -3\} |-p_4, -p_5 \text{ in}\rangle \\ &= -i^3M_1M_2M_3\langle 0| R\{+1, -3, +2\} |-p_4, -p_5 \text{ in}\rangle \\ &\quad + i^2M_1M_3 \langle 0| R\{+1, -3\} \\ &\quad \times |Q^{p_2}_{\text{out}} - p_4 - p_5 \text{ in}\rangle \quad \text{[using (22d)]} \\ &= -i^3M_1M_2M_3\langle 0| R\{+1, -3, +2\}Q^{-p_4*}_{\text{in}} \\ &\quad \times |-p_5 \text{ in}\rangle \quad \text{as} \quad p_2 + p_4 + p_5 \in V_+ \\ &= i^4M_1M_2M_3M_4\langle 0| R\{+1, -3, +2, -4\} |-p_5 \text{ in}\rangle \\ &= -i^5M_1M_2M_3M_5\langle 0|R\{+1,-3,+2,-4,-5\} |0\rangle, \end{aligned}$$

as required. Clearly no complications arise when $n > 5$.

Next we consider the case when $p \in C_\alpha$ and C_α is of the type (21b) with $j = n$. Clearly, if this can be reduced to the proper retarded function, the case $j < n$ will present no difficulties. We cannot apply the above method directly as the final result is expected to be r_α and this can not be obtained from any of the $(n - 1)$ point functions. It seems plausible that using Steinmann relations, we can show that the linear combination of the retarded functions obtained as the result of reduction is just equal to r_α. This can be verified in the 5-point case, where all the cones have been classified and all the Steinmann relations are known.[12] Let us take the example of

$$S_5 = \langle p_1p_2 \text{ out} \mid -p_3, p-_4, -p_5 \text{ in}\rangle,$$

where $p \in C'_{12}$, i.e.,

$$\begin{matrix} p_1 + p_3, & p_1 + p_4, & p_1 + p_5 \\ p_2 + p_3, & p_2 + p_4, & p_2 + p_5 \end{matrix} \in V_+.$$

We have

$$\begin{aligned} S_5 &= \langle p_1 \text{ out}| Q^{p_2}_{\text{out}} |-p_3, -p_4, -p_5 \text{ in}\rangle \\ &= -iM_2\langle p_1 \text{ out}| Q(x_2) |-p_3, -p_4, -p_5 \text{ in}\rangle \\ &= i^2M_2M_3\langle p_1 \text{ out}| R\{+2, -3\} |-p_4, -p_5 \text{ in}\rangle \\ &\quad - iM_2\langle p_1 \text{ out}| Q^{-p_3*}_{\text{in}} Q(x_2) |-p_4, -p_5 \text{ in}\rangle \\ &= -i^3M_2M_3M_4\langle p_1 \text{ out}| R\{+2, -3, -4\} |-p_5 \text{ in}\rangle \\ &\quad + i^2M_2M_3\langle p_1 \text{ out}| Q^{-p_4*}_{\text{in}}R\{+2, -3\} |-p_5 \text{ in}\rangle \\ &\quad + i^2M_2M_4\langle p_1 \text{ out}| Q^{-p_3*}_{\text{in}}R\{+2, -4\} |-p_5 \text{ in}\rangle - iM_2 \\ &\quad \times \langle p_1 \text{ out}| Q^{-p_3*}_{\text{in}}Q^{-p_4*}_{\text{in}}Q(x_2) |-p_5 \text{ in}\rangle. \end{aligned}$$

The last term vanishes by the lemma as

$$-p_3 - p_4 - p_1 \in V_-,$$

so that

$$\begin{aligned} S_5 &= -iM_2M_3M_4\langle p_1 \text{ out}| R\{+2, -3, -4\} |-p_5 \text{ in}\rangle \\ &\quad + i^2M_2M_4\langle p_1 \text{ out}| Q^{-p_3*}_{\text{in}}R\{+2, -4\} |-p_5 \text{ in}\rangle \\ &\quad + i^2M_2M_3\langle p_1 \text{ out}| Q^{-p_4*}_{\text{in}}R\{+2, -3\} |-p_5 \text{ in}\rangle. \end{aligned}$$

Now

$$\begin{aligned} &\langle p_1 \text{ out}| Q^{-p_3*}_{\text{in}}R\{+2, -4\} |-p_5 \text{ in}\rangle \\ &= -iM_5\langle p_1 \text{ out}| Q^{p_3*}_{\text{in}}R\{+2, -4, -5\} |0\rangle \\ &\quad + \langle p_1 \text{ out}| Q^{-p_4*}_{\text{in}}Q^{-p_3*}_{\text{in}}R\{+2, -4\} |0\rangle \\ &= -iM_5\langle p_1 \text{ out}| Q^{-p_3*}_{\text{in}}R\{+2, -4, -5\} |0\rangle, \\ &\qquad\qquad \text{as} \quad -p_3 - p_5 - p_2 \in V_-, \\ &= -i^2M_5M_3\langle p_1 \text{ out}| R\{+2, -4, -5, -3\} |0\rangle \\ &\quad - iM_5\langle p_1 \text{ out}| R\{+2, -4, -5\} |-p_3 \text{ in}\rangle, \\ &= i^3M_5M_3M_1\langle 0| R\{+2, -4, -5, -3,\} |0\rangle \\ &\quad - i^{-3}M_5M_1M_3\langle 0 R\{+2, -4, -5, +1, -3\} |0\rangle \\ &= i^3M_5M_3M_1\langle 0| R_{21} - R_{21,3} |0\rangle, \end{aligned}$$

in Burgoyne's[10] notation. Similarly,

$$\begin{aligned} &\langle p_1 \text{ out}| Q^{-p_4*}_{\text{in}}R\{+2, -3\} |-p_5 \text{ in}\rangle \\ &\qquad = i^3M_1M_4M_5\langle 0| R_{21} - R_{21,4} |0\rangle. \end{aligned}$$

Also,

$$\begin{aligned} &\langle p_1 \text{ out}| R\{+2, -3, -4\} |-p_5 \text{ in}\rangle \\ &\qquad = i^2M_1M_5\langle 0| R\{+2, -3, -4, +1, -5\} |0\rangle \\ &\qquad = i^2M_1M_5\langle 0| R_{21,5} |0\rangle. \end{aligned}$$

[12] See reference 10.

We, therefore, get

$$S_5 = -i^5 M_1 M_2 M_3 M_4 M_5 \times \langle 0 | R_{21,5} - R_{21} + R_{21,3} - R_{21} + R_{21,4} | 0 \rangle.$$

Now we have the Steinmann relations

$$R_{jk} - R_{jk,n} + R'_{jk,m} - R_{jk,t} = 0,$$

$$R'_{jk} - R'_{jk,n} + R_{jk,m} - R'_{jk,t} = 0,$$

where (j, k, m, n, t) is a permutation of $(1, 2, 3, 4, 5)$. These give

$$R'_{jk} = R'_{kj} = R_{jk,m} + R_{jk,n} + R_{jk,t} - 2R_{jk},$$

$$R'_{12} = R_{21,3} + R_{21,4} + R_{21,5} - 2R_{21}.$$

Hence, we finally reach the result

$$S_5 \equiv \langle p_1, p_2 \text{ out} | -p_3, -p_4, -p_5 \text{ in} \rangle = -i^5 M_1 M_2 M_3 M_4 M_5 \langle 0 | R'_{12}(x_1, x_2, x_3, x_4, x_5) | 0 \rangle,$$

as required.

Until now, we have considered only the case when all the p_J's are timelike. We have seen that in the reductions already carried out, at matrix element of the type

$$\langle \cdots \text{ out} | R(\cdots) Q^{p_1}_{\text{out}} \cdots Q^{p_j}_{\text{out}} \times | -p_{j+1}, p_{j+2}, \cdots, -p_l \text{ in} \rangle,$$

vanishes when

$$\sum_{i=1}^{l} p_i \in V_+.$$

But the lemma states that it will aslo vanish when $\sum p_i$ is spacelike, so that the same process of reduction can be carried out in this case also. Thus, if a matrix element can be reduced to a retarded function when $p \in C_\alpha$, it can also be reduced (to the same retarded function) when p is such that, for each J, p_J is either in V_+ according as $q_J \in V_+$ for $q \in C_\alpha$, or is spacelike.

It should be noted that in our calculations we have not used the assumption of stability of one-particle states

$$|p \text{ in}\rangle = |p \text{ out}\rangle.$$

A consequence of this assumption is that on the mass shell

$$p^2_{n+1} = \mu^2,$$

we have

$$M_{n+1} \langle 0 | R\{\alpha_n, +(n+1)\} | 0 \rangle = M_{n+1} \langle 0 | R\{\alpha_n, -(n+1)\} | 0 \rangle, \qquad (35)$$

where R_{α_n} is an arbitrary generalized retarded operator. This follows from

$$\langle 0 | R_{\alpha_n}(x_1, \cdots, x_n) | p_{n+1} \text{ out} \rangle = \langle 0 | R_{\alpha_n}(x_1, \cdots, x_n) | p_{n+1} \text{ in} \rangle.$$

ACKNOWLEDGMENTS

I wish to thank Dr. R. F. Streater for suggesting this problem and for his continued guidance and encouragement. I am also grateful to the British Department of Technical Cooperation and the Pakistan Atomic Energy Commission for the grant of a Fellowship under the Colombo Plan.

Theorem on Energy Shifts Due to a Potential in a Large Box*

FUMIAKI IWAMOTO†
Department of Physics, Brandeis University, Waltham, Massachusetts
(Received 21 January 1963)

A theorem conjectured by Cohen, which gives a concise form to the Rayleigh–Schrödinger perturbation series for energy shifts due to a potential in a large box, is proved.

IN the many-body problem, especially in the study of low-density systems, we often reduce the problem to a one-body problem in a large box with a potential of finite range at the center. The energy of the total system may be approximately expressed as the sum of small individual level shifts caused by the potential. The theorem which has been conjectured by Cohen,[1] which we prove below, is concerned with the perturbation expression for this level shift.

Let us consider a particle of mass m in a large spherical box with radius R_0. We denote an eigenfunction and an eigenvalue of the Hamiltonian $H_0 = p^2/2m$ by $|a\rangle$ and E_a;

$$H_0 |a\rangle = E_a |a\rangle, \tag{1}$$

$$|a\rangle = N_l j_l(p_a r) Y_{lm}(\theta, \phi),$$

$$N_l \text{ is a normalization factor,} \tag{2}$$

$$E_a = p_a^2/2m. \tag{3}$$

Under the usual zero boundary condition at the surface, one has the following relation valid for large R_0 and finite p_a:

$$p_a R_0 = (n_a + \tfrac{1}{2}l)\pi. \tag{4}$$

The energy separations near the level E_a are $(\pi p_a/mR_0)n$, $n = \pm 1, \pm 2, \cdots$, for a given l value.

We introduce a spherical potential v at the origin. The level E_a shifts by $\Delta E_a \equiv (\pi p_a/mR_0)z$, which is usually of the same order as the level separation. The level shift is the diagonal matrix element of the R_a matrix defined by the nonlinear integral equation

$$R_a = v + v \frac{Q_a}{E_a + \langle a| R_a |a\rangle - H_0} R_a, \tag{5}$$

$$\Delta E_a = \langle a| R_a |a\rangle, \tag{6}$$

where Q_a is a projection operator which excludes the state $|a\rangle$ in the intermediate states, i.e., $Q_a |a\rangle = 0$, $Q_a |b\rangle = |b\rangle$ for $b \neq a$. This leads to the usual Brillouin–Wigner perturbation series, which is an implicit equation for ΔE_a. On the other hand, the Rayleigh–Schrödinger perturbation series, which is an explicit form of the level shift, contains more and more complicated terms as the order increases. Cohen's theorem states that for large R_0 the R_a matrix can be expressed by the following series which has a simple regularity[2]:

$$\langle b| R_a |a\rangle = \langle b| v + v \frac{1}{E_a - H_0} v + v \frac{1}{E_a - H_0} v \frac{1}{E_a - H_0} v + \cdots |a\rangle_{\text{no repetition}}, \tag{7}$$

where the symbol $\langle \cdots \rangle_{\text{no repetition}}$ means that all the intermediate states are different from the initial state $|a\rangle$ and from each other. For example, the fourth-order term $R_a^{(4)}$ is given by

$$\langle b| R_a^{(4)} |a\rangle = \sum_{c \neq a} \sum_{d \neq a,c} \sum_{e \neq a,c,d} \frac{\langle b| v |c\rangle\langle c| v |d\rangle\langle d| v |e\rangle\langle e| v |a\rangle}{(E_a - E_c)(E_a - E_d)(E_a - E_e)}.$$

In particular, the energy shift $\langle a| R_a |a\rangle$ is expressed as

$$\Delta E_a = \sum_{n=0}^{\infty} \langle a| v\left(\frac{1}{E_a - H_0} v\right)^n |a\rangle_{\text{no repetition}}. \tag{8}$$

The theorem may be useful when one encounters such an expression as the right-hand side of (7) or (8) in a complete many-body perturbation series; one can simply identify the expression as an energy shift. Cohen[1] conjectured that the relations (7) and (8) may be true from the analysis of some lower-order terms. However, he was unable to give a general proof since his analysis became highly intractable in the case of higher-order terms.

Let us prove the theorem in general. First we notice that utmost care should be taken in the

* Work supported by the National Science Foundation.
† On leave of absence from College of General Education, University of Tokyo, Tokyo, Japan.
[1] M. Cohen, Phys. Rev. **118**, 27 (1960).

[2] *Note added in proof:* General perturbation formula without repeated summation was presented by Feenberg. In the general case the formula is not an explicit one, containing rather involved energy denominators. [E. Feenberg, Phys. Rev. **74**, 206 (1948).]

summation of those intermediate states whose energies lie close to E_a, since in the energy denominator of (5) both the level shift $\langle a|\ R_a\ |a\rangle$ and those level distances $E_b - E_a$ are of the same order of magnitude, $\pi p_a/mR_0$. Therefore, we separate the intermediate states into two parts, I and $\bar{I}$. I consists of those states whose energies E_b are close to the initial energy E_a, i.e., $|E_b - E_a| < \Delta$. $\bar{I}$ is the rest of the states, i.e., $|E_b - E_a| \geq \Delta$. The magnitude of Δ is so chosen that it is much larger than the level distance $\pi p_a/mR_0$, but still it tends to zero in the limit of large R_0. For example, we can choose $\Delta = \pi p_a/mR_0^{\frac{1}{2}}$. Then, for a function $f(E)$, which is continuous at $E = E_a$, we have following formulas:

$$\lim_{R_0\to\infty} \frac{\pi p_a}{mR_0} \sum_{\substack{b\in I \\ b\neq a}} \frac{f(E_b)}{E_a + \Delta E_a - E_b} = f(E_a) \lim_{N\to\infty} \sum_{\substack{|n|<N \\ n\neq 0}} \frac{1}{z-n} = (\pi \cot \pi z - z^{-1}) f(E_a), \tag{9}$$

$$\lim_{R_0\to\infty} \frac{\pi p_a}{mR_0} \sum_{b\in \bar{I}} \frac{f(E_b)}{E_a + \Delta E_a - E_b} = \int \frac{\mathcal{P}}{E_a - E} f(E)(p_a/p)\, dE, \tag{10}$$

where $\mathcal{P}$ indicates a principal value integral.

Before going to the proof of Cohen's theorem, we need the relation between the R_a matrix and reaction matrix G_a, which was discussed by several workers[3] in connection with the original theory of nuclear matter by Brueckner, Levinson, and Mahmoud.[4] Applying formulas (9) and (10) to (5), one can derive the following equation:

$$R_a\ |a\rangle = \pi z \cot \pi z\, v\ |a\rangle + v[\mathcal{P}/(E_a - H_0)]R_a\ |a\rangle. \tag{11}$$

On the other hand, the reaction matrix G_a is defined by

$$G_a = v + v[\mathcal{P}/(E_a - H_0)]G_a. \tag{12}$$

Its diagonal element is related to the phase shift δ by

$$\langle a|\ G_a\ |a\rangle = -(p_a/mR_0) \tan \delta(E_a). \tag{13}$$

On comparing (11) and (12), one finds an important relation between R_a and G_a;

$$R_a\ |a\rangle = \pi z \cot \pi z\, G_a\ |a\rangle. \tag{14}$$

By considering the diagonal element of (14) and using the relation (13) one can express the energy shift in terms of the phase shift[3];

$$\pi z = -\delta,$$

or

$$\Delta E_a = \langle a|\ R_a\ |a\rangle = -(p_a/mR_0)\ \delta(E_a). \tag{15}$$

From (13), (14), (15), and the series

$$\pi z \cot \pi z = 1 - (\tfrac{1}{3}) \tan^2 \pi z + (\tfrac{1}{5}) \tan^4 \pi z - \cdots, \tag{16}$$

the power-series expansion of the matrix element of R_a with respect to that of G_a may be written as

$$\langle b|\ R_a\ |a\rangle = \langle b|\ G_a\ |a\rangle \sum_{\lambda=0}^{\infty} (-1)^{\lambda}(2\lambda + 1)^{-1} \times [(mR_0/p_a)\langle a|\ G_a\ |a\rangle]^{2\lambda}. \tag{17}$$

Now we will show that Cohen's relation (7) can be rewritten in the form (17). First we make use of a formula similar to (9) and (10);

$$v \frac{1}{E_a - H_0} v = \sum_{b\in I} \frac{v\ |b\rangle\langle b|\ v}{E_a - E_b} + \sum_{b\in \bar{I}} \frac{v\ |b\rangle\langle b|\ v}{E_a - E_b} = \frac{mR_0}{\pi p_a} \left(\sum \frac{-1}{n} \right) v\ |a\rangle\langle a|\ v + v \frac{\mathcal{P}}{E_a - H_0} v, \tag{18}$$

where the summation $\sum(1/n)$ is formally retained since, in the following application of this formula, certain terms are to be omitted due to the restriction of no repeated summation. Further we notice that the restriction of no repeated summation over the intermediate states needs only to be imposed for the states in I, since in $\bar{I}$ the energy denominator is larger than Δ while each matrix element of v, $\langle c|\ v\ |d\rangle$, is of the order of $1/R_0$. Therefore, applying the formula (18) to every intermediate state summation of the nth-order term $R_a^{(n)}$ in (7), we obtain

$$\langle b|\ R_a^{(n)}\ |a\rangle = \sum_{p=1}^{\infty} (-mR_0/\pi p_a)^{p-1} C_p \times \sum_{k_1=1}^{\infty} \sum_{k_2=1}^{\infty} \cdots \sum_{k_p=1}^{\infty} Y_{k_1} X_{k_2} \cdots X_{k_p}\, \delta_{k_1+k_2+\cdots+k_p,\,n}, \tag{19}$$

where $\delta_{k,n}$ is Kronecker δ symbol, X_k and Y_k are defined by

$$X_k = \langle a|\ v\left(\frac{\mathcal{P}}{E_a - H_0} v\right)^{k-1}\ |a\rangle, \qquad Y_k = \langle b|\ v\left(\frac{\mathcal{P}}{E_a - H_0} v\right)^{k-1}\ |a\rangle, \tag{20}$$

and C_p is the quantity

[3] N. Fukuda and R. G. Newton, Phys. Rev. **103**, 1558 (1956), B. S. DeWitt, Phys. Rev. **103**, 1565 (1956); W. B. Riesenfeld and K. M. Watson, Phys. Rev. **104**, 492 (1956).

[4] K. A. Brueckner, C. A. Levinson, and H. M. Mahmoud, Phys. Rev. **95**, 217 (1954).

$$C_p = \sum_{\substack{-\infty<m_i<\infty,\, m_i=0 \\ \text{all } m_i \text{ different}}} \sum \cdots \sum (m_1 m_2 \cdots m_{p-1})^{-1}$$

$$= (p-1)! \sum_{\substack{-\infty<m_1<m_2<\cdots<m_{p-1}<\infty \\ m_i \neq 0}} \sum \cdots \sum (m_1 m_2 \cdots m_{p-1})^{-1}; \quad (21)$$

the restricted summation implies no repeated intermediate states. By observing the coefficient of the ξ^{p-1} term in the identity

$$\sin(\pi\xi)/(\pi\xi) = \prod_{n=1}^{\infty} (1 - \xi/n)(1 + \xi/n), \quad (22)$$

we have simply

$$C_p = \begin{cases} 0 & p \text{ even}, \\ (-1)^{(p-1)/2}(\pi^{p-1}/p) & p \text{ odd}. \end{cases} \quad (23)$$

From (19) and (23) we finally obtain

$$\langle b|\, R_a\, |a\rangle = \left(\sum_{k=1}^{\infty} Y_k\right) \sum_{\lambda=0}^{\infty} (-1)^{\lambda}(2\lambda+1)^{-1} \times \left[(mR_0/p_a)\left(\sum_{k=1}^{\infty} X_k\right)\right]^{2\lambda}. \quad (24)$$

This is just the relation (17), since X_k and Y_k are the kth-order term of $\langle a|\, G_a\, |a\rangle$ and $\langle b|\, G_a\, |a\rangle$. Thus we have established the theorem.

The author would like to express his sincere thanks to Professor R. Rockmore for calling his attention to the theorem and for his encouragement.

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 4, NUMBER 6 JUNE 1963

Generalized Dielectric Function for Classical Many-Body Systems*

AMIRAM RON†
Plasma Physics Laboratory, Princeton University, Princeton, New Jersey

A dielectric function is introduced for classical systems of many interacting particles in terms of the N-particle distribution function in phase space. The system is considered as a whole using the Liouville equation instead of the canonical equations or the B-B-G-K-Y hierarchy. To facilitate the calculation, the diagrammatic technique of Prigogine and Balescu is introduced in terms of which a systematic asymptotic analysis is carried out for two simple cases: (1) a system of weak interaction and (2) a simple model of plasma. The treatment is restricted to systems in thermal equilibrium.

I. INTRODUCTION

RECENTLY, methods have been developed to express many important physical properties of systems of many particles in terms of one so-called "dielectric function." (We employ this terminology even if the interaction is not Coulombic.) Originally the method of the dielectric function was formulated by Lindhard[1] for systems of charged particles, in both the classical and quantum cases. Lindhard's approach was based on the self-consistent field, while the recent more sophisticated *quantum* treatments have been based on perturbation expansions.[2-7] The aim of the present work is the derivation of the dielectric function for *classical systems* of many particles using the perturbation expansion of Prigogine and Balescu[8] and their group.[9-12]

The method of the dielectric function is based on the introduction of a frequency and wave-vector dependent "dielectric constant," in terms of which a description of the properties of systems composed of many interacting particles is expressed. The formulation of this function is based on very general assumptions about the linear response of a system to small applied fields. This formulation includes, in principle, a complete theoretical treatment of both single-particle and collective effects. We restrict ourself only to the formulation of the dielectric function. The relationship of the dielectric function to prop-

* This work was accomplished under the auspices of the United States Atomic Energy Commission.

† On leave of absence from Technion–Israel Institute of Technology, Haifa, Israel.

[1] J. Lindard, Kgl. Danske Videnskab. Selskab, Mat.-Fys. Medd. **28**, No. 8 (1954).

[2] P. Nozieres and D. Pines, Nuovo Cimento **9**, 470 (1958); A. J. Glick, Ann. Phys. **17**, 61 (1961).

[3] F. Englert, J. Phys. Chem. Solids **11**, 78 (1959).

[4] F. Englert and R. Brout, Phys. Rev. **120**, 1085 (1960).

[5] D. Pines, Physica **26**, 103-S (1960).

[6] J. Hubbard, Proc. Roy. Soc. (London) **A240**, 539 (1957); **A243**, 336 (1958).

[7] D. F. DuBois, Ann. Phys. (N. Y.) **7**, 174 (1959); **8**, 24 (1959).

[8] I. Prigogine and R. Balescu, Physica **25**, 281, 302 (1959); **26**, 145 (1960).

[9] R. Balescu, Phys. Fluids **3**, 52 (1960).

[10] R. Balescu and H. S. Taylor, Phys. Fluids **4**, 85 (1961).

[11] R. Balescu, in *Lectures In Theoretical Physics* (Interscience Publishers, Inc., New York, 1961), Vol. III, pp. 382–444.

[12] P. Résibois, F. Henin, and F. C. Andrews, J. Math. Phys. **2**, 68 (1961).

$$C_p = \sum_{\substack{-\infty<m_i<\infty,\, m_i=0 \\ \text{all } m_i \text{ different}}} \sum \cdots \sum (m_1 m_2 \cdots m_{p-1})^{-1}$$

$$= (p-1)! \sum_{\substack{-\infty<m_1<m_2<\cdots<m_{p-1}<\infty \\ m_i \neq 0}} \sum \cdots \sum (m_1 m_2 \cdots m_{p-1})^{-1}; \quad (21)$$

the restricted summation implies no repeated intermediate states. By observing the coefficient of the ξ^{p-1} term in the identity

$$\sin(\pi\xi)/(\pi\xi) = \prod_{n=1}^{\infty} (1 - \xi/n)(1 + \xi/n), \quad (22)$$

we have simply

$$C_p = \begin{cases} 0 & p \text{ even}, \\ (-1)^{(p-1)/2}(\pi^{p-1}/p) & p \text{ odd}. \end{cases} \quad (23)$$

From (19) and (23) we finally obtain

$$\langle b|\, R_a\, |a\rangle = \left(\sum_{k=1}^{\infty} Y_k\right) \sum_{\lambda=0}^{\infty} (-1)^{\lambda}(2\lambda+1)^{-1} \times \left[(mR_0/p_a)\left(\sum_{k=1}^{\infty} X_k\right)\right]^{2\lambda}. \quad (24)$$

This is just the relation (17), since X_k and Y_k are the kth-order term of $\langle a|\, G_a\, |a\rangle$ and $\langle b|\, G_a\, |a\rangle$. Thus we have established the theorem.

The author would like to express his sincere thanks to Professor R. Rockmore for calling his attention to the theorem and for his encouragement.

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 4, NUMBER 6 JUNE 1963

Generalized Dielectric Function for Classical Many-Body Systems*

AMIRAM RON†
Plasma Physics Laboratory, Princeton University, Princeton, New Jersey

A dielectric function is introduced for classical systems of many interacting particles in terms of the N-particle distribution function in phase space. The system is considered as a whole using the Liouville equation instead of the canonical equations or the B-B-G-K-Y hierarchy. To facilitate the calculation, the diagrammatic technique of Prigogine and Balescu is introduced in terms of which a systematic asymptotic analysis is carried out for two simple cases: (1) a system of weak interaction and (2) a simple model of plasma. The treatment is restricted to systems in thermal equilibrium.

I. INTRODUCTION

RECENTLY, methods have been developed to express many important physical properties of systems of many particles in terms of one so-called "dielectric function." (We employ this terminology even if the interaction is not Coulombic.) Originally the method of the dielectric function was formulated by Lindhard[1] for systems of charged particles, in both the classical and quantum cases. Lindhard's approach was based on the self-consistent field, while the recent more sophisticated *quantum* treatments have been based on perturbation expansions.[2-7] The aim of the present work is the derivation of the dielectric function for *classical systems* of many particles using the perturbation expansion of Prigogine and Balescu[8] and their group.[9-12]

The method of the dielectric function is based on the introduction of a frequency and wave-vector dependent "dielectric constant," in terms of which a description of the properties of systems composed of many interacting particles is expressed. The formulation of this function is based on very general assumptions about the linear response of a system to small applied fields. This formulation includes, in principle, a complete theoretical treatment of both single-particle and collective effects. We restrict ourself only to the formulation of the dielectric function. The relationship of the dielectric function to prop-

* This work was accomplished under the auspices of the United States Atomic Energy Commission.

† On leave of absence from Technion–Israel Institute of Technology, Haifa, Israel.

[1] J. Lindard, Kgl. Danske Videnskab. Selskab, Mat.-Fys. Medd. **28**, No. 8 (1954).

[2] P. Nozieres and D. Pines, Nuovo Cimento **9**, 470 (1958); A. J. Glick, Ann. Phys. **17**, 61 (1961).

[3] F. Englert, J. Phys. Chem. Solids **11**, 78 (1959).

[4] F. Englert and R. Brout, Phys. Rev. **120**, 1085 (1960).

[5] D. Pines, Physica **26**, 103-S (1960).

[6] J. Hubbard, Proc. Roy. Soc. (London) **A240**, 539 (1957); **A243**, 336 (1958).

[7] D. F. DuBois, Ann. Phys. (N. Y.) **7**, 174 (1959); **8**, 24 (1959).

[8] I. Prigogine and R. Balescu, Physica **25**, 281, 302 (1959); **26**, 145 (1960).

[9] R. Balescu, Phys. Fluids **3**, 52 (1960).

[10] R. Balescu and H. S. Taylor, Phys. Fluids **4**, 85 (1961).

[11] R. Balescu, in *Lectures In Theoretical Physics* (Interscience Publishers, Inc., New York, 1961), Vol. III, pp. 382–444.

[12] P. Résibois, F. Henin, and F. C. Andrews, J. Math. Phys. **2**, 68 (1961).

erties of many-particle systems such as free energy, stopping power, etc., has been defined in terms of the dielectric function by other authors.[2-4,13-18]

Section II recalls the definition of the dielectric function and gives a formal exact expression for it in terms of the classical N-particle equilibrium distribution function. The treatment is restricted to systems which are in thermal equilibrium, and are of infinite extent. This implies that the system is considered as invariant under translation in space and time. Several formal properties of the dielectric function are discussed including the relation to density fluctuations.

In Sec. III we trace the calculation of the dielectric function by a method developed by Prigogine and Balescu (PB).[8-12] We give a brief review of their method as far as it suits our problem. The main features of their method are: (1) Fourier expansion in the phase space, (2) a formal perturbation expansion of the resolvent operator in terms of the interaction factor λ^2, and (3) a diagram technique corresponding to the latter expansion, which enables one to classify the terms of the expansion in powers of λ^2, N, and V, where N and V are the number of particles and the volume of the system, respectively. Our present work represents a generalization to the PB procedure in that we do not classify diagrams according to their time scale, so that we are not restricted to asymptotic time behavior in order to fit the Bogoliubov hypothesis.[19]

Some examples of calculations of the dielectric function for specific systems are given in Sec. IV, The cases discussed are: (1) systems with weak and short-range forces between the particles, and (2) systems of charged particles. In each case the infinite series developed in Sec. III is regrouped into partial sums on the basis of an ordering assigned to physical parameters characterizing the system under discussion. Section V consists of a summary of present work.

II. FORMULATION OF THE DIELECTRIC FUNCTION

We consider a classical system of N identical particles in a volume V (A generalization to systems of more components is straightforward.) The quantities N and V are both assumed to be very large and the limit

$$N \to \infty \qquad V \to \infty$$

is to be taken in such a manner that

$$\lim_{N\to\infty V\to\infty} N/V \equiv n \tag{1}$$

is finite, with n the average concentration. The particles are assumed to interact through a two-body force, which depends only on the distance between the particles. The interaction potential $V(\mathbf{x})$ at a point $\mathbf{x}$ in space, due to the presence of a particle at $\mathbf{x}'$, is given by

$$V(\mathbf{x}) = \lambda^2\phi(|\mathbf{x} - \mathbf{x}'|), \tag{2}$$

where λ^2 measures the strength of the interaction energy.[20] If we denote by $\rho(\mathbf{x}, t)$ the number of particles per unit volume in the vicinity of $\mathbf{x}$ at the time t, we have for the potential,

$$V(\mathbf{x}, t) = \lambda \int d\mathbf{x}'\phi(\mathbf{x} - \mathbf{x}')\rho(\mathbf{x}', t). \tag{3}$$

A. Definition of the Dielectric Function

Let us now apply an external potential $V_{\text{ext}}(\mathbf{x}, t)$ of the same nature as the particle interaction potential. This potential will induce density perturbation $\rho_{\text{ind}}(\mathbf{x}, t)$ and the total potential at $\mathbf{x}$ will be given by

$$V_{\text{tot}}(\mathbf{x}, t) = V_0(\mathbf{x}, t) + V_{\text{ext}}(\mathbf{x}, t) + \lambda \int d\mathbf{x}'\phi(\mathbf{x} - \mathbf{x}')\rho_{\text{ind}}(\mathbf{x}', t), \tag{4}$$

where $V_0(\mathbf{x}, t)$ is the potential when the external field is absent. If we assume the system responds linearly to the external disturbance, we can express the change of the potential as

$$V_{\text{tot}}(\mathbf{x}, t) - V_0(\mathbf{x}, t) = \int d\mathbf{x}'\, dt' R(\mathbf{x}, t; \mathbf{x}'; t')V_{\text{ext}}(\mathbf{x}', t'), \tag{5}$$

where $R(\mathbf{x}, t; \mathbf{x}', t')$ is the response function of the system. R is not a simple δ function in space and time, as can be seen from Eq. (4). If we consider only systems which are homogeneous in space and time, Eqs. (4) and (5) can be combined to give

$$\int d\mathbf{x}'\, dt' R(\mathbf{x} - \mathbf{x}'; t - t')V_{\text{ext}}(\mathbf{x}', t') = V_{\text{ext}}(\mathbf{x}, t) + \lambda \int d\mathbf{x}'\phi(\mathbf{x} - \mathbf{x}')\rho_{\text{ind}}(\mathbf{x}', t). \tag{6}$$

[13] J. J. Quinn and R. A. Ferrell, Phys. Rev. **112**, 812 (1958).
[14] R. H. Ritchie, Phys. Rev. **114**, 644 (1959).
[15] A. J. Glick and R. A. Ferrell, Ann. Phys. **11**, 359 (1960).
[16] J. Neufeld and R. H. Ritchie, Phys. Rev. **98**, 1632 (1955).
[17] W. B. Thompson and J. Hubbard, Rev. Mod. Phys. **32**, 714 (1960).
[18] A. A. Rukhadze and V. P. Silin, Usp. Fiz. Nauk **74**, 223 (1961) [English transl.: Soviet Phys.—Uspekhi **4**, 459 (1961)].
[19] N. N. Bogoliubov, *Studies in Statistical Mechanics*, edited by J. de Boer and G. E. Uhlenbeck (North-Holland Publishing Company, Amsterdam, 1962).

[20] The parameter λ is used in this paper in the sense of a charge, mass, etc., namely, it corresponds to a property of the particles participating in the interaction.

We now define the dielectric function $\epsilon(\mathbf{k}, \omega)$ as the inverse of the Fourier transform of the response function, namely,

$$[\epsilon(\mathbf{k}, \omega)]^{-1} = \iint d\xi\, d\tau R(\xi, \tau) \times \exp(+i\mathbf{k}\cdot\xi + i\omega\tau). \quad (7)$$

Obviously, in the case of systems of charged particles, $\epsilon(\mathbf{k}, \omega)$ agrees, for long wavelengths, with the usual macroscopic dielectric constant.[1]

In systems which obey the principle of causality, i.e., that an applied field have no effect before it is applied,

$$R(\mathbf{x} - \mathbf{x}', t - t') = 0 \quad \text{for} \quad t < t',$$

and $1/\epsilon(\mathbf{k}, \omega)$ is analytic in the upper half-plane of the complex frequency ω. Moreover, since the particles of the system cannot respond to very high frequencies $\epsilon(\mathbf{k}, \omega)$ enjoys the property

$$\lim_{\omega\to\infty} 1/\epsilon(\mathbf{k}, \omega) = 1.$$

Thus, one is able to derive the Kramers–Kronig dispersion relations

$$\mathrm{Re}\,\frac{1}{\epsilon(\mathbf{k}, \omega)} = 1 + \frac{2}{\pi} P \times \int_0^\infty d\omega' \frac{\omega'}{\omega'^2 - \omega^2}\,\mathrm{Im}\,\frac{1}{\epsilon(\mathbf{k}, \omega')}, \quad (8)$$

where the symbol P denotes the principal value and Re and Im stand for the real and imaginary parts, respectively. Although the integration in Eq. (8) is performed only over the positive part of the real axis, the entire dielectric function is determined once Im $[1/\epsilon(\mathbf{k}, \omega)]$ is known for real positive values of ω. Finally, the roots of the equation $\epsilon(\mathbf{k}, \omega) = 0$ are of great importance, as they permit the existence of oscillatory modes when $V_{\mathrm{ext}} = 0$. In systems of charged particles the equation $\epsilon = 0$ is known as the "dispersion relation" for the plasma oscillations.

To conclude, we express the dielectric function in terms of the Fourier transforms of V_{ext} and ρ_{ind}. It turns out that this expression can serve for direct calculation of $\epsilon(\mathbf{k}, \omega)$ if ρ_{ind} is defined in terms of V_{ext} and the parameters of the system. If we define

$$f(\mathbf{k}, \omega) = \iint d\mathbf{x}\, dt \exp(i\mathbf{k}\mathbf{x} + i\omega t) f(\mathbf{x}, t), \quad (9)$$

and take the Fourier transform of Eq. (6), we readily find the basic relation

$$\frac{1}{\epsilon(\mathbf{k}, \omega)} = 1 + \lambda\phi(k)\frac{\rho_{\mathrm{ind}}(\mathbf{k}, \omega)}{V_{\mathrm{ext}}(\mathbf{k}, \omega)}. \quad (10)$$

Obviously $\epsilon(\mathbf{k}, \omega)$ is independent of $V_{\mathrm{ext}}(\mathbf{k}, \omega)$ because $\rho_{\mathrm{ind}}(\mathbf{k}, \omega)$ depends linearly on the latter in cases of small disturbances. Thus, V_{ext} enters into the theory only as an auxiliary function. However, the resulting $\epsilon(\mathbf{k}, \omega)$ is exact although first-order perturbation theory is used.

B. Formal Exact Expression for The Dielectric Function

If we denote by $\mathbf{x}_i$ and $\mathbf{p}_i$ the position and the momentum of the ith particle, and by m the mass of a particle, the Hamiltonian of the system is given by

$$H = \sum_i p_i^2/2m + \lambda^2 \sum_{i<j} \phi(|\mathbf{x}_i - \mathbf{x}_j|), \quad (11)$$

where the sum runs over all the N particles. When an external field $V_{\mathrm{ext}}(\mathbf{x}, t)$ is introduced, a term

$$H_{\mathrm{ext}} = \lambda \int d\mathbf{x} V_{\mathrm{ext}}(\mathbf{x}, t) \sum_i \delta(\mathbf{x} - \mathbf{x}_i) \quad (12)$$

is added to H to form the complete Hamiltonian H' of the system.

We consider an ensemble of systems all of the same N, V, and $H' = H + H_{\mathrm{ext}}$. The Liouville equation

$$\partial f/\partial t = [H, f] \quad (13)$$

describes the time development of the ensemble's density, $f(\mathbf{x}_1 \cdots \mathbf{x}_N, \mathbf{P}_1 \cdots \mathbf{P}_N, t)$, in phase space. Equation (13) can be written as

$$\partial f/\partial t = -iL'f \quad (14)$$

in terms of the *Hermitian* Liouville operator L'. It is convenient to separate L' into an "internal" time-independent part

$$L = L_0 + L_I, \quad (15)$$

with

$$L_0 = -\frac{i}{m}\sum_i \mathbf{p}_i\cdot\frac{\partial}{\partial\mathbf{x}_i}, \quad (16)$$

and

$$L_I = i\lambda^2 \sum_{i<j} \frac{\partial\phi(|\mathbf{x}_i - \mathbf{x}_j|)}{\partial\mathbf{x}_i}\left(\frac{\partial}{\partial\mathbf{p}_i} - \frac{\partial}{\partial\mathbf{p}_j}\right), \quad (17)$$

and an "external" time-dependent part

$$L_{\mathrm{ext}}(t) = i\lambda \int d\mathbf{x} V_{\mathrm{ext}}(\mathbf{x}, t) \sum_i \frac{\partial\delta(\mathbf{x} - \mathbf{x}_i)}{\partial\mathbf{x}_i}\cdot\frac{\partial}{\partial\mathbf{p}_i}. \quad (18)$$

The Liouville equation (14) will be solved *formally*, as an initial-value problem along the following lines. We write Eq. (14) as

$$\partial f/\partial t = -iLf - iL_{\mathrm{ext}}f, \quad (19)$$

and substitute

$$\tilde{f}(t) = e^{iLt}f(t)$$

and

$$\tilde{L}_{\text{ext}}(t) = e^{iLt}L_{\text{ext}}(t)\epsilon^{-iLt}, \tag{20}$$

to obtain

$$\partial\tilde{f}/\partial t = -i\tilde{L}_{\text{ext}}\tilde{f}, \tag{21}$$

and consequently

$$\tilde{f}(t) = \tilde{f}(t_0) - i\int_{t_0}^{t} dt'\tilde{L}_{\text{ext}}(t')\tilde{f}(t'). \tag{22}$$

If we now use Eq. (20), we finally have an integral equation for $f(t)$, which reads,

$$f(t) = e^{-iL(t-t_0)}f(t_0) - i\int_{t_0}^{t} dt'e^{-iL(t-t')}L_{\text{ext}}(t')f(t'), \tag{23}$$

where $f(t_0)$ is the initial value of f. To continue, we make two assumptions:

(a) The system is initially ($t_0 \to \infty$) in thermal equilibrium at temperature $T \equiv 1/k\beta$ (k is the Boltzmann constant), namely,

$$f(-\infty) \equiv f_0 = C\exp(-\beta H), \tag{24}$$

where C is a normalization factor, and consequently

$$e^{-iL(t-t_0)}f_0 \to f_0. \tag{25}$$

(b) The applied external force is very weak so that Eq. (23) can be iterated, taking only the linear term in L_{ext} into account. Thus, the formal solution of Eq. (23) is

$$f(t) = f_0 - i\int_{-\infty}^{t} dt'e^{-iL(t-t')} \times L_{\text{ext}}(t')f_0 = f_0 + f_1. \tag{26}$$

Once $f(t)$ is known, the induced density $\rho_{\text{ind}}(\mathbf{x}, t)$ is easily found by the equation

$$\rho_{\text{ind}}(\mathbf{x}, t) = \int d\mathbf{x}_1 \cdots d\mathbf{x}_N\, d\mathbf{p}_1 \cdots d\mathbf{p}_N \times \sum_i \delta(\mathbf{x} - \mathbf{x}_i)f_1(\mathbf{x}_1 \cdots \mathbf{x}_N\mathbf{p}_1 \cdots \mathbf{p}_N; t). \tag{27}$$

Now $\rho_{\text{ind}}(\mathbf{k}, \omega)$ can be given in terms of $V_{\text{ext}}(\mathbf{k}, \omega)$. To this end we rewrite Eq. (18) as

$$L_{\text{ext}}(t)f_0 = \lambda\frac{1}{2\pi V}\int d\omega e^{-i\omega t}\sum_{\mathbf{k}} V_{\text{ext}}(\mathbf{k}, \omega) \times \sum_l e^{-i\mathbf{k}\cdot\mathbf{x}_l}\mathbf{k}\cdot\frac{\partial f_0}{\partial \mathbf{p}_l}, \tag{28}$$

substitute it into Eq. (26), and carry out the time integration; taking into account that Im $\omega > 0$, we obtain

$$f(t) = f_0 - \lambda\frac{1}{2\pi V}\int d\omega e^{-i\omega t} \times \sum_{\mathbf{k}} V_{\text{ext}}(\mathbf{k}, \omega)\frac{1}{L - \omega}\sum_l \epsilon^{-i\mathbf{k}\cdot\mathbf{x}_l}\mathbf{k}\cdot\frac{\partial f_0}{\partial \mathbf{p}_l}. \tag{29}$$

If we take the Fourier transform of Eq. (27) one finds

$$\rho_{\text{ind}}(\mathbf{k}, \omega) = -\frac{\lambda}{V}V_{\text{ext}}(\mathbf{k}, \omega)\int d\mathbf{x}_1 \cdots d\mathbf{p}_N \times \sum_j e^{i\mathbf{k}\cdot\mathbf{x}_j}\frac{1}{L - \omega}\sum_l e^{-i\mathbf{k}\cdot\mathbf{x}_l}\mathbf{k}\cdot\frac{\partial f_0}{\partial \mathbf{p}_l} \tag{30}$$

and, finally, the dielectric function reads

$$\frac{1}{\epsilon(\mathbf{k}, \omega)} = 1 - \frac{\lambda^2}{V}\phi(k)\int d\mathbf{x}_1 \cdots d\mathbf{p}_N \times \sum_j e^{i\mathbf{k}\cdot\mathbf{x}_j}\frac{1}{L - \omega}\sum_l e^{-i\mathbf{k}\cdot\mathbf{x}_l}\mathbf{k}\cdot\frac{\partial f_0}{\partial \mathbf{p}_l}. \tag{31}$$

The last equation gives the exact formal expression for the dielectric function in terms of the system parameters and is independent of the applied field.

Equation (31) can be brought into another form which is more appropriate for physical interpretation, and which is linked to Kubo's theory of transport coefficients[21,22] We denote by

$$n(\mathbf{x}.\ t) \equiv n[\mathbf{x}; \mathbf{x}_i(t)] = \sum_i \delta[\mathbf{x} - \mathbf{x}_i(t)],$$
$$n(\mathbf{k}, t) \equiv n[\mathbf{k}; \mathbf{x}_i(t)] = \sum_i e^{i\mathbf{k}\cdot\mathbf{x}_i(t)}, \tag{32}$$

the particle density "operator" in phase space, and write Eq. (31) as

$$\frac{1}{\epsilon(\mathbf{k}, \omega)} = 1 + \frac{\lambda^2}{V}\phi(k)g(\mathbf{k}, \omega), \tag{33}$$

where $g(\mathbf{k}, \omega)$ is the Fourier transform of

$$g(\mathbf{k}, t - t') = \int d\mathbf{x}_1 \cdots d\mathbf{p}_N \times \sum_j e^{i\mathbf{k}\cdot\mathbf{x}_j}e^{-iL(t-t')}[\sum_l e^{-i\mathbf{k}\cdot\mathbf{x}_l}, f_0] = \int d\mathbf{x}_1 \cdots d\mathbf{p}_N n(\mathbf{k}, t - t')[n(-\mathbf{k}, 0), f_0], \tag{34}$$

or, in terms of the average symbol $\langle\cdots\rangle$,

$$g(\mathbf{k}, t) = \langle[n(-\mathbf{k}, 0), n(\mathbf{k}, t)]\rangle. \tag{35}$$

The function $g(t)$ is known as the after-effect function[21] and it describes a correlation between

[21] R. J. Kubo, J. Phys. Soc. Japan **12**, 570 (1957).
[22] R. Balescu, Physica **27**, 693 (1961); S. F. Edwards and J. J. Sanderson, Phil. Mag. **6**, 71 (1961).

the density fluctuation at two successive times. This autocorrelation function is directly related to the dynamics of the system in the absence of any external disturbance.

C. Simple Example: Systems of Free Particles

Before we go on to treat systems of interacting particles, it is interesting to discuss briefly a model system where the particles do not interact with each other. In this case

$$L \to L_0 = -\frac{i}{m}\sum_i \mathbf{p}_i\cdot\frac{\partial}{\partial \mathbf{x}_i}, \tag{36}$$

$$f_0 \to (Va^3)^{-N}\exp\left(-\beta\sum_i p_i^2/2m\right), \tag{37}$$

where

$$a = \int_{-\infty}^{\infty} dp\,\exp(-\beta p^2/2m), \tag{38}$$

and consequently,

$$\frac{1}{L_0-\omega}e^{-i\mathbf{k}\cdot\mathbf{x}_l} = \frac{1}{-\mathbf{k}\cdot\mathbf{p}_l/m-\omega}e^{-i\mathbf{k}\cdot\mathbf{x}_l}. \tag{39}$$

Thus, Eq. (31) reads

$$\frac{1}{\epsilon(\mathbf{k},\omega)} = 1+\frac{\lambda^2}{V}\phi(k)\int d\mathbf{x}_1\cdots d\mathbf{x}_N\sum_{j,l}e^{i\mathbf{k}\cdot(\mathbf{x}_j-\mathbf{x}_l)} \times (Va^3)^{-N}\int d\mathbf{p}_1\cdots d\mathbf{p}_N\frac{1}{\mathbf{k}\cdot\mathbf{p}/m+\omega} \times \mathbf{k}\cdot\frac{\partial}{\partial\mathbf{p}_l}\exp\left(-\beta\sum_i P_i^2/2m\right), \tag{40}$$

or, after some simple integrations,

$$\frac{1}{\epsilon(\mathbf{k},\omega)} = 1+\frac{\lambda^2 N}{mV}\phi(k)I(k,\omega), \tag{41}$$

where

$$I(k,\omega) = \int_{-\infty}^{\infty}\frac{du}{u+\omega/k}\frac{\partial F}{\partial u}, \tag{42}$$

and $F(u)$ is the normalized Maxwellian distribution

$$F(u) = \exp(-\beta m u^2/2) \times\left[\int_{-\infty}^{\infty}du\exp(-\beta m u^2/2)\right]^{-1}. \tag{43}$$

The function $I(\mathbf{k},\omega)$ is the same one occurring in the ordinary plasma oscillation theory. For real frequencies one has

$$I(k,\omega) = \lim_{\epsilon\to 0}\int_{-\infty}^{\infty}du\frac{1}{u+\omega/k+i\epsilon}\frac{\partial F}{\partial u},$$

or

$$I(k,\omega) = -\beta[1+i(\pi m\beta/2)^{\frac{1}{2}}\omega/k \times\exp(-m\beta\omega^2/2k^2)\{1+\Phi[i\omega/k(m\beta/2)^{\frac{1}{2}}]\}], \tag{45}$$

with

$$\Phi(\mathbf{x}) = \pi^{-\frac{1}{2}}\int_0^x dt\exp(-t^2).$$

It is easy to verify that for high frequencies or long wavelengths

$$I(k,\omega)\to k^2/\omega^2 \tag{46}$$

and the dielectric function takes the form

$$1/\epsilon(k,\omega) = 1+\lambda^2 n\phi(k)k^2/m\omega^2. \tag{47}$$

III. PERTURBATION ANALYSIS

The dielectric function, as given by Eq. (31), is a complicated formal expression which is not mathematically tractable. Prigogine and Balescu (PB) in their theory of irreversible processes[8,11] have developed a systematic procedure by which this problem may be treated. The application of PB theory to our problem is carried out along the following lines:

(a) A Fourier expansion of the intergand of Eq. (31) in phase space.

(b) A formal perturbation expansion of the operator $(L-\omega)^{-1}$ in terms of the interaction factor λ^2.

(c) A classification of the terms of the expansion according to their asymptotic dependence on λ^2, N, and V. In order to classify the various terms, a diagram technique is developed and this enables one to choose, without calculation, those terms which are important to a given problem.

It is important to remark here that, the difficulty of PB theory with respect to time scale of evolution is not met here, because we have made no assumptions about asymptotic time dependence.

A. Fourier Expansion

We begin with writing the integrand of Eq. (31) in terms of the Fourier transforms of its components. First we define the Fourier representation[11] in the phase space

$$|\mathbf{k}_1\cdots\mathbf{k}_N\rangle \equiv |\{\mathbf{k}\}\rangle = \exp(i\sum_j \mathbf{k}_j\cdot\mathbf{x}_j), \tag{48}$$

the Fourier transform of a function $f(\{\mathbf{x}\}) \equiv f(\mathbf{x}_1\cdots\mathbf{x}_N)$,

$$f(\{\mathbf{k}\}) \equiv f(\mathbf{k}_1\cdots\mathbf{k}_N) = \int d\mathbf{x}_1\cdots d\mathbf{x}_N\exp(i\sum_j\mathbf{k}_j\cdot\mathbf{x}_j)f(\mathbf{x}_1\cdots\mathbf{x}_N), \tag{49}$$

and the Fourier matrix of an operator $A(\mathbf{x}_1\cdots\mathbf{x}_N)$

$$\langle \mathbf{k}_1 \cdots \mathbf{k}_N | A | \mathbf{k}_1' \cdots \mathbf{k}_N' \rangle = \langle \{\mathbf{k}\} | A | \{\mathbf{k}'\} \rangle$$
$$= \int d\mathbf{x}_1 \cdots d\mathbf{x}_N \exp(+i \sum_i \mathbf{k}_i \cdot \mathbf{x}_i) \times A(\mathbf{x}_1 \cdots \mathbf{x}_N) \exp(-i \sum_l \mathbf{k}_l' \cdot \mathbf{x}_l). \quad (50)$$

Now the equilibrium N-body distribution function can be written as

$$f_0(\mathbf{x}_1 \cdots \mathbf{x}_N; \mathbf{p}_1 \cdots \mathbf{p}_N) = F(\mathbf{x}_1 \cdots \mathbf{x}_N) G(\mathbf{p}_1 \cdots \mathbf{p}_N), \quad (51)$$

with

$$F(\mathbf{x}_1 \cdots \mathbf{x}_N) = [Z(\lambda^2, N, V)]^{-1} \times \exp\{-\beta\lambda^2 \sum_{i<j} \phi(|\mathbf{x}_i - \mathbf{x}_j|)\}, \quad (52)$$

and

$$Z(\lambda^2, N, V) = \int d\mathbf{x}_1 \cdots d\mathbf{x}_N \times \exp\{-\beta\lambda^2 \sum_{i<j} \phi(|\mathbf{x}_i - \mathbf{x}_j|), \quad (53)$$

and

$$G(\mathbf{p}_1 \cdots \mathbf{p}_N) = a^{-3N} \exp(-\beta \sum_i p_i^2/2m), \quad (54)$$

where

$$a = \int_{-\infty}^{\infty} dp \exp(-\beta p^2/2m). \quad (55)$$

If we now insert Eqs. (50)–(55) into Eq. (31) we obtain

$$1/\epsilon(\mathbf{q}, \omega) = 1 - \lambda^2 (N/V) \phi(q) J(\mathbf{q}, \omega), \quad (56)$$

where

$$J(q, \omega) = \int d\mathbf{p}_1 \cdots d\mathbf{p}_N \sum_{\mathbf{k}_1 \cdots \mathbf{k}_N} \sum_l \langle 0, \cdots, -\mathbf{q}, \cdots 0 | (L - \omega)^{-1} | \mathbf{k}_1 \cdots \mathbf{k}_N \rangle \times \tilde{F}(\mathbf{k}_1, \cdots, \mathbf{k}_l + \mathbf{q}, \cdots \mathbf{k}_N) \mathbf{q} \cdot \frac{\partial}{\partial \mathbf{p}_l} G(\mathbf{p}_1 \cdots \mathbf{p}_N). \quad (57)$$

In transforming Eq. (31) we have used the fact that the summation over j can be replaced by N and the "wave vector" $\langle 0 \cdots, -\mathbf{q}, \cdots 0|$ represents a "state" where all the k's are zero but one, which is equal to $-\mathbf{q}$. Sometimes we shall refer to this "state" as a state where one particle (the αth, say) has a wave-number $-\mathbf{q}$ and the other particles have zero wavenumbers. One should notice that the Fourier transform of $F(\mathbf{x}_1 \cdots \mathbf{x}_N)$ as it appears in Eq. (57) is "shifted" by $\mathbf{q}$ due to the presence of $\exp\{-i\mathbf{q} \cdot \mathbf{x}_l\}$ in Eq. (31).

One of the main reasons for using the Fourier transformation in the phase space is that the components, $\tilde{F}(\mathbf{k}_1 \cdots \mathbf{k}_N)$, of the expansion of $F(\mathbf{x}_1 \cdots \mathbf{x}_N)$ are easily classified in terms of λ^2, N, and V, while $F(\mathbf{x}_1 \cdots \mathbf{x}_N)$ itself is not. This feature of $\tilde{F}(\mathbf{k}_1 \cdots \mathbf{k}_N)$ is of importance in what follows.

We define $\tilde{F}$ by

$$F(\mathbf{x}_1 \cdots \mathbf{x}_N) = V^{-N} \sum_{\mathbf{k}_1 \cdots \mathbf{k}_N} \tilde{F}(\mathbf{k}_1 \cdots \mathbf{k}_N) \exp(-i \sum_i \mathbf{k}_i \cdot \mathbf{x}_i). \quad (58)$$

On the other hand, one can expand $F(\mathbf{x}_1 \cdots \mathbf{x}_N)$ formally in Eq. (52) in terms of λ^2, yielding

$$V^N F(\mathbf{x}_1 \cdots \mathbf{x}_N) = 1 + \lambda^2\beta\Big[- \sum_{i<j} \phi(|\mathbf{x}_i - \mathbf{x}_j|) + V^{-N} \int d\mathbf{x}_1 \cdots d\mathbf{x}_N \sum_{i<j} \phi(|\mathbf{x}_i - \mathbf{x}_j|)\Big] + (\lambda^4\beta^2/2!)\Big\{ \sum_{i<j} \phi(|\mathbf{x}_i - \mathbf{x}_j| \sum_{l<m} \phi(|\mathbf{x}_l - \mathbf{x}_m|) + 2!\, V^{-2N}\Big[\int d\mathbf{x}_1 \cdots d\mathbf{x}_N \sum_{i<j} \phi(|\mathbf{x}_i - \mathbf{x}_j|)\Big]^2 - V^{-N} \int d\mathbf{x}_1 \cdots d\mathbf{x}_N \sum_{i<j} \phi(|\mathbf{x}_i - \mathbf{x}_j|) \times \sum_{l<m} \phi(|\mathbf{x}_l - \mathbf{x}_m|) - 2V^{-N} \sum_{i<j} \phi(|\mathbf{x}_i - \mathbf{x}_j|) \times \int d\mathbf{x}_1 \cdots d\mathbf{x}_N \sum_{l<m} \phi(|\mathbf{x}_l - \mathbf{x}_m|)\Big\} + O(\lambda^6\beta^3) + \cdots. \quad (59)$$

In Eq. (59) the terms which include integrals are derived from the expansion of Z^{-1}, and those without integrals are due to $\exp(-\beta H_I)$. It is of importance to notice that the termination of the latter series on the basis of the smallness of λ^2 alone is not justified due to the summation involved in each term, i.e., N is large. The way out of this difficulty is to equate Eqs. (58) and (59) and to evaluate term by term the Fourier transforms to $F(\mathbf{x}_1 \cdots \mathbf{x}_N)$ which removes the explicit dependence on N. In order to demonstrate this fact we first integrate both sides over all the $\mathbf{x}$'s, yielding

$$\tilde{F}(0 \cdots 0) = 1, \quad (60)$$

then we integrate over all $\mathbf{x}$'s but one, to get

$$\tilde{F}(0 \cdots \mathbf{k} \cdots 0) = 0. \quad (61)$$

Next we integrate over all $\mathbf{x}$'s but two, and find for $\mathbf{k} \neq 0$

$$\tilde{F}(0 \cdots \mathbf{k} \cdots -\mathbf{k} \cdots 0) = -(\lambda^2\beta/V)\phi(\mathbf{k}) + (\lambda^4\beta^2/2!\, V^2) \sum_{k'} \phi(\mathbf{k} - \mathbf{k}')\phi(\mathbf{k}') + \cdots, \quad (62)$$

and

$$\tilde{F}(0, \cdots \mathbf{k} \cdots \mathbf{k}' \cdots 0) = 0, \quad \text{for } \mathbf{k}' \neq \mathbf{k}. \tag{63}$$

Equations (61) and (63) are special cases of a general rule, which homogeneous systems obey,

$$\tilde{F}(\mathbf{k}_1 \cdots \mathbf{k}_N) = 0, \quad \text{for } \sum_i \mathbf{k}_i \neq 0. \tag{64}$$

The procedure outlined above can be continued to yield a hierarchy of terms with clear dependence on λ^2 and V, e.g.,

$$\begin{aligned} &\tilde{F}(0 \cdots \mathbf{k} \cdots \mathbf{k}' \cdots -\mathbf{k} \cdots -\mathbf{k}' \cdots 0) \\ &\qquad = O(\lambda^4/V^2), \\ &\tilde{F}(0 \cdots \mathbf{k} \cdots \mathbf{k}' \cdots -\mathbf{k} - \mathbf{k}' \cdots 0) \\ &\qquad = O(\lambda^4/V^2), \end{aligned} \tag{65}$$

and so on.

We conclude this section with some remarks:

(a) The Fourier transforms $\tilde{F}(\mathbf{k}_1 \cdots \mathbf{k}_N)$ have simple physical interpretation in terms of the n-body distribution function. For example, the density of the particles is given by

$$n(\mathbf{x}) = n[\tilde{F}(0 \cdots 0) + V^{-1} \times \sum_{\mathbf{k}} \tilde{F}(0 \cdots \mathbf{k} \cdots 0) \exp(i\mathbf{k}\mathbf{x})]. \tag{66}$$

This aspect has been discussed at length by PB, and we refer the reader to their work.[8,11]

(b) The dependence of the $\tilde{F}$'s on V is very simple, although it is formally written as a power series in V^{-1}. To illustrate this point we take Eq. (62) as an example. The first term of the right-hand side of Eq. (62) is of order V^{-1}, while the second one is apparently of order V^{-2}, and so on. But, in fact, one of the V's in the second term is "canceled" against the summation over $\mathbf{k}'$ and the same "cancellation" occurs in the higher terms of Eq. (62), yielding

$$\begin{aligned} &\tilde{F}(0 \cdots \mathbf{k} \cdots -\mathbf{k} \cdots 0) \\ &\quad = V^{-1}\{-\lambda^2\beta\phi(k) + \text{term of order } \lambda^4 \\ &\qquad + \text{term of order } \lambda^6 + \cdots\}. \end{aligned} \tag{67}$$

One easily sees that this rule can be applied to all other $\tilde{F}$'s.

(c) In the case of a plasma, the $\tilde{F}$' sare classified "naturally" in terms of the plasma dimensionless factor $e^2n^{\frac{1}{3}}/kT$ and N. Taking $\lambda = e$, where e is the charge of a particle and $\phi(k) = 4\pi/k^2$, we can write, for example,

$$\begin{aligned} &F(0 \cdots \mathbf{q} \cdots -\mathbf{q} \cdots 0) \\ &\quad = N^{-1}\{e^2n^{\frac{1}{3}}/kT)4\pi n^{\frac{2}{3}}/q^2 + \cdots\}, \end{aligned} \tag{68}$$

and so on. Thus, the Fourier decomposition of the N-particle distribution function in the phase space is a "natural" perturbation expansion.

B. Formal Expansion of $(L - \omega)^{-1}$

We shall now investigate thoroughly the expansion into infinite series of the operator

$$R(\omega) \equiv (L - \omega)^{-1} = (L_0 - \omega + L_I)^{-1} \tag{69}$$

in terms of the interaction operator L_I. If we use the identity

$$\begin{aligned} (L_0 - \omega + L_I)^{-1} &= (L_0 - \omega)^{-1} \\ &\quad - (L_0 - \omega)^{-1}L_I(L_0 - \omega + L_I)^{-1}, \end{aligned} \tag{70}$$

or, with $R_0(\omega) = (L_0 - \omega)^{-1}$,

$$R(\omega) = R_0(\omega) - R_0(\omega)L_IR(\omega), \tag{71}$$

we obtain, by iteration, an infinite series

$$R(\omega) = \sum_{n=0}^{\infty} R_0(\omega)[-L_IR_0(\omega)]^n. \tag{72}$$

The convergence of this series is assumed *a priori*, while it should be justified *a posteriori* when it applies to the function considered in any given problem.

Expressing Eq. (72) in the Fourier representation $|\{\mathbf{k}\}\rangle$, we get

$$\begin{aligned} \langle\{\mathbf{k}\}|\, R(\omega)\, |\{\mathbf{k}'\}\rangle &= \sum_{n=0}^{\infty} \sum_{\{\mathbf{k}''\}} \langle\{\mathbf{k}\}|\, R_0(\omega)\, |\{\mathbf{k}''\}\rangle \\ &\quad \times \langle\{\mathbf{k}''\}|\, [-L_IR_0(\omega)]^n\, |\{\mathbf{k}'\}\rangle. \end{aligned} \tag{73}$$

The advantages of this representation are twofold:

(a) $|\{\mathbf{k}\}\rangle$ is an eigenfunction of the operator $R_0(\omega)$ and, thus,

$$\begin{aligned} &\langle\{\mathbf{k}\}|\, R_0(\omega)\, |\{\mathbf{k}'\}\rangle \\ &\quad = +(\sum_i \mathbf{k}_i \cdot \mathbf{p}_i/m - \omega)^{-1}\, \delta_{\mathrm{Kr}}(\{\mathbf{k}\} - \{\mathbf{k}'\}) \end{aligned} \tag{74}$$

is diagonal. Here δ_{Kr} stands for the Kronecker's symbol.

(b) The matrix elements of the interaction operator L_I exhibit the property of the so-called *law of conservation of the wave vectors*. To see this, we first Fourier analyze the potential

$$\phi(|\mathbf{x}_i - \mathbf{x}_l|) = V^{-1} \sum_l \phi(l)e^{i\mathbf{l}\cdot(\mathbf{x}_l - \mathbf{x}_j)}; \tag{75}$$

then the "Fourier matrix" of $L_I = \sum_{i<j} L_I(ij)$ becomes

$$\begin{aligned} &\langle\{\mathbf{k}\}|\, L_I(ij)\, |\{\mathbf{k}'\}\rangle \\ &\quad = \frac{\lambda^2}{V} \sum_l \phi(l)\, \delta_{\mathrm{Kr}}(\mathbf{k}_i - \mathbf{k}_i' + \mathbf{l})\, \delta_{\mathrm{Kr}}(\mathbf{k}_j - \mathbf{k}_j' - \mathbf{l}) \end{aligned}$$

$$\times [\prod_{m\neq i\,j} \delta_{\mathrm{Kr}}(\mathbf{k}_m - \mathbf{k}'_m)]1\cdot\left(\frac{\partial}{\partial \mathbf{p}_i} - \frac{\partial}{\partial \mathbf{p}_j}\right)$$

$$= -\frac{\lambda^2}{V}\phi(|\mathbf{k}_i - \mathbf{k}'_i|)\ \delta_{\mathrm{Kr}}(\mathbf{k}_i + \mathbf{k}_j - \mathbf{k}'_i - \mathbf{k}'_j)$$

$$\times [\prod_{m\neq i\,j} \delta_{\mathrm{Kr}}(\mathbf{k}_m - \mathbf{k}'_m)](\mathbf{k}_i - \mathbf{k}'_i)\cdot\left(\frac{\partial}{\partial \mathbf{p}_i} - \frac{\partial}{\partial \mathbf{p}_j}\right). \tag{76}$$

It is clear from Eq. (76) that the wave vectors are restricted: The particles i and j, which interact through $\phi(|\mathbf{x}_i - \mathbf{x}_j|)$, can only exchange their wave vectors, namely,

$$\mathbf{k}_i + \mathbf{k}_j = \mathbf{k}'_i + \mathbf{k}'_j,$$

while the other particles preserve the same wave vectors

$$\mathbf{k}_m = \mathbf{k}'_m, \qquad m \neq i, j.$$

It should be noted that the matrix elements of Eq. (76) are still operators in the momentum space.

We shall now employ the expansion of $(L - \omega)^{-1}$ by substituting Eq. (72) into Eq. (57) and use Eqs. (74) and (76). We write

$$J(\mathbf{q}, \omega) = \sum_{n=0}^{\infty} J_n(\mathbf{q}, \omega), \tag{77}$$

where the index n corresponds to the nth power of the right-hand side of Eq. (72). The first terms of this series are

$$J_0(\mathbf{q}, \omega) = -\int d\mathbf{p}_1 \cdots d\mathbf{p}_N$$

$$\times (\omega + \mathbf{q}\cdot\mathbf{p}_\alpha/m)^{-1}\left[\tilde{F}(0 \cdots 0)\mathbf{q}\cdot\frac{\partial G(\mathbf{p}_1 \cdots \mathbf{p}_N)}{\partial \mathbf{p}_\alpha}\right.$$

$$\left. + \sum_{l\neq\alpha} \tilde{F}(0 \cdots \overset{\alpha}{-\mathbf{q}} \cdots \overset{l}{\mathbf{q}} \cdots 0)\mathbf{q}\cdot\frac{\partial G(\mathbf{p}_1 \cdots \mathbf{p}_N)}{\partial \mathbf{p}_l}\right]; \tag{78}$$

$$J_1(\mathbf{q}, \omega) = \int d\mathbf{p}_1 \cdots d\mathbf{p}_N(\omega + \mathbf{q}\cdot\mathbf{p}_\alpha/m)^{-1}$$

$$\times \sum_{i<j} \sum_{\mathbf{k}_1 \cdots \mathbf{k}_N} \langle 0 \cdots \overset{\alpha}{-\mathbf{q}} \cdots 0|\ L_I(ij)\ |\mathbf{k}_1 \cdots \mathbf{k}_N\rangle$$

$$\times (\omega - \sum_i \mathbf{k}_i\cdot\mathbf{p}_i/m)^{-1}$$

$$\times \sum_l \tilde{F}(\mathbf{k}_1 \cdots \mathbf{k}_l + \mathbf{q} \cdots \mathbf{k}_N)\mathbf{q}\cdot\frac{\partial G(\mathbf{p}_1 \cdots \mathbf{p}_N)}{\partial \mathbf{p}_l}; \tag{79}$$

$$J_2(\mathbf{q}, \omega) = -\int d\mathbf{p}_1 \cdots d\mathbf{p}_N(\omega + \mathbf{q}\cdot\mathbf{p}_\alpha/m)^{-1}$$

$$\times \sum_{i<j} \sum_{\mathbf{k}_1 \cdots \mathbf{k}_N} \langle 0 \cdots \overset{\alpha}{-\mathbf{q}} \cdots 0|\ L_I(ij)\ |\mathbf{k}_1 \cdots \mathbf{k}_N\rangle$$

$$\times (\omega - \sum_i \mathbf{k}_i\cdot\mathbf{p}_i/m)^{-1}$$

$$\times \sum_{l<m} \sum_{\mathbf{k}_1' \cdots \mathbf{k}_N'} \langle \mathbf{k}_1 \cdots \mathbf{k}_N|\ L_I(lm)\ |\mathbf{k}'_1 \cdots \mathbf{k}'_N\rangle$$

$$\times (\omega - \sum_i \mathbf{k}_i\cdot\mathbf{p}_i/m)^{-1}$$

$$\times \sum_l \tilde{F}(\mathbf{k}'_1 \cdots \mathbf{k}'_l + \mathbf{q} \cdots \mathbf{k}'_N)\mathbf{q}\cdot\frac{\partial G(\mathbf{p}_1 \cdots \mathbf{p}_N)}{\partial \mathbf{p}_l}, \tag{80}$$

and so on. In the above equations α indicates one "known" (fixed) particle, $\tilde{F}(0 \cdots \overset{\alpha}{-\mathbf{q}} \cdots \overset{l}{\mathbf{q}} \cdots 0)$ stands for the Fourier transform of $F(\mathbf{x}_1 \cdots \mathbf{x}_N)$ with

$$\mathbf{k}_\alpha = -\mathbf{q}, \quad \mathbf{k}_l = \mathbf{q}, \quad \text{and} \quad \mathbf{k}_i = 0 \quad \text{for} \quad i \neq \alpha, l,$$

and the "state" $\langle 0 \cdots \overset{\alpha}{-\mathbf{q}} \cdots 0|$ stands for the Fourier vector $|\{k\}\rangle$ with

$$\mathbf{k}_\alpha = \mathbf{q} \quad \text{and} \quad \mathbf{k}_i = 0 \quad \text{for} \quad i \neq \alpha.$$

C. Diagrams

The study of Eq. (57) becomes much easier if one uses the diagram technique developed by PB. The advantage of this method is that the terms of the series of Eqs. (57) and (77) can be classified with respect to their dependence on λ^2, N, and V, just by applying some simple topological arguments to the diagrams representing those terms. Once the diagrams (terms) are classified they can be resummed according to an expansion parameter, which is determined by the actual physical problem under consideration. We shall give here a brief review of PB's general method, just enough for the reader to follow the treatment of the present problem. For a more general discussion of the method, the reader is referred to PB's original work.

We begin with a diagram representation of Eq. (77). Any term of the series is represented by means of a diagram. The diagram is composed of lines and vertices, where each line corresponds to a wave vector $\mathbf{k} \neq 0$ and each vertex to an interaction matrix element. As the term is read from right to left, the corresponding diagram is drawn from right to left. With each expression

$$-(\omega - \sum_i \mathbf{k}_i\cdot\mathbf{p}_i/m)^{-1},$$

often called a "free propagator," a number of lines are associated equal to the number of nonvanishing $\mathbf{k}$'s in that expression. With each matrix element of $L_I(ij)$ a vertex is associated, representing the exchange of the wave vectors of the ith and the jth particles. Since only six independent processes

of exchanging wave vectors are consistent with Eq. (76), there are only six "basic" diagrams to represent the interaction. Thus, the general diagram is composed of those six "basic diagrams" connected by "free propagator" lines.

First, we study briefly the basic diagrams and summarize their properties. It is convenient to associate with each basic diagram the $\mathbf{k}$ summation preceding its corresponding matrix element [Eq. (76)] and to *prescribe* the particles and their wave numbers on the left side of the diagram. Now a "dictionary" is constructed with which one is able to translate back and forth between a diagram and its corresponding expression.

Greek letters are used to denote "known" particles (either "fixed" particles or ones occurring in the summation of preceding basic diagram) and *Roman* letters for particles over which a summation has to be performed. The "prescribed" wave vectors (in the same sense as "known" particles) are denoted by $\mathbf{k}_1$ and $\mathbf{k}_2$, while $\mathbf{k}$ corresponds to the dummy wave vector. From Eq. (76) we get the following "dictionary" for the basic diagrams of Fig. 1.

$$-(\lambda^2/V)\phi(k_1)\sum_{j\neq\alpha}\mathbf{k}_1\cdot\left(\frac{\partial}{\partial\mathbf{p}_\alpha}-\frac{\partial}{\partial\mathbf{p}_j}\right), \qquad \text{(A)}$$

$$(\lambda^2/V)\sum_{\mathbf{k}}\phi(k)\sum_{i<j}\mathbf{k}\cdot\left(\frac{\partial}{\partial\mathbf{p}_i}-\frac{\partial}{\partial\mathbf{p}_j}\right), \qquad \text{(B)}$$

$$(\lambda^2/V)\phi(k_1)\mathbf{k}_1\cdot\left(\frac{\partial}{\partial\mathbf{p}_\alpha}-\frac{\partial}{\partial\mathbf{p}_\beta}\right), \qquad \text{(C)}$$

$$-(\lambda^2/V)\sum_{k}\phi(k)\sum_{j\neq\alpha}\mathbf{k}\cdot\left(\frac{\partial}{\partial\mathbf{p}_\alpha}-\frac{\partial}{\partial\mathbf{p}_j}\right), \qquad \text{(D)}$$

$$(\lambda^2/V)\phi(k_2)\mathbf{k}_2\cdot\left(\frac{\partial}{\partial\mathbf{p}_\alpha}-\frac{\partial}{\partial\mathbf{p}_\beta}\right), \qquad \text{(E)}$$

$$(\lambda^2/V)\sum_{\mathbf{k}}\phi(|\mathbf{k}_1-\mathbf{k}|)(\mathbf{k}_1-\mathbf{k})\cdot\left(\frac{\partial}{\partial\mathbf{p}_\alpha}-\frac{\partial}{\partial\mathbf{p}_\beta}\right). \qquad \text{(F)}$$

It is now very easy to classify the basic diagrams (A)–(F) according to their asymptotic dependence on λ^2, N, and V. For convenience, Table I gives a "topological index" corresponding to r, s, t of $(\lambda^2)^r(N)^s(V)^t$.

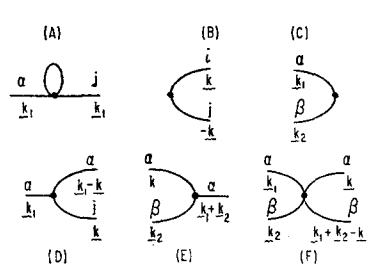

FIG. 1. The basic diagrams.

TABLE I. Topological indexes of the basic diagrams.

	r	s	t
A	1	1	−1
B	1	2	0
C	1	0	−1
D	1	1	0
E	1	0	−1
F	1	0	0

In Table I we considered the summation over $\mathbf{k}$ as another V factor. It should be noted here that our topological indexes differ slightly from those of PB, due to the fact that PB included the "boundaries" of the diagrams in their asymptotic arguments, while we find it advantageous to choose another convention. This will be explained later.

Once Table I is given, one can determine the asymptotic λ^2, N, and V dependence of any diagram just by adding the topological indexes of the basic diagrams composing the given diagram.[23] Now, in order to determine the asymptotic behavior of the terms of Eq. (77) one should also consider the "boundaries" of the diagrams corresponding to each term. By "boundaries" we mean the $\tilde{F}(\mathbf{k}_1\cdots\mathbf{k}_N)$ terminating each of the expressions $J_n(\mathbf{q},\omega)$ of Eq. (77). Because the $\tilde{F}$'s have their own asymptotic λ^2, N, and V dependence, the asymptotic behavior of each term is obtained by adding the topological indexes of the corresponding diagram to the indexes of the "boundaries."

As an illustration, we shall study one of the diagrams corresponding to $J_2(\mathbf{q},\omega)$ of Eq. (80) [Fig. 2]. The "body" of the diagram [Fig. 2(a)] contributes $\lambda^4N^2V^{-1}(r=1+1=2, s=1+1=2, t=-1+0=-1)$, while the contribution of the "boundary" depends on the actual $\tilde{F}$ fitting the right "end" of the diagram. Two of the possibilities are given by

$$\tilde{F}(0\cdots-\mathbf{q}-\overset{i}{\mathbf{k}}\cdots\overset{l=m}{\mathbf{k}}+\mathbf{q}\cdots 0)$$

and

$$\tilde{F}(0\cdots-\overset{\alpha}{\mathbf{q}}-\mathbf{k}\cdots\overset{l}{\mathbf{q}}\cdots\overset{m}{\mathbf{k}}\cdots 0),$$

corresponding to Figs. 2(b) and 2(c), respectively. The first $\tilde{F}$ is of order $\lambda^{2+2n}N/V(n=0,1,\cdots)$, where N is due to the summation over l. We conclude that the diagram of Fig. 2 is asymptotically of order

$$\lambda^{6+2n}N^3/V^2, \qquad n=0,1\cdots.$$

To complete the treatment of the asymptotic λ^2, N, V dependence of the terms of Eq. (77), we

[23] There is one exception [see footnote on p. 292 of PBI (reference 8)], which is of no importance in our case.

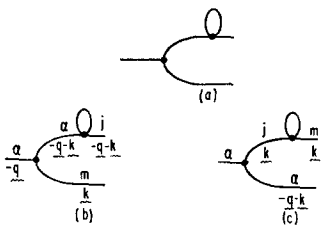

FIG. 2. An example of a second-order diagram.

add the first few terms of a "dictionary" for the contribution of the "boundaries."

(1) A diagram ending with one line corresponds to a term terminated with either

$$\tilde{F}(0 \cdots 0) \sum_l \mathbf{q} \frac{\partial G}{\partial \mathbf{p}_l} \to (\lambda^2)^0 N^1 V^0, \qquad (81a)$$

or

$$\tilde{F}(0 \cdots \mathbf{q} \cdots -\mathbf{q} \cdots 0) \sum_l \mathbf{q} \cdot \frac{\partial G}{\partial \mathbf{p}_l} \to (\lambda^2)^n N^{-1} N^{-1} \qquad n = 1, 2 \cdots . \qquad (81b)$$

(2) A diagram ending with two lines corresponds to either

$$\tilde{F}(0 \cdots \mathbf{q} + \mathbf{k} \cdots -\mathbf{q} - \mathbf{k} \cdots 0) \sum_l \mathbf{q} \cdot \frac{\partial G}{\partial \mathbf{p}_l} \to (\lambda^2)^n N^1 V^{-1} \qquad n = 1, 2 \cdots , \qquad (82a)$$

or

$$F(0 \cdots -\mathbf{q} - \mathbf{k} \cdots \mathbf{q} \cdots -\mathbf{k} \cdots 0) \sum_l \mathbf{q} \cdot \frac{\partial G}{\partial \mathbf{p}_l} \to (\lambda^2)^{2+n} N^1 V^{-2} \qquad n = 0, 1 \cdots , \qquad (82b)$$

and so on.

We conclude this section with some comments on the diagrams corresponding to $J(\mathbf{q}, \omega)$.

(a) Each diagram begins with *only* one external line on its left. This line represents a *"fixed"* particle (α, say) with a given wave vector $-\mathbf{q}$.

(b) The law of conservation of wave vectors implies that if a cut is made across a diagram somewhere, the sum of the wave vectors is $-\mathbf{q}$.

(c) Diagrams that can be divided into two or more parts without cutting lines are called *disconnected*. All other diagrams are *connected*. The connected diagrams [satisfying (a) and (b) above] contribute to $J(\mathbf{q}, \omega)$, while only those disconnected diagrams, in which each of their parts has a particle in common with another part (or has a "fixed" particle), give a contribution to $J(\mathbf{q}, \omega)$. This can be seen as follows. In a disconnected diagram only one part has the external line of the diagram. The other parts begin with a vertex of type (B) associated with an operator $\partial/\partial \mathbf{p}_i - \partial/\partial \mathbf{p}_j$ acting on its right. But due to the integration over all the $\mathbf{p}$'s the integrals over $\mathbf{p}_i$ and $\mathbf{p}_j$ are vanished unless the operator $\partial/\partial \mathbf{p}_i - \partial/\partial \mathbf{p}_j$ is preceded (on its left) by a function of $\mathbf{p}_i$ or $\mathbf{p}_j$. The latter diagrams are called *semiconnected*. Thus, the topological index, corresponding to N, of a semiconnected part of a diagram reduces by one.

(d) The momentum operator $\partial/\partial \mathbf{p}_i - \partial/\partial \mathbf{p}_j$, corresponding to a vertex, sometimes reduces to $\partial/\partial \mathbf{p}_i$ or $-\partial/\partial \mathbf{p}_j$ if the vertex is not preceded by a line with i or j. The reasons for this are the same as in (c).

(e) The same arguments can be applied to the summation of l at the end of each term of Eq. (77). This implies that the contribution of the "boundary" [Eqs. (81) and (82)] is N independent. Thus, the asymptotic index of $\tilde{F}$ is always $t = 0$.

(f) The above comments lead to the important result that all diagrams satisfy asymptotically the relation

$$s + t \leq 0,$$

which means that the diagrams are not divergent with respect to the limits $N \to \infty$, $V \to \infty$, $N/V =$ finite. There is always a compensating V^{-1} to any vertex which generates an N. This fact may be considered as one of the main advantages of the Fourier representation in the phase space.

IV. PARTICULAR CASES

We now apply the method developed in the last section to some examples of systems of many interacting particles. We shall obtain the dominant contribution to the dielectric function in a systematic expansion in terms of the systems parameters. It is just for the right approximations that the PB method was developed. To this end, we can classify our systems according to certain expansion parameters, which are combinations of λ^2, N, and V, and then regroup and resum the dominant diagrams (the leading asymptotic terms). It is clear that the suitable expansion parameters should be decided on physical ground. Here we discuss two simple particular cases:(1) a system where the interparticle forces are very weak and (2) a plasma with Coulomb interaction.

A. Weak Interaction

Under the category of weak interaction we classify systems of many particles satisfying the following conditions:

(a) The average potential energy of interaction between two particles is much smaller than the

average kinetic energy per particle. A dimensionless parameter λ^2 is defined to measure the strength of the interaction, so that

$$\lambda^2 \ll 1.$$

(b) The weakness of the interaction is independent of the density. This corresponds to the assumption that even if the system becomes denser and denser (and the particles become closer and closer), the total effect of the interaction is still small. Formally this condition is formulated as

$$\lambda^2 n \ll 1;$$

thus, λ^2 is a legitimate expansion parameter of the system.

In order to get a systematic expansion of the dielectric function for systems of weak interaction, in terms of λ^2, we regroup the terms (diagrams) of Eqs. (57) or (77) with respect to their λ^2 dependence. We rewrite

$$1/\epsilon(\mathbf{q}, \omega) - 1 = -\lambda^2(N/V)\phi(q) \sum_{n=0}^{\infty} J^{(n)}(\mathbf{q}, \omega), \quad (83)$$

where $J^{(n)}$ indicates the regrouped terms of the series with $(\lambda^2)^n$ dependence. The first few terms of Eq. (83) will be calculated here.

(a) The lowest-order diagram is the one without any vertex [Fig. 3(a)], and its contribution is given by the first term of Eq. (78), namely,

$$J^{(0)}(\mathbf{q}, \ ) = -\int d\mathbf{p}_1 \cdots d\mathbf{p}_N(\omega + \mathbf{q}\cdot\mathbf{p}_\alpha/m)^{-1} \times \tilde{F}(0 \cdots 0)\mathbf{q}\cdot\frac{\partial G}{\partial \mathbf{p}_\alpha}, \quad (84)$$

or using Eqs. (60), (54), and (41), we get

$$J^0(q, \omega) = -I(q, \omega)\frac{1}{m}. \quad (85)$$

This result is clearly expected as it represents the "free particle" approximation of Sec. II.

(b) First-order approximation. Without regard to the "boundaries" there are only two possible diagrams, which contribute to $J^{(1)}$, namely, Figs. 3(b) and 3(c). But if we consider the "ends" [Eqs. (81) and (82)] we see that only Fig. 3(b) contributes to this order while Fig. 3(c) contributes only to the next order. Thus, we have

$$\begin{aligned} J^{(1)}(q, \omega) &= \int d\mathbf{p}_1 \cdots d\mathbf{p}_N(\omega + \mathbf{q}\cdot\mathbf{p}_\alpha/m)^{-1} \\ &\times \sum_{i \neq \alpha} (\lambda^2/V)\phi(q)\mathbf{q}\cdot\left(\frac{\partial}{\partial \mathbf{p}_\alpha} - \frac{\partial}{\partial \mathbf{p}_i}\right) \\ &\times (\omega + \mathbf{q}\cdot\mathbf{p}_i/m)^{-1}\tilde{F}(0 \cdots 0)\mathbf{q}\cdot\frac{\partial G}{\partial \mathbf{p}_i} \\ &= \lambda^2(N/V)\phi(q)[I(\mathbf{q}, \omega)]^2. \end{aligned} \quad (86)$$

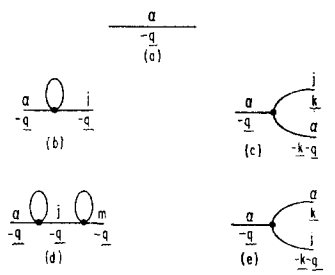

FIG. 3. The first diagrams contribute to the "weak interaction" case. (a) is a zero-order diagram, (b) first-order diagram, and (c)–(e) second-order diagrams.

We should like to call attention to a point which could be misleading. It would seem that Fig. 3(a) contributes to this order, due to its "boundary" [see Eq. (81)],

$$-\int d\mathbf{p}_1 \cdots d\mathbf{p}_N(\omega + \mathbf{q}\cdot\mathbf{p}_\alpha/m)^{-1} \times \tilde{F}(0 \cdots \mathbf{q} \cdots -\mathbf{q} \cdots 0) \sum_{l \neq \alpha} \mathbf{q}\cdot\frac{\partial G}{\partial \mathbf{p}_l}. \quad (87)$$

But, in fact, this integral vanishes for $l \neq \alpha$ [the term $l \neq \alpha$ is given in Eq. (84)], as was explained at the end of the last section.

(c) Second-order approximation. We proceed to the next order and get contributions from the two diagrams of Figs. 3(d) and 3(e). The first one gives

$$J_1^{(2)}(\mathbf{q}, \omega) = -\lambda^4(N/V)^2[\phi(q)]^2[I(q, \omega)]^3\frac{1}{m^3}, \quad (88)$$

and the second yields

$$\begin{aligned} J_2^{(2)}(\mathbf{q}, \ ) &= 2\lambda^4\beta\frac{N}{V}\frac{1}{V}\sum_k \phi(k)a^{-6} \\ &\times \int d\mathbf{p}_1\, d\mathbf{p}_2(\omega + \mathbf{q}\cdot\mathbf{p}_1/m)^{-1} \\ &\times k\cdot\frac{\partial}{\partial \mathbf{p}_1}[\omega + (\mathbf{k} + \mathbf{q})\cdot\mathbf{p}_1/m - \mathbf{k}\cdot\mathbf{p}_2/m]^{-1} \\ &\times \left[\phi(k)\mathbf{q}\cdot\frac{\partial}{\partial \mathbf{p}_1} + \phi(|\mathbf{k} + \mathbf{q}|)\mathbf{q}\cdot\frac{\partial}{\partial \mathbf{p}_2}\right] e^{-\beta(p_1^2 + p_2^2)/2m}. \end{aligned} \quad (89)$$

The integral of Eq. (89) is not easy to perform, but it is important to notice that it is asymptotically of order $\lambda^4 n$, while Eq. (88) is of order $\lambda^4 n^2$. Thus, in the weak interaction case, the series of Eq. (83) is not simply a geometric series as in the plasma case. This fact is of importance when a comparison is made between systems of weak interaction and Coulomb gas.

One can continue in this way to find higher terms of the series of Eq. (83). It is easy to see that, the higher the term, the more diagrams are to be summed, and their calculation is more involved.

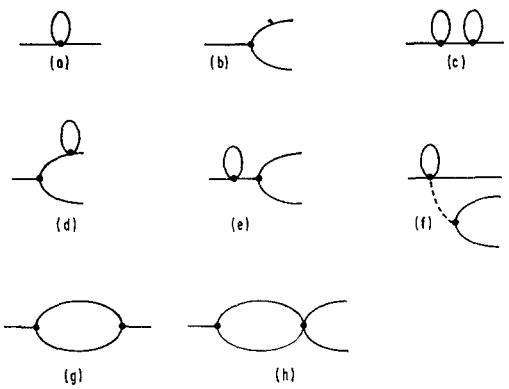

FIG. 4. Some diagrams and their dependence on $(e^2)^l(e^2n)^m$. (a) $l = 0$, $m = 1$; (b) $l = 1$, $m = 1$; (c) $l = 0$, $m = 2$; (d)–(f) $l = 1$, $m = 2$; (g) $l = 1$, $m = 1$; and (h) $l = 2$, $m = 1$.

B. Fully Ionized Gases

We consider now a system of charged particles embedded in a uniform neutralizing media and interacting via *long-range* Coulomb forces. In this case the dielectric function retains its usual meaning and it is easily interpreted. We give here only a brief treatment and illustrate the application of PB's method to this problem. The dielectric function of a plasma has been studied lately by many authors[4,17] using other methods, but we believe that better approximations can be achieved by following the method outlined above.

In a plasma we write $\lambda \equiv e$ and $\phi(k) = 4\pi/k^2$, where e denotes the charge of a particle. It is a well-known fact that in a plasma neither e^2 nor the density n are significant expansion parameters, but rather the combination e^2n is the right parameter.[9,24] Thus, one cannot limit himself to the first powers of e^2 or of n, in any expansion (even if these quantities are sometimes relatively small), but should rather retain all terms proportional to e^2n. Strictly speaking, e^2n is not a legitimate expansion parameter, as it is not dimensionless, but one can be convenienced easily that it can be used as such for shortness. The reason is simply that e^2n always appears in the combination

$$e^2n\beta\phi(k) = \frac{4\pi e^2 n\beta}{n^{\frac{2}{3}}}\frac{n^{\frac{2}{3}}}{k^2} \equiv \left(\frac{d}{h}\right)^2 \frac{1}{d^2k^2}, \qquad (90)$$

where d is the interparticle distance (a natural unit of length) and h is the radius of the Debye sphere. Thus, we come to the conclusion that our expansion parameter is essentially the usual plasma parameter d/h (which is inversely proportional to the number of particles in the Debye sphere).

[24] N. Rostoker and M. N. Rosenbluth, Phys. Fluids **3**, 1 (1960).

Once the physical expansion parameter is decided, the classification of the diagrams with respect to this parameter is a simple task. Using Table I and taking into account the "boundaries effects" [Eqs. (81), (82), etc.] we see that each diagram is classified asymptotically in terms of

$$(e^2)^l V^{-k}(e^2n)^m, \qquad (91)$$

where k, l, and m are positive integers. In the limit of an infinite system all the diagrams with $k > 0$ can be neglected, and only the diagrams with arbitrary $(e^2)^l(e^2n)^m$ are to be considered. In the present paper, we restrict ourself only to the lowest-order contribution to the dielectric function, namely, to

$$1/\epsilon(\mathbf{q}, \omega) = 1 + e^2n\phi(q) \times \sum \text{all diagrams of } (e^2n)^m. \qquad (92)$$

To facilitate our calculation let us examine some small-order diagrams of the expansion in e^2. If we apply the rules of Sec. III to the diagrams of Fig. 4, we see immediately that only the diagrams 4(a) and 4(c) belong to the right-hand side of Eq. (92), being of order $(e^2n)^1$ and $(e^2n)^2$, respectively. The other diagrams of Fig. 4 have at least one extra e^2 attached to e^2n [Fig. 4(b) is of order $e^2(e^2n)$; Figs. 4(d), 4(e), and 4(f) are of order $e^2(e^2n)^2$; and Figs. 4(g) and 4(h) are of order $e^2(e^2n)$ and $e^4(e^2n)$, respectively]. If one examines the diagrams in detail, one finds that diagrams with two or more parallel lines contribute to a higher order than those with only one horizontal line. Moreover, the only diagrams which contribute to the right-hand side of Eq. (92) are those composed of "loops" of type (A) along one horizontal line; the contribution of a chain of m "loops" is $(e^2n)^m$. Thus, we can rewrite Eq. (92) as

$$1/\epsilon(\mathbf{q}, \omega) = 1 + e^2n\phi(q) \sum \text{all chains of "loops"}, \qquad (93)$$

where the sum is represented diagrammatically by Fig. 5.

The calculation of the general term of Eq. (93) is quite simple. The term which corresponds to a

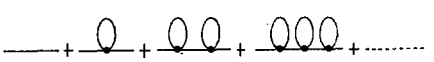

FIG. 5. Diagrams correspond to the right-hand side of Eq. (93)—the dominant contribution to the dielectric function for a plasma.

chain of m "loops" is given by

$$I_m(\mathbf{q},\omega) = (-1)^{m+1}\int d\mathbf{p}_1\cdots d\mathbf{p}_N \times(\omega+\mathbf{q}\cdot\mathbf{p}_\alpha/m)^{-1}\sum_{j\neq\alpha}\frac{e^2}{V}\phi(q)\mathbf{q}\cdot\frac{\partial}{\partial\mathbf{p}_\alpha} \times(\omega+\mathbf{q}\cdot\mathbf{p}_j/m)^{-1}\sum_{s\neq j}\frac{e^2}{V}\phi(q)\mathbf{q}\cdot\frac{\partial}{\partial\mathbf{p}_j} \times(\omega+\mathbf{q}\cdot\mathbf{p}_s/m)^{-1}\cdots \times(\omega+\mathbf{q}\cdot\mathbf{p}_l/m)\tilde{F}(0\cdots 0)\mathbf{q}\cdot\frac{\partial G}{\partial\mathbf{p}_l}. \quad (94)$$

If we substitute $G(\mathbf{p}_1\cdots\mathbf{p}_N)$ and $\tilde{F}(0\cdots 0)$ from Eqs. (54) and (60), we obtain

$$I_s(\mathbf{q},\omega) = \left(\frac{-1}{m}\right)^{s+1} I(q,\omega)[e^2n\phi(q)I(q,\omega)]^s, \quad (95)$$

where $I(q,\omega)$ is given by Eq. (42). It should be noted that passing from Eq. (94) to Eq. (95) we exclude the possibility of repeated particles along the chain [see the calculation of Eq. (88) and the discussion preceding this equation].

Now, Eq. (93) can be read as

$$1/\epsilon(\mathbf{q},\omega) = 1+\sum_{s=1}^{\infty}\left[-\frac{1}{m}e^2n\phi(q)I(q,\omega)\right]^s; \quad (96)$$

or, denoting

$$\alpha(q,\omega) = -\frac{1}{m}e^2n\phi(q)I(q,\omega), \quad (97)$$

and performing the summation in Eq. (96), we get

$$\frac{1}{\epsilon(q,\omega)} = 1-\frac{\alpha(q,\omega)}{1+\alpha(q,\omega)}, \quad (98)$$

or

$$\epsilon(q,\omega) = 1+\alpha(q,\omega) = 1-\frac{1}{m}e^2n\phi(q)I(q,\omega) = 1-\frac{1}{m}4\pi e^2nk^{-2}\int_{-\infty}^{\infty}du\,\frac{1}{u+\omega/k}\frac{\partial F(u)}{\partial u}, \quad (99)$$

where $F(u)$ is defined by Eq. (43).

Equation (99) is the well-known result of the linearized Vlasov equation.[17] As was shown by Balescu,[9] the chain of "loops" corresponds to the "random phase approximation" of Bohm and Pines,[25] as in any intermediate state we have only one line. Many aspects of Eq. (99) were studied in recent years (see, e.g., reference 26) and we are not concerned with these problems here.

To conclude this section we want to point out that, in principle, one can calculate the dielectric function to any order, by using the above method, and in doing so, get much more insight into those processes which are intimately related to the dielectric function. We hope to come back to this problem in the future.

V. DISCUSSION

In this paper we attempt to pave the way toward a systematic perturbation expansion for the dielectric function of systems of interacting particles. The introduction of this function enables one to investigate many properties of many-body systems. Those properties can be studied to the same degree of approximation as the calculated dielectric function. In the method developed above, the system is considered as a whole, using the Liouville equation, instead of the canonical equations or the hierarchy of integrated Liouville equations. The definition of the dielectric function as an integral in the whole phase space of the system and the introduction of the diagrammatic technique of Prigogine and Balescu permits a detailed analysis of the interactions between the particles of the system.

The dielectric function was defined by finding the response to a vanishingly small external field. However, the resulting dielectric function is a property of the equilibrium of the closed system alone and can serve to describe exactly some important properties of the system itself. The possibility of exploiting the properties of non-equilibrium systems along similiar lines seems feasible.[27]

ACKNOWLEDGMENTS

The author wishes to thank Dr. E. A. Frieman and Dr. Carl Oberman for valuable discussions and for reading the manuscript.

[25] D. Pines and D. Bohm, Phys. Rev. **85**, 338 (1952).
[26] J. D. Jackson, J. Nucl. Energy **1**, 171 (1960).
[27] N. Rostoker, Nucl. Fusion **1**, 101 (1960).

Lorentz–Lorenz and Sellmeier Formulas in Irregular Gases

LEONARD S. TAYLOR
General Electric Space Sciences Laboratory, Valley Forge, Pennsylvania
(Received 28 September 1962)

The integral equation expressing the condition of dynamical equilibrium of dipoles in an applied electromagnetic field is solved to determine the index of refraction for the ensemble-average wave in a media in which the polarizability varies irregularly. The solution closely follows that employed in the usual derivation of the Ewald–Oseen theorem for isotropic media. A Lorentz–Lorenz formula, modified by the appearance of an additional term, is determined. The additional term is real when the scale length of the irregularities is small compared to the wavelength. For larger irregularities, however, the correction leads to a complex index of refraction, expressing the attenuation due to scattering. Applying the concept of the depolarizing effect of electron–ion collisions, these results are extended to the case of an ionized gas. In this latter instance, a similar modification of the Sellmeier formula for the index of refraction is determined. Explicit formulas are given for the case of irregularities with scale large compared to a wavelength.

INTRODUCTION

In many investigations involving the propagation of electromagnetic waves in irregular media, an important aspect is the characteristics of the average wave. In general, this problem has been treated by assuming that the statistical structure of the variations of dielectric permittivity is known, and solving Maxwell's equations in various approximations. In this paper, which is closely related to treatments[1] of the permittivity of solid mixtures, the point of view is that the statistical structure of the dipole number density is known and by the method used in the proof of the Ewald–Oseen extinction theorem, the index of refraction (square root of the dielectric permittivity) will be obtained for the ensemble-average wave in an irregular media. (This method, which employs an integral equation representing the condition of dynamical equilibrium, has been described in detail by Born and Wolf,[2] where the ordinary Lorentz–Lorenz formula is obtained for a medium in which the molecular dipoles are evenly distributed throughout.) An explicit formula is given for the case of large-scale irregularities. Because the Lorentz local field approximation is used, the direct application is to gases. The results are extended to ionized gases by introducing the depolarizing effect first described by Darwin[3,4] and a modified Sellmeier formula is obtained.

The point of view of the analysis in this paper is that an index of refraction exists which describes the propagation of a wave in the ensemble average medium. This concept of a "scattering medium" appears to have first been introduced for scalar waves by Foldy[5] who suggested some years ago the extension to the Lorentz–Lorenz formula. To the author's knowledge, however, this program was never carried out.

GENERALIZED LORENTZ–LORENZ FORMULA

The basic integral equation for the effective electric field acting on the dipole at $\mathbf{r}$ in the medium is

$$\mathbf{E}'(\mathbf{r}, t) = \mathbf{E}^{(i)} + \int_{\sigma}^{\Sigma} \nabla \times \nabla \times \frac{N\alpha \mathbf{E}'(\mathbf{r}', t - R/c)}{R}\, dv', \qquad \mathbf{R} = \mathbf{r}' - \mathbf{r}, \quad R = |\mathbf{R}|, \tag{1}$$

where $\mathbf{E}^{(i)}$ is the incident field (propagated with velocity c) and $N\alpha$ is the polarizability. The integration is carried out over the medium Σ, excluding the Lorentz sphere σ of vanishing radius, a, about $\mathbf{r}$. The medium is assumed to be nonmagnetic, so that the magnetic field need not be considered in what follows. The dipole moment per unit volume is by definition:

$$\mathbf{P}(\mathbf{r}', t) = N\alpha \mathbf{E}'(\mathbf{r}', t). \tag{2}$$

In this equation, the polarizability $N\alpha$ is regarded as a function of $\mathbf{r}'$ but not of t. [With no real loss in generality, it shall henceforth be assumed in this section that α = constant, $N = N(\mathbf{r}')$.] The incident field is assumed to be monochromatic with angular

[1] W. F. Brown, Jr., J. Chem. Phys. **23**, 1514 (1955).
[2] M. Born and E. Wolf, *The Principles of Optics* (Pergamon Press, Inc., New York, 1959), Chap. 2, pp. 97–103.
[3] C. G. Darwin, Proc. Roy. Soc. (London) **A182**, 152 (1943).
[4] O. Theimer and L. S. Taylor, J. Geophys. Res. **66**, 3157 (1961).
[5] L. L. Foldy, Phys. Rev. **67**, 107 (1945).

frequency $\omega = c/k_0$. The dipole moment at $(\mathbf{r}, t)$ for a specific configuration of irregularities is written as

$$\mathbf{P}(\mathbf{r}, t) = \langle \mathbf{P}(\mathbf{r}, t) \rangle + \delta\mathbf{P}(\mathbf{r}, t). \tag{3}$$

In this expression, $\langle \mathbf{P}(\mathbf{r}, t) \rangle$ is the ensemble-average dipole moment per unit volume at $\mathbf{r}$. In what follows, an equation for $\langle \mathbf{P}(\mathbf{r}, t) \rangle$ shall be obtained from $\mathbf{P}(\mathbf{r}, t)$ by first averaging over configurations of dipole moments relative to a fixed dipole moment at $\mathbf{r}$, and then completing the average by taking the mean with respect to the dipole moment at $\mathbf{r}$. The trial solution for $\langle \mathbf{P} \rangle$ is, similar to the procedure in reference 2, a wave with frequency ω and velocity of propagation c/n, where n is regarded as the quantity to be determined. Thus

$$\nabla^2 \langle \mathbf{P}_0 \rangle + n^2 k_0^2 \langle \mathbf{P}_0 \rangle = 0, \tag{4}$$

where the subscript has been introduced to indicate that the factor $\exp(-i\omega t)$ has been removed. It is also assumed that $\langle \mathbf{P}_0 \rangle$ has no sources in the medium

$$\nabla \cdot \langle \mathbf{P}_0 \rangle = 0. \tag{5}$$

As a result of the lengthy analysis in reference 2, Eqs. (1)–(5) lead to the following expression:

$$\mathbf{E}_0'(\mathbf{r}) = \mathbf{E}_0^{(i)} + \mathbf{E}_0^{(1)} + \mathbf{E}_0^{(2)} + \mathbf{E}_0^{(3)}, \tag{6}$$

where

$$\mathbf{E}_0^{(1)}(\mathbf{r}) = \frac{4\pi}{3}\left(\frac{n^2+2}{n^2-1}\right)\langle \mathbf{P}_0(\mathbf{r}) \rangle, \tag{7}$$

$$\mathbf{E}_0^{(2)}(\mathbf{r}) = \frac{1}{k_0^2(n^2-1)} \nabla \times \nabla \times \int_\Sigma \left[\mathbf{P}_0(\mathbf{r}') \frac{\partial G(R)}{\partial \nu'} - G(R) \frac{\partial \mathbf{P}_0(\mathbf{r}')}{\partial \nu'} \right] dS', \tag{8}$$

$$\mathbf{E}_0^{(3)}(\mathbf{r}) = \int_\sigma^\Sigma \nabla \times \nabla \times \delta\mathbf{P}_0(\mathbf{r}') G(R)\, d\nu', \tag{9}$$

$$G(R) = e^{ik_0 R}/R, \tag{10}$$

and $\partial/\partial\nu'$ denotes differentiation along the outward normal to the boundary Σ. The only difference between these expressions and those in reference 2 is the presence of the term $\mathbf{E}_0^{(3)}(\mathbf{r})$.

In reference 2 it is shown that in the limit $a \to 0$,

$$\int_\sigma^\Sigma \nabla \times \nabla \times \mathbf{Q}(\mathbf{r}') G(R)\, d\nu' = \nabla \times \nabla \times \int_\sigma^\Sigma \mathbf{Q}(\mathbf{r}') G(R)\, d\nu' - \frac{8\pi}{3}\mathbf{Q}(\mathbf{r}), \tag{11}$$

where $\mathbf{Q}(\mathbf{r}')$ is an arbitrary vector function of position. Thus, applying this result, Eq. (9) is

$$\mathbf{E}_0^{(3)}(\mathbf{r}) = \nabla \times \nabla \times \int_\sigma^\Sigma \delta\mathbf{P}_0(\mathbf{r}') G(R)\, d\nu' - \frac{8\pi}{3}\delta\mathbf{P}_0(\mathbf{r}). \tag{12}$$

Thus far, a procedure has been followed which is mathematically exact and the only physical assumptions which have been employed are those normally used in the Lorentz–Lorenz derivation. Continuing, the next step is to obtain the equation of dynamical equilibrium of the ensemble average, $\langle \mathbf{P}_0(\mathbf{r}) \rangle$. The procedure followed in this paper is to first calculate the quantity $[\mathbf{E}_0^{(3)}(\mathbf{r})]_{\mathrm{AV}}$, the "configuration" average of $\mathbf{E}_0^{(3)}(\mathbf{r})$ *for a given value of* $\delta\mathbf{P}_0(\mathbf{r})$. Thus, Eq. (12) is written

$$[\mathbf{E}_0^{(3)}(\mathbf{r})]_{\mathrm{AV}} = \nabla \times \nabla \times \int_\sigma^\Sigma \langle \delta\mathbf{P}_0(\mathbf{r}) | \delta\mathbf{P}_0(\mathbf{r}') \rangle G(R)\, d\nu' - \frac{8\pi}{3}\delta\mathbf{P}_0(\mathbf{r}), \tag{13}$$

where $\langle \delta\mathbf{P}_0(\mathbf{r}) | \delta\mathbf{P}_0(\mathbf{r}') \rangle$ is the mean of $\delta\mathbf{P}_0(\mathbf{r}')$, *given* $\delta\mathbf{P}_0(\mathbf{r})$. The replacing of $\delta\mathbf{P}_0(\mathbf{r}')$ within the derivative operations in Eq. (9) by this averaged quantity is justified by the fact that the derivatives of fluctuations are uncorrelated and yield zero contribution.

Obviously,

$$[\mathbf{E}_0^{(1)}(\mathbf{r})]_{\mathrm{AV}} = \mathbf{E}_0^{(1)}(\mathbf{r}). \tag{14}$$

The assumption shall now be made that there exists a function $\mu(\mathbf{r}, \mathbf{R})$ such that

$$\langle \delta\mathbf{P}_0(\mathbf{r}) | \delta\mathbf{P}_0(\mathbf{r}') \rangle = \mu(\mathbf{r}, \mathbf{R}) \langle \mathbf{P}_0(\mathbf{r}) \rangle, \tag{15}$$

$$\lim_{R \to \infty} \mu(\mathbf{r}, \mathbf{R}) = 0. \tag{16}$$

This assumption will be justified later. Meanwhile, proceeding on by substitution, and restricting attention to points $\mathbf{r}$ which do not lie close to the surface, Eq. (8) yields

$$[\mathbf{E}_0^{(2)}(\mathbf{r})]_{\mathrm{AV}} = \frac{1}{k_0^2(n^2-1)} \nabla \times \nabla \times \int_\Sigma \left[\langle \mathbf{P}_0(\mathbf{r}') \rangle \times \frac{\partial G(R)}{\partial \nu'} - G(R) \frac{\partial \langle \mathbf{P}_0(\mathbf{r}') \rangle}{\partial \nu'} \right] dS'. \tag{17}$$

In order to obtain Eq. (17), it has been assumed that the subensemble of configurations with fixed dipole moment at $\mathbf{r}$ is equivalent, as regards the dipole moments on the surface of Σ, to the complete ensemble.

Collecting results, and multiplying by $\alpha N(\mathbf{r})$ yields

$$[\mathbf{P}_0(\mathbf{r})]_{AV} = \alpha N(\mathbf{r})\Big\{\mathbf{E}_0^{(i)}(\mathbf{r}) + \frac{4\pi}{3}\left(\frac{n^2+2}{n^2-1}\right)\langle\mathbf{P}_0(\mathbf{r})\rangle + \frac{1}{k_0^2(n^2-1)}\nabla\times\nabla\times\int_\Sigma\left[\langle\mathbf{P}_0(\mathbf{r}')\rangle\frac{\partial G(R)}{\partial\nu'} - G(R)\frac{\partial\langle\mathbf{P}_0(\mathbf{r}')\rangle}{\partial\nu'}\right]dS' + [\mathbf{E}_0^{(3)}(\mathbf{r})]_{AV}\Big\}. \quad (18)$$

Defining $\bar{N}$, the number density corresponding to the ensemble-average dipole moment per unit volume, and averaging Eq. (18) with respect to the dipole moment per unit volume at $\mathbf{r}$,

$$\langle\mathbf{P}_0(\mathbf{r})\rangle = \alpha\bar{N}\mathbf{E}_0^{(i)}(\mathbf{r}) + \frac{4}{3}\pi\alpha\bar{N}\left(\frac{n^2+2}{n^2-1}\right)\langle\mathbf{P}_0(\mathbf{r})\rangle + \frac{\alpha\bar{N}}{k_0^2(n^2-1)}\nabla\times\nabla\times\int_\Sigma\left[\langle\mathbf{P}_0(\mathbf{r}')\rangle\frac{\partial G(R)}{\partial\nu'} - G(R)\frac{\partial\langle\mathbf{P}_0(\mathbf{r}')\rangle}{\partial\nu'}\right]dS' + \alpha\langle N(\mathbf{r})[\mathbf{E}_0^{(3)}(\mathbf{r})]_{AV}\rangle. \quad (19)$$

As in reference 2, it is now noted that the first and third terms on the right represent waves travelling with velocity c. The left side and the second term on the right of Eq. (19) are waves with velocity c/n. In a uniform medium, the last term does not appear and each group must vanish separately. The relation

$$\mathbf{E}_0^{(i)}(\mathbf{r}) + \frac{1}{k_0^2(n^2-1)}\nabla\times\nabla\times\int_\Sigma\left[\langle\mathbf{P}_0(\mathbf{r}')\rangle\frac{\partial G(R)}{\partial\nu'} - G(R)\frac{\partial\langle\mathbf{P}_0(\mathbf{r}')\rangle}{\partial\nu'}\right]dS' = 0 \quad (20)$$

represents the extinction of the incident wave by part of the dipole field. It is not necessary to determine whether the last term in Eq. (19) contributes to Eq. (20). Instead, attention is turned to the dipole field in the medium which, from Eq. (19), satisfies

$$\langle\mathbf{P}_0(\mathbf{r})\rangle = \frac{4\pi}{3}\alpha\bar{N}\left(\frac{n^2+2}{n^2-1}\right)\langle\mathbf{P}_0(\mathbf{r})\rangle + \alpha\langle N(\mathbf{r})[\mathbf{E}_0^{(3)}(\mathbf{r})]_{AV}\rangle'. \quad (21)$$

A prime has been attached to the last term as a reminder that only waves with velocity c/n are included.

Writing

$$\zeta = \langle N(\mathbf{r})[\mathbf{E}_0^{(3)}(\mathbf{r})]_{AV}\rangle'\cdot\langle\mathbf{P}_0(\mathbf{r})\rangle/(4\pi/3)\bar{N}\,|\langle\mathbf{P}_0(\mathbf{r})\rangle|^2, \quad (22)$$

it follows that

$$1 = \frac{4\pi}{3}\alpha\bar{N}\left(\frac{n^2+2}{n^2-1} + \zeta\right). \quad (23)$$

Equation (22) is the desired generalization of the Lorentz–Lorenz formula. In order to employ this formula, it is necessary to evaluate the form factor ζ. It is evident that, for irregularities whose scale is small compared to a wavelength, ζ is real and Eq. (23) leads to a simple correction for the index of refraction. For irregularities with larger scale length, ζ will be complex, however, leading to a complex index of refraction which expresses the loss of energy due to scattering.

Attention is now turned to media in which the scale length of the irregularities is much greater than a wavelength. An iteration procedure in which use is made of the approximation

$$\delta\mathbf{P}_0(\mathbf{r}') \approx \Delta N(\mathbf{r}')\langle\mathbf{P}_0(\mathbf{r}')\rangle/\bar{N} \quad (24)$$

is immediately suggested. This equation must be supplemented by the relations

$$N(\mathbf{r}) = \bar{N} + \Delta N(\mathbf{r}), \quad (25)$$

$$\langle\Delta N(\mathbf{r})\rangle = 0. \quad (26)$$

That is, $\bar{N}$ is taken to be the average dipole number density. Of course, it is necessary that ζ represent only a small correction to the index of refraction if the iteration procedure is to be valid. For a plane wave in the $\mathbf{k}_0$ direction, therefore,

$$\delta\mathbf{P}_0(\mathbf{r}') \approx \Delta N(\mathbf{r}')\langle\mathbf{P}_0(\mathbf{r})\rangle\exp(in\mathbf{k}_0\cdot\mathbf{R})/\bar{N}. \quad (27)$$

As a result,

$$\langle\delta\mathbf{P}_0(\mathbf{r})\,|\,\delta\mathbf{P}_0(\mathbf{r}')\rangle \approx \langle\Delta N(\mathbf{r})\,|\,\Delta N(\mathbf{r}')\rangle\langle\mathbf{P}_0(\mathbf{r})\rangle\exp(in\mathbf{k}_0\cdot\mathbf{R})/\bar{N}. \quad (28)$$

For a Gaussian distribution

$$p_1(N) = [1/(2\pi)^{\frac{1}{2}}\sigma]\exp[-(N-\bar{N})^2/2\sigma^2], \quad (29)$$

the conditional probability $p_2(N_1|N_2)$ is obtained from the two-dimensional Gaussian distribution

$$p_2(N_1, N_2) = \frac{1}{2\pi\sigma^2(1-\rho^2)^{\frac{1}{2}}}\exp\left\{\frac{-1}{2\sigma^2(1-\rho^2)}\times[(N_1-\bar{N})^2 + (N_2-\bar{N})^2 - 2\rho(N_1-\bar{N})(N_2-\bar{N})]\right\}. \quad (30)$$

Thus,

$$p_2(N_1|N_2) = p_2(N_1, N_2)/p_1(N_1) = \frac{1}{[2\pi\sigma^2(1-\rho)^2]^{\frac{1}{2}}}\times\exp\left\{\frac{-1}{2\sigma^2(1-\rho^2)}[N_2 - \bar{N} - \rho(N_1-\bar{N})]^2\right\}. \quad (31)$$

As a result, the mean of N_2, given N_1, is

$$\langle N_1|N_2\rangle = \bar{N} + \rho(N_1 - \bar{N}), \quad (32)$$

or

$$\langle \Delta N(\mathbf{r}) | \Delta N(\mathbf{r}') \rangle = \rho\, \Delta N(\mathbf{r}). \tag{33}$$

In these formulas, since $\langle (N_1 - \bar{N})(N_2 - \bar{N}) \rangle = \sigma^2 \rho$, ρ is a normalized correlation coefficient. Of course, $\rho = \rho(\mathbf{R})$.

With this justification, the special assumption is made that Eq. (28) may be written

$$\langle \delta \mathbf{P}_0(\mathbf{r}) | \delta \mathbf{P}_0(r') \rangle \approx \rho(\mathbf{R})\, \Delta N(\mathbf{r}) \langle \mathbf{P}_0(\mathbf{r}) \rangle \exp(in\mathbf{k}_0 \cdot \mathbf{R})/\bar{N}, \tag{34}$$

where $\rho(\mathbf{R})$ is a correlation function (which, for Gaussian distribution of $\Delta N(\mathbf{r})$, is identical to the usual correlation factor). Equation (34) justifies Eqs. (15) and (16). Of course,

$$\lim_{\mathbf{R}\to\infty} \rho(\mathbf{R}) = 0.$$

Applying these results, it is found that

$$\zeta \approx \langle (\Delta N/\bar{N})^2 \rangle \Big[-2 + \frac{3}{4\pi} \times \exp(-in\mathbf{k}_0 \cdot \mathbf{r}) \hat{p} \cdot \nabla \times \nabla \times \int_\sigma^\Sigma \rho(\mathbf{R}) \times \exp(in\mathbf{k}_0 \cdot \mathbf{r}') G(R)\, dv' \Big], \tag{35}$$

where $\hat{p}$ is a unit vector in the direction of $\langle \mathbf{P}_0(\mathbf{r}) \rangle$. In writing Eq. (35), the terms containing derivatives of $\Delta N(\mathbf{r})$ have been dropped because they are uncorrelated.

In the remainder of the present work, the medium will be assumed to have zero conductivity and isotropic irregularities. In this case, $\rho(\mathbf{R}) = \rho(R)$, and Eq. (35) may be written

$$\zeta \approx \langle (\Delta N/\bar{N})^2 \rangle \Big[-2 + \frac{3}{4\pi} \exp(-in\mathbf{k}_0 \cdot \mathbf{r}) \hat{p} \cdot \nabla \times \nabla \times \hat{p} \times \exp(in\mathbf{k}_0 \cdot \mathbf{r}) \int_\sigma^\Sigma \rho(R) G(R) \exp(in\mathbf{k}_0 \cdot \mathbf{R})\, dv' \Big]. \tag{36}$$

The integral is independent of $\mathbf{r}$, and in these equations, $\mathbf{k}_0$ is a constant vector, which for a homogeneous plane wave satisfies

$$\hat{p} \cdot \mathbf{k}_0 = 0. \tag{37}$$

Thus using

$$\nabla \times \nabla \times \hat{p} f = (\hat{p} \cdot \nabla) \nabla f - \hat{p} \nabla^2 f, \tag{38}$$

it is found that

$$\hat{p} \cdot \nabla \times \nabla \times \hat{p} \exp(in\mathbf{k}_0 \cdot \mathbf{r}) = n^2 k_0^2 \exp(in\mathbf{k}_0 \cdot \mathbf{r}), \tag{39}$$

whence,

$$\zeta \approx \langle (\Delta N/\bar{N})^2 \rangle \Big[-2 + \frac{3}{4\pi} n^2 k_0^2 \int_\sigma^\Sigma \rho(R) G(R) \times \exp(in\mathbf{k}_0 \cdot \mathbf{R})\, dv' \Big]. \tag{40}$$

The integral in Eq. (40) is readily carried out in the angular coordinates, choosing the z axis in the $\mathbf{k}_0$ direction. Thus, passing to the limits, it follows that, for "interior points,"

$$\zeta \approx \langle (\Delta N/\bar{N})^2 \rangle \Big[-2 + 3nk_0 \int_0^\infty \rho(R) \times \exp(ik_0 R) \sin(nk_0 R)\, dR \Big]. \tag{41}$$

The appearance of an imaginary component in the index of refraction is attributed to loss of energy from the wave by scattering. Equations (41) and (23) determine the complete index of refraction. The imaginary component may be checked against the attenuation coefficient (determined by ordinary scattering theory by integrating over the scattering cross section). Thus, writing $n_0 = n_0 + \delta n$, it follows from Eq. (23), that ($n_0^2 - 1 \ll 1$)

$$\mathrm{Im}\,\{\delta n\} \approx \tfrac{1}{6}(4\pi\alpha\bar{N})^2\, \mathrm{Im}\,\{\zeta\} \tag{42}$$

However, for a uniform medium of density $\bar{N} + \Delta N$ ($n = n_0 + \Delta n$),

$$\Delta n \approx 2\pi\alpha \Delta N. \tag{43}$$

The power attenuation coefficient is

$$\alpha' = 2k_0\, \mathrm{Im}\,\{\delta n\}. \tag{44}$$

Applying all these results, it is readily shown that, as $n_0 \to 1$,

$$\alpha' \approx 2k_0^2 \langle (\Delta n)^2 \rangle \int_0^\infty \rho(R)[1 - \cos 2k_0 R]\, dR. \tag{45}$$

Equation (45) is identical to that obtained[6] from the scattering cross section. [In reference 6, this formula is derived only for scalar waves; it is readily observed, however, that for large scale fluctuations, the scalar (pressure) and vector (electromagnetic) cases lead to identical results.] In order to obtain Eq. (45) from Eq. (41), it is necessary to employ the relation $|\delta n|^2 k_0^2 l^2 \ll 1$, where l is the scale length of the irregularities, i.e., the range in which $\rho(R)$ is appreciable. Thus, Eq. (45) is not valid in the limit $k_0 \to \infty$. This restriction is also obtained in the previous derivation.[6] However, Eq. (41) is not subject to this limitation as shall be observed in the following.

[6] L. A. Chernov, *Wave Propagation in a Random Medium*, translated from the Russian by R. A. Silverman (McGraw-Hill Book Co., Inc., New York, 1960), p. 54.

In general, Eqs. (23) and (41) will lead to a complicated formula for the index of refraction. For example, for the Booker–Gordon model, $\rho(R) = \exp(-k_c R)$ with $k_c \ll k_0$, it follows from Eq. (41) that

$$\zeta \approx \langle(\Delta N/\bar{N})^2\rangle[n^2 + 2 + 6n^2 i(k_c/k_0)/(n^2 - 1)](n^2 - 1)^{-1}. \quad (46)$$

Thus, as $k_0 \to \infty$,

$$1 = \frac{4\pi}{3}\alpha\bar{N}[1 + \langle(\Delta N/\bar{N})^2\rangle]\left(\frac{n^2 + 2}{n^2 - 1}\right). \quad (47)$$

Moreover, writing $n = n_0 + \delta n$, it can be assumed that for cases of practical interest, $n_0 - 1 \ll 1$ and $\delta n/(n_0 - 1) \ll 1$, and it is found that Eq. (46) yields

$$n \approx n_0 + \langle(\Delta N/\bar{N})^2\rangle[n_0 - 1 + ik_c/k_0]. \quad (48)$$

In these formulas, n_0 is the index of refraction corresponding to the uniform medium.

It may be noted that the present analysis does not provide the angular distribution of scattered energy. On the other hand, it does yield the small real correction to the refractive index. Moreover, because analysis is based upon the microscopic properties of the media, it is particularly simple to consider those effects which must be included when considering ionized gases.

GENERALIZED SELLMEIER FORMULA

The preceding section provides a suitable background for the rapid generation of a formula for the index of refraction of an irregular ionized gas. Neglecting the effects of dynamical friction for the moment, the polarizability of the free electrons is taken as $\alpha = -e^2/m\omega^2$. In addition, it is necessary to include on the left side of Eq. (1), the depolarizing effect of electron–ion collisions, $\frac{4}{3}\pi\mathbf{P}(\mathbf{r}, t)$. (It is emphasized that this is the only logically consistent procedure by which the Sellmeier formula may be obtained.[7]) After removing the time-dependent factor, averaging over configurations, multiplying by $\alpha N(\mathbf{r})$ [see Eq. (18)], taking the ensemble average, and taking the scalar product with $\langle\mathbf{P}_0(\mathbf{r})\rangle/|\langle\mathbf{P}_0(\mathbf{r})\rangle|^2$ [see Eq. (22)], it follows that, instead of Eq. (23), for a slightly ionized gas

$$1 + \tfrac{4}{3}\pi\langle N(\mathbf{r})[\mathbf{P}_0(\mathbf{r})]_{\mathrm{AV}}\rangle\cdot\langle\mathbf{P}_0(\mathbf{r})\rangle/|\langle\mathbf{P}_0(\mathbf{r})\rangle|^2 \approx \frac{4}{3}\pi\alpha\bar{N}\left(\frac{n^2 + 2}{n^2 - 1} + \zeta\right). \quad (49)$$

[7] V. L. Ginzburg, *Propagation of Electromagnetic Waves in Plasma*, translated from the Russian by Royer and Roger (Gordon and Breach Science Publishers, Inc., New York, 1961) Chap. 2, pp. 27–32.

Thus,

$$1 + \frac{4}{3}\pi\alpha\bar{N} \approx \frac{4}{3}\pi\alpha\bar{N}\left(\frac{n^2 + 2}{n^2 - 1} - \zeta'\right), \quad (50)$$

where, by definition,

$$\zeta' = -\zeta - \langle\Delta N(\mathbf{r})[\mathbf{P}_0(\mathbf{r})]_{\mathrm{AV}}\rangle\cdot\langle\mathbf{P}_0(\mathbf{r})\rangle/\bar{N}\,|\langle\mathbf{P}_0(\mathbf{r})\rangle|^2. \quad (51)$$

It is consistent to employ Eq. (24) to evaluate Eq. (51). Thus, it follows immediately that

$$\zeta' \approx -\zeta - \langle(\Delta N/\bar{N})^2\rangle \approx 3\langle(\Delta N/\bar{N})^2\rangle \times\left[1 - nk_0\int_0^\infty \rho(R) \times \exp(ik_0R)\sin(nk_0R)\,dR\right]. \quad (52)$$

In general, it is necessary to consider the effects of a varying polarizability α. (It is assumed here that the irregularities do not extend down to the dimensions of the mean free path for electron–ion collisions so that the local values of the effective collision frequency ν_c are obtained directly from the local number densities. It is also pointed out that ν_c depends upon long-range collisions with ions, in distinction to the depolarizing effect,[4] so that it is legitimate to treat these separately.) Thus writing $\eta = \alpha N$, it follows immediately that it is only necessary to replace $\bar{N}$ by $\bar{\eta}$, and ΔN by $\Delta\eta$, removing the α's to obtain the desired generalization. These procedures add nothing new to the present argument, so that it shall be continued to be assumed that $\nu_c = 0$.

Equations (51) and (52) yield ($\omega_p^2 = 4\pi\bar{N}e^2/m$)

$$1 - \frac{1}{3}\left(\frac{\omega_p}{\omega}\right)^2 \approx -\frac{1}{3}\left(\frac{\omega_p}{\omega}\right)^2\left[\frac{n^2 + 2}{n^2 - 1} - \zeta'\right]. \quad (53)$$

Thus

$$n^2 \approx 1 - \left(\frac{\omega_p}{\omega}\right)^2\left[1 + \zeta'\left(\frac{\omega_p}{\omega}\right)^2\right], \quad (54)$$

where the relation $|(\omega_p/\omega)^2\zeta'| \ll 1$ has been employed. As a matter of interest, a formula obtained[8] by a different method for the index of refraction of an irregular ionized gas when the wavelength is much greater than the scale length of the irregularities is displayed here:

$$n^2 \approx 1 - \left(\frac{\omega_p}{\omega}\right)^2\left[1 + \frac{2}{3}\langle(\Delta N/\bar{N})^2\rangle\left(\frac{\omega_p}{\omega}\right)^2\right], \quad (55)$$

where it has been assumed that $(\omega_p/\omega)^2 \ll 1$.

[8] L. S. Taylor, J. Geophys. Res. **67**, 3843, (1962).

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 4, NUMBER 6 JUNE 1963

Zeros of Hankel Functions and Poles of Scattering Amplitudes*

JOSEPH B. KELLER, S. I. RUBINOW,† AND MAX GOLDSTEIN
Courant Institute of Mathematical Sciences New York University, New York, New York

The complex zeros $\nu_n(z)$, $n = 1, 2, \cdots$ of $H_\nu^{(1)}(z)$, $dH_\nu^{(1)}(z)/dz$ and $dH_\nu^{(1)}(z)/dz + iZH_\nu^{(1)}(z)$ are investigated. These zeros determine the poles in the scattering amplitudes resulting from scattering of various kinds of waves by spheres and cylinders. Formulas for $\nu_n(z)$ are obtained for both large and small values of $|z|$ and for large values of n. In addition, for $H_\nu^{(1)}(z)$ and $dH_\nu^{(1)}(z)/dz$, numerical solutions are found for real z in the interval $0.01 \leq z \leq 7$ and $n = 1, 2, 3, 4, 5$. The resulting loci of $\nu_n(z)$ in the complex ν plane are presented. These loci are the trajectories of the so-called Regge poles for scattering by spheres and cylinders.

1. INTRODUCTION

IN 1918 Watson[1] discovered that a certain scattering amplitude in electromagnetic theory had poles at the values of ν for which $H_\nu^{(1)}(z) = 0$. Here $H_\nu^{(1)}(z)$ is the Hankel function of the first kind of order ν and argument z. Similar poles have since been found in other scattering amplitudes at the zeros of other transcendental functions. Recently Regge[2] has examined them in quantum-mechanical potential scattering and this has stimulated many other investigations. Because of the importance of these poles and their trajectories, we have considered some special cases in detail and have obtained asymptotic formulas and numerical results for them.

Mathematically our investigation concerns the roots $\nu_n(z)$, $n = 1, 2, \cdots$ of the following three equations:

$$H_\nu^{(1)}(z) = 0, \tag{1}$$

$$(d/dz)H_\nu^{(1)}(z) = 0, \tag{2}$$

$$(d/dz)H_\nu^{(1)}(z) + iZH_\nu^{(1)}(z) = 0. \tag{3}$$

In (3) Z is either a given constant or a given function of z and ν. Each root $\nu_n(z)$ of each equation is a complex function of the complex argument z. We present some old and some new expansions of $\nu_n(z)$ for both large and small values of $|z|$ as well as for large values of n for any z. In addition, with the aid of an electronic computer, we have computed the first five roots of (1) and (2) for real z in the range $.01 \leq z \leq 7$ and have plotted graphs of them. [See Figs. (1) and (2).] We have also compared these "exact" numerical values with the expansions for large and small values of $|z|$, thus determining the accuracy and range of validity of these expansions.

Equation (1) determines the poles in the quantum-mechanical scattering by a rigid sphere or cylinder, i.e., by a potential which is infinite within a sphere or cylinder and zero outside it. It also determines the poles in the scattering of an acoustic wave by an acoustically soft sphere or cylinder. In addition, it determines some of the poles in the scattering of an electromagnetic wave by a perfectly conducting sphere or cylinder. Equation (2) determines the poles in the scattering of an acoustic wave by a rigid cylinder and some of the poles in electromagnetic scattering by a perfectly conducting cylinder. Equation (3) determines the poles in acoustic or electromagnetic scattering by a cylinder of surface impedance Z. In all cases $z = ka$, where a is the radius of the sphere or cylinder and $k = 2\pi/\lambda$, with λ being the incident wavelength.

Because of the importance of the scattering problems just mentioned, some studies have been made of the Eqs. (1)–(3). The most complete study of (1) is that of Magnus and Kotin,[3] which led to the present work. In part, our analysis is similar to theirs. However, we succeeded in obtaining expansions of $\nu_n(z)$ for $|z|$ small which they did not find. These expansions show that Theorems 6.1 and 6.2 of reference 3 are false, and it is then not difficult to locate the flaws in the proofs. Fortunately none of their subsequent results depend upon these theorems. In addition, we have found a number of misprints in their formulas on p. 243 for $\nu_n(z)$ for large n. The correct formulas are given below.

2. ZEROS OF $H_\nu^{(1)}(z)$

Let us begin by expressing $H_\nu^{(1)}(z)$ in terms of Bessel functions by the formula

* This research was supported by the National Science Foundation under Grant No. G-19671. Reproduction in whole or in part is permitted for any purpose of the U. S. Government.

† Also, Physics Department, Stevens Institute of Technology, Hoboken, New Jersey.

[1] G. N. Watson, Proc. Roy. Soc. (London) **A95**, 83 (1918).

[2] T. Regge, Nuovo Cimento **14**, 951 (1958).

[3] W. Magnus and L. Kotin, Numerische Math. **2**, 228 (1960).

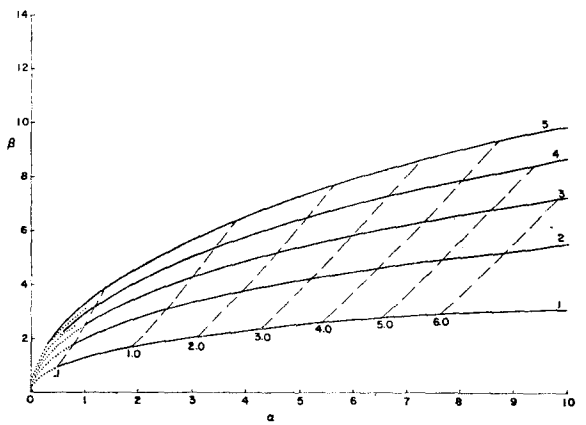

FIG. 1. The zeros $\nu_n(z)$ of $H_\nu^{(1)}(z)$ in the complex ν plane for z real. Re ν_n is plotted horizontally and Im ν_n is plotted vertically for $n = 1, 2, 3, 4, 5$, and $0 \leq z \leq 7$. The solid lines are the loci of $\nu_n(z)$ for fixed n as functions of z. The dashed lines connect values of $\nu_n(z)$ for fixed z and different values of n. The zeros are symmetric about $\nu = 0$ so there is a similar set of curves in the third quadrant.

$$i \sin \nu\pi H_\nu^{(1)}(z) = J_{-\nu}(z) - J_\nu(z)e^{-i\nu\pi}. \tag{4}$$

The power series for $J_\nu(z)$ is

$$J_\nu(z) = \sum_{m=0}^{\infty} (-1)^m \left(\frac{z}{2}\right)^{\nu+2m} [m!\,\Gamma(\nu + 1 + m)]^{-1}. \tag{5}$$

Upon using (5) for J_ν and $J_{-\nu}$ in (4), it becomes for $0 < |z| \ll 1$ or $|\nu| \gg 1 + |z|^2$,

$$i \sin \nu\pi\Gamma(\nu + 1)\left(\frac{z}{2}\right)^\nu H_\nu^{(1)}(z) = \frac{\Gamma(\nu + 1)}{\Gamma(\nu - 1)}\left[1 + O\left(\frac{z^2}{\nu + 1}\right)\right] - \left(\frac{z}{2}\right)^\nu e^{-i\nu\pi}\left[1 + O\left(\frac{z^2}{\nu + 1}\right)\right]. \tag{6}$$

When $H_\nu^{(1)}(z) = 0$ we transpose the quotient of gamma functions in (6) and take logarithms of the two sides of the resulting equation, obtaining

$$2\nu\left(\log \frac{z}{2} - i\frac{\pi}{2}\right) = -2\pi i n + \log \frac{\Gamma(\nu + 1)}{\Gamma(1 - \nu)} + O\left(\frac{z^2}{\nu + 1}\right). \tag{7}$$

Here n is an integer.

For $|z| \ll 1$ it is convenient to use the following series for the logarithm of the quotient of gamma functions:

$$\log \frac{\Gamma(1 + \nu)}{\Gamma(1 - \nu)} = -2\gamma\nu - 2\sum_{m=1}^{\infty} \frac{\zeta(2m + 1)}{2m + 1} \nu^{2m+1}. \tag{8}$$

Here γ is Euler's constant and ζ is the Riemann zeta function. With the aid of (8), (7) can be rewritten as

$$\log \frac{z}{2} = -\frac{i\pi n}{\nu} + \frac{i\pi}{2} - \gamma - \sum_{m=1}^{\infty} \frac{\zeta(2m + 1)}{2m + 1} \nu^{2m} + O\left(\frac{z^2}{\nu}\right). \tag{9}$$

Upon reverting the series (9) for $n \neq 0$ we obtain ν as a power series in $[\log (z/2)]^{-1}$. For $n = 0$ there is no root of (9) for which $|\nu| \ll 1$. Denoting the value of ν by ν_n and setting $z = re^{i\varphi}$, we may write the result as the following series in $[\log (r/2)]^{-1}$

$$\nu_n = -i\pi n[\log (r/2)]^{-1}\left\{1 + \left[i\left(\frac{\pi}{2} - \varphi\right) - \gamma\right] \times [\log (r/2)]^{-1} + \left[i\left(\frac{\pi}{2} - \varphi\right) - \gamma\right]^2 [\log (r/2)]^{-2} + \left(\left[i\left(\frac{\pi}{2} - \varphi\right) - \gamma\right]^3 - \zeta(3)\pi^2 n^2/3\right)[\log (r/2)]^{-3} + O(z^2/\log (z/2))\right\}, \quad |z| \ll 1. \tag{10}$$

This result for ν_n, which appears to be new, shows that all the roots ν_n tend to zero as z tends to zero. For z real this was shown to be true by Magnus and Kotin.[3] However, their theorems 6.1 and 6.2, which describe the manner in which ν_n tends to zero, are in disagreement with (10) and are incorrect.

To determine ν_n for n large and z fixed, we again proceed from (7). We assume that $|\nu| \gg 1$ and use Stirling's formula for the gamma functions in (7),

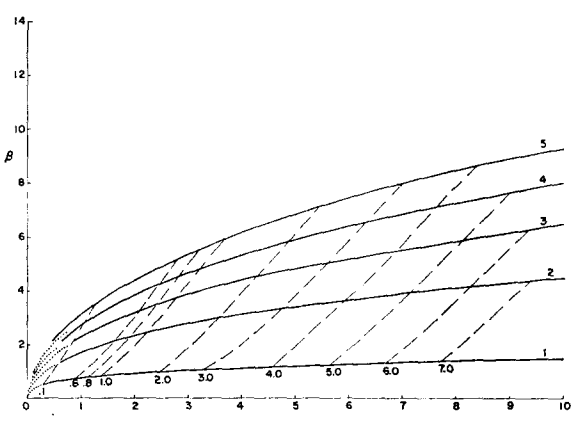

FIG. 2. The zeros $\nu_n(z)$ of $dH_\nu^{(1)}(z)/dz$ in the complex ν plane for z real. Re ν_n is plotted horizontally and Im ν_n is plotted vertically for $n = 1, 2, 3, 4, 5$, and $0 \leq z \leq 7$. The solid lines are the loci of $\nu_n(z)$ for fixed n as functions of z. The dashed lines connect values of $\nu_n(z)$ for fixed z and different values of n. The zeros are symmetric about $\nu = 0$ so there is a similar set of curves in the third quadrant.

which yields

$$\log\frac{\Gamma(1+\nu)}{\Gamma(1-\nu)} = 2\nu(\log\nu - 1 + i\pi/2) - i\pi/2 + O(\nu^{-1}). \quad (11)$$

Upon using (11) in (7), and assuming that $|\nu| \gg 1 + |z|^2$, we readily find for $n \gg 1$,

$$\operatorname{Re}\nu_n = \pi\left(\frac{\pi}{2}-\varphi\right)(n-\tfrac{1}{4})\left[\log\frac{2\pi(n-\frac{1}{4})}{er}\right]^{-2} \times [1 + O(\log\log n/\log n)],$$

$$\operatorname{Im}\nu_n = \pi(n-\tfrac{1}{4})\left[\log\frac{2\pi(n-\frac{1}{4})}{er}\right]^{-1} \times [1 + O(\log\log n/\log n)], \quad n \gg 1. \quad (12)$$

This result was obtained by Magnus and Kotin, Theorem 8.1, but their formulas contain a number of misprints. From (12) we see that both Re ν_n and Im ν_n become infinite as n becomes infinite, but that Im ν_n increases more rapidly than does Re ν_n. Consequently, arg ν_n tends to $\pi/2$ as n increases. This fact has led numerous authors to the false conclusion that ν_n approaches the imaginary axis of the ν plane as n increases.

When $|z|$ is large, $|\nu_n|$ is also large. Then for fixed n, ν_n is given by the well known formula obtained by van der Pol and Bremmer[4] with the aid of the Debye expansion for the Bessel function, and refined by Franz[5]:

$$\nu_n = z + 6^{-\frac{1}{3}}e^{i\pi/3}q_n z^{\frac{1}{3}} + \frac{1}{180}6^{\frac{1}{3}}e^{i2\pi/3}q_n^2 z^{-\frac{1}{3}} + O(z^{-1}), \quad |z| \gg n > 0. \quad (13)$$

Here q_n is the nth zero of the Airy function $A(q)$,

$$A(q_n) = \int_0^\infty \cos(t^3 - q_n t)\,dt = 0. \quad (14)$$

The first five zeros, as given by Franz,[5] are listed in Table I. For large values of n, q_n is given by the asymptotic formula

$$q_n \sim [3\pi(n+\tfrac{3}{4})]^{\frac{2}{3}}6^{\frac{1}{3}}/2, \quad n \gg 1. \quad (15)$$

Equations (10) and (13) give $\nu_n(z)$ for both large and small values of $|z|$. To obtain $\nu_n(z)$ for intermediate values of z we have solved (1) numerically for z real in the range $0.01 \leq z \leq 7$ and $n = 1, 2, 3, 4, 5$. The resulting values of ν_n are shown in Fig. 1, which shows the locus of each of the first five roots in the complex ν plane. As z increases from zero each root moves from the origin upward and to the right. Such loci have recently been called "Regge trajectories" in quantum mechanics.

[4] H. Bremmer, *Terrestrial Radio Waves* (Elsevier Publishing Company, New York, 1949).
[5] W. Franz, Z. Naturforsch. **9a**, 705 (1954).

TABLE I. The first five zeros q_n and q_n' of the Airy function and its derivative, respectively. In terms of them the zeros of $H_\nu^{(1)}(z)$ and $dH_\nu^{(1)}(z)/dz$ can be expressed by (13) and (18).

n	q_n	q_n'
1	3.372134	1.469354
2	5.895843	4.684712
3	7.962025	6.951786
4	9.788127	8.889027
5	11.457423	10.632519

We have also compared the values of ν_n given by (10) and (13) with the numerical results. For $z = 0.01$, (10) yields Re $\nu_1 = 0.205$, Im $\nu_1 = 0.613$, while the numerical solution is Re $\nu_1 = 0.184$, Im $\nu_1 = 0.592$. For larger values of z and n the disagreement is greater. Thus, we conclude that (10) is accurate only for $|z| < 0.01$. On the other hand, for $z = 1$, (13) yields Re $\nu_1 = 1.871$, Im $\nu_1 = 1.706$, while the numerical solution is Re $\nu_1 = 1.880$, Im $\nu_1 = 1.708$. This agreement is very good, and becomes better as $|z|$ increases, but worse as n increases. However, even for $n = 5$ the error in Re ν_5 is only 4% and that in Im ν_5 is only 1% at $z = 2$. At $z = 7$, (13) yields Re $\nu_1 = 8.745$, Im $\nu_1 = 3.126$, while the numerical solution is Re $\nu_1 = 8.746$, Im $\nu_1 = 3.127$.

We have restricted n to positive values in (12) and (13) and have given only the roots $\nu_n(z)$ with Re $\nu_n \geq 0$ in Fig. 1 because the roots are symmetric about the origin. This follows from the relation $H_{-\nu}^{(1)}(z) = e^{i\nu\pi}H_\nu^{(1)}(z)$.

3. ZEROS OF $dH_\nu^{(1)}(z)/dz$

The zeros of $dH_\nu^{(1)}(z)/dz$ can be found by exactly the same methods as were used in the preceding section. Therefore, we shall give only the results. Since the zeros are symmetric about the origin, we shall again give some formulas only for positive n, which corresponds to zeros in the half plane Re $\nu \geq 0$.

When $|z|$ is small we find

$$\nu_n = -i\pi(n-\tfrac{1}{2})[\log(r/2)]^{-1}\left\{1 + \left[i\left(\frac{\pi}{2}-\varphi\right)-\gamma\right] \times [\log(r/2)]^{-1} + \left[i\left(\frac{\pi}{2}-\varphi\right)-\gamma\right]^2[\log(r/2)]^{-2} + \left(\left[i\left(\frac{\pi}{2}-\varphi\right)-\gamma\right]^3 - \zeta(3)\pi^2n^2/3\right)[\log(r/2)]^{-3} + O[|z|^2/\log(z/2)]\right\}, \quad |z| \ll 1. \quad (16)$$

For $|\nu| \gg 1 + |z|^2$ we obtain

$$\operatorname{Re} \nu_n = \pi\left(\frac{\pi}{2} - \varphi\right)(n - \tfrac{3}{4})\left[\log \frac{2\pi(n - \frac{3}{4})}{er}\right]^{-2} \times \left[1 + O\left(\frac{\log\log n}{\log n}\right)\right], \quad (n \gg 1) \qquad (17)$$

$$\operatorname{Im} \nu_n = \pi(n - \tfrac{3}{4})\left[\log \frac{2\pi(n - \frac{3}{4})}{er}\right]^{-1} \times \left[1 + O\left(\frac{\log\log n}{\log n}\right)\right].$$

Both (16) and (17) are apparently new. The former shows that all ν_n tend to zero as z tends to zero. The latter shows that both Re ν_n and Im ν_n become infinite as n does and that arg ν_n tends to $\pi/2$.

When $|z|$ is large and n is fixed, ν_n is given by the formula[5]

$$\nu_n = z + 6^{-\frac{1}{3}}e^{i\pi/3}q_n' z^{\frac{1}{3}} + 6^{\frac{1}{3}}e^{i2\pi/3} \times \left[\frac{(q_n')^2}{180} + \frac{1}{10q_n'}\right]z^{-\frac{1}{3}} + O(z^{-1}), \quad |z| \gg n > 0. \qquad (18)$$

Here q_n' is the nth zero of $A'(q) = 0$ where $A(q)$ is the Airy function defined in (14). The first five zeros, as given by Franz, are listed in Table I. For n large, q_n' is given by

$$q_n' \sim [3\pi(n + \tfrac{1}{4})]^{\frac{2}{3}}6^{\frac{1}{3}}/2, \qquad n \gg 1. \qquad (19)$$

In Fig. 2 are shown the loci of values of $\nu_n(z)$ obtained by solving (2) numerically for $n = 1, 2, 3, 4, 5$, and $.01 \leq z \leq 7$. Each root moves from the origin, upward and to the right, as z increases from zero. Comparison of the numerical solutions with the results given by (16) and (19) shows about the same agreement as in the preceding case.

4. ZEROS OF $dH_\nu^{(1)}(z)/dz + iZH_\nu^{(1)}(z) = 0$

To solve (3) we use (6) to obtain for $|z| \ll 1$ or $|\nu| \gg 1 + |z|^2$,

$$2i \sin \nu\pi\,\Gamma(\nu)(z/2)^{\nu+1}\left[\frac{dH_\nu^{(1)}(z)}{dz} + iZH_\nu^{(1)}(z)\right] = -\frac{\Gamma(1+\nu)}{\Gamma(1-\nu)}\left(1 - i\frac{Zz}{\nu}\right)\left[1 + O\left(\frac{z^2}{\nu}\right)\right] - \left(\frac{z}{2}\right)^{2\nu}e^{-i\nu\pi}\left(1 + i\frac{Zz}{\nu}\right)\left[1 + O\left(\frac{z^2}{\nu}\right)\right]. \qquad (20)$$

Upon equating to zero the right side of (20) and taking logarithms of the resulting equation, we find

$$2\nu[\log(z/2) - i\pi/2] = -i2\pi(n - \tfrac{1}{2}) + \log\frac{\Gamma(1+\nu)}{\Gamma(1-\nu)} + O\left(\frac{Zz}{\nu}\right) + O\left(\frac{z^2}{\nu}\right). \qquad (21)$$

Let us first suppose that the impedance Z is a finite constant, independent of z and ν. Then it follows from (21) that the zeros of (3) are asymptotically the same as those of (1) in the two cases $|z| \ll 1$ and $|\nu| \gg 1 + |z|^2$. Thus in these two cases the zeros ν_n of (3) are given by (16) if $|z| \ll 1$ and by (17) if $|\nu| \gg 1 + |z|^2$, with an additional error term $O(Zz/\nu)$. The same result (16) applies if Z is a function of z and ν such that Zz/ν tends to zero as z tends to zero with ν given by (16). Similarly (17) applies if Zz/ν tends to zero for fixed z as ν becomes infinite through the sequence (17).

When $|z|$ is large and n is fixed, ν_n is given by the following formula, obtained by Levy and Keller[6]:

$$\nu_n = z + 6^{-\frac{1}{3}}e^{i\pi/3}q_n(Zz^{\frac{1}{2}})z^{\frac{1}{2}} + O(z^{-\frac{1}{2}}), \quad |z| \gg n > 0. \qquad (22)$$

Here $q_n(Zz^{\frac{1}{2}})$ is the nth root of the equation

$$A'(q) = A(q)e^{5\pi i/6}6^{-\frac{1}{3}}Zz^{\frac{1}{2}}. \qquad (23)$$

If $|Zz^{\frac{1}{2}}|$ is large, q_n is given by

$$q_n(Zz^{\frac{1}{2}}) = q_n(\infty) + e^{-5\pi i/6}6^{\frac{1}{3}}(Zz^{\frac{1}{2}})^{-1} + O(|Zz^{\frac{1}{2}}|^{-2}). \qquad (24)$$

The number $q_n(\infty)$ is the nth root of (14), to which (23) reduces when $Zz^{\frac{1}{2}}$ becomes infinite. If $|Zz^{\frac{1}{2}}|$ is small, q_n is given by

$$q_n(Zz^{\frac{1}{2}}) = q_n(0) - e^{5\pi i/6}[3Z/q_n(0)](z/6)^{\frac{1}{2}} + O(|Zz^{\frac{1}{2}}|^2). \qquad (25)$$

Here $q_n(0)$ is the nth root of $A'(q) = 0$, to which (23) reduces when $Zz^{\frac{1}{2}} = 0$.

[6] B. R. Levy and J. B. Keller, Commun. Pure Appl. Math. **12**, 159 (1959).

Charged Spheroid in Cylinder

W. R. SMYTHE
California Institute of Technology, Pasadena, California
(Received 1 March 1963)

The problem of a charged conducting spheroid within a coaxial conducting cylinder is solved by a slight variation of the method in J. Appl. Phys. **31**, 553 (1960). Errors in the terminal digits of Table I in that paper have been corrected and the table extended. The charge density on the spheroid, the potential between it and the cylinder, and its capacitance are given for ratios of the spheroid equatorial radius to cylinder radius of 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9, and 0.95. Tables give numerical results for five cases: the disk, the oblate spheroid with 2 to 1 axial ratio, the sphere, and the prolate spheroid with axial ratio 1 to 2. The thin prolate spheroid requires special treatment.

1. INTRODUCTION

ELECTRON guns, Faraday cylinders, Geiger point counters, and similar devices often use electrodes such as disks or wire stubs mounted coaxially in a tube. All these are special cases of a charged spheroid in a coaxial conducting cylinder. Ordinary potential-theory methods fail when the boundaries belong to coordinate systems having nothing in common but an axis of symmetry. The method used here yields essentially exact results in useful form. The special case of a charged sphere in a conducting coaxial cylinder was first treated by Knight[1] who found the potential to 0.1% for sphere-to-cylinder radius ratios of 0.1, 0.2, 0.3, 0.4, and 0.5. The present author[2] obtained very similar formulas by a different approach and gave results for the range 0.1 to 0.95 to eight places, only five of which are correct in some cases, as will be seen later. The crude treatment of the special case of the coaxial disk was given by the author[3] and a much more elegant one by Cooke and Tranter[4] who used dual Fourier–Bessel series but gave only two numerical examples.

2. PROCEDURE

The "angular" charge density assumed for oblate (o), spherical (s), and prolate (p) spheroids is a series of Legendre polynomials with constant coefficients multiplied by a factor proportional to the charge density on the focal conducting disk (o), point (s) or line segment (p) which alone would give spheroidal equipotentials. The resultant potential is a series of spheroidal harmonics, each of which is expressed in cylindrical coordinates by a Fourier cosine transform. The interior potential of the induced charge on the cylinder is then written down. Inside the spheroid, the sum of the potentials of the spheroidal and cylindrical charges must be made constant by adjustment of the spheroidal density so that all but the zero-order term in the expansion of the total potential in powers of ρ vanish in the $z = 0$ plane. This is the same as making all z derivatives of the potential zero at the origin. Each coefficient of ρ^{2p} except $p = 0$ is equated to zero giving a linear relation between the constants in the original charge density expansion. The solution of $p + 1$ of these linear algebraic simultaneous equations yields $p + 1$ of the constants. Only a few harmonics suffice for small spheroids but as they expand, p must be increased. In the following analysis, oblate (ζ, ξ), spherical (r, θ), and prolate (η, ξ) spheroids will be carried along together and identified by subscripts o, s, and p. Superimposed sections of the four cases treated when the equatorial radius ρ_0 is four fifths of the cylinder radius a is shown in Fig. 1.

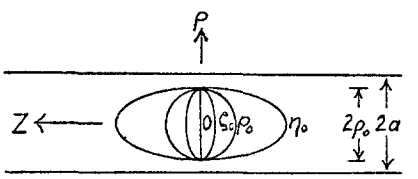

FIG. 1. Coordinate systems.

3. THEORY

The cylindrical coordinates are expressed in terms of the oblate, spherical and prolate ones by[5]

[1] R. C. Knight. Quart. J. of Math (Oxford Series) **7**, 124 (1936).
[2] W. R. Smythe. J. Appl. Phys. **31**, 553 (1960)
[3] W. R. Smythe. J. Appl. Phys. **24**, 773 (1953).
[4] J. C. Cooke and C. J. Tranter, Quart. J. Mech. Appl. Math. **12**, 379 (1959).
[5] W. R. Smythe, *Static and Dynamic Electricity* (McGraw-Hill Book Company, Inc., New York, 1950), pp. 158–169.

TABLE I. Values of $(n!)^{-2}\, I(2n) = (n!)^{-2} \int_0^\infty t^{2n} I_0(t)^{-2}\, dt$.

n	$(n!)^{-2}I(2n)$	n	$(n!)^{-2}I(2n)$	n	$(n!)^{-2}I(2n)$
0	1.3676764	28	9.419373	56	13.292934
1	1.9406673	29	9.584731	57	13.410590
2	2.5873244	30	9.747283	58	13.527222
3	3.1438618	31	9.907164	59	13.642857
4	3.6201965	32	10.064501	60	13.757518
5	4.0383488	33	10.219415	61	13.871233
6	4.4147566	34	10.372015	62	13.984021
7	4.7600615	35	10.522400	63	14.095909
8	5.0811442	36	10.670664	64	14.206914
9	5.3826426	37	10.816897	65	14.317060
10	5.6678448	38	10.961177	66	14.426362
11	5.9391878	39	11.103581	67	14.534845
12	6.1985432	40	11.244179	68	14.642522
13	6.4473885	41	11.383040	69	14.749413
14	6.6869175	42	11.520225	70	14.855535
15	6.9181111	43	11.655797	71	14.960904
16	7.1417880	44	11.789808	72	15.065536
17	7.3586412	45	11.922312	73	15.169446
18	7.5692619	46	12.053358	74	15.272651
19	7.7741601	47	12.182995	75	15.375161
20	7.9737799	48	12.311266	76	15.476991
21	8.1685105	49	12.438214	77	15.578157
22	8.3586955	50	12.563878	78	15.678670
23	8.5446398	51	12.688298	79	15.778544
24	8.7266123	52	12.811509	80	15.877786
25	8.9048631	53	12.933546	81	15.976415
26	9.0796134	54	13.054442	82	16.074436
27	9.2510554	55	13.174228	83	16.171864

$$\begin{aligned} z &= c\xi\zeta, & \rho &= c(1-\xi^2)^{\frac{1}{2}}(1+\zeta^2)^{\frac{1}{2}}, \\ z &= r\cos\theta, & \rho &= r\sin\theta, \\ z &= c\xi\eta, & \rho &= c(1-\xi^2)^{\frac{1}{2}}(\eta^2-1)^{\frac{1}{2}}. \end{aligned} \tag{1}$$

The focus spacing is $2c$. The coordinate ranges are: -1 to $+1$ for ξ, 0 to π for θ, 0 to ∞ for ζ and r, and 1 to ∞ for η. The charge densities assumed on the spheroids $\zeta = \zeta_0$, $r = c$, and $\eta = \eta_0$ are

$$\begin{aligned} \sigma_0 &= [c^2(\zeta_0^2+\xi^2)]^{-\frac{1}{2}} \sum_{n=0}^{\infty} A_n P_{2n}(\xi), \\ \sigma_s &= c^{-1} \sum_{n=0}^{\infty} A_n P_{2n}(\cos\theta), \\ \sigma_p &= [c^2(\eta_0^2-\xi^2)]^{-\frac{1}{2}} \sum_{n=0}^{\infty} A_n P_{2n}(\xi). \end{aligned} \tag{2}$$

Calculation of $\epsilon\partial V'/\partial n - \epsilon\partial V''/\partial n$ and simplification by use of the Wronskian verify that the internal and external potentials of σ_0, σ_s, and σ_p are

$$\begin{aligned} \epsilon V_0' &= j(1+\zeta_0^2)^{\frac{1}{2}} \sum_{n=0}^{\infty} A_n Q_{2n}(j\zeta_0) P_{2n}(j\zeta) P_{2n}(\xi), \\ \epsilon V_s' &= \sum_{n=0}^{\infty} A_n (4n+1)^{-1} c^{-2n} r^{2n} P_{2n}(\cos\theta), \\ \epsilon V_p' &= (\eta_0^2-1)^{\frac{1}{2}} \sum_{n=0}^{\infty} A_n Q_{2n}(\eta_0) P_{2n}(\eta) P_{2n}(\xi), \\ \epsilon V_0'' &= j(1+\zeta_0^2)^{\frac{1}{2}} \sum_{n=0}^{\infty} A_n P_{2n}(j\zeta_0) Q_{2n}(j\zeta) P_{2n}(\xi), \end{aligned} \tag{3}$$

$$\begin{aligned} \epsilon V_s'' &= \sum_{n=0}^{\infty} A_n (4n+1)^{-1} c^{2n+1} r^{-(2n+1)} P_{2n}(\cos\theta), \\ \epsilon V_p'' &= (\eta_0^2-1)^{\frac{1}{2}} \sum_{n=0}^{\infty} A_n P_{2n}(\eta_0) Q_{2n}(\eta) P_{2n}(\xi). \end{aligned} \tag{4}$$

When $z = 0$ and $\rho < c$ so $P_{2n}(\xi)$ is a polynomial in $(\rho/c)^2$, or $\cos\theta = 0$ so $P_{2n}(\cos\theta)$ is $(-1)^n(n-\frac{1}{2})!/n!$, or $\xi = 0$ so $P_{2n}(\eta)$ is a polynomial in $(\rho/c)^2$, then the nth terms of the coefficients S_o, S_s, and S_p of $(\rho/c)^{2p}$ satisfy

$$\begin{aligned} \frac{S_o(n,p)}{A_n} &= \frac{j(1+\zeta_0^2)^{\frac{1}{2}} Q_{2n}(j\zeta_0)(n+p-\frac{1}{2})!}{(-1)^{n+p}(n-p)!\,(p!)^2}, \\ \frac{S_s(n,p)}{A_n} &= \frac{(-1)^p(p-\frac{1}{2})!}{(4p+1)p!} \qquad (0 \text{ if } n \neq p), \\ \frac{S_p(n,p)}{A_n} &= \frac{(\eta_0^2-1)^{\frac{1}{2}} Q_{2n}(\eta_0)(n+p-\frac{1}{2})!}{(-1)^n(n-p)!\,(p!)^2}. \end{aligned} \tag{5}$$

Spheroidal and cylindrical solutions of LaPlace's equation are related by the Fourier transforms

$$\begin{aligned} G_0(\rho,z) &= \int_0^\infty t^{-\frac{1}{2}} K_0(t\rho/c) I_{2n+\frac{1}{2}}(t) \cos(tz/c)\, dt \\ &= j(\tfrac{1}{2}\pi)^{\frac{1}{2}} Q_{2n}(j\zeta) P_{2n}(\xi), \\ G_s(\rho,z) &= \int_0^\infty t^{2n} K_0(t\rho/c) \cos(tz/c)\, dt \\ &= (-1)^2(2n)!\,(c/r)^{2n+1} \tfrac{1}{2}\pi P_{2n}(\cos\theta), \\ G_p(\rho,z) &= \int_0^\infty t^{-\frac{1}{2}} K_0(t\rho/c) J_{2n+\frac{1}{2}}(t) \cos(tz/c)\, dt \\ &= (-1)^n(\tfrac{1}{2}\pi)^{\frac{1}{2}} Q_{2n}(\eta) P_{2n}(\xi). \end{aligned} \tag{6}$$

Thus the nth terms of the spheroidal potentials are

$$\begin{aligned} \epsilon V_{on}'' &= G_o(\rho,z)(\tfrac{1}{2}\pi)^{-\frac{1}{2}}(1+\zeta_0^2)^{\frac{1}{2}} P_{2n}(j\zeta_0) A_n, \\ \epsilon V_{sn}'' &= G_s(\rho,z)(-1)^n[\tfrac{1}{2}\pi(2n)!\,(4n+1)]^{-1} A_n, \\ \epsilon V_{pn}'' &= G_p(\rho,z)(-1)^n(\tfrac{1}{2}\pi)^{-\frac{1}{2}}(\eta_0^2-1)^{\frac{1}{2}} P_{2n}(\eta_0) A_n. \end{aligned} \tag{7}$$

Substitution of $-K_0(ta/c)I_0(t\rho/c)/I_0(ta/c)$ for $K_0(t\rho/c)$ in $G(\rho, z)$ gives the potential, when $\rho < a$, of the induced charges on the cylinder $\rho = a$ which cancel V_n'' there. In the plane $z = 0$, $I_0(t\rho/c)$ is expanded in powers of ρ/c and (for V_{on}'' or V_{pn}'') $t^{-\frac{1}{2}} I_{2n+\frac{1}{2}}(t)$ or $t^{-\frac{1}{2}} J_{2n+\frac{1}{2}}(t)$ in powers of t. The coefficients $C_o(n,p)/A_n$, $C_s(n,p)/A_n$, and $C_p(n,p)/A_n$ of $(\rho/c)^{2p}$ are

$$\begin{aligned} &\frac{-(1+\zeta_0^2)^{\frac{1}{2}} P_{2n}(j\zeta_0)}{\frac{1}{2}\pi\epsilon(p!)^2} \sum_{r=0}^{\infty} \left(\frac{c}{2a}\right)^{2n+2p+2r+1} \Phi_r, \\ &\frac{-(-4)^n}{\frac{1}{2}\pi\epsilon(p!)^2} \left(\frac{c}{2a}\right)^{2n+2p+1} \int_0^\infty \frac{u^{2n+2p} K_0(u)\, du}{(2n)!\,(2n+\frac{1}{2}) I_0(u)}, \\ &\frac{-(\eta_0^2-1)^{\frac{1}{2}} P_{2n}(\eta_0)}{(-1)^n \frac{1}{2}\pi\epsilon(p!)^2} \sum_{r=0}^{\infty} \left(\frac{c}{2a}\right)^{2n+2p+2r+1} \frac{\Phi_r}{(-1)^r}. \end{aligned} \tag{8}$$

In the integral, cu/a has been substituted for t so

$$\Phi_r = \int_0^\infty \frac{u^{2n+2p+2r} K_0(u)\, du}{r!\,(2n + r + \tfrac{1}{2})!\, I_0(u)}. \tag{9}$$

Integration by parts simplifies the integral; thus

$$\int_0^\infty \frac{u^{2m} K_0(u)\, du}{I_0(u)} = \int_0^\infty \frac{u^{2m}\, du}{(2m + 1) I_0^2(u)} = \frac{I(2m)}{2m + 1}. \tag{10}$$

Values of $(m!)^{-2} I(2m)$ are tabulated in Table I. The potential inside the spheroid will be V_0 everywhere if the coefficients of all powers of ρ/c except $p = 0$ are zero. From (5) and (8), the $N + 1$ equations, one for each value of p from 0 to N, to be solved for A_n are

$$\delta_p^0 V_0 = \sum_{n=0}^{N} [S(n, p) + C(n, p)], \tag{11}$$

where $\delta_p^0 = 1$ if $p = 0$, and zero if $p \neq 0$.

4. INTEGRATION

The integral $I(2m)$ of (1), tabulated in reference 2 for n values from 0 to 37, was evaluated by numerical integration using Weddle's rule with intervals of 0.1. Less than seven digits were correct in the range 3 to 17, and only four at $n = 10$. The integration has been repeated with intervals of 0.01 from 0 to 40, and an accurate asyptotic formula was used for $n > 37$. The three overlapping values agree to eight places, so the eighth digit in Table I is probably correct when the first digit is one or two. It is possible, but not easy, to evaluate this integral analytically.[6] The precision with which the boundary condition $V = 0$ on the cylinder surface is met depends entirely on the accuracy of Table I and is quite independent of the values of A_n.

5. SOLUTION FOR CHARGE DISTRIBUTION

Eight digits were carried through in all calculations so that no values of A_n which contribute less than $10^{-8} V_0$ to the spheroid need be found. The accuracy at equator and pole and the number of contributing digits in A_n is indicated by the sum and the individual terms in the formulas

$$V_{eq} = \sum_{p=0}^{\infty} [S(n, p) + C(n, p)](\rho_0/c)^{2p}, \tag{12}$$

$$V_{p0} = \sum_{p=0}^{\infty} [S(n, p) + C(n, p)] \frac{(-1)^p p!}{(p - \tfrac{1}{2})!} \left(\frac{z_0}{c}\right)^{2p}, \tag{13}$$

where ρ_0 is the equatorial semiaxis and z_0 the polar semiaxis. Large spheroids require more A_n values. Charge densities for ρ_0/a values of 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9, and 0.95 were found for all spheroids except the prolate for which the distance $2c$ between the foci may greatly exceed the diameter $2a$ of the cylinder, so that the r summation in (8) diverges. With a 2 to 1 axial ratio, the solution blows up at $\rho_0 = 0.7a$, although A_0 appears to be stable to five places for this value where (8) should still converge. Only 100 values of $I(2m)$ were provided, the terms alternate in sign, and may exceed the sum in size; therefore more than eight places and more values of $I(2m)$ should be carried. The following tables give the charge density coefficients $A_0, A_1, \cdots A_n$ of (2). The pole and equator errors are shown by symbols such as $+1(6)P$ and $-7(4)E$ which mean that the pole and equator potentials from (12) and (13) are 1.000001 V_0 and 0.9993 V_0. The disk has zero pole error throughout, but the error at the edge gets very large when $c = 0.95a$.

[6] G. N. Watson, Proc. Roy. Soc. (London) **A130**, 29 (1930).

TABLE II. Disk charge density coefficients A_n.

n	$b = 0.1a$	n	$b = 0.6a$	n	$b = 0.9a$
0	0.67404096	0	0.98739140	0	1.6463610
1	−0.00011871	1	−0.05129962	1	−0.5735855
2	0.00000015	2	0.00324102	2	0.1684656
	$+1(7)E$	3	−0.00017520	3	−0.0459833
		4	0.00000904	4	0.0122054
n	$b = 0.2a$	5	−0.00000045	5	−0.0031163
			$-1(8)E$	6	0.0007442
0	0.71656810			7	−0.0001599
1	−0.00103341	n	$b = 0.7a$	8	0.0000295
2	0.00000522			9	−0.0000044
	$+2(8)E$	0	1.1114951	10	0.0000005
		1	−0.1062954		$-2(4)E$
		2	0.0107524		
n	$b = 0.3a$	3	−0.0009578	n	$b = 0.95a$
		4	0.0000827		
0	0.76585448	5	−0.0000070	0	2.0230351
1	−0.00387960	6	0.0000006	1	−1.0931716
2	0.00004623		$-2(8)E$	2	0.4665000
3	−0.00000044			3	−0.1842942
	$+1(7)E$			4	0.0689906
		n	$b = 0.8a$	5	−0.0240151
				6	0.0075682
n	$b = 0.4a$	0	1.2981660	7	−0.0020970
		1	−0.2282870	8	0.0004950
0	0.82435971	2	0.0376865	9	−0.0000960
1	−0.01049675	3	−0.0056167	10	0.0000146
2	0.00023792	4	0.0008039	11	−0.0000016
3	−0.00000439	5	−0.0001074	12	0.0000001
	$+7(8)E$	6	0.0000123		$-3(3)E$
		7	−0.0000010		
			$-2(6)E$		
n	$b = 0.5a$				
0	0.89597682				
1	−0.02415880				
2	0.00093811				
3	−0.00003036				
4	0.00000092				
	$+1(7)E$				

TABLE III. Oblate spheroid: axial ratio 2 to 1; coefficients A_n.

n	$\rho_0 = 0.1a$	n	$\rho_0 = 0.6a$	n	$\rho_0 = 0.9a$
0	0.89124399	0	1.5134060	0	3.3628572
1	−0.00044114	1	−0.2039024	1	−2.5608492
2	0.00000121	2	0.0262279	2	1.2820136
	$+1(7)E$ $-1(8)P$	3	−0.0028433	3	−0.5759987
		4	0.0002915	4	0.2496183
		5	−0.0000293	5	−0.1061917
n	$\rho_0 = 0.2a$	6	0.0000029	6	0.0446301
			$+1(8)E$ $+2(8)P$	7	−0.0185842
0	0.96690610			8	0.0076715
1	−0.00389588			9	−0.0031308
2	0.00004354	n	$\rho_0 = 0.7a$	10	0.0012524
3	−0.00000038			11	−0.0004825
	$+5(8)E$ $-3(8)P$	0	1.8017630	12	0.0001738
		1	−0.4302339	13	−0.0000561
		2	0.0851600	14	0.0000153
n	$\rho_0 = 0.3a$	3	−0.0145422	15	−0.0000032
		4	0.0023774	16	0.0000005
0	1.0580652	5	−0.0003829		$-1(6)E$ $-3(8)P$
1	−0.0148229	6	0.0000613		
2	0.0003859	7	−0.0000098		
3	−0.0000080	8	0.0000015	n	$\rho_0 = 0.95a$
4	0.0000002	9	−0.0000002		
	$+4(8)E$ $+5(8)P$		$+3(8)E$ $+4(8)P$	0	4.8491544
				1	−5.3882024
				2	3.6841919
n	$\rho_0 = 0.4a$	n	$\rho_0 = 0.8a$	3	−2.2675053
				4	1.3401280
0	1.1710668	0	2.2839489	5	−0.7740432
1	−0.0406189	1	−0.9517740	6	0.4401314
2	0.0019772	2	0.2914529	7	−0.2472183
3	−0.0000779	3	−0.0787566	8	0.1372507
4	0.0000028	4	0.0205295	9	−0.0751038
5	−0.0000001	5	−0.0052759	10	0.0402010
	$+4(8)E$ $+5(8)P$	6	0.0013452	11	−0.0207629
		7	−0.0003411	12	0.0101369
		8	0.0000859	13	−0.0045533
n	$\rho_0 = 0.5a$	9	−0.0000214	14	0.0018202
		10	0.0000052	15	−0.0006219
0	1.3164529	11	−0.0000012	16	0.0001725
1	−0.0946855		$-1(8)E$ $+0(8)P$	17	−0.0000361
2	0.0077157			18	0.0000050
3	−0.0005185				$-1(8)E$ $+3(8)P$
4	0.0000326				
5	−0.0000020				
	$+5(8)E$ $+4(8)P$				

TABLE IV. Sphere charge density coefficients, A_n.

n	$c = 0.1a$	n	$c = 0.6a$		$c = 0.9a$
0	1.0953730	0	2.1028947	0	5.4049470
1	−0.0011278	1	−0.4802449	1	−5.2715178
2	0.0000054	2	0.0862206	2	2.9084072
	$+4(8)E$ $+2(8)P$	3	−0.0123210	3	−1.3779971
		4	0.0015912	4	0.6179115
		5	−0.0001953	5	−0.2696932
n	$c = 0.2a$	6	0.0000233	6	0.1157088
		7	−0.0000027	7	−0.0490241
0	1.2108599	8	0.0000003	8	0.0205694
1	−0.0099742		$+6(8)E$ $+4(8)P$	9	−0.0085640
2	0.0001915			10	0.0035435
3	−0.0000029			11	−0.0014588
	$+1(8)E$ $+1(8)P$	n	$c = 0.7a$	12	0.0005981
				13	−0.0002444
		0	2.5965817	14	0.0000996
n	$c = 0.3a$	1	−0.9733744	15	−0.0000405
		2	0.2493095	16	0.0000164
0	1.3536157	3	−0.0517044	17	−0.0000066
1	−0.0376580	4	0.0098373	18	0.0000027
2	0.0016286	5	−0.0017998		$+6(8)E$ $+2(7)P$
3	−0.0000552	6	0.0003231		
4	0.0000017	7	−0.0000574		
	$+5(8)E$ $+4(8)P$	8	0.0000101	n	$c = 0.95a$
		9	−0.0000018		
		10	0.0000003	0	8.236134
n	$c = 0.4a$		$+2(8)E$ $+3(8)P$	1	−10.829254
				2	7.829408
0	1.5348028	n	$c = 0.8a$	3	−4.943960
1	−0.1014744			4	2.963928
2	0.0078308	0	3.4435550	5	−1.727846
3	−0.0004745	1	−2.0573935	6	0.988951
4	0.0000265	2	0.7528915	7	−0.558645
5	−0.0000013	3	−0.2290124	8	0.312483
	$+4(8)E$ $+4(8)P$	4	0.0650554	9	−0.173470
		5	−0.0179459	10	0.095724
n	$c = 0.5a$	6	0.0048761	11	−0.052568
		7	−0.0013121	12	0.028753
0	1.7731044	8	0.0003504	13	−0.015674
1	−0.2305890	9	−0.0000930	14	0.008520
2	0.0280898	10	0.0000245	15	−0.004620
3	−0.0026983	11	−0.0000064	16	0.002499
4	0.0002321	12	0.0000017	17	−0.001349
5	−0.0000188		$-2(8)E$ $+4(8)P$	18	0.000727
6	0.0000015			19	−0.000391
	$+6(8)E$ $+6(8)P$				$0(6)E$ $0(6)P$

6. POTENTIAL

The potential between spheroid and cylinder can be written

$$V = \epsilon^{-1} \sum_{m=0}^{\infty} B_m e^{-k_m z} J_0(k_m \rho), \tag{14}$$

where k_m is chosen so that $J_0(k_m a) = 0$. The B_m is obtained by finding the charge distribution on the focal disk, point or line segment which yields the same external potential as the given oblate, spherical, or prolate spheroidal distribution. This source can be described in cylindrical coordinates and B_m determined in the usual way from the Green's function (reference 5, pp. 177–179). The results are, if ρ_0 is the equatorial radius,

$$B_{mo} = \sum_{n=0}^{N} \frac{(-1)^n P_{2n}(j\zeta_0) J_{2n+\frac{1}{2}}(k_m c) A_n}{(2\pi k_m \rho_0)^{-\frac{1}{2}} [k_m a J_1(k_m a)]^2},$$

$$B_{ms} = \sum_{n=0}^{N} \frac{2(k_m \rho_0)^{2n+1} A_n}{(2n)!\,(4n+1)[k_m a J_1(k_m a)]^2}, \tag{15}$$

$$B_{mp} = \sum_{n=0}^{N} \frac{P_{2n}(\eta_0) I_{2n+\frac{1}{2}}(k_m c) A_n}{(2\pi k_m \rho_0)^{-\frac{1}{2}} [k_m a J_1(k_m a)]^2},$$

where B_{mp} is useful only when $z > c$.

7. CAPACITANCE

The capacitance between spheroid and cylinder, found by integration of (2) to get the total charge and division by the potential, depends only on the ratio of the equatorial radius ρ_0 to the cylinder radius a, on a, and on A_0.

TABLE V. Prolate spheroid: axial ratio 1 to 2; coefficients A_n.

n	$\rho_0 = 0.1a$	n	$\rho_0 = 0.3a$	n	$\rho_0 = 0.5a$
0.	1.4843771	0	1.9648054	0	2.7601823
1	−0.0042531	1	−0.1294283	1	−0.6680829
2	0.0000442	2	0.0100067	2	0.1104591
3	−0.0000004	3	−0.0005570	3	−0.0116244
	+3(8)E +1(8)P	4	0.0000244	4	0.0007431
		5	−0.0000008	5	−0.0000044
			+3(8)E +0(8)P	6	−0.0000052
n	$\rho_0 = 0.2a$			7	0.0000005
					+1(8)E 0(8)P
0	1.6968580	n	$\rho_0 = 0.4a$		
1	−0.0364688				
2	0.0014053	0	2.3077698	n	$\rho_0 = 0.6a$
3	−0.0000410	1	−0.3214858		
4	0.0000010	2	0.0386581	0	3.387760
	+4(8)E +2(8)P	3	−0.0031302	1	−1.274119
		4	0.0001743	2	0.275713
		5	−0.0000043	3	−0.037899
		6	0.0000004	4	0.003455
			+3(7)E +1(8)P	5	−0.000218
				6	0.000004
					−1(5)E +8(6)P

$$C = 4\pi\epsilon_v\rho_0 A_0 = 10^7 c^{-2}\rho_0 A_0 , \qquad (16)$$

where the velocity of light c is taken to be 299,792.5 ± 0.3 m/sec. The uncertainty in c limits the accuracy of the capacitance in farads to about six places.

The capacitance when one or both ends of the cylinder is closed can be found with good accuracy if the end is not too close to the spheroid by the method of reference 2.

8. THIN PROLATE SPHEROID

Examination of the $n = 0$ terms of (2) and (4) shows that, in unbounded space, a uniform linear charge of length $2c$ produces a potential independent of ξ so that all equipotentials are confocal prolate spheroids in which the charge occupies the axis between the foci. When the total charge Q is not zero, the potential at the line charge is infinite. Thus, if an infinitely thin charged conducting wire of length $2c$ is placed on the axis of a conducting cylinder, the potential of the induced charges on the wall will not disturb the uniform charge density and the equipotentials sufficiently near the wire will remain spheroidal. The potential of such a line charge, found by integration of the Green's function for a cylinder (reference 5, p. 178) is

TABLE VI. Capacitance is C F; $C \times 10^{10}/a$ is tabulated.

ρ^0/a	Disk	Oblate (2/1)	Sphere	Prolate (1/2)
0.1	0.0749972	0.0991642	0.121877	0.165159
0.2	0.159458	0.215166	0.269454	0.377602
0.3	0.255638	0.353177	0.451830	0.655842
0.4	0.366889	0.521195	0.683079	1.027096
0.5	0.498454	0.732375	0.986422	1.53556
0.6	0.659173	1.010334	1.403871	2.26163
0.7	0.865693	1.403312	2.022360	3.3740
0.8	1.15553	2.032988	3.065176	—
0.9	1.6486	3.36752	5.412431	—
0.95	2.139	5.1468	8.70574	—

$$V = \frac{Q}{2\pi\epsilon c}\sum_{r=0}^{\infty}\frac{\sinh k_r c J_0(k_r\rho)}{[k_r a J_1(k_r a)]^2}\, e^{-k_r|z|} \qquad |z| > c, \qquad (17)$$

$$= \frac{Q}{2\pi\epsilon c}\sum_{r=0}^{\infty}\frac{J_0(k_r\rho)(1 - e^{-k_r c}\cosh k_r z)}{[k_r a J_1(k_r a)]^2} \qquad |z| < c. \qquad (18)$$

For a given focal distance $2c$, the equipotentials given by (17) and (18) will have a greater polar-to-equatorial axis ratio than a true spheroid because the negative potential of the induced charge on the cylinder has less influence on the high field near the pole than on the weaker one near the equator. Thus if the pole of one of these surfaces coincides with that of a given spheroid, the surface will lie entirely inside and will have a smaller capacitance but if the equators coincide, it will lie entirely outside and will have a greater capacitance This provides a means of getting close limit[8] on the capacitance between a thin prolate spheroid of equatorial radius ρ_0 and polar radius z_0 and a coaxial cylinder of radius a when $\rho_0 \ll a$. The semifocal distance is $c = (z_0^2 - \rho_0^2)^{\frac{1}{2}}$. If insertion of c, $z = z_0$ and $\rho = 0$ in (17) gives $V = V_1$, and insertion of c, $z = 0$ and $\rho = \rho_0$ in (18) gives V_2, then

$$Q/V_1 < C < Q/V_2. \qquad (19)$$

The calculation of V_1 is straightforward but tedious because of the slow convergence of (17) for very thin spheroids when z_0 is close to c.

Note added in proof: In a recent article in Proc. Cambridge Phil. Soc. **57**, 623 (1961), W. D. Collins solves the disk case by an integral equation method. He gives additional references, but no numerical results.

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 4, NUMBER 6 JUNE 1963

Quadratic Invariants of Surface Deformations and the Strain Energy of Thin Elastic Shells*

HYMAN SERBIN
Hughes Aircraft Company, El Segundo, California
(Received 22 June 1962)

A conceptual theory of thin, elastic shells is developed. A general form for the strain energy is postulated in terms of parameters which characterize, in the sense of Euclid, the geometry of the shell. The strain energy permits a formal definition of generalized stresses and then a statement of equilibrium. The physical interpretation of the stresses is obtained from the boundary conditions. The energy coefficients, which are not derivable within the scope of this theory, are related to material constants by using classical results derived from the three-dimensional theory of elasticity. The result is a complete, self-contained theory of shells which can be used for prediction. The methods used depend on a combination of Cartesian tensors and exterior differential forms.

INTRODUCTION

THERE have been occasions in the development of modern physics when the existing physical theory failed to predict the outcome of experiments, and it became necessary to reformulate the physical laws at the foundation of the theory. Again, in a number of cases, the first step was to consider the *formal* requirements of the theory, and the second step was to postulate only the simplest formalism which could satisfy these requirements. Without any further justification, there was no *a priori* assurance that such simple laws were inherently more descriptive of natural phenomena than more complex statements. The true justification awaited the execution of further experiments which presented *a posteriori* the vindication of the simple laws, at least to the first order of accuracy.

There can hardly be any alternative to this program when the foundations of the science are in question. On the other hand, the situation seems quite different in continuum mechanics. Here the physical laws dealing with the equilibrium, dynamics, and thermodynamics of media in familiar situations have never remained at variance with experiment, and they are accepted as the correct foundations for all problems of continuum mechanics. Ideally, every such problem is solvable by deductive methods starting from these laws.

It is perhaps less appreciated than it should be that, even within the range of classical physics, progress has been made by departing from the deductive pattern in favor of the formal or conceptual pattern described above. These situations occur because the deductive methods have been difficult to carry out, whereas the formal methods have led to predictions which could be compared with experiments. The most immediate instance, from our standpoint, is the concept that a collection of atoms and molecules, themselves complex systems, can be regarded as a continuum. A formal consequence is that a description of the displacement of this collection can be furnished by a continuous vector field. This is a potent simplification which, so far as the writer knows, has never been formally justified on the basis of any more fundamental theory. Yet one accepts this description on the pragmatic grounds that rigorously derived consequences are in agreement with experiment.

Continuing in the same vein, one could, in principle, develop a theory of elasticity for solid bodies by considering the forces between the molecules of the body, or between the atoms or their component particles. In fact, the classical work of Navier (Love[1]) was in this spirit. To bring the work up to date would require a tremendous improvement in our knowledge of molecular forces. On the other hand, the presentation of the theory of three-dimensional elasticity is today almost always made on *formal* lines depending on the construction of invariants. Again it must be emphasized that there is no *a priori* proof of the theory. It is quite possible that a more refined theory of elasticity derived from the physical properties of atoms will show that our conceptions of elasticity (whether for small or large displacements) are only approximate. This would not alter the viewpoint that the present theory of three-dimensional elasticity is correct in a pragmatic way for many materials.

* Principal results were announced in a Letter; J. Aerospace Sci. 29, 1271 (1962).

[1] A. E. H. Love, *A Treatise on the Mathematical Theory of Elasticity*, (Cambridge University Press, London, 1934). 4th ed.

STATEMENT OF THE PROBLEM

We find it necessary to dwell on these aspects of science, stressing the variety of philosophical attitudes according to which a theory must be judged, as a prelude to our discussion of the thin elastic shells and their displacements. The theory of such continua can, in principle, be subsumed under three-dimensional elasticity. The difficulties and the inconsistencies between the results of various scholars in this field are well known (see for instance, Green[2]). Therefore an alternative method, complementing the deductive method, is desirable. We pose the question: If a thin shell is regarded as a mathematical surface S, what is the most general expression $\mathfrak{F}$ in a class of functions, depending on S and on an infinitesimal displacement field of points on S, such that $\mathfrak{F}$ is nonnegative valued, vanishing if and only if the displacement field is a Euclidean transformation? Such an expression might then represent the strain energy per unit area of the thin shell.

The question is stated somewhat loosely because of the indefiniteness associated with the "class of functions". What this class should be is a matter of personal taste. A mathematician might favor as wide a class as possible. From the standpoint of constructing a physical theory to be compared with tests, it is best that the class should be as narrow as possible, just large enough to permit a nonvacuous solution and large enough to contain functions of types expected in other comparable physical problems; the important point is that a solution will permit decisive prediction of and comparison with test results.

In order to specify a class of functions, we postulate first that the function $\mathfrak{F}$ should take the same value for two deformations which differ only by a Euclidean transformation. Then, if under a Euclidean motion S goes into S', an arbitrary deformation of S into a final configuration can be regarded as a deformation from S' into the same final configuration. Therefore, the function $\mathfrak{F}$ should be unchanged under the Euclidean motion $S \rightarrow S'$ of the undeformed surface.

It is known (Eisenhart[3]) that a surface (in three dimensions) is completely specified up to Euclidean motion[4] by the coefficients of its first and second fundamental forms. That is, given these coefficients, the equations defining the surface coordinates are unique except for constants where values define Euclidean motions of the surface. Any function of these coordinaties which is independent of the constants is therefore a function (in the broad sense) of the coefficients of the first and second fundamental forms associated with the surface. When two surfaces are concerned, then the function must depend on the first and second fundamental forms of both surfaces, or, what is equivalent, on the first and second fundamental forms of the first surface and the differences $\epsilon_{\alpha\beta}$, $\kappa_{\alpha\beta}(\alpha, \beta = 1, 2)$ of corresponding coefficients of these forms. Since these differences are small, we shall restrict the class of functions in our problem to homogeneous quadratic forms of $\epsilon_{\alpha\beta}$ and $\kappa_{\alpha\beta}$. The coefficients of these forms, in turn, are functions of the first and second fundamental forms of the undeformed surface S. We shall be interested in the case where the principal radii of curvature are very large.[5] When the radii are both infinite, as in the case where S is a plane, the function $\mathfrak{F}$ must certainly exist. Hence when these radii are large, we shall suppose that the coefficients of the quadratic forms can be approximated by functions only of the coefficients of the first fundamental form of S.

At each point P of S, consider the tangent plane and the normal. *A Cartesian coordinate system* or *frame* consists of associating to each point P as origin, a triad of mutually orthogonal unit vectors of which the first two, denoted by 1 and 2, are in the tangent plane, and the third is directed along normal in such a direction that the transformation from the triad at P to the triad at a nearby point P' is properly orthogonal. We include coordinate systems such that the transformation from one triad of one system to another triad of a second system are improper. A vector P can be decomposed in each coordinate system along the local unit vectors into tangential components u_α, $\alpha = 1, 2$, and a normal component w. Under a change from one coordinate system to another, the components u_α transform like the components of a two-vector under the orthogonal group whereas w is invariant; the matrix of coefficients of the first fundamental form reduces to the unit matrix. With this choice of axes, the function $\mathfrak{F}$ depends only on the changes $\epsilon_{\alpha\beta}$ and $\kappa_{\alpha\beta}$ in the coefficients of the first and second fundamental forms.

We can now state the fundamental problem in

[2] A. E. Green, Proc. Roy. Soc. (London), **A226**, 143 (1962).

[3] L. P. Eisenhart, *Introduction To Differential Geometry*, (Princeton University Press, Princeton New Jersey, 1947).

[4] The motions referred to include both proper and improper motions. Since we are concerned only with infinitesimal displacements, the Euclidean motions are proper (determinant = 1).

[5] Strictly, large in comparison with the thickness of the shell.

precise terms. Let u_α, w represent the components of a displacement field of the surface S. It is required to exhibit the most general function $\mathfrak{F}$ of $\epsilon_{\alpha\beta}$ and $\kappa_{\alpha\beta}$ satisfying the requirements

(a) $\mathfrak{F}$ is homogeneous and quadratic in $\epsilon_{\alpha\beta}$ and $\kappa_{\alpha\beta}$ with constant coefficients;

(b) $\mathfrak{F}$ is invariant under a change in coordinate system;

(c) $\mathfrak{F}$ is positive definite.

The solution of this problem will be discussed in the next section. For our purposes, it is necessary to give an explicit representation of $\epsilon_{\alpha\beta}$ and $\kappa_{\alpha\beta}$ in terms of u_α and w. Finally, in order to complete the theory within the scope set forth, we shall consider the variation in $\mathfrak{F}$ induced by a variation in the deflection field and present the integrated consequences as the equations of equilibrium.

It must be emphasized for the mathematically oriented reader, that the problem as stated above is closely related to but different from the classical result[3] which characterizes surfaces by a number of functions which are the coefficients of fundamental forms. The difference is that we ask for only *one* form of specific (second) degree in certain variables.

INVARIANCE

If $\mathfrak{F}(\epsilon_{\alpha\beta}, \kappa_{\alpha\beta})$ is a polynomial in $\epsilon_{\alpha\beta}$ and $\kappa_{\alpha\beta}$ which is invariant under an orthogonal transformation of coordinates in the tangent plane, then it remains invariant under the formal substitutions

$$\epsilon_{\alpha\beta} \to x_\alpha x'_\beta, \qquad \kappa_{\alpha\beta} \to y_\alpha y'_\beta,$$

where x_α, x'_α, y_α, y'_α are coordinates of four vectors respectively. If $\mathfrak{F}$ is of degree 2 in $\epsilon_{\alpha\beta}$ and $\kappa_{\alpha\beta}$, then $\mathfrak{F}$ is of degree 4 in the components of these vectors, but of the same degree in x_α and x'_α on the one hand, and in y_α and y'_α on the other. An invariant function of this type becomes, by restitution of the variables $\epsilon_{\alpha\beta}$ and $\kappa_{\alpha\beta}$, an invariant function in these variables. A formal construction of all polynomial functions of vectors is known (Weyl[6]). First, it is necessary to mention that because the orthogonal group includes both proper and improper rotations, then one must restrict the discussion only to *even* invariants (those which keep the same sign under improper transformations). With this restriction, all formal invariants can be derived by using a basis of all possible inner products of x_α, x'_α, y_β, y'_β. First consider only those of fourth degree which depend on x_α and x'_α. Then these variables must occur to second degree and the only possibilities[7] are

$$(x_\alpha x_\alpha)(x'_\beta x'_\beta) = x_\alpha x'_\beta x_\alpha x'_\beta \to \epsilon_{\alpha\beta}\epsilon_{\alpha\beta},$$
$$(x_\alpha x'_\alpha)(x_\beta x'_\beta) \to \epsilon_{\alpha\alpha}\epsilon_{\beta\beta}.$$

Similar consideration of the invariants depending on y_β and y'_β gives rise to the invariants $\kappa_{\alpha\beta}\kappa_{\beta\alpha}$ and $\kappa_{\alpha\alpha}\kappa_{\beta\beta}$. The invariants which depend linearly on x_α, x'_α, y_β, y'_β are of the form

$$(x_\alpha x'_\alpha)(y_\beta y'_\beta) \to \epsilon_{\alpha\alpha}\kappa_{\beta\beta},$$
$$(x_\alpha y_\alpha)(x'_\beta y'_\beta) \to \epsilon_{\alpha\beta}\kappa_{\alpha\beta},$$
$$(x_\alpha y'_\alpha)(x'_\beta y_\beta) \to \epsilon_{\alpha\beta}\kappa_{\beta\alpha}.$$

Because of the symmetry of the tensors $\epsilon_{\alpha\beta}$ and $\kappa_{\alpha\beta}$, it follows that the most general quadratic invariant is a linear function of

$$\epsilon^2_{\alpha\alpha}, \quad \epsilon_{\alpha\beta}\epsilon_{\beta\alpha}, \quad \kappa^2_{\alpha\alpha}, \quad \kappa_{\alpha\beta}\kappa_{\beta\alpha}, \quad \epsilon_{\alpha\alpha}\kappa_{\beta\beta}, \quad \epsilon_{\alpha\beta}\kappa_{\beta\alpha}. \tag{1}$$

It will be shown below that, under an inversion of the normal to the surface, the terms $\epsilon_{\alpha\beta}$ remain invariant whereas the terms $\kappa_{\alpha\beta}$ change sign. Therefore a general linear function of the invariants in (1) must be independent of the last two. Hence $\mathfrak{F}$ must have the form

$$\mathfrak{F} = \tfrac{1}{2}(F_1\epsilon^2_{\alpha\alpha} + F_2\epsilon_{\alpha\beta}\epsilon_{\beta\alpha} + F_3\kappa^2_{\alpha\alpha} + F_4\kappa_{\alpha\beta}\kappa_{\beta\alpha}). \tag{2}$$

Since $\epsilon_{\alpha\beta}$ and $\kappa_{\alpha\beta}$ are regarded as independent, $\mathfrak{F}$ is positive definite if and only if the two forms

$$F_1\epsilon^2_{\alpha\alpha} + F_2\epsilon_{\alpha\beta}\epsilon_{\beta\alpha}, \tag{3}$$

$$F_3\kappa^2_{\alpha\alpha} + F_4\kappa_{\alpha\beta}\kappa_{\beta\alpha} \tag{4}$$

are positive definite. Considering first (3), choose a coordinate system in which $\epsilon_{12} = \epsilon_{21} = 0$. Then

$$F_1\epsilon^2_{\alpha\alpha} + F_2\epsilon_{\alpha\beta}\epsilon_{\beta\alpha} = (F_1 + F_2)\epsilon^2_{11} + 2F_1\epsilon_{11}\epsilon_{22} + (F_1 + F_2)\epsilon^2_{22}.$$

Since the discriminant of the quadratic in ϵ_{11}, ϵ_{22} is

$$\begin{vmatrix} F_1 + F_2 & F_1 \\ F_1 & F_1 + F_2 \end{vmatrix} = (F_1 + F_2)^2 - F_1^2,$$

the quadratic form is positive definite if and only if

$$F_1 + F_2 \geqq |F_1| \geqq 0. \tag{5}$$

Similarly, considering (4),

$$F_3 + F_4 \geqq |F_3| \geqq 0. \tag{6}$$

Theorem 1. *The most general invariant $\mathfrak{F}$ satisfying*

[6] H. Weyl, *The Classical Groups*, (Princeton University Press, Princeton, New Jersey, 1946). p. 53.

[7] The tensor summation convention is followed, i.e., a repeated Greek index represents summation on that index on the range 1, 2. A repeated Latin index represents summation over the number of dimensions of the space.

conditions a–c of Sec. 2 is of the form Eq. (2) where the coefficients satisfy inequalities (5) and (6).

CARTESIAN TENSORS

We have already described, in Sec. 2, a method of attaching to each point P of S a local Cartesian coordinate system such that two axes 1 and 2 are in the plane, tangent to S at P, and the third axis 3 is normal. Any function on S which is unchanged when the coordinate system is unchanged is a *scalar*; the normal displacement w along the axis 3 is an example. On the other hand, the components u_α are the components of a Cartesian vector. One can regard vectors as special kinds of *tensors* in the sense that a vector is a tensor of rank one. A tensor $u_{\alpha\beta}$ of rank two, for instance, is defined as a set of quantities ($\alpha, \beta = 1, 2$) which transform like $u_\alpha u'_\beta$, respectively, where u_α, u'_α are vectors.

From another standpoint, one may wish to regard Cartesian tensors as a special case of general tensors, special in the sense that the components of the metric tensor $g_{\alpha\beta}$ reduce to the Kronecker $\delta_{\alpha\beta}$. The representation of the surface geometry by means of general tensors, a practice which is favored by many researchers in shell theory, comes about by coordinatizing the points on S by means of a system of curvilinear coordinates. The local coordinate system at a point P is then attached to the local oblique coordinate system defined by the curvilinear coordinates. In contrast, the Cartesian frames can be positioned arbitrarily relative to one another. Moreover, the Cartesian frames have the advantage that a double-valued (covariant, contravariant) representation of vectors is not required. Finally, any tensor equation in Cartesian coordinates is, *ipso facto*, valid in curvilinear coordinates.

Two important tensors are defined by

$$\begin{aligned} \delta_{\alpha\beta} &= 0, \qquad \alpha \neq \beta, \\ \delta_{\alpha\beta} &= 1, \qquad \alpha = \beta, \end{aligned} \tag{7}$$

and

$$\begin{aligned} e_{\alpha\beta} &= 0, \qquad \alpha = \beta, \\ e_{12} &= -e_{21} = 1. \end{aligned} \tag{8}$$

The use of $e_{\alpha\beta}$ is predicated upon the restriction of the orthogonal group to *proper* rotations.

A useful vector transformation is given by

$$u'_\alpha = e_{\alpha\beta} u_\beta,$$

which results in the transformation

$$\begin{aligned} u'_1 &= u_2, \\ u'_2 &= -u_1. \end{aligned} \tag{9}$$

Let $\mathbf{e}_\alpha$, $\alpha = 1, 2$, represent unit vectors directed along axes 1 and 2 of the local coordinate system at P, and let $\mathbf{e}(= \mathbf{e}_3)$ represent the unit vector along axis 3, normal to the tangent plane. The vector $\mathbf{PP'}$ issuing from P to a nearby point P' on S may be represented by

$$\mathbf{PP'} = \omega_\alpha \mathbf{e}_\alpha, \tag{10}$$

where ω_α are the components. At the point P', the corresponding unit vectors may be regarded as having been obtained[8] by small rotations ω_{12} about axis $\mathbf{e}$, ω_{23} about $\mathbf{e}_1$, and ω_{31} about axis $\mathbf{e}_2$. Each of the components of rotation are infinitesimals of the first order in the displacement along S, hence are representable as linear combinations of the ω_α introduced in Eq. (10). However, before writing these relations, it is necessary to identify the tensorial character of the quantities ω_{ij}. It is clear that ω_{12} is an invariant; it can therefore be written

$$\omega_{12} = r_\alpha \omega_\alpha, \tag{11}$$

where r_α are components of a vector. The triad ω_{23}, ω_{31}, ω_{12} are, in general, components of a three-vector; however, when the transformations are restricted so that ω_{12} is a scalar, then the pair ω_{23}, ω_{31} transform like a two-vector. According to Eq. (9), it follows that $-\omega_{31}$, ω_{23} are also components of a two-vector. Then one can write

$$\begin{aligned} \omega_{13} = -\omega_{31} &= p_{1\beta}\omega_\beta, \\ \omega_{23} &= p_{2\beta}\omega_\beta, \end{aligned} \tag{12}$$

where the $p_{\alpha\beta}$ are components of a tensor of rank two.

Let u_α be components of a vector at P and u'_α components of another vector at the nearby point P'. Then u'_2 is said to be obtained from u_α by *parallel displacement* if the quantities u'_α are related to u_α by the rotation of axes given by Eq. (11), i.e., if

$$\begin{aligned} u'_1 &= u_1 + \omega_{12} u_2, \\ u'_2 &= u_2 - \omega_{12} u_1. \end{aligned}$$

This may be written more compactly in the form

$$u'_\alpha = u_\alpha + \omega_{\alpha\beta} u_\beta. \tag{13}$$

For more general u'_α, the difference between u'_α and the vector represented by Eq. (13) is again a vector whose components are infinitesimal to the first order in ω_α. First write the small changes in

[8] It is convenient to describe the general infinitesimal rotation of an n-dimensional Cartesian coordinate system by writing the change in the unit vector $d\mathbf{e}_i$ in the form $d\mathbf{e}_i = \omega_{ij}\mathbf{e}_j$. The ω_{ij} form the components of a skew symmetrical tensor.

the components u_α in the form

$$u'_\alpha - u_\alpha = u_{\alpha|\beta}\omega_\beta. \tag{14}$$

Then the vector difference is

$$u'_\alpha - (u_\alpha + \omega_{\alpha\beta}u_\beta) = u_{\alpha|\beta}\omega_\beta - \omega_{\alpha\beta}u_\beta. \tag{15}$$

The components of $\omega_{\alpha\beta}$ are zero unless α and β are different, in which case,

$$\omega_{12} = -\omega_{21} = r_\alpha\omega_\alpha,$$

according to Eq. (11). Hence $\omega_{\alpha\beta}$ is, except for a multiplicative factor, equal to $e_{\alpha\beta}$;

$$\omega_{\alpha\beta} = \omega_{12}e_{\alpha\beta}.$$

The right side of Eq. (15) can then be put into the form

$$u_{\alpha|\beta}\omega_\beta - \omega_{12}e_{\alpha\beta}u_\beta = u_{\alpha,\beta}\omega_\beta, \tag{16}$$

where $u_{\alpha,\beta}$ are the components of a tensor of rank two, called the *derivative* of u_α, given by

$$u_{\alpha,\beta} = u_{\alpha|\beta} + e_{\gamma\alpha}r_\beta u_\gamma. \tag{17}$$

Just as in the usual treatment of tensor analysis, one can extend the concept of differentiation to tensors of higher rank, first by considering tensors formed by multiplying components of vectors and defining differentiation so as to maintain the product rule, and then by applying the result directly to arbitrary tensors of the same rank. Restricting the discussion only to tensors of rank $\leqq 2$, one has the following result:

Theorem 2. *The components of the derivative of the vector u_α and the tensor $u_{\alpha\beta}$ are given by*

$$u_{\alpha,\beta} = u_{\alpha|\beta} + e_{\gamma\alpha}r_\beta u_\gamma,$$

$$u_{\alpha\beta,\gamma} = u_{\alpha\beta|\gamma} + e_{\epsilon\alpha}r_\gamma u_{\epsilon\beta} + e_{\epsilon\beta}r_\gamma u_{\alpha\epsilon}.$$

Explicitly,

$$(u_{\alpha,\beta}) = \begin{bmatrix} u_{1|1} - r_1u_2 & u_{1|2} - r_2u_2 \\ u_{2|1} + r_1u_1 & u_{2|2} + r_2u_1 \end{bmatrix},$$

$$(u_{\alpha\beta,1}) = \begin{bmatrix} u_{11|1} - r_1(u_{12}+u_{21}) & u_{12|1} + r_1(u_{11}-u_{22}) \\ u_{21|1} + r_1(u_{11}-u_{22}) & u_{22|1} + r_1(u_{12}+u_{21}) \end{bmatrix},$$

$$(u_{\alpha\beta,2}) = \begin{bmatrix} u_{11|2} - r_2(u_{12}+u_{21}) & u_{12|2} + r_2(u_{11}-u_{22}) \\ u_{21|2} + r_2(u_{11}-u_{22}) & u_{22|2} + r_2(u_{12}+u_{21}) \end{bmatrix}.$$

It is of interest to apply the theorem to the tensors $\delta_{\alpha\beta}$ and $e_{\alpha\beta}$. One readily finds the following result:

Corollary. *The tensors $\delta_{\alpha\beta}$ and $\epsilon_{\alpha\beta}$ are constants in the sense that the components of the derivatives vanish.*

It is desirable to extend the notation to scalars. Let B be a scalar field defined on S. Write the change in B between points P and P' in the form

$$B(P') - B(P) = B_{|\alpha}\omega_\alpha.$$

Since this change is an invariant, it follows that the coefficients of ω_α are components of a vector. To emphasize this, we write

$$B_\alpha = B_{|\alpha}. \tag{18}$$

DEFORMATION OF SURFACES

Let u_α, w represent the deformation δ of a surface S (in the manner described in Sec. 2) into a surface S_1. We shall henceforth suppose the components to be small (more precisely, infinitesimal). Let P denote a point on S and P_1 the point on S_1 into which P is carried under δ.

To the surface S_1, one can attach a system of local Cartesian coordinate frames in the same manner as for S, with axes 1 and 2 in the tangent plane. One can think of the deformation as inducing a rigid-body motion of the coordinate system initially at P on S into that at P_1 on S_1. It is convenient, in some respects, to consider instead the Euclidean motion of 3-space caused by the deformation relative to an observer attached to coordinate frame originally attached to S at P. If the directions of the coordinate axes are reversed, then u_α, w again represent the coordinates of displacement.

In general, a Euclidean motion of 3-space can be characterized by giving the three components ω_1, ω_2, ω_3 of the displacement of the origin, and the three components ω_{13}, ω_{23}, ω_{12}, of the angular rotation of the axes. To emphasize the dependence on the deformation, we will occasionally display the deformation as an argument. Thus,

$$(\omega_1(\delta), \omega_2(\delta), \omega_3(\delta)) = (u_1, u_2, \omega). \tag{19}$$

Introduce the notation

$$(\omega_{13}(\delta), \omega_{23}(\delta), \omega_{12}(\delta)) = (P_1, P_2, R). \tag{20}$$

The relation between two nearby points P and P' on S has already been discussed (Cartesian Tensors). The relation between the coordinate frames attached to P and P', respectively, can also be regarded as a Euclidean motion of 3-space. Let d represent the motion induced by the transformation $P \to P'$. Write

$$(\omega_1(d), \omega_2(d), \omega_3(d)) = (\omega_1, \omega_2, 0), \tag{21}$$

$$(\omega_{13}(d), \omega_{23}(d), \omega_{12}(d)) = (\Omega_1, \Omega_2, \Omega). \tag{22}$$

According to Eqs. (11) and (12),

$$\Omega_\alpha = P_{\alpha\beta}\omega_\beta(d), \tag{23}$$

$$\Omega = r_\beta \omega_\beta(d). \tag{24}$$

The deformation δ can be decomposed orthogonally into a displacement u_α along S and a displacement w normal to S. Therefore the components of axis rotation can accordingly be written in the form

$$P_\alpha = P_{\alpha\beta}\omega_\beta(\delta) + p_\alpha \omega_3(\delta),$$

$$R = r_\beta \omega_\beta(\delta) + r\omega_3(\delta).$$

Using Eq. (19), one can write

$$P_\alpha = p_{\alpha\beta} u_\beta + p_\alpha w, \tag{25}$$

$$R = r_\beta u_\beta + rw. \tag{26}$$

Consider, now from a general standpoint, a set of points in 3-space and a set of Cartesian frames (x_1, x_2, x_3), one attached to each point of the set. Let d represent an infinitesimal displacement field, and let $\omega_i(d)$, $\omega_{ij}(d)$ represent the components of displacement and axis rotation at each point *referred to the coordinate system at that point.* If, however, one chooses a single coordinate system, say that attached to a point P, then the components of displacement at a point P' with coordinates x_1, x_2, x_3 can be represented by dx_1, dx_2, dx_3 and the differentials $\omega_i(d)$ and $\omega_{ij}(d)$ can be put in the form[9]

$$\omega(x, dx) = a_i(x)\, dx_i, \tag{27}$$

where $a_i(x)$ are functions of the coordinates x_i of P'.

Expressions such as shown on the right side of Eq. (27) are special kinds of *exterior differential forms.* We shall give below a short introduction to the subject of exterior differential forms and then indicate its application to geometric problems of structure. Our notation and presentation will follow closely the treatment given in Cartan's book[10] (henceforth be denoted by C). This is somewhat intuitive and the reader who is oriented to the mathematically formal treatment will find more modern presentations in other sources (e.g., Nomizu[11]).

EXTERIOR DIFFERENTIAL FORMS

The algebra of exterior forms was introduced by H. Grassman.[12] The subject received an impetus at the hands of E. Cartan[10] who developed the theory of exterior differential forms and applied it to geometry. One may describe one kind of contribution which the theory of exterior differential forms has provided as a generalization of the principal theorem of integral calculus. Such generalizations already have existed in the form of Gauss', Green's, and Stokes' theorems. The theory of exterior differential forms (edf) however, brings all of these together in a unified form. In a certain sense, one may regard this only as an improvement in notation; however, the use of this tool affords a sharper prespective of the essential geometric aspects of the problem.

Consider a rectangular space in n dimensions with coordinates $x_1, \cdots, x_n$. An edf of zero rank (or degree) is simply a function of the coordinates. An edf of rank 1 is a linear combination of $dx_1, \cdots, dx_n$ with coefficients which are functions of $x_1, \cdots, x_n$.

$$\text{Rank } 0\text{: } f(x_1, \cdots, x_n),$$

$$\text{Rank } 1\text{: } \sum_i f_i(x)\, dx_i. \tag{28}$$

A differential of rank two is defined by:

$$\text{Rank } 2\text{: } \sum_{i,j} f_{ij}(x) \begin{vmatrix} \delta dx_i & \delta dx_j \\ \delta x_i & \delta x_j \end{vmatrix}, \tag{29}$$

where $(dx_1, \cdots, dx_n)$ and $(\delta x_1, \cdots, \delta x_n)$ are two independent differential vectors. One can proceed to define differentials up to and including rank n. Since our application is concerned only with $n = 2$, the above definitions suffice.

Two edf's of the same rank are *equal* if, regarded as functions of the independent variables $x_1, \cdots, x_n, dx_1, \cdots, dx_n, \delta x_1, \cdots, \delta x_n$, etc., they are identically equal. It is therefore often not necessary to display the symbols $d, \delta, \cdots$ of the displacements, and one can write expression (29), for instance, in the form (reference 7)

$$f_{ij}[dx_i\ dx_j].$$

Two edf's of the same rank can be added by adding corresponding coefficients as, for instance,

$$f_{ij}[dx_i\ dx_j] + g_{ij}[dx_i\ dx_j] = (f_{ij} + g_{ij})[dx_i\ dx_j]. \tag{30}$$

Exterior multiplication of any two edf's is represented by brackets as, for instance,

$$[\![A(x)\ f_{ij}[dx_i\ dx_j]]\!] = Af_{ij}[dx_i\ dx_j],$$

$$[a_i\, dx_i\ b_j\, dx_j] = a_i b_j [dx_i\ dx_j]. \tag{31}$$

Exterior differentiation is defined as follows for

---

[9] The tensor summation convention is here used to mean that a repeated Latin index indicates summation on the range 1, 2, 3.

[10] E. Cartan, *Les Systémes Différentials et leurs Applications Géométriques,* (Hermann & Cie., Paris, 1945).

[11] K. Nomizu, *Lie Groups and Differential Geometry,* (Mathematical Society of Japan, 1956).

[12] H. Grassman, *Ausdehnungslehre* (Leipzig, 1862).

forms of degrees zero, one, and two, respectively,

$$da(x) = \frac{\partial a}{\partial x_i}\, dx_i, \tag{32}$$

$$d(a_i(x)\, dx_i) = [da_i\ dx_i] = \frac{\partial a_i}{\partial x_j}\, [dx_j\ dx_i], \tag{33}$$

$$d(a_{ij}[dx_i\ dx_j]) = [da_{ij}\ dx_i\ dx_j]$$

$$= \frac{\partial a_{ij}}{\partial x_k}\, [dx_k\ dx_i\ dx_j]. \tag{34}$$

Let R_{k+1} be a $(k+1)$-dimensional region in the rectangular space with boundary R'_{k+1}. Then for suitable definition of the integrals, one has the following generalization of the fundamental theorem of integral calculus:

Theorem 2(C, p.40). *If ω is an edf of degree k, then*

$$\int_{R_{k+1}} d\omega = \int_{R'_{k+1}} d\omega.$$

Additional theorems useful for our purpose are:

Theorem 3 (C, p. 35). *The exterior differential of the exterior product of two edf's ω_1 and ω_2 of ranks p and q, respectively, is*

$$d(\omega_1\ \omega_2) = [d\omega_1\ \omega_2] + (-1)^p[\omega_1\ d\omega_2].$$

Theorem 4 (C, p. 37). *The exterior differential of the exterior differential of a differential form is zero.*

Theorem 5 (C, p. 38). *Given an edf of degree k whose exterior differential is zero in a neighborhood of space, there exists an edf of degree $k-1$ in this region whose exterior differential coincides with the given edf.*

The application of edf's to geometrical problems has been discussed by Cartan (C. Chap. VII). He derived the *equations of structure* connecting the six edf's $\omega_1, \omega_2, \omega_3, \omega_{23}, \omega_{31}, \omega_{12}$ which characterize the displacements and angular rotations of a set of Cartesian frames in 3-space.

Theorem 6 (C, pp. 122–123). *A necessary and sufficient condition that the forms ω_i and ω_{ij} represent the components of displacement and rotation is*

$$d\omega_i = [\omega_{ij}\ \omega_j],$$

$$d\omega_{ij} = [\omega_{ik}\ \omega_{kj}].$$

It should be noted that there are two independent displacements involved in these equations because the edf's on the right side are of rank two. One displacement is represented by the symbol d. A second δ, not explicit, should be understood as an argument of the edf's ω_i and ω_{ij}. Moreover, it is implicit in the derivation of the theorem that the displacements d and δ are *commutative*, i.e., a point P displaced first by d and then by δ is to be found at the same terminal position as it would have if displaced first by δ and then by d.

Theorem 7. *Let u_α be a vector in 2-space, and let*

$$\omega = u_\alpha \omega_\alpha.$$

Then

$$d\omega = u_{\alpha,\beta}[\omega_\beta\ \omega_\alpha].$$

The result follows by applying the rules for exterior differentiation to ω. By Theorem 3 and Theorem 6,

$$d\omega = [du_\alpha\ \omega_\alpha] + [u_\alpha\ d\omega_\alpha],$$

$$= [u_{\alpha|\beta}\omega_\beta\ \omega_\alpha] + u_\alpha[\omega_{\alpha\beta}\ \omega_\beta].$$

But

$$\omega_{\alpha\beta} = e_{\alpha\beta}r_\gamma\omega_\gamma.$$

Substituting this expression and distributing the result,

$$d\omega = u_{\alpha|\beta}[\omega_\beta\ \omega_\alpha] + e_{\alpha\beta}r_\gamma u_\alpha[\omega_\gamma\ \omega_\beta],$$

$$= u_{\alpha|\beta}[\omega_\beta\ \omega_\alpha] + e_{\gamma\alpha}r_\beta u_\gamma[\omega_\beta\ \omega_\alpha],$$

$$= u_{\alpha,\beta}[\omega_\beta\ \omega_\alpha],$$

using Eq. (17). This demonstrates the Theorem.

Corollary. *If w is a scalar with gradient $w_\alpha \equiv w_{|\alpha}$, then $w_{\alpha,\beta} = w_{\beta,\alpha}$.*

Form the edf dw

$$dw = w_\alpha\omega_\alpha.$$

The exterior differential of dw vanishes according to Theorem 4. Therefore, from Theorem 7,

$$w_{1,2} - w_{2,1} = 0,$$

which proves the Corollary.

DEFORMATION TENSORS

Consider a Cartesian frame x_1, x_2, x_3 in 3-space. Let d and δ represent two infinitesimal displacement fields, so that under d and δ, respectively,

$$d\colon x_i \to x_i + dx_i,$$

$$\delta\colon x_i \to x_i + \delta x_i.$$

Under δ, the point with coordinates $x_i + dx_i$ is transformed to a point with coordinates

$$(x_i + dx_i) + \delta(x_i + dx_i)$$

$$= x_i + dx_i + \delta x_i + \delta\, dx_i, \tag{35}$$

which defines δdx_i. Similarly, under d, the point with

coordinates $x_i + \delta x_i$ is transformed into one with coordinates

$$(x_i + \delta x_i) + d(x_i + \delta x_i) = x_i + \delta x_i + dx_i + d\,\delta x_i. \quad (36)$$

The two points represented by Eqs. (35) and (36) are the same, i.e., δ and d commute, if and only if

$$d\,\delta x_i = \delta\, dx_i. \quad (37)$$

Let f be a function of x_i. Introduce the *difference operator* Δ_δ, associated with the displacement δ, and define this operator acting on the scalar function $f(x_i)$ as follows:

$$\Delta_\delta f = f(x_i + \delta x_i) - f(x_i) = (\partial f/\partial x_i)\,\delta x_i. \quad (38)$$

Define the operator acting on the edf of rank one,

$$\omega(d) = a_i(x)\, dx_i,$$

according to the formula

$$\Delta_\delta\omega(d) = a_i(x + \delta x)(dx_i + \delta\, dx_i) - a_i(x)\, dx_i, = (\partial a_i/\partial x_j)\,\delta x_j\, dx_i + a_i\,\delta\, dx_i. \quad (39)$$

The definition can be extended to edf's of higher rank but this is not necessary for our purpose.

Interchanging d and δ in Eq. (39) and subtracting the two resulting equations, one finds

$$\Delta_\delta\omega(d) - \Delta_d\omega(\delta) = (\partial a_i/\partial x_j)(\delta x_j\, dx_i - dx_j\,\delta x_i) + a_i(\delta\, dx_i - d\,\delta x_i). \quad (40)$$

Now suppose the displacement fields d and δ to be commutative. Then the right side of Eq. (40) can be identified, according to Eq. (33), as the exterior differential of the edf ω. Therefore one has the result

Theorem 8. *The exterior differential of an edf of rank one can be expressed as the difference between two difference operators, i.e.,*

$$d\omega(\delta) = \Delta_d\omega(\delta) - \Delta_\delta\omega(d).$$

Theorem 8 is the principal tool by which we will describe the properties of a surface S_1 obtained by deformation δ from a given surface S. Let d represent an infinitesimal displacement field of S into itself. Suppose under d, the point P is carried into P', and suppose, under the deformation δ, P and P' are carried into P_1 and P_1', respectively, of S_1. Then define d on S_1 to be the displacement field $P_1 \to P_1'$. With this definition, d and δ commute. Now consider an edf $\omega(d)$ of rank 1, associated with the displacement field d, and defined for points on both S and S_1. Then the change in $\omega(d)$ under Δ will be expressible, according to Theorem 7, in the form

$$\Delta_\delta\omega(d) = \Delta_d\omega(\delta) - d\omega(\delta). \quad (41)$$

We shall next apply this result to the components of displacement ω_i and the components of rotation ω_{ij}. In each case, it will be conveneint to use the equations of structure (Theorem 6) in order to simplify the second term on the right side of Eq. (41).

First consider $\omega_3(d)$. This differential vanishes both on S and on the deformed surface S_1. Hence,

$$\Delta_\delta\omega_3(d) = 0. \quad (42)$$

Next consider the first term on the right side of Eq. (41). This is a difference operator under the displacement d applied to

$$\omega_3(\delta) \equiv w.$$

Therefore,

$$\Delta_d\omega_3(\delta) = \Delta_d(w), = w_\beta\omega_\beta. \quad (43)$$

The second term on the right side of Eq. (41) can be first simplified using the appropriate equation of structure from Theorem 6. The exterior differential of $\omega_3(\delta)$ is

$$d\omega_3(\delta) = [\omega_{31}\ \omega_1] + [\omega_{32}\ \omega_2], = \begin{vmatrix}\omega_{31}(d) & \omega_1(d)\\ \omega_{31}(\delta) & \omega_1(\delta)\end{vmatrix} + \begin{vmatrix}\omega_{32}(d) & \omega_2(d)\\ \omega_{32}(\delta) & \omega_2(\delta)\end{vmatrix}.$$

The quantities appearing on the right side can be expressed in terms of the notation, introduced in Eqs. (19)–(26), which is more suitable for two-dimensional tensor terminology:

$$d\omega_3(\delta) = \begin{vmatrix}-\Omega_1 & \omega_1\\ -P_1 & u_1\end{vmatrix} + \begin{vmatrix}-\Omega_2 & \omega_2\\ -P_2 & u_2\end{vmatrix}, = \begin{vmatrix}-p_{1\beta}\omega_\beta & \omega_1\\ -p_{1\beta}u_\beta - p_1 w & u_1\end{vmatrix} + \begin{vmatrix}-p_{2\beta}\omega_\beta & \omega_2\\ -p_{2\beta}u_\beta - p_2 w & u_2\end{vmatrix}, = p_\beta\omega_\beta w - p_{12}\begin{vmatrix}\omega_2 & \omega_1\\ u_2 & u_1\end{vmatrix} - p_{21}\begin{vmatrix}\omega_1 & \omega_2\\ u_1 & u_2\end{vmatrix}.$$

The sum of the last two terms on the right vanishes because the matrix $p_{\alpha\beta}$ is symmetric (C, p. 126). Hence,

$$d\omega_3(\delta) = p_\beta\omega_\beta w. \quad (44)$$

Substituting the results of Eqs. (42)–(44) into Eq. (41), one has

$$0 = w_\beta\omega_\beta - p_\beta\omega_\beta w,$$

whence, comparing the coefficients of ω_β,

$$p_\beta w = w_\beta, \qquad (\beta = 1, 2). \tag{45}$$

Equation (25) may then be written, after eliminating p_β, as

$$P_\alpha = p_{\alpha\beta}u_\beta + w_\beta. \tag{46}$$

Now apply Eq. (41) to $\omega \equiv \omega_1$. The first term on the right is

$$\begin{aligned}\Delta_d\omega_1(\delta) &= \Delta_d u_1, \\ &= u_{1\,|\beta}\omega_\beta.\end{aligned} \tag{47}$$

The second term on the right may be transformed using the equation of structure (Theorem 6), and then simplifying using Eqs. (19)–(26):

$$\begin{aligned}d\omega_1(\delta) &= [\omega_{12} \quad \omega_2] + [\omega_{13} \quad \omega_3], \\ &= \begin{vmatrix}\omega_{12}(d) & \omega_2(d)\\ \omega_{12}(\delta) & \omega_2(\delta)\end{vmatrix} + \begin{vmatrix}\omega_{13}(d) & \omega_3(d)\\ \omega_{13}(\delta) & \omega_3(\delta)\end{vmatrix}, \\ &= \begin{vmatrix}r_\beta\omega_\beta & \omega_2\\ R & u_2\end{vmatrix} + \begin{vmatrix}p_{1\beta}\omega_\beta & 0\\ p_{1\beta}u_\beta + p_1 w & w\end{vmatrix}, \\ &= -R\omega_2 + r_\beta u_2\omega_\beta + p_{1\beta}w\omega_\beta, \\ &= -Re_{1\beta}\omega_\beta + e_{1\gamma}r_\beta u_\gamma\omega_\beta + p_{1\beta}w\omega_\beta.\end{aligned} \tag{48}$$

Similar results can be obtained for $\omega \equiv \omega_2$. In both cases, substitution into the right side of Eq. (41) gives results of the form

$$\Delta_\delta\omega_\alpha = \epsilon_{\alpha\beta}\omega_\beta, \qquad \alpha, \beta = 1, 2,$$

where, using the notation of tensor derivatives [see Eq. (17)],

$$\begin{aligned}\epsilon_{\alpha\beta} &= u_{\alpha\,|\beta} - e_{\alpha\gamma}r_\beta u_\gamma - p_{\alpha\beta}w + Re_{\alpha\beta}, \\ &= u_{\alpha,\beta} - p_{\alpha\beta}w + Re_{\alpha\beta}.\end{aligned} \tag{49}$$

So far, the local reference frames for the deformed surfaces S_1 have not been related to those of S. A small rotation of the axes 1 and 2 about a normal to the surface S_1 induces a linear transformation of the components 1 and 2 whose matrix differs from the unit matrix by a small skew-symmetric matrix. The corresponding matrix $(\epsilon_{\alpha\beta})$ is increased by this skew-symmetric matrix. Therefore, the axes 1 and 2 in the local reference frame at S_1 can be so chosen that the tensor $\epsilon_{\alpha\beta}$ is symmetric; such frames will be called *cannoically connected* to the reference frames of S. Then for these frames

$$0 = \epsilon_{12} - \epsilon_{21} = u_{1,2} - u_{2,1} + 2R.$$

Therefore, using Eq. (26), one finds, for R and rw,

$$R = r_\alpha u_\alpha + rw = -\tfrac{1}{2}(u_{1,2} - u_{2,1}) = -\tfrac{1}{2}e_{\alpha\beta}u_{\alpha,\beta}. \tag{50}$$

One finally finds

$$\epsilon_{\alpha\beta} = \tfrac{1}{2}(u_{\alpha,\beta} + u_{\beta,\alpha}) - p_{\alpha\beta}w. \tag{51}$$

Since

$$\omega_\alpha + \Delta\omega_\alpha = (\delta_{\alpha\beta} + \epsilon_{\alpha\beta})\omega_\beta, \tag{52}$$

the tensor $\epsilon_{\alpha\beta}$ describes to first order the change of the first fundamental form under the deformation $S \to S_1$. In a similar way, the equation for S_1 which is equivalent to Eq. (23) for S, can be put in the form

$$\Omega_\alpha + \Delta\Omega_\alpha = (p_{\alpha\beta} + \kappa_{\alpha\beta})(\omega_\beta + \Delta\omega_\beta). \tag{53}$$

The tensor $\kappa_{\alpha\beta}$ describes the change in the second fundamental form under the deformation $S \to S_1$.

In order to obtain expressions for $\Delta\Omega_\alpha$, again apply Eq. (41) to $\omega \equiv \omega_{13}(d) = \Omega_1$:

$$\begin{aligned}\Delta_d\omega_{13}(\delta) &= \Delta_d P_1 \\ &= P_{1\,|\beta}\omega_\beta.\end{aligned} \tag{54}$$

$$\begin{aligned}d\omega_{13} &= [\omega_{12} \quad \omega_{23}], \\ &= \begin{vmatrix}\omega_{12}(d) & \omega_{23}(d)\\ \omega_{12}(\delta) & \omega_{23}(\delta)\end{vmatrix}, \\ &= \begin{vmatrix}\Omega & \Omega_2\\ R & P_2\end{vmatrix}, \\ &= \begin{vmatrix}r_\beta\omega_\beta & p_{2\beta}\omega_\beta\\ R & P_2\end{vmatrix}, \\ &= r_\beta P_2\omega_\beta - p_{2\beta}R\omega_\beta.\end{aligned}$$

Then,

$$\begin{aligned}\Delta_\delta\Omega_1 &= P_{1\,|\beta}\omega_\beta - r_\beta P_2\omega_\beta + p_{2\beta}R\omega_\beta, \\ \Delta_\delta\Omega_1 &= (P_{1,\beta} + e_{1\gamma}p_{\gamma\beta}R)\omega_\beta.\end{aligned} \tag{55}$$

The tensor derivative $P_{\alpha,\beta}$ can be expressed in terms of the displacements u_α and w, using Eq. (46):

$$\begin{aligned}P_{\alpha,\beta} &= (p_{\alpha\gamma}u_\gamma + w_\alpha)_{,\beta}, \\ &= p_{\alpha\gamma,\beta}u_\gamma + p_{\alpha\gamma}u_{\gamma,\beta} + w_{\alpha,\beta}.\end{aligned}$$

Substitute into Eq. (55), and then find

$$\begin{aligned}\Omega_\alpha + \Delta\Omega_\alpha = (&p_{\alpha\beta} + p_{\alpha\gamma}u_{\gamma,\beta} + p_{\alpha\gamma,\beta}u_\gamma \\ &+ w_{\alpha,\beta} + e_{\alpha\gamma}p_{\gamma\beta}R)\omega_\beta.\end{aligned}$$

Substitute both this result and the expression for $\omega_\alpha + \Delta\omega_\alpha$ shown in Eq. (52) into Eq. (53). The first-order terms (in ω_α) cancel and the terms bilinear in ω_α and u_α can be retained. Comparing coefficients of ω_α, one finds the desired expressions

for $\kappa_{\alpha\beta}$:

$$\kappa_{\alpha\beta} = p_{\alpha\gamma,\beta}u_\gamma + R(e_{\alpha\gamma}p_{\gamma\beta} + e_{\beta\gamma}p_{\gamma\alpha}) + w_{\alpha,\beta} + p_{\alpha\gamma}p_{\gamma\beta}w. \quad (56)$$

The tensor $\kappa_{\alpha\beta}$ is symmetric in the indices α and β. The symmetry of the first term follows from the symmetry

$$p_{\alpha\gamma,\beta} = p_{\beta\gamma,\alpha}, \quad (57)$$

which will be proved later (Theorem 10). The symmetry of the second and fourth terms follows by inspection. The symmetry of the third term is a consequence of Theorem 7.

We can now consider the question of inversion of the normal (axis 3) of the local reference frame. The significance of the inversion was raised in Sec. 3 where it was stated without proof that $\epsilon_{\alpha\beta}$ is invariant whereas $\kappa_{\alpha\beta}$ changes sign. Note first that the components of the tensor u_α are unchanged whereas w changes sign; the components $p_{\alpha\beta}$ change sign whereas the components r_α are unchanged. Hence the components $\epsilon_{\alpha\beta}$ shown in Eq. (49) are unchanged, and the components of $\kappa_{\alpha\beta}$ shown in Eq. (56) change sign.

Theorem 9. *In a canonically connected set of reference frames, the changes $\epsilon_{\alpha\beta}$ and $\kappa_{\alpha\beta}$ in the first and second fundamental forms, respectively, associated with the deformation δ: (u_α, w) are given to first order by Eqs. (51) and (56). The tensor $\epsilon_{\alpha\beta}$ is symmetric and unchanged under inversion of the normal axis; the tensor $\kappa_{\alpha\beta}$ is symmetric and changes sign under inversion.*

The tensors $\epsilon_{\alpha\beta}$ and $\kappa_{\alpha\beta}$ are called the *strain tensors.*

VIRTUAL DEFORMATIONS

Let $\mathfrak{F}$ represent the strain energy per unit area of an infinitesimally thin shell which is deformed under an external load. Suppose $\mathfrak{F}$ has the form given in Eq. (2), or, what is equivalent,

$$\mathfrak{F} = \tfrac{1}{2}(f_1\epsilon_{\alpha\alpha}^2 + f_2\,|\epsilon_{\alpha\beta}| + f_3\kappa_{\alpha\alpha}^2 + f_4\,|\kappa_{\alpha\beta}|). \quad (58)$$

The coefficients f_i are related to the F_i in Eq. (2) according to the equations

$$f_1 = F_1 + F_2, \quad f_2 = -2F_2, \quad f_3 = F_3 + F_4, \quad f_4 = -2F_4, \quad (59)$$

$$F_1 = f_1 + \tfrac{1}{2}f_2, \quad F_2 = -\tfrac{1}{2}f_2, \quad F_3 = f_3 + \tfrac{1}{2}f_4, \quad F_4 = -\tfrac{1}{2}f_4. \quad (60)$$

Regard $\mathfrak{F}$ as a function of the independent variables $\epsilon_{\alpha\beta}$ and $\kappa_{\alpha\beta}$. Introduce the symmetric tensors

$$s_{\alpha\beta} = \partial\mathfrak{F}/\partial\epsilon_{\alpha\beta}, \quad m_{\alpha\beta} = \partial\mathfrak{F}/\partial\kappa_{\alpha\beta}. \quad (61)$$

The tensors $\epsilon_{\alpha\beta}$, $\kappa_{\alpha\beta}$, $s_{\alpha\beta}$, and $m_{\alpha\beta}$ are linear functions of the displacement vector (u_α, w); to show this functional dependence, we shall write $\epsilon_{\alpha\beta}(u)$, $\kappa_{\alpha\beta}(u)$, $s_{\alpha\beta}(u)$, and $m_{\alpha\beta}(u)$.

When (u_α, w) is replaced by $(u_\alpha + \delta u_\alpha, w + \delta w)$, then $\epsilon_{\alpha\beta}$ and $\kappa_{\alpha\beta}$ are changed accordingly. The change $\delta\mathfrak{F}$ in $\mathfrak{F}$ can be written, to first order in δu_α and δw, in the form

$$\delta\mathfrak{F} = s_{\alpha\beta}(u)\epsilon_{\alpha\beta}(\delta u) + m_{\alpha\beta}(u)\kappa_{\alpha\beta}(\delta u). \quad (62)$$

The first term on the right can be transformed by substituting from Eq. (51) and noting that $s_{\alpha\beta}$ is symmetric;

$$s_{\alpha\beta}(u)\epsilon_{\alpha\beta}(\delta u) = \tfrac{1}{2}s_{\alpha\beta}(\delta u_{\alpha,\beta} + \delta u_{\beta,\alpha}) - s_{\alpha\beta}p_{\alpha\beta}\,\delta w, = s_{\alpha\beta}\,\delta u_{\alpha,\beta} - s_{\alpha\beta}p_{\alpha\beta}\,\delta w. \quad (63)$$

Similarly, the second term on the right becomes, using Eqs. (50) and (56),

$$m_{\alpha\beta}(u)\kappa_{\alpha\beta}(\delta u) = m_{\alpha\beta}p_{\alpha\gamma,\beta}\,\delta u_\gamma + m_{\alpha\beta}\,\delta w_{\alpha,\beta} + m_{\alpha\beta}p_{\alpha\gamma}p_{\gamma\beta}\,\delta w - Ae_{\alpha\beta}\,\delta u_{\alpha,\beta}, \quad (64)$$

where

$$A = \tfrac{1}{2}m_{\alpha\beta}(e_{\alpha\gamma}p_{\gamma\beta} + e_{\beta\gamma}p_{\gamma\alpha}). \quad (65)$$

EQUATIONS OF EQUILIBRIUM

The analysis of this section is devoted to finding an expression for the integral of $\delta\mathfrak{F}$ over a portion of the surface S bounded by a contour L. Let l denote the coordinate along L, equal to the arc length from some fixed point when the contour is traversed in the sense for which Theorem 2 is applicable.[13] At each point of L, regard this positive sense as the direction of axis 2 of a Cartesian 2 frame, and construct axis 1 in the plane, tangent to S at that point, so that the axes 1 and 2 are coherent with axes in the interior of L.

In order to carry out the integration, we will start by forming the exterior differential of an edf, and then identify the result with some of the terms in Eqs. (63) and (64). Using Theorem 7, consider the edf

$$d(\delta u_\alpha s_{\alpha\beta}e_{\beta\gamma}\omega_\gamma) = (\delta u_\alpha s_{\alpha\beta}e_{\beta\gamma})_{,\delta}\ [\omega_\delta\,\omega_\gamma]. \quad (66)$$

Since $e_{\beta\gamma}$ is constant under the operation of tensor differentiation, it may be taken outside the parenthesis. Also, for $\gamma, \delta = 1, 2$,

$$[\omega_\delta\ \omega_\gamma] = e_{\delta\gamma}[\omega_1\ \omega_2].$$

[13] Otherwise there is a difference in sign.

Therefore, the edf [Eq. (66)] can be put in the form

$$\begin{aligned} d(\delta u_\alpha s_{\alpha\beta} e_{\beta\gamma} \omega_\gamma) &= (\delta u_\alpha s_{\alpha\beta})_{,\delta} e_{\beta\gamma} e_{\delta\gamma} \quad [\omega_1\ \omega_2], \\ &= (\delta u_\alpha s_{\alpha\beta})_{,\delta}\ \delta_{\beta\delta} [\omega_1\ \omega_2], \\ &= (\delta u_\alpha s_{\alpha\beta})_{,\beta} [\omega_1\ \omega_2]. \end{aligned} \tag{67}$$

The tensor differentiation can be distributed:

$$(\delta u_\alpha s_{\alpha\beta})_{,\beta} = \delta u_{\alpha,\beta} s_{\alpha\beta} + \delta u_\alpha s_{\alpha\beta,\beta}. \tag{68}$$

Combining Eqs. (67) and (68), and rearranging terms, one finds

$$\begin{aligned} &s_{\alpha\beta}\ \delta u_{\alpha,\beta} [\omega_1\ \omega_2] \\ &\quad = -s_{\alpha\beta,\beta}\ \delta u_\alpha [\omega_1\ \omega_2] + d(s_{\alpha\beta} e_{\beta\gamma}\ \delta u_\alpha \omega_\gamma). \end{aligned} \tag{69}$$

In a similar fashion,

$$\begin{aligned} &m_{\alpha\beta}\ \delta w_{\alpha,\beta} [\omega_1\ \omega_2] \\ &\quad = -M_\alpha\ \delta w_\alpha [\omega_1\ \omega_2] + d(m_{\alpha\beta} e_{\beta\gamma}\ \delta w_\alpha \omega_\gamma), \end{aligned} \tag{70}$$

where the vector M_α is defined by

$$M_\alpha = m_{\alpha\beta,\beta}. \tag{71}$$

Also,

$$\begin{aligned} &M_\alpha\ \delta w_\alpha [\omega_1\ \omega_2] \\ &\quad = -M_{\alpha,\alpha}\ \delta w [\omega_1\ \omega_2] + d(M_\alpha e_{\alpha\beta}\ \delta w \omega_\beta). \end{aligned} \tag{72}$$

Substituting this result in Eq. (70), one finds

$$\begin{aligned} m_{\alpha\beta}\ \delta w_{\alpha,\beta} [\omega_1\ \omega_2] &= M_{\alpha,\alpha}\ \delta w [\omega_1\ \omega_2] \\ &\quad + d(m_{\alpha\beta} e_{\beta\gamma}\ \delta w_\alpha \omega_\gamma - M_\alpha e_{\alpha\beta}\ \delta w \omega_\beta). \end{aligned} \tag{73}$$

In Eq. (69), replace $s_{\alpha\beta}$ by $Ae_{\alpha\beta}$. Then,

$$\begin{aligned} &Ae_{\alpha\beta}\ \delta u_{\alpha,\beta} [\omega_1\ \omega_2] \\ &= -(Ae_{\alpha\beta})_{,\beta}\ \delta u_\alpha [\omega_1\ \omega_2] + d(Ae_{\alpha\beta} e_{\beta\gamma}\ \delta u_\alpha \omega_\gamma), \\ &= -A_{,\beta} e_{\alpha\beta}\ \delta u_\alpha [\omega_1\ \omega_2] - d(A\ \delta u_\alpha \omega_\alpha). \end{aligned} \tag{74}$$

[The quantity $A_{,\beta}$ represents the gradient of the scalar A. The use of the comma, which differs from the usage defined in Eq. (18), is in order to distinguish from the symbols A_β introduced below in Sec. 10.]

The terms within parentheses on the right sides of Eqs. (69) and (72) are invariants. The integral over the surface S of the differentials reduces, according to Theorem 2, to the integral of these invariant differentials on the contour L. In order to carry out the integration conveniently, first evaluate the invariants in the coordinate system, described above, based on the tangent to the curve L. In this coordinate system, an element of arc dl has components

$$\omega_1 = 0, \qquad \omega_2 = dl. \tag{75}$$

Instead of the subscripts 1 and 2 to represent axes normal and tangent to the contour, respectively, we shall now use subscripts n and l, respectively. Then the following changes in notations must be made:

$$\begin{aligned} (\delta u_1,\ \delta u_2) &\to (\delta u_n,\ \delta u_l), \\ (\delta w_1,\ \delta w_2) &\to (\delta w_n,\ \delta w_l), \\ (s_{\alpha\beta}) &\to \begin{bmatrix} s_{nn} & s_{nl} \\ s_{ln} & s_{ll} \end{bmatrix}, \\ (m_{\alpha\beta}) &\to \begin{bmatrix} m_{nn} & m_{nl} \\ m_{ln} & m_{ll} \end{bmatrix}, \end{aligned} \tag{76}$$

and so forth. Then applying Theorem 2 to the integrals of the exact edf on the right side of Eq. (69),

$$\int_S d(s_{\alpha\beta} e_{\beta\gamma}\ \delta u_\alpha \omega_\gamma) = \int_L s_{\alpha\beta} e_{\beta\gamma}\ \delta u_\alpha \omega_\gamma.$$

Since $\omega_1 = 0$ by Eq. (75), then the only nonzero contribution to the integrand is for $\beta = 1$, $\gamma = 2$. Changing to the notation described in (76),

$$\int_S d(s_{\alpha\beta} e_{\beta\gamma}\ \delta u_\alpha \omega_\gamma) = \int_L (s_{nn}\ \delta u_n + s_{ln}\ \delta u_l)\ dl. \tag{77}$$

In a similar fashion, transform the integral of the exact differential on the right side of Eq. (73):

$$\begin{aligned} &\int_S d(m_{\alpha\beta} e_{\beta\gamma}\ \delta w_\alpha \omega_\gamma - M_\alpha e_{\alpha\beta}\ \delta w \omega_\beta) \\ &\quad = \int_L (m_{\alpha\beta} e_{\beta\gamma}\ \delta w_\alpha \omega_\gamma - M_\alpha e_{\alpha\beta}\ \delta w \omega_\beta), \\ &\quad = \int_L (m_{nn}\ \delta w_n + m_{ln}\ \delta w_l - M_n\ \delta w)\ dl. \end{aligned} \tag{78}$$

The middle term on the right is then transformed by integrating by parts around the closed contour L:

$$\begin{aligned} \int_L m_{ln}\ \delta w_l\ dl &= \int_L m_{ln} \frac{d}{dl}\ \delta w\ dl, \\ &= -\int_L \left(\frac{d}{dl} m_{ln}\right) \delta w\ dl. \end{aligned} \tag{79}$$

Substituting this into Eq. (78),

$$\begin{aligned} &\int_S d(m_{\alpha\beta} e_{\beta\gamma}\ \delta w_\alpha \omega_\gamma - M_\alpha e_{\alpha\beta}\ \delta w \omega_\beta) \\ &\quad = \int_L \left\{ m_{nn}\ \delta w_n - \left(\frac{d}{dl} m_{ln} + M_n\right) \delta w \right\} dl. \end{aligned} \tag{80}$$

Transform the integral of the exact differential

on the right side of Eq. (74):

$$\int_S d(A\ \delta u_\alpha \omega_\alpha) = \int_L A\ \delta u_\alpha \omega_\alpha,$$

$$= \int_L A\ \delta u_l\ dl. \tag{81}$$

Multiply Eq. (62) through by $[\omega_1\omega_2]$ and integrate over S. Use Eqs. (63) and (64) to express the integrand in terms of δu_α and δw:

$$\int_S \delta\mathfrak{F}[\omega_1\ \omega_2] = \int_S \{s_{\alpha\beta}\ \delta u_{\alpha,\beta} - p_{\alpha\beta}s_{\alpha\beta}\ \delta w + p_{\alpha\gamma,\beta}m_{\alpha\beta}\ \delta u_\gamma + m_{\alpha\beta}\ \delta w_{\alpha,\beta} + p_{\alpha\gamma}p_{\gamma\beta}m_{\alpha\beta}\ \delta w - Ae_{\alpha\beta}\ \delta u_{\alpha,\beta}\}[\omega_1\ \omega_2]. \tag{82}$$

Use Eqs. (69) and (77) to transform the first term in the integrand on the right:

$$\int_S s_{\alpha\beta}\ \delta u_{\alpha,\beta}[\omega_1\ \omega_2] = -\int_S s_{\alpha\beta,\beta}\ \delta u_\alpha[\omega_1\ \omega_2] + \int_L (s_{nn}\ \delta u_n + s_{ln}\ \delta u_l)\ dl. \tag{83}$$

Use Eqs. (73) and (80) to transform the fourth term in the integrand of Eq. (82):

$$\int_S m_{\alpha\beta}\ \delta w_{\alpha,\beta}[\omega_1\ \omega_2] = \int_S M_{\alpha,\alpha}\ \delta w[\omega_1\ \omega_2] + \int_L \left\{m_{nn}\ \delta w_n - \left(\frac{d}{dl}\ m_{ln} + M_n\right)\delta w\right\} dl. \tag{84}$$

Use Eqs. (74) and (81) to transform the last term in the integrand of Eq. (82):

$$\int_S Ae_{\alpha\beta}\ \delta u_{\alpha,\beta}[\omega_1\ \omega_2] = -\int_S A_{,\beta}e_{\alpha\beta}\ \delta u_\alpha[\omega_1\ \omega_2] - \int_L A\ \delta u_l\ dl. \tag{85}$$

The result of these transformations on Eq. (82) need not be shown explicitly for our purpose.

Now suppose the shell loaded on the surface and boundary. Let (X_α, Z) denote the components of surface loading; let (s_n, s_l) denote[14] the components of force per unit length applied to the boundary in the tangent plane, let s denote the component of force per unit length applied normal to the tangent plane; let Q denote applied moment per unit length of boundary, positive in the direction of rotation indicated by positive normal derivative dw/dn. Then, under a virtual displacement $(\delta u_\alpha, \delta w)$, the work done by these externally applied forces is

$$\int_S (X_\alpha\ \delta u_\alpha + Z\ \delta w)[\omega_1\ \omega_2] + \int_L (s_n\ \delta u_{nn} + s_l\ \delta u_l + s\ \delta w + Q\ \delta w_n)\ dl. \tag{86}$$

When the virtual displacements take place from the equilibrium position established under the loading, the theorem of virtual work states that the contribution due to Eq. (86) equals the internal strain energy [Eq. (82)] for arbitrary virtual displacements. Allow the virtual displacements to take place first in the interior of the contour L. Then the contributions of the surface integrals must be separately equal; therefore the contributions of the contour integrals must also be equal. Moreover, since the components δu_α and δw are independent within the contour, corresponding coefficients must be separately equal. Comparing these coefficients between Eqs. (82), as modified by Eqs. (83)–(85), and Eq. (86), one arrives at the field equations

$$X_\alpha = -s_{\alpha\beta,\beta} + p_{\gamma\alpha,\beta}m_{\gamma\beta} + e_{\alpha\beta}A_{,\beta}, \tag{87}$$

$$Z = M_{\alpha,\alpha} - p_{\alpha\beta}s_{\alpha\beta} + p_{\alpha\gamma}p_{\gamma\beta}m_{\alpha\beta}. \tag{88}$$

On the contour L, suppose that the virtual deformations δu_n, δu_l, δw, and δw_n are independent.[15] Making a comparison of coefficients, one finds

$$s_n = s_{nn}, \tag{89}$$

$$s_l = s_{ln} + A, \tag{90}$$

$$s = -(d/dl)m_{ln} - M_n, \tag{91}$$

$$Q = m_{nn}. \tag{92}$$

A PHYSICAL INTERPRETATION

The tensors $s_{\alpha\beta}$ and $m_{\alpha\beta}$ were introduced only formally in the previous section. One may interpret these tensors physically by referring to Eqs. (89)–(92) which, in effect, identify them with the external stresses imposed by the environment. Thus, Eq. (89) shows that one component of the tensor $s_{\alpha\beta}$ in the (n, l) coordinate system is equal to the externally applied lineal normal stress. Hence we identify $s_{\alpha\beta}$ as the components of lineal stress[16] within the shell surface S. In the same way, from Eq. (92), the externally applied lineal moment Q is seen to be equal to one component of the tensor $m_{\alpha\beta}$.

[14] Note that the single subscript suffices to distinguish the externally applied lineal stress from the formally defined tensor $s_{\alpha\beta}$.

[15] To the degree that the shell is constrained, the boundary conditions are geometric rather than physical. We leave to the reader the description of the intermediate cases.

[16] It follows, however, from Eq. (90), that s_{12} is not the lineal shear stress but differs from it by the quantity A.

We therefore identify the tensor $m_{\alpha\beta}$ as the lineal bending moment within the shell surface S.

SPECIALIZATION TO PRINCIPAL COORDINATES

Curvilinear coordinates α_1, α_2 on S are orthogonal if, for an infinitesimal displacement ω_1, ω_2 along the surface, the corresponding change in α_β, $\beta = 1, 2$, is given by

$$d\alpha_\beta = (1/A_\beta)\omega_\beta, \tag{93}$$

where A_β is a function of α_1, α_2. Any edf of rank unity can be regarded as an edf of the independent variables α_β. Thus,

$$\begin{aligned} \omega &= u_\beta \omega_\beta, \\ &= A_\beta u_\beta \, d\alpha_\beta. \end{aligned} \tag{94}$$

Form the differential and apply Theorem 7:

$$\begin{aligned} u_{\beta,\gamma}[\omega_\gamma \ \omega_\beta] &= d(A_\beta u_\beta \, d\alpha_\beta), \\ &= [\partial(A_\beta u_\beta)/\partial\alpha_\gamma][d\alpha_\gamma \ d\alpha_\beta]. \end{aligned}$$

Restore ω_β and ω_γ by substitution in the right side from Eq. (93), and equate the coefficients of $[\omega_\beta\omega_\gamma]$:

$$e_{\gamma\beta}u_{\beta,\gamma} = e_{\gamma\beta}[\partial(A_\beta u_\beta)/\partial\alpha_\gamma].$$

Replace u_β by $e_{\beta\gamma}u_\gamma$. After simplification, one finds

$$u_{\gamma,\gamma} = \frac{1}{A_1A_2} \sum_{(\beta,\gamma)} \frac{\partial(A_\beta u_\gamma)}{\partial\alpha_\gamma}. \tag{95}$$

where the summation is on $(\beta, \gamma) = (1, 2), (2, 1)$.

One can find, by forming the differential of both sides of Eq. (93) and applying Theorem 6, a relation between r_β and the quantities A_β. For brevity, introduce the notation

$$(\ )' = \frac{1}{A_1}\frac{\partial}{\partial\alpha_1}(\ ), \qquad (\ )'' = \frac{1}{A_2}\frac{\partial}{\partial\alpha_2}(\ ). \tag{96}$$

Then the formulas for r_1 and r_2 are

$$r_1 = -(\log A_1)'', \qquad r_2 = (\log A_2)'. \tag{97}$$

The orthogonal coordinates are *principal* if

$$p_{12} = p_{21} = 0. \tag{98}$$

The absolute values of p_{11} and p_{22} are then the *principal curvatures*. The Gauss–Codazzi equations take the form

$$(A_1p_{11})'' = p_{22}A_1', \tag{99}$$

$$(A_2p_{22})' = p_{11}A_2', \tag{100}$$

$$p_{11}p_{22} = -A_1^{-1}(A_1')'' - A_2^{-1}(A_2')'. \tag{101}$$

The components of a tensor derivative can be expressed in terms of the quantities A_β. For instance, the formulas of Theorem 2 can be rewritten replacing the subscripts by operators:

$$|1 \to (\ )', \qquad |2 \to (\ )''. \tag{102}$$

In particular, when the formulas are applied to the tensor $p_{\alpha\beta}$ and the coordinates are principal, then one finds

$$\begin{aligned} (p_{\alpha\beta,1}) &= \begin{bmatrix} p_{11}' & r_1(p_{11} - p_{22}) \\ r_1(p_{11} - p_{22}) & r_2(p_{11} - p_{22}) \end{bmatrix}, \\ (p_{\alpha\beta,2}) &= \begin{bmatrix} r_1(p_{11} - p_{22}) & r_2(p_{11} - p_{22}) \\ r_2(p_{11} - p_{22}) & p_{22}'' \end{bmatrix}. \end{aligned} \tag{103}$$

It follows that $p_{\alpha\beta,\gamma}$ is symmetric under interchange of any two subscripts, hence is a completely symmetric tensor. This property subsists under a transformation of coordinates. Hence,

Theorem 10. *The tensor $p_{\alpha\beta,\gamma}$ is symmetric under any permutation of the subscripts.*

The expressions for the components of strain tensors in principal coordinates can be found by substitution into Eqs. (51) and (56) in accordance with the discussion above. The results are ($\epsilon_{12} = \epsilon_{21}$, $\kappa_{12} = \kappa_{21}$)

$$\begin{aligned} \epsilon_{11} &= u_1' - r_1u_2 - p_{11}w, \\ \epsilon_{12} &= \tfrac{1}{2}[u_1'' + u_2' + r_1u_1 - r_2u_2], \\ \epsilon_{22} &= u_2'' + r_2u_1 - p_{22}w, \end{aligned} \tag{104}$$

$$\begin{aligned} \kappa_{11} &= p_{11}'u_1 + p_{11}'u_2 + (w')' - r_1w'' + p_{11}^2w, \\ \kappa_{12} &= \tfrac{1}{2}(p_{11} - p_{22})(u_1'' - u_2') + \tfrac{1}{2}[p_{11}''u_1 + p_{22}'u_2] \\ &\quad + \tfrac{1}{2}[(w')'' + (w'')'] + \tfrac{1}{2}[r_1w' - r_2w''], \\ \kappa_{22} &= p_{22}'u_1 + p_{22}''u_2 + (w'')'' + r_2w' + p_{22}^2w. \end{aligned} \tag{105}$$

The strain energy $\mathcal{E}$ per unit area is put in the form

$$\begin{aligned} \mathcal{E} = \tfrac{1}{2}Eh[F_1\epsilon_{\alpha\alpha}^2 + F_2\epsilon_{\alpha\beta}\epsilon_{\beta\alpha} + h^2F_3\kappa_{\alpha\alpha}^2 \\ + h^2F_4\kappa_{\alpha\beta}\kappa_{\beta\alpha}], \end{aligned} \tag{106}$$

where

E is Young's modulus, h is the half-thickness of the shell, and $F_1, \cdots, F_4$ are nondimensional, material constants. Define the stress $s_{\alpha\beta}$ and the bending moment tensor $m_{\alpha\beta}$ as in Eq. (61) but with $\mathfrak{T}$ replaced by $\mathcal{E}$. Then,

$$s_{\alpha\beta} = Eh(F_1\epsilon_{\gamma\gamma}\,\delta_{\alpha\beta} + F_2\epsilon_{\alpha\beta}), \tag{107}$$

$$m_{\alpha\beta} = Eh^3(F_3\kappa_{\gamma\gamma}\,\delta_{\alpha\beta} + F_4\kappa_{\alpha\beta}). \tag{108}$$

The equations of equilibrium (87) and (88)

become, in principal coordinates,

$$X_1 = -s_{11}' - s_{12}'' + 2r_1 s_{12} - r_2(s_{11} - s_{22}) + p_{11}' m_{11} + 2p_{11}'' m_{12} + p_{22}' m_{22} + A'',$$

$$X_2 = -s_{12}' - s_{22}'' - r_1(s_{11} - s_{22}) - 2r_2 s_{12} + p_{11}'' m_{11} + 2p_{22}' m_{12} + p_{22}'' m_{22} - A',$$

$$Z = M_1' + r_2 M_1 + M_2'' - r_1 M_2 + p_{11}^2 m_{11} + p_{22}^2 m_{22} - (p_{11}s_{11} + p_{22}s_{22}). \quad (109)$$

The auxiliary functions A, M_1, and M_2 defined in Eqs. (65) and (71) are now given by

$$A = -m_{12}(p_{11} - p_{22}),$$

$$M_1 = m_{11}' + m_{12}'' - 2r_1 m_{12} + r_2(m_{11} - m_{22}),$$

$$M_2 = m_{12}' + m_{22}'' + r_1(m_{11} - m_{22}) + 2r_2 m_{12}. \quad (110)$$

The boundary conditions at a completely unconstrained edge are again given by Eqs. (89)–(92).

When the surface S has positive Gauss curvature, then it is convenient to choose the third axis of the local Cartesian frame in the direction toward the concave side of the surface. In this frame, both p_{11} and p_{22} are positive and one can write

$$p_{11} = 1/R_1, \qquad p_{22} = 1/R_2, \quad (111)$$

where R_1 and R_2 are the *principal radii of curvature.* When the Gauss curvature is zero, then a similar choice of normal axis can be made so that

$$p_{11} = 1/R_1, \qquad p_{22} = 0. \quad (112)$$

On the other hand, when the Gauss curvature is negative, one can choose the normal axis so that

$$p_{11} = 1/R_1, \qquad p_{22} = -1/R_2, \quad (113)$$

where R_1 and R_2 are positive.

PHYSICAL CONSTANTS

The theory so obtained is complete except in one detail—the identification of the physical constants $F_1, \cdots, F_4$ which appear in Eqs. (106)–(108). These constants could conceivably be obtained by experiment; however, it is more satisfactory for completing the theory to obtain them by connecting the theory derived above with the classical three dimensional theory of small strains.

The theory of thin plates has been investigated from the standpoint of the three-dimensional theory, and expressions have been derived for the strain energy in stretching and bending (Reference 1, pp. 469, 503). These are of the same forms, up to principal orders of thickness, as Eq. (58) except for a factor Eh. Making the comparison, one finds

$$f_1 = \frac{2}{1-\nu^2}, \qquad f_2 = -\frac{4}{1+\nu},$$

$$f_3 = \frac{2}{3}\frac{1}{1-\nu^2}, \qquad f_4 = -\frac{2}{3}\frac{2}{1+\nu}. \quad (114)$$

The quantity ν is Poisson's ratio. Solving for the constants $F_1, \cdots, F_4$ according to Eq. (60), one has, for Eq. (106),

$$F_1 = \frac{2\nu}{1-\nu^2}, \qquad F_2 = \frac{2}{1+\nu},$$

$$F_3 = \frac{2}{3}\frac{\nu}{1-\nu^2}, \qquad F_4 = \frac{2}{3}\frac{1}{1+\nu}. \quad (115)$$

It is not the intention here to state that the demonstration of these expressions is completely satisfactory. We believe that an analysis independent of Love's work would be desirable. This aspect of the theory, together with an analysis of the limitations of our results, will be developed in a subsequent paper.

Statistical Behavior of Linear Systems with Randomly Varying Parameters

M. A. LEIBOWITZ*
International Business Machines Corporation, Thomas J. Watson Research Center, Yorktown Heights, New York
(Received 11 October 1962)

We consider an nth-order system of ordinary linear differential equations whose coefficients are random functions of the time. In particular, we discuss the case where each of these coefficients is a random noise. A differential equation for the probability distribution of the solutions of the random D. E. is derived and from this the moments can be calculated. Special attention is given to the case of Gaussian noise but the treatment is applicable to any type of noise. Finally, various conditions for stability are discussed.

INTRODUCTION

THE statistical properties of linear systems subject to random external forces has been extensively treated.[1] There are, however, relatively few discussions available of the case where the parameters characterizing the linear systems themselves are random functions of the time.[2] Such a system, with n degrees of freedom, may be described analytically by a set of ordinary differential equations:

$$\frac{dx_i}{dt} = \sum_{j=1}^{n} [c_{ij}(t) + r_{ij}(t)]x_j. \tag{1}$$

The $c_{ij}(t)$ are known (i.e. deterministic) functions and the $r_{ij}(t)$ are random functions. The difficulty in treating a set of equations such as (1) is that notwithstanding their linearity, the solutions depend in a nonlinear way on the coefficients. Indeed, for the simple first-order equation, $x(t) = x_0 \exp [\int_0^t (c(t) + r(t))\, dt]$, while for higher-order equations this dependence cannot even be explicitly represented, at least in a closed form.

A consequence of this fact is that the statistics of the solutions of (1) are not related in any simple way to those of the random functions $r_{ij}(t)$. This is in contrast to the apparently similar set:

$$\frac{dx_i}{dt} = \sum_{j=1}^{n} c_{ij}(t)x_j + r_i(t). \tag{2}$$

For the system (2) a knowledge of all the kth and preceding moments of $r_i(t)$ is sufficient to determine the kth moments of $x_i(t)$.

Thus, it is necessary in order to find, say, the moments of the solutions of Eq. (1) to specify the random process governing $r_{ij}(t)$ in far more detail than is required for the case of (2). In this paper we assume that each $r_{ij}(t)$ is a random (white) noise, though no restriction will be made on the type of noise (e.g. it need not be Gaussian or stationary). This assumption will allow us to obtain a differential equation for the probability density of the solutions x_i; and from this, systems of ordinary linear differential equations for the moments can be derived. It should be noted that the average solution $\langle x_i \rangle$ does not coincide with the solution of the average equation; i.e., the equation obtained by setting $r_{ij}(t) = \langle r_{ij}(t) \rangle$.

Finally, we will define and derive some conditions for the stability of solutions of Eq. (1). It is shown (as may be expected) that Gaussian noise coefficients act as destabilizing influences rather than as stabilizing ones.

1. THE CASE OF A FIRST-ORDER SYSTEM

Consider first the simplest case

$$dx/dt = r(t)x, \tag{3}$$

where $r(t)$ is a stationary *shot* noise.[3] Such a noise is composed of a series of identical impulses of very short duration occurring at random times. Analytically, $r(t)$ may be represented in an idealized form by a sum of delta functions:

$$r(t) = S\,\delta(t - t_1) + S\,\delta(t - t_2) + \cdots + S\,\delta(t - t_n) + \cdots, \tag{4}$$

where the strength S is the integral of $r(t)$ over a single impulse, and the distribution of the random times $t_1, t_2, \cdots$ at which impulses occur, is characterized by specifying that the probability of an im-

* Present address: Bellcomm Inc., Washington, D. C.

[1] For example, J. Laning and R. Batten, *Random Processes in Automatic Control* (McGraw-Hill Book Company, Inc., New York, 1956).

[2] J. P. Samuels, J. Acoust. Soc. Am. 32, 594 (1960).

[3] Reference 1, p. 147.

pulse in a time dt is $\lambda\, dt$. λ is thus the *rate* at which impulses appear.

Let us find the change in x due to an impulse. If t_a, t_b ($t_a > t_b$) denote two times between which only one impulse takes place, say at time t^*, then on solving Eq. (2), we have

$$x_a = x_b \exp\left[\int_{t_b}^{t_a} r(t)\, dt\right] = x_b \exp\left[S \int_{t_b}^{t_a} \delta(t - t^*)\, dt\right],$$

where x_b and x_a are the values of x before and after an impulse, respectively. Hence,

$$x_a = x_b e^S. \tag{5}$$

Equation (5) is fundamental for what follows. It is of considerable interest to observe that it cannot be obtained by simply integrating both sides of (3) over the interval $[t_b, t_a]$ and replacing $r(t)$ by $S\,\delta(t - t^*)$. For then we would have

$$x_a - x_b = \int_{t_b}^{t_a} r(t)x(t)\, dt = S \int_{t_b}^{t_a} \delta(t - t^*)x(t)\, dt = Sx(t^*).$$

But $x(t^*)$, the value of x while an impulse is occurring, has no unambiguous meaning since during an impulse, x is varying rapidly between x_a and x_b. This follows from the fact that the derivative $dx/dt = r(t)x$ is very large during an impulse. In particular, it is not permissible to write $x(t^*) = x_b$, since this gives $x_a = (1 + S)x_b$ instead of the correct Eq. (5). The significance of this last remark will be apparent when the case of Gaussian noise is considered.

Now at any time t, let $x(t)$ be distributed according to the probability density $p(x, t)$. We will derive an equation describing how $p(x, t)$ changes with time. To do so, it will be convenient to adopt the following point of view: We assume that there are a large number N of particles on a line, all having the same equation of motion, namely (3), and such that the number of particles in the small interval $(x, x + dx)$ at time t, is just $Np(x, t)\, dx$. During the subsequent short time dt, $\lambda Np(x, t)\, dx\, dt$ particles will leave the interval $(x, x + dx)$ due to an impulse in $r(t)$ while $\lambda Np(e^{-S}x, t)e^{-S}\, dx\, dt$ particles will enter. To see this, recall that the probability of an impulse in time dt is $\lambda\, dt$: and furthermore that particles in the interval $(x, x + dx)$ *after* an impulse, must have been in the interval $(x_b, x_b + dx_b)$ *before*, where by Eq. (5),

$$x_b = e^{-S}x, \qquad dx_b = e^{-S}\, dx.$$

Hence,

$$Np(x, t + dt)\, dx - Np(x, t)\, dx = \lambda Np(xe^{-S}, t)e^{-S}\, dx\, dt - \lambda Np(x, t)\, dx\, dt,$$

or

$$\partial p(x, t)/\partial t = \lambda p(xe^{-S}, t)e^{-S} - \lambda p(x, t). \tag{6}$$

For the slightly more general equation

$$dx/dt = [c + r(t)]x, \tag{7}$$

where c is a known (deterministic) constant, we find by an immediate extension of the argument leading to (6) that

$$\frac{\partial p(x, t)}{\partial t} = -c\,\frac{\partial [xp(x, t)]}{\partial x} + \lambda p(xe^{-S}, t)e^{-S} - \lambda p(x, t). \tag{8}$$

The moments $\langle x^n \rangle$ of $p(x, t)$ can be obtained easily by multiplying (8) by x^n and integrating. We get

$$d\langle x^n \rangle/dt = nc\langle x^n \rangle + \lambda(e^{nS} - 1)\langle x^n \rangle, \tag{9}$$

or

$$\langle x^n \rangle = \langle x_0^n \rangle \exp\,[nc + \lambda(e^{nS} - 1)]t, \tag{10}$$

where $\langle x_0^n \rangle$ is the expected value of x^n at time $t = 0$.

The result (10) may also be found by solving Eq. (3) directly. Then

$$x = x_0 \exp\left[ct + \int_0^t r(t')\, dt'\right], \tag{11}$$

and

$$\langle x^n \rangle = \langle x_0^n \rangle e^{nct}\left\langle \exp\left[n \int_0^t r(t')\, dt'\right]\right\rangle. \tag{12}$$

Now, $\int_0^t r(t')\, dt' = k(t)S$ where $k(t)$ is the number of impulses occurring in the time t. Since impulses appear at random with average rate λ, $k(t)$ has a Poisson distribution: i.e., the probability of $k(t) = j$ is $e^{-\lambda t}\,(\lambda t)^j/j!$. Hence,

$$\left\langle \exp\left[n \int_0^t r(t')\, dt'\right]\right\rangle = \langle e^{nk(t)S}\rangle = \sum_{j=0}^{\infty} e^{njS}e^{-\lambda t}(\lambda t)^j/j! = \exp\,[\lambda t(e^{nS} - 1)],$$

which gives Eq. (12) for $c = 0$.

The use of Eq. (6) to calculate the moments has, however, two advantages over solving the original differential equation (3) directly. First, and more important, it is applicable to an nth-order system containing shot-noise coefficients. Secondly, the case where c, λ, and S are functions of the time can be

readily included. Indeed, in Eq. (6) we need only replace these constants by $c(t)$, $\lambda(t)$, and $S(t)$ without making any other changes. Then, the quantities $\langle x^n \rangle$ are given by

$$\langle x^n \rangle = \langle x_0^n \rangle \exp\left[\int_0^t \{nc(t') + \lambda(t')e^{nS(t')}\}\, dt'\right]. \quad (13)$$

Suppose that, in Eq. (2), $r(t)$ was actually the sum of two *independent* shot noises $r_1(t)$ and $r_2(t)$ with impulse rates λ_1 and λ_2 and strengths S_1 and S_2 respectively. Then, since $r_1(t)$ and $r_2(t)$ are independent, and noting that the probability of an impulse in *both* $r_1(t)$ and $r_2(t)$ during a short time dt is $\lambda_1\lambda_2\ (dt)^2$ and hence negligible, we find that

$$\frac{\partial p(x,t)}{\partial t} = -c\frac{\partial[xp(x,t)]}{\partial x} + \lambda_1 p(xe^{-S_1}, t)e^{-S_1} + \lambda_2 p(xe^{-S_2}, t)e^{-S_2} - (\lambda_1 + \lambda_2)p(x,t), \quad (14)$$

which gives the equations

$$\frac{d\langle x^n \rangle}{dt} = c\langle x^n \rangle + [\lambda_1(e^{nS_1} - 1) + \lambda_2(e^{nS_2} - 1)]\langle x^n \rangle. \quad (15)$$

The extension to the case where $r(t)$ is composed of an arbitrary number of shot noises of different strengths and rates is obvious. Since any type of noise can be regarded as the superposition of a finite or infinite number of shot noises, the formalism developed here is generally applicable to a linear differential equation with random noise coefficients.[4]

In particular let us consider the case of Gaussian noise[3] $g(t)$ with zero mean and let σ^2 denote the *power* of the noise: $\langle g(t)g(t')\rangle = \sigma^2\,\delta(t - t')$.

Let $r_1(t)$ and $r_2(t)$ be shot noises with the same rate λ, and with strengths S, and $-S$. Then if we allow λ to approach infinity and S zero, in such a way that $\lambda S^2 \to \frac{1}{2}\sigma^2$, the sum $r_1(t) + r_2(t)$ tends to $g(t)$.[5] Intuitively, we can regard the Gaussian noise $g(t)$ as composed of a very large number of very small impulses of equal magnitude, half of which are positive and the other half negative.

In order to find the probability density of the solutions of the equation

$$dx/dt = [c + g(t)]x. \quad (16)$$

first set $\lambda = \lambda_1$, $\lambda = \lambda_2$, $S = S_1$, $-S = S_2$, in Eq. (14). Thus

$$\frac{\partial p(x,t)}{\partial t} = -c\frac{\partial[xp(x,t)]}{\partial x} + \lambda[p(xe^{-S}, t)e^{-S} + p(xe^{S}, t)e^{S}] - 2\lambda p(x,t). \quad (17)$$

Now expand the right-hand member of (17) in a power series in S through terms in S^2. We have, e.g., $p(xe^{S}\ t) \approx p(x + x(S + \frac{1}{2}S^2)t)$

$$= p(x,t) + (S + \tfrac{1}{2}S^2)x\frac{\partial p(x,t)}{\partial x} + \tfrac{1}{2}S^2x^2\frac{\partial^2 p(x,t)}{\partial x^2} + \cdots.$$

Similarly, expanding the other terms in (17), we find that all terms of zero- and first-order in S cancel out, and we are left with

$$\frac{\partial p}{\partial t} = \lambda S^2\left[p + 3x\frac{\partial p}{\partial x} + 2\frac{\partial^2 p}{\partial x^2}\right] - c\frac{\partial(xp)}{\partial x} + 0(S^2),$$

where $p = p(x,t)$. Hence, allowing $\lambda \to \infty$, $S \to 0$, and $\lambda S^2 \to \frac{1}{2}\sigma^2$, we finally obtain the desired equation

$$\frac{\partial p}{\partial t} = -c\frac{\partial(xp)}{\partial x} + \tfrac{1}{2}\sigma^2\left[p + 3x\frac{\partial p}{\partial x} + x^2\frac{\partial^2 p}{\partial x^2}\right], \quad (18)$$

or

$$\frac{\partial p}{\partial t} = -c\frac{\partial(xp)}{\partial x} + \tfrac{1}{2}\sigma^2\left[-\frac{\partial(xp)}{\partial x} + \frac{\partial^2(x^2p)}{\partial x^2}\right], \quad (19)$$

Equation (19) may also be found directly from (16). Let

$$y(t) = ct + \int_0^t g(t')\, dt'. \quad (20)$$

Then if $w(y,t)$ is the probability density of $y(t)$, we have

$$w(y,t) = (2\pi\sigma^2 t)^{-\frac{1}{2}} \exp\left[-(y - ct)^2/2\sigma^2 t\right], \quad (21)$$

and this satisfies the equation

$$\frac{\partial w}{\partial t} = -c\frac{\partial w}{\partial y} + \tfrac{1}{2}\sigma^2\frac{\partial^2 w}{\partial y^2}. \quad (22)$$

Since $x(t) = x_0e^{y(t)}$, it follows that $p(x,t) = w(y,t)\,dy/dx = w(y,t)/x$, $\partial/\partial y = (x)\,\partial/\partial x$, and from (22),

$$\frac{\partial p}{\partial t} = -c\frac{\partial}{\partial x}(xp) + \tfrac{1}{2}\sigma^2\frac{\partial}{\partial x}x\frac{\partial(xp)}{\partial x} = -c\frac{\partial(xp)}{\partial x} + \tfrac{1}{2}\sigma^2\left[-\frac{\partial(xp)}{\partial x} + \frac{\partial^2(x^2p)}{\partial^2 x}\right], \quad (23)$$

which agrees with Eq. (19).

[4] M. S. Bartlett, *An Introduction to Stochastic Processes*, (Cambridge University Press, London, 1955), Chap. 3. Note that here, the *integral* of noise process, rather than the noise itself is discussed. An *additive* process or a process with *independent increments* is then said to be a superposition of Poisson processes.

[5] The easiest way to show this is to prove that the characteristic function of $\int_0^t [r_1(t') + r_2(t')]\, dt'$ tends to that of $\int_0^t g(t')\, dt'$ for any t, as $S \to 0$, $\lambda \to \infty$, $\lambda^2 S \to \frac{1}{2}\sigma^2$. But this follows at once on noting that $\langle \exp i\,\delta[\int_0^t r_1(t') + r_2(t')]\, dt'\rangle = \exp[\lambda t(e^{i\delta S} - 1) + \lambda t(e^{i\delta S} - 1)]$, and $\langle \exp i\,\delta \int_0^t g(t')\, dt'\rangle = \exp[-\frac{1}{2}\delta^2\sigma^2 t]$.

The moments, $\langle x^n \rangle$ may be calculated by multiplying Eq. (19) by x^n and integrating over x. To evaluate

$$\int_{-\infty}^{\infty} x^n \frac{\partial(xp)}{\partial x}\, dx,$$

we integrate by parts. Thus,

$$\int_{-\infty}^{\infty} x^n \frac{\partial(xp)}{\partial x}\, dx = x^{n+1}p\Big|_{-\infty}^{\infty} - n\int_{-\infty}^{\infty} x^n p\, dx = -n\langle x^n \rangle,$$

the vanishing of $x^{n+1}p$ at $\pm\infty$ being a consequence of the existence of the moments. Similarly, we find that

$$\int_{-\infty}^{\infty} x^n \frac{\partial^2(x^2p)}{\partial x^2}\, dx = n(n-1)\langle x^n \rangle.$$

Hence,

$$d\langle x^n \rangle/dt = (c + \tfrac{1}{2}\sigma^2 n^2)\langle x^n \rangle. \tag{24}$$

We call attention to the result that even if $c = 0$, $\langle x \rangle$ does not in general vanish; indeed, except for the special case $\langle x_0 \rangle = 0$, $\langle x \rangle$ increases exponentially as $e^{\frac{1}{2}\sigma^2 t}$. This may seem surprising since $\langle g(t) \rangle = 0$; the reason is that $g(t)$ enters into Eq. (16) in a nonlinear way.

Equation (24) may also be derived by setting $\lambda = \lambda_1, \lambda_2$, and $S = S_1, -S_2$, in (15), expanding in powers of S, and passing to the limit $\lambda S^2 = \frac{1}{2}\sigma^2$.

Equation (19) is a special case of the Fokker–Planck equation which plays a significant role in statistical physics.[6] The usual derivation of the Fokker–Planck equation begins with the relation

$$p(x, t + \Delta t) = \int_{-\infty}^{\infty} p(x - \Delta x, t)\psi(\Delta x, t \mid x - \Delta x)\, d(\Delta x), \tag{25}$$

where $\psi(\Delta x, \Delta t \mid x)$, "the transition function" is the probability density of the change Δx in x during a time Δt. (25) is valid provided that $x(t)$ constitutes a Markoff Process. This will certainly be the case if $x(t)$ is generated by the random differential equation

$$dx/dt = g(t)x. \tag{26}$$

In fact, since $g(t)$ is completely independent of its past history, the distribution of $\Delta x = x(t+\Delta t) - x(t)$ can depend only on $x(t)$. However in finding $\psi(\Delta x, \Delta t \mid x)$, we must caution against the following argument which seems to have appeared at least implicitly, in a number of places.[7] Namely, that if one integrates (26) over a very short time interval Δt, then

$$\Delta x = x(t + \Delta t) - x(t) = \int_t^{t+\Delta t} g(t')x\, dt' \approx x(t)\int_t^{t+\Delta t} g(t')\, dt'. \tag{27}$$

Hence for sufficiently small Δt, Δx has approximately the mean *zero* and variance

$$x^2(t)\left\langle \left(\int_t^{t+\Delta t} g(t')\, dt'\right)^2 \right\rangle = \sigma^2 x^2(t)\, \Delta t.$$

Using this in (25), expanding in powers of Δx, and passing to limit $\Delta t \to 0$, one is led not to the correct Eq. (19) (with $c = 0$), but rather to

$$\partial p/\partial t = \tfrac{1}{2}\sigma^2[\partial^2(x^2p)/\partial x^2], \tag{28}$$

where $p = p(x, t)$.

The source of the discrepancy lies in the use of the approximation

$$x(t)\int_t^{t+\Delta t} g(t')\, dt' \quad \text{for} \quad \int_t^{t+\Delta t} x(t')g(t')\, dt' \quad \text{in (27)}.$$

It has been seen that such an approximation is not valid for the case of shot noise; since $g(t)$ can be regarded as the limit of a sum of shot noises, it will also not be valid for the case of Gaussian noise.[8] In order to find $\psi(\Delta x, t \mid x)$, it is necessary to determine the distribution of Δx from the integrated equation

$$\Delta x = x(t)\left[\exp\left(\int_t^{t+\Delta t} g(t')\, dt'\right) - 1\right],$$

which shows that

$$\langle \Delta x \rangle = \tfrac{1}{2}\sigma^2 x(t)\, \Delta t, \qquad \langle (\Delta x)^2 \rangle = \sigma^2 x^2(t)\, \Delta t.$$

The use of this in (25) results in the correct Eq. (19).

2. GENERALIZATION TO AN *N*TH-ORDER SYSTEM

The preceding analysis may be readily extended to the case of an nth-order system. Let this system be

$$\frac{dx_i}{dt} = \sum_{j=1}^{n} (c_{ij} + r_{ij}(t))x_j \quad (i = 1, 2, \cdots, n), \tag{29}$$

where the c_{ij} are known constants and the r_{ij} *mutually independent* random functions of time. First suppose the r_{ij} are all shot noises with impulse rates

[6] S. Chandrasekhar, Rev. Mod. Phys. **15**, 1 (1943).

[7] J. L. Bogdanoff and F. Kozin, J. Acoust. Soc. Am. **34**, 1063 (1960).

[8] It is worthwhile noting that if instead of using the correct relation $x_a = x_b e^S$ for the jump due an impulse in a shot noise of strength S, we used $x_a = (1 + S)x_b$, and then passed to the limit of a Gaussian noise, we would obtain Eq. (28).

λ_{ij} and strengths S_{ij}. Just as before, we can find the change in x_i, say, due to an impulse in one of the coefficients r_{ij}. However, we must distinguish between the case where this impulse occurs in the diagonal coefficients r_{ii} or in off-diagonal coefficient r_{ij} $(i \neq j)$. Consider first the effect of an impulse in r_{ii}. Then while this impulse is taking place, the differential equation for x_i reduces to

$$dx_i/dt = r_{ii}(t)x_i,$$

This follows from the fact that during an impulse, $\sum_{j=1}^{n} c_{ij}x_j$ is negligibly small compared with the very large values attained by $r_{ii}(t)$; and furthermore, that because of the very short duration of an impulse, the possibility of an impulse occurring simultaneously in any of the other coefficients r_{ij} $(i \neq j)$ can be ignored. Thus, just as in Sec. 1, we have

$$(x_i)_a = e^{S_{ii}}(x_i)_b. \tag{30}$$

On the other hand, if j is different than i, we have, during an impulse r_{ij},

$$dx_i/dt = r_{ij}x_j. \tag{31}$$

Here, since dx_j/dt is finite, x_j may be assumed constant during the short time of an impulse; and so

$$\begin{aligned}(x_i)_a - (x_i)_b &= \int_{t_b}^{t_a} r_{ij}x_j\,dt = x_j\int_{t_b}^{t_a} r_{ij}\,dt \\ &= S_{ij}x_j,\end{aligned} \tag{32}$$

where we need not distinguish between $(x_j)_a$ and $(x_j)_b$

Let $p(x_1, \cdots, x_n, t)$ be the joint probability density of $x_1, \cdots, x_n$ at time t. By considering the number of particles leaving and entering the n-dimensional interval $[x_i, x_i + dx_i]$ $(i = 1, 2, \cdots, n)$ during a short time dt, using Eqs. (30) and (32), and paralleling the argument leading to (6), we find

$$\begin{aligned}\frac{\partial p(x_1, \cdots, x_n, t)}{\partial t} &= -\sum_{i,j=1}^{n} c_{ij}\frac{\partial(x_j p(x_1 \cdots, x_n, t))}{\partial x_i} \\ &+ \sum_{i=1}^{n} \lambda_{ii}p(x_1, \cdots, x_i e^{-S_{ii}}, \cdots x_n, t)e^{-S_{ii}} \\ &+ \sum_{\substack{i,j=1 \\ i\neq j}}^{n} \lambda_{ij}p(x_1, \cdots, x_i - S_{ij}x_j, \cdots, x_j \cdots, x_n, t) \\ &- \sum_{i,j=1}^{n} \lambda_{ij}p(x_1, \cdots, x_n t).\end{aligned} \tag{33}$$

The notation $p(x_1, \cdots, x_i - S_{ij}x_j, \cdots, x_j, \cdots, x_n, t)$ means that the variable in the probability density corresponding to the ith coordinate is evaluated at $x_i - S_{ij}\,x_j$. As we have written it, the index j appears to be greater than i, but actually the sum is taken over all values of i and j with $i \neq j$.

We now derive a system of equations for the expectations $\langle x_i \rangle$. We have, on multiplying (33) by x_i and integrating,

$$\begin{aligned}\frac{d\langle x_i\rangle}{dt} &= \sum_{j=1}^{n} c_{ij}\langle x_j\rangle \\ &+ \lambda_{ii}(e^{S_{ii}} - 1)\langle x_i\rangle + \sum_{\substack{j=1 \\ j\neq i}}^{n} \lambda_{ij}S_{ij}\langle x_j\rangle.\end{aligned} \tag{34}$$

The integral

$$\begin{aligned}\int x_i p(x_1, \cdots, x_i &\\ - S_{ij}x_j, \cdots, x_j, \cdots, x_n)\,dx_1 \cdots dx_n &\\ = \langle x_i\rangle + S_{ij}\langle x_j\rangle &\end{aligned}$$

is evaluated by making the change of variables $x_i - S_{ij}x_j = z_i$ and $x_j = z_j$. In a similar way, we have for the second moments $\langle x_i x_j\rangle$,

$$\begin{aligned}\frac{d\langle x_i^2\rangle}{dt} &= 2\sum_{j=1}^{n} c_{ij}\langle x_i x_j\rangle + \lambda_{ii}(e^{2S_{ii}} - 1)\langle x_i^2\rangle \\ &+ \sum_{\substack{j=1 \\ i\neq j}}^{n} \lambda_{ij}[2S_{ij}\langle x_i x_j\rangle + S_{ij}^2\langle x_j^2\rangle],\end{aligned} \tag{35}$$

$$\begin{aligned}\frac{d\langle x_i x_j\rangle}{dt} &= \sum_{k=1}^{n} c_{ik}\langle x_i x_k\rangle + \sum_{k=1}^{n} c_{kj}\langle x_i x_k\rangle \\ &+ [\lambda_{jj}(e^{S_{jj}} - 1) + \lambda_{ii}(e^{S_{ii}} - 1)]\langle x_{ij}\rangle \\ &+ \sum_{k=1}^{n} \lambda_{ik}S_{ik}\langle x_j x_k\rangle + \sum_{k=1}^{n} \lambda_{jk}S_{jk}\langle x_i x_k\rangle \quad (i \neq j).\end{aligned} \tag{36}$$

In deriving Eqs. (35) and (36) we have used the integrals

$$\begin{aligned}\int x_i^2 p(x_1, \cdots, x_i &\\ - S_{ik}x_k, \cdots, x_k \cdots, x_n)\,dx_1 \cdots dx_n &\\ = \langle x_i^2\rangle + 2S_{ik}\langle x_i x_k\rangle + S_{ik}^2\langle x_k^2\rangle, &\end{aligned}$$

and

$$\begin{aligned}\int x_i x_j p(x_1 \cdots, x_k &\\ - S_{kl}x_l, \cdots, x_l, \cdots, x_n)\,dx_1 \cdots dx_n &\\ (i \neq j) &\\ = \begin{cases}\langle x_i x_j\rangle & (i, j \neq k) \\ \langle x_i x_j\rangle + S_{il}\langle x_i x_l\rangle & (i = k) \\ \langle x_i x_j\rangle + S_{jl}\langle x_i x_l\rangle & (j = k)\end{cases}. &\end{aligned}$$

Similar, but more complicated differential equations can be derived for the higher moments in a like manner. Extensions to the case where the parameters S_{ij}, λ_{ij}, c_{ij} are time-dependent are immediate; indeed, this case is actually included in the preceding analysis.

It is also clear that an equation similar to (33) can easily be written when *actually* each r_{ij} is a sum of shot noises. We need only consider separately the effect of impulses in each of the component noises in r_{ij}, just as was done in deriving Eq. (14).

Let us, in particular, determine the equation for $p(x_1, \cdots x_n, t)$ when each of the r_{ij} are actually Gaussian noises. We will use the notation $g_{ij}(t)$ instead of r_{ij}, and σ_{ij}^2 will denote the power of the noise: $\langle g_{ij}(t)g_{ij}(t')\rangle = \sigma_{ij}^2\,\delta(t-t')$.

We follow, almost word for word, the procedure used in Sec. 1, i.e. we suppose that each r_{ij} is the sum of two shot noises with identical impulse rates λ_{ij}, one having strength S_{ij} and the other strength $-S_{ij}$. In the equation for $p(x_1, \cdots, x_n, t)$ for this case, we expand in powers of S_{ij} up to terms of the second order, and pass to the limit as $\lambda_{ij} \to \infty$, $S_{ij} \to 0$, $\lambda_{ij}S_{ij}^2 \to \frac{1}{2}\sigma_{ij}^2$. Then

$$\frac{\partial p}{\partial t} = -\sum_{i,j=1}^{n} c_{ij}\frac{\partial(x_j p)}{\partial x_i} + \frac{1}{2}\sum_{i=1}^{n}\sigma_{ii}^2\left[-\frac{\partial(x_i p)}{\partial x_i} + \frac{\partial^2(x_i^2 p)}{\partial x_i^2}\right] + \frac{1}{2}\sum_{\substack{i,j=1\\ i\neq j}}^{n}\sigma_{ij}^2\frac{\partial^2(x_i^2 p)}{\partial x_i^2}. \tag{37}$$

The moments can be obtained, as before, by multiplying (37) by the appropriate quantity and integrating. We find that

$$\frac{d\langle x_i\rangle}{dt} = \sum_{j=1}^{n} c_{ij}\langle x_j\rangle + \tfrac{1}{2}\sigma_{ii}^2\langle x_i\rangle, \tag{38}$$

$$\frac{d\langle x_i^2\rangle}{dt} = 2\sum_{j=1}^{n} c_{ij}\langle x_i x_j\rangle + 2\sigma_{ii}^2\langle x_i^2\rangle + \sum_{\substack{j=1\\ j\neq i}}^{n}\sigma_{ij}^2\langle x_j^2\rangle, \tag{39}$$

$$\frac{d\langle x_i x_j\rangle}{dt} = \sum_{k=1}^{n} c_{ik}\langle x_i x_k\rangle + \sum_{k=1}^{n} c_{kj}\langle x_i x_k\rangle + \tfrac{1}{2}(\sigma_{ij}^2 + \sigma_{ii}^2)\langle x_i x_j\rangle \qquad (i \neq j). \tag{40}$$

When the coefficients c_{ij} all vanish, the equations for $\langle x_i\rangle$ and $\langle x_i x_j\rangle$ $(i \neq j)$ become remarkably simple. They can be immediately integrated to give

$$\langle x_i\rangle = \langle x_i\rangle_0 \exp\,[\sigma_{ii}^2 t/2],$$

$$\langle x_i x_j\rangle = \langle x_i x_j\rangle_0 \exp\,[\tfrac{1}{2}(\sigma_{ii}^2 + \sigma_{jj}^2)t].$$

As an important special case, consider the nth-order differential equation

$$\frac{d^n x}{dt^n} + [c_{n-1} + r_{n-1}(t)]\frac{d^{n-1}x}{d^{n-1}t} + \cdots + [c_0 + r_0(t)]x = 0, \tag{41}$$

where the c_i are constants and the $r_i(t)$ are random functions. If, in particular, the $r_i(t)$ are Gaussian noises with power σ_i^2, it follows from (38) that

$$\frac{d^n\langle x\rangle}{dt^n} + [c_{n-1} - \tfrac{1}{2}\sigma_{n-1}^2]\frac{d^{n-1}\langle x\rangle}{dt^{n-1}} + \sum_{i=0}^{n-2} c_i\frac{d^i\langle x\rangle}{dt^i} = 0. \tag{42}$$

Thus, the equation satisfied by the "average" solution *does not*, generally, coincide with the equation obtained by replacing each random function by its average value.

3. STABILITY

A system of linear ordinary (deterministic) differential equations

$$\frac{dx_i}{dt} = \sum_{j=1}^{n} c_{ij}x_j \qquad (i = 1, 2, \cdots, n) \tag{43}$$

is said to be *stable* if for any solution, the sum $\sum_{i=1}^{n}|x_i|$ tends to zero as t approaches infinity. For the system

$$\frac{dx_i}{dt} = \sum_{j=1}^{n}[c_{ij} + r_{ij}(t)]x_j, \tag{44}$$

involving the random functions $r_{ij}(t)$, the preceding definition of stability must be generalized. We will say that the system (44) is *stable in the* kth *mean* $(k > 0)$ if $\langle\sum_{i=1}^{n}|x_i|^k\rangle$ tends to zero as t approaches infinity for every choice of the initial (i.e. at time $t = 0$) probability distribution of the x_i.

Similarly, we will say that the system (44) is *stable* in *probability* if for every $\epsilon > 0$, the probability of the inequality $\sum_{i=1}^{n}|x_i| > \epsilon$ tends to zero as t approaches infinity. Well-known inequalities of probability theory imply the following order among these criteria for stability[9]: If a system is stable in the kth mean, it is stable in the k'th mean for every $k' < k$. If the system is stable in the kth mean for any k, then it is stable in probability.

The case of the first-order equation

$$dx/dt = [c + g(t)]x \tag{45}$$

[9] The basic result is that for any random variable y, $\langle|y|^k\rangle^{1/k}$ is a nondecreasing function of k; cf M. Loeve, *Probability Theory* (D. Van Nostrand, Inc., Princeton, New Jersey, 1960).

illustrates this. According to Eq. (24), (45) will be stable in the kth mean if and only if $c + \frac{1}{2}\sigma^2 k^2 < 0$, while it will be stable in probability if and only if $c < 0$. It is tempting to conjecture that for the nth-order system,

$$\frac{dx_i}{dt} = \sum_{j=1}^{n} [c_{ij} + g_{ij}(t)]x_j, \tag{46}$$

with Gaussian noise coefficients the condition for stability in probability is that the deterministic system (43) be stable.

In general, the condition easiest to apply is that of stability in the second mean. The equations for $\langle x_i^2 \rangle$ (35) or (38) are linear ordinary differential equations, and for stationary processes, have constant coefficients. Their stability may therefore be determined by standard methods (e.g. the Routh–Hurwitz procedure).

We will conclude by showing that a Gaussian noise coefficient never acts as a stabilizing influence; indeed, it may even destabilize an otherwise stable system.[10] For $\sum_{i=1}^{n} \langle x_i^2 \rangle$ to tend to zero with increasing t, it is necessary that $\sum_{i=1}^{n} \langle |x_i| \rangle$ and hence $\sum_{i=1}^{n} \langle x_i \rangle$ should as well.

Now let $\mathbf{x}$ be the vector with components x_i, C the matrix with elements c_{ij}, and D the diagonal matrix with elements $\frac{1}{2}\sigma_{ii}^2$. Then the Eqs. (43) for the deterministic system can be written as

$$dx/dt = C\mathbf{x}, \tag{47}$$

and Eqs. (38) for $\langle x_i \rangle$ as

$$d\langle x \rangle/dt = C\langle \mathbf{x} \rangle + D\langle \mathbf{x} \rangle. \tag{48}$$

If (48) is stable, the expression $\mathbf{v}C\mathbf{v}$ must be negative for any n component vector v; instability means that there is a vector, $\mathbf{v}_0$ say, such that $\mathbf{v}_0 C \mathbf{v}_0 \geq 0$. But for this same vector, $\mathbf{v}_0 C \mathbf{v}_0 + \mathbf{v}_0 D \mathbf{v}_0 \geq 0$. This implies that if the deterministic system (48) is unstable, then so will be the random system. On the other hand, since D has all positive elements, it is possible for (48) to be unstable even though (47) may be stable.

As an illustration consider the second-order system

$$\frac{d^2x}{dt^2} + [c_1 + g_1(t)]\frac{dx}{dt} + [c_0 + g_0(t)] \qquad x = 0. \tag{49}$$

Then by (42),

$$\frac{d^2\langle x \rangle}{dt^2} + [c_1 - \tfrac{1}{2}\sigma_1^2]\frac{d\langle x \rangle}{dt} + c_0\langle x \rangle = 0. \tag{50}$$

If we suppose that (49) describes a mechanical system with x being the displacement of a particle from equilibrium, then by (50), the random coefficient $g_1(t)$ on the average acts as a *negative* friction even though $\langle g_1(t) \rangle = 0$.

[10] This is in constrast to the result claimed in reference 2.