Complex potential formulation of the axially symmetric gravitational field problem*

Frederick J. Ernst

Department of Physics, Illinois Institute of Technology, Chicago, Illinois 60616
(Received 15 March 1973)

Spin-coefficients and null tetrad components of the Ricci tensor and the Weyl conform tensor are evaluated in terms of a single complex gravitational potential ϵ, while null tetrad components of the electromagnetic stress energy tensor are evaluated in terms of a second complex potential Φ. All the results are expressed elegantly in terms of a differential operator $\eth$, similar to the "thop" of Newman and Penrose. The problem of finding physically pertinent stationary axially symmetric Einstein–Maxwell fields is reduced to the search for a complex solution $\xi_0(x, y)$ of one nonlinear differential equation subject to simple subsidiary conditions.

I. INTRODUCTION

Our objective is the systematic description of those space-times which may be regarded as representing the gravitational fields of bounded uniformly rotating axially symmetric sources. Until recently the only known exact solution of Einstein's field equations having these attributes was the Kerr metric.[1] However, Tomimatsu and Sato, using this author's complex potential formalism, have now discovered new exact solutions which belong to the class in which we are interested.[2]

In the complex potential formulation of the axial symmetry problem one seeks a solution of the nonlinear differential equation

$$(\xi_0\xi_0^* - 1)\nabla^2\xi_0 = 2\xi_0^*\nabla\xi_0 \cdot \nabla\xi_0, \tag{1}$$

where ∇ is the gradient operator and ∇^2 is the three-dimensional Laplacian operator.[3] Instead of the canonical (ρ, z) coordinates of Weyl we use prolate spheroidal coordinates (x, y) defined by

$$\rho = [(x^2 - 1)(1 - y^2)]^{1/2}, \qquad z = xy. \tag{2}$$

Then the basic field equation (1) assumes the highly symmetrical form

$$\frac{\xi_0\xi_0^* - 1}{x^2 - y^2}\left(\frac{\partial}{\partial x}(x^2 - 1)\frac{\partial\xi_0}{\partial x} + \frac{\partial}{\partial y}(1 - y^2)\frac{\partial\xi_0}{\partial y}\right) = \frac{2\xi_0^*}{x^2 - y^2}\left[(x^2 - 1)\left(\frac{\partial\xi_0}{\partial x}\right)^2 + (1 - y^2)\left(\frac{\partial\xi_0}{\partial y}\right)^2\right]. \tag{3}$$

The fairly obvious solution

$$\xi_0 = x\cos\lambda - iy\sin\lambda \tag{4}$$

may be used to generate the Kerr metric (or indeed the charged Kerr metric), while the less obvious solution

$$\xi_0 = \frac{x^4\cos^2\lambda + y^2\sin^2\lambda - 2ixy(x^2 - y^2)\cos\lambda\sin\lambda - 1}{2x(x^2 - 1)\cos\lambda - 2iy(1 - y^2)\sin\lambda} \tag{5}$$

gives rise to the simplest of the new metrics published by Tomimatsu and Sato.

We have utilized the same complex potential (5) to generate a new solution of the coupled Einstein-Maxwell equations, corresponding to a uniformly rotating axially symmetric *charged* source.[4] All of the basic equations which we employed in our analysis will be written in terms of a covariant differential operator "thop" similar (but not identical) to that introduced by Newman and Penrose in another context.[5] This simplifies immensely the task of generating from the complex potential the metric tensor components, the spin coefficients, and the Weyl tensor components, when a coordinate system other than the Weyl canonical coordinate system is employed.

It is hoped that the availability of this concise formalism will encourage others to pursue the quest for physically pertinent solutions of the basic field equation (3). In particular, it is certain that there exist other rational functions $\xi_0(x, y)$ which satisfy this field equation.

II. NULL TETRAD FORMALISM

As in so many other applications of general relativity, a null tetrad formalism is advantageous in dealing with problems involving stationary axially symmetric fields. The symbols $\underset{\sim}{k}, \underset{\sim}{m}, \underset{\sim}{t}$, and $\underset{\sim}{t}^*$ will denote two real and two complex null vector fields, among which the only non-vanishing inner products are[6]

$$\underset{\sim}{k} \rceil \underset{\sim}{m} = \underset{\sim}{t} \rceil \underset{\sim}{t}^* = 1. \tag{6}$$

The corresponding differential forms will be denoted by the symbols k, m, t and t^*, while the non-vanishing inner products will be

$$k \lceil m = t \lceil t^* = 1. \tag{7}$$

A basis for 2-forms can be constructed by forming exterior products of the basic 1-forms. In particular, we shall employ the basic 2-forms

$$B_+ = kt, \qquad B_0 = km + tt^*, \qquad \text{and} \qquad B_- = mt^*, \tag{8}$$

all of which correspond to the same eigenvalue of the duality operator. (The symbol $\wedge$, which is so often used in connection with exterior calculus, will be suppressed between differential forms).

The 1-forms defined by

$$u = d\underset{\sim}{k} \rceil \underset{\sim}{m} + d\underset{\sim}{t} \rceil \underset{\sim}{t}^*, \qquad v = d\underset{\sim}{k} \rceil \underset{\sim}{t}, \qquad \text{and } w = d\underset{\sim}{m} \rceil \underset{\sim}{t}^* \tag{9}$$

are evaluated in practice by solving the equations

$$dB_+ = B_+u - B_0v, \tag{10a}$$

$$dB_0 = 2B_+w - 2B_-v, \tag{10b}$$

and

$$dB_- = B_0w - B_-u. \tag{10c}$$

A knowledge of u, v, and w is important both for considerations of the equations of motion and for proceed-

ing to the Riemann tensor. The twelve complex Newman-Penrose "spin-coefficients" are obtained from u, v, and w by projecting these 1-forms onto the basic 1-forms k, m, t, and t^*. For example, the spin coefficient commonly called the "shear" is given by

$$v_t = t \lrcorner v = t \lrcorner d\underline{k} \lrcorner \underline{t} = t^\nu k^\mu{}_{;\nu} t_\mu. \tag{11}$$

The Weyl conform tensor may be characterized completely in terms of five complex fields c_i ($i = -2, \ldots, +2$), where the index denotes the "spin-weight" of the field,[7] while the Ricci tensor may be characterized completely in terms of the Ricci scalar R and the null tetrad components of the "reduced Ricci tensor"

$$S_{\mu\nu} = R_{\mu\nu} - \tfrac{1}{4} g_{\mu\nu} \mathrm{R}. \tag{12}$$

In practice, all of these fields may be evaluated by observing that

$$dv + vu = c_2 B_- + c_1 B_0 + (c_0 + \mathrm{R}/12) B_+ + \tfrac{1}{2} S_{kk} B_-{}^* + \tfrac{1}{2} S_{kt} B_0{}^* + \tfrac{1}{2} S_{tt} B_+{}^*, \tag{13a}$$

$$du - 2wv = -2[c_1 B_- + (c_0 - \mathrm{R}/24) B_0 + c_{-1} B_+ + \tfrac{1}{2} S_{kt^*} B_-{}^* + \tfrac{1}{2} S_{tt^*} B_0 - \tfrac{1}{2} S_{mt} B_+{}^*], \tag{13b}$$

$$dw - wu = (c_0 + \mathrm{R}/12) B_- + c_{-1} B_0 + c_{-2} B_+ + \tfrac{1}{2} S_{t^*t^*} B_-{}^* - \tfrac{1}{2} S_{mt^*} B_0{}^* + \tfrac{1}{2} S_{mm} B_+{}^*, \tag{13c}$$

where $S_{kt} = k^\mu S_{\mu\nu} t^\nu$, etc.

III. STATIONARY AXIALLY SYMMETRIC SPACE-TIMES

In the case of a stationary axially symmetric space-time under a wide variety of circumstances coordinates ρ, z, ϕ, and T may be introduced so that the 1-forms

$$\sigma^1 = f^{-1/2} P^{-1} d\rho, \quad \sigma^2 = f^{-1/2} P^{-1} dz, \quad \sigma^3 = f^{-1/2} R d\phi, \quad \sigma^4 = f^{1/2}(dT - \omega d\phi) \tag{14}$$

constitute an orthonormal tetrad system.[8] Here f, ω, P, and R are real functions of ρ and z only. Our null tetrad system will in turn be defined by

$$t = (1/\sqrt{2})(\sigma^1 + i\sigma^2), \quad t^* = (1/\sqrt{2})(\sigma^1 - i\sigma^2), \quad k = (1/\sqrt{2})(\sigma^3 - \sigma^4), \quad m = (1/\sqrt{2})(\sigma^3 + \sigma^4). \tag{15}$$

A fairly simple calculation involving Eqs. (10) yields six nonvanishing spin coefficients, which we here express in manifestly covariant form:[9]

$$u_t = -(1/\sqrt{2}) f^{1/2}\, \eth\, (\ln P) - (1/2\sqrt{2}) f^{-1/2} G_+, \tag{16a}$$

$$u_{t^*} = (1/\sqrt{2}) f^{1/2} \eth^* (\ln P) + (1/2\sqrt{2}) f^{-1/2} G_-, \tag{16b}$$

$$v_k = -(1/2\sqrt{2}) f^{1/2} R^{-1} \eth R + (1/2\sqrt{2}) f^{-1/2} G_+, \tag{16c}$$

$$v_m = -(1/2\sqrt{2}) f^{1/2} R^{-1} \eth R, \tag{16d}$$

$$w_k = -(1/2\sqrt{2}) f^{1/2} R^{-1} \eth^* R, \tag{16e}$$

$$w_m = -(1/2\sqrt{2}) f^{1/2} R^{-1} \eth^* R + (1/2\sqrt{2}) f^{-1/2} G_-. \tag{16f}$$

The quantities

$$G_+ = \eth f - R^{-1} f^2 \eth \omega, \quad G_- = \eth^* f + R^{-1} f^2 \eth^* \omega \tag{17}$$

have, respectively, spin weight plus one and minus one. The operators $\eth$ and $\eth^*$ are defined as follows for a field η of spin weight s:

$$\eth \eta = P^{1-s} \nabla (P^s \eta), \qquad \eth^* \eta = P^{1+s} \nabla^* (P^{-s} \eta), \tag{18}$$

where (in the ρ-z coordinate system)

$$\nabla = \frac{\partial}{\partial \rho} + i \frac{\partial}{\partial z}. \tag{19}$$

The operators $\eth$ and $\eth^*$ raise and lower spin weight by one, respectively.

Utilizing Eq. (13), one may show that in our case

$$S_{kt} = S_{kt^*} = S_{mt} = S_{mt^*} = 0, \tag{20}$$

while the algebraically independent nonvanishing components of the Ricci tensor are given by the following manifestly covariant expressions.

$$R^{-1} \eth^* \eth R = \tfrac{1}{2} f^{-1} (\mathrm{R} - 4S_{tt^*}), \tag{21a}$$

$$\eth G_- - \eth^* G_+ = 2(S_{kk} - S_{mm}), \tag{21b}$$

$$\tfrac{1}{2} f [R^{-1} \eth^* (R G_+) + R^{-1} \eth (R G_-)] - G_+ G_- = \tfrac{1}{2} f [\mathrm{R} - 4S_{tt^*} - 2(S_{kk} + S_{mm})], \tag{21c}$$

$$R^{-1} \eth \eth R + \tfrac{1}{2} f^{-2} G_+ G_-{}^* = 2 f^{-1} S_{tt}, \tag{21d}$$

$$\eth^* \eth (\ln P) - \tfrac{1}{8} f^{-2} (G_+ G_+{}^* + G_- G_-{}^*) = -\tfrac{1}{4} f^{-1} [\mathrm{R} + 4S_{tt^*} - 2(S_{kk} + S_{mm})]. \tag{21e}$$

The last equation is, however, derivable from the others by virtue of the Bianchi identities.

We also find that

$$c_1 = c_{-1} = 0, \tag{22}$$

while the nonvanishing components of the Weyl conform tensor are given by the following manifestly covariant expressions:

$$c_2 - \tfrac{1}{2} S_{tt} = -\tfrac{1}{8} [2 \eth G_+ + f^{-1} G_+^2], \tag{23a}$$

$$c_0 + \mathrm{R}/12 - \tfrac{1}{4} (S_{kk} + S_{mm}) = -\tfrac{1}{8} [-\eth^* G_+ - \eth G_- + f^{-1} G_+ G_-], \tag{23b}$$

$$c_{-2} - \tfrac{1}{2} S_{t^*t^*} = -\tfrac{1}{8} [2 \eth^* G_- + f^{-1} G_-^2]. \tag{23c}$$

For an axially symmetric stationary space-time described by Eq. (14), only three Petrov types are possible:

(1) Degenerate type N—here only c_2 or c_{-2} (but not both) are nonvanishing. This case is not particularly interesting if you seek fields which are asymptotically Schwarzschild.

(2) Degenerate type D—here either:
   (a) only c_0 is nonvanishing, or
   (b) $c_2 c_{-2} = 9 c_0^2$. The Kerr metric has this form of conform tensor when our null tetrad system is employed.

(3) Algebraically general space-times—the new Tomimatsu-Sato solution falls in this class. Hopefully, in the future more thought will be given to the refined classification of algebraically general space-times.

Equations (17), (21), and (23) may be written directly in terms of the prolate spheroidal coordinates defined by

the transformation (2). It suffices to note that in this system of units the differential operator ∇ becomes

$$\nabla = \sqrt{x^2-1}\,\frac{\partial}{\partial x} + i\sqrt{1-y^2}\,\frac{\partial}{\partial y}, \tag{24}$$

while the line element assumes the form

$$ds^2 = f^{-1}\left[P^{-2}\left(\frac{dx^2}{x^2-1} + \frac{dy^2}{1-y^2}\right) + R^2 d\phi^2\right] - f(dT - \omega d\phi)^2. \tag{25}$$

IV. COMPLEX POTENTIAL DESCRIPTION OF ELECTROVAC FIELDS

In those regions of space-time where the stress-energy is purely electromagnetic, one has

$$\mathbf{R} = 0 \quad \text{and} \quad S_{\mu\nu} = 8\pi T_{\mu\nu}, \tag{26}$$

where

$$T_{\mu\nu} = 1/4\pi[F_\mu{}^\alpha F_{\alpha\nu} + \tfrac{1}{4} g_{\mu\nu}(F^{\alpha\beta}F_{\alpha\beta})], \tag{27}$$

and the electromagnetic field tensor $F_{\mu\nu}$ is given in terms of the 4-potential A_μ by

$$F_{\mu\nu} = \frac{\partial}{\partial x^\mu} A_\nu - \frac{\partial}{\partial x^\nu} A_\mu. \tag{28}$$

Our attention will be directed toward static fields for which $A_1 = A_2 = 0$ and for which A_3 and A_4 are functions of ρ and z alone. In this case it is advantageous to introduce a magnetic scalar potential A_3' such that

$$R^{-1}(\eth A_3 + \omega \eth A_4) = if^{-1}\eth A_3'. \tag{29}$$

In fact, the electromagnetic field tensor and the stress-energy tensor may be expressed entirely in terms of the complex potential

$$\Phi = A_4 + iA_3'. \tag{30}$$

In particular, the nonvanishing null tetrad components of the reduced Ricci tensor are given by

$$S_{kk} = -(\eth\Phi)(\eth^*\Phi^*), \tag{31a}$$

$$S_{mm} = -(\eth^*\Phi)(\eth\Phi^*), \tag{31b}$$

$$S_{tt} = (\eth\Phi)(\eth\Phi^*) = (S_{t^*t^*})^*. \tag{31c}$$

A "duality rotation" corresponds to a transformation $\Phi \to e^{i\alpha}\Phi$, under which the stress tensor and the gravitational field remain unchanged.

In view of the vanishing of $\mathbf{R} - 4S_{tt^*}$, Eq. (21a) implies that $R(\rho, z)$ is a harmonic function of ρ and z. As a result one may introduce the Weyl canonical coordinate system in which $R = \rho$. The fact that little progress has been made in the analysis of stationary axially symmetric interior solutions may be attributed to the nonvanishing of the right side of Eq. (21a) when matter is present.

Equation (21b) implies the existence of a complex scalar potential ϵ such that

$$G_+ = \eth\epsilon + 2\Phi^*\eth\Phi, \qquad G_- = \eth^*\epsilon + 2\Phi^*\eth^*\Phi. \tag{32}$$

Equation (21c), which may be cast into the form

$$f[\rho^{-1}\eth(\rho\eth^*\epsilon) + \rho^{-1}\eth^*(\rho\eth\epsilon)] = G_+\eth^*\epsilon + G_-\eth\epsilon, \tag{33a}$$

assumes with the Maxwell field equation

$$f[\rho^{-1}\eth(\rho\eth^*\Phi) + \rho^{-1}\eth^*(\rho\eth\Phi)] = G_+\eth^*\Phi + G_-\eth\Phi, \tag{33b}$$

the role of principal field equation in the complex potential formalism. The remaining gravitational field equation (21d), which in the present case assumes the form

$$\rho^{-1}\eth\eth\rho + \tfrac{1}{2} f^{-2} G_+ G_-^* = 2f^{-1}(\eth\Phi)(\eth\Phi^*), \tag{34}$$

may be employed in order to evaluate P (which appears in the definition of "thop") once Eqs. (33) have been solved for ϵ and Φ. Finally, comparing Eqs. (17) and (32), we find that f and ω may be evaluated using the equations

$$f = \mathrm{Re}\,\epsilon + \Phi^*\Phi, \tag{35a}$$

$$-\rho^{-1}f^2\eth\omega = i\eth(\mathrm{Im}\,\epsilon) + \Phi^*\eth\Phi - \Phi\eth\Phi^*. \tag{35b}$$

V. UTILITY OF THE ξ_0 POTENTIAL

As suggested in Paper II, we may consider electrovac fields for which ϵ is an analytic function of Φ. In this case we may write both ϵ and Φ in terms of a new complex potential ξ such that

$$\epsilon = (\xi - 1)/(\xi + 1) \quad \text{and} \quad \Phi = q/(\xi + 1), \tag{36}$$

where q is a constant, which we may temporarily regard as real, since the electromagnetic fields appropriate for complex values of q may be obtained later by a duality rotation.

The principal field equation (33) may now be written

$$(\xi_0^*\xi_0 - 1)[\rho^{-1}\eth^*(\rho\eth\xi_0) + \rho^{-1}\eth(\rho\eth^*\xi_0)] = 4\xi_0^*(\eth\xi_0)(\eth^*\xi_0), \tag{37}$$

where

$$\xi = \xi_0(1 - q^2)^{1/2}. \tag{38}$$

Since Eq. (37) makes no reference to q, we have reduced the electrovac problem to the vacuum problem.

We should now like to discuss certain general features of the ξ_0 potential and of the prolate spheroidal coordinate system, which make them particularly appropriate for the study of stationary axially symmetric fields.

The Schwarzschild solution corresponds to $\xi_0 = x$; so does the Reissner-Nordstrom solution. Since we wish to center our attention upon solutions of Einstein's field equations which are asymptotically Schwarzschild, we shall demand that, for large values of x, ξ_0 should behave as some real constant times x.

Equation (37), when expressed in terms of prolate spheroidal coordinates, assumes the form (3). It should be observed that if $\xi_0(x, y)$ is a solution of Eq. (3), then new solutions can be constructed by combinations of the following procedures:

(1) complex conjugation,
(2) replacement of ξ_0 by its reciprocal,
(3) multiplication by a constant number of modulus 1,
(4) substitution of $x \to -x$,
(5) substitution of $y \to -y$,
(6) interchanging x and y. (39)

Of course, these procedures can be expected to yield physically uninteresting solution most of the time. In order to select physically interesting solutions with a

plane of symmetry orthogonal to the axis of symmetry, we propose that the following additional requirements be imposed upon ξ_0:

$$\xi_0(-x,y) = -\xi_0^*(x,y), \tag{40a}$$

$$\xi_0(x,-y) = +\xi_0^*(x,y). \tag{40b}$$

The field equation (34) may be replaced by

$$\rho^{-1}\eth\eth\rho + [2/(\xi_0\xi_0^* - 1)^2](\eth\xi_0)(\eth\xi_0^*) = 0, \tag{41}$$

which implies immediately that the metric function P does not depend upon the choice of q. Once one has evaluated it for the vacuum field, one may carry it over for the charged version of the field.

The metric function

$$f = (\xi_0\xi_0^* - 1)^2/|\xi_0 + (1-q^2)^{-1/2}|^2 \tag{42}$$

is easily evaluated, so we need only consider the evaluation of the ω field.

If the field $\xi_0(x,y)$ satisfies the condition (40a), then from Eq. (35b) it may be shown that

$$(1-q^2)^{1/2}\,\mathrm{Even}(\omega) = \tfrac{1}{2}[(1-q^2)^{1/2} + (1-q^2)^{-1/2}]\,\mathrm{Even}(\omega_0), \tag{43a}$$

$$(1-q^2)^{1/2}\,\mathrm{Odd}(\omega) = \mathrm{Odd}(\omega_0), \tag{43b}$$

where "Even" and "Odd" refer to the parts of the function which are even and odd in x, respectively. Thus, once ω_0 has been constructed for the vacuum field, ω may be inferred immediately for the electrovac field. The function ω_0 is, however, governed by the differential equation

$$\rho^{-1}\eth\omega_0 = [(\xi_0+1)^2\eth\xi_0^* - (\xi_0^*+1)^2\eth\xi_0]/(\xi_0\xi_0^* - 1)^2. \tag{44}$$

VI. CONCLUSIONS

While the study of stationary axially symmetric fields remains considerably more difficult than the study of static axially symmetric fields initiated by H. Weyl in 1917, the complex potential formalism has at least isolated certain key aspects of the problem from the complexities of the four-dimensional space-time. One does not even have to understand relativity theory in order to attack the problem of finding new solutions of Eq. (3).

It is clear that an infinity of solutions remain to be discovered. If one introduces a new complex potential ψ such that

$$\xi_0 = -\coth\psi, \tag{45}$$

then this field satisfies the equation

$$\nabla^2\psi = 2i\tan(2\,\mathrm{Im}\psi)\nabla\psi\cdot\nabla\psi, \tag{46}$$

where we have reverted to the notation employed in the Introduction. In the static case this reduces to Laplace's equation, so that the general solution can be found easily. In our case Eq. (46) provides a convenient vehicle for the application of approximation techniques.

For fields satisfying conditions (40) we infer that

$$\psi(-x,y) = -\psi^*(x,y), \tag{47a}$$

$$\psi(x,-y) = +\psi^*(x,y). \tag{47b}$$

The solutions of Eq. (46) having these attributes may be constructed explicitly to any desired order in $1/x$. Thus, one finds

$$\psi(x,y) = -x^{-1}\sum_{k=0}^{\infty}\sum_{l=0}^{k} Q_{kl}P_l(y)(ix^{-1})^k, \tag{48}$$

where the real constants Q_{kk} are *arbitrary* and the real constants Q_{kl} for $l<k$ are *determined* once $Q_{00}, Q_{11}, \ldots, Q_{kk}$ are specified. In particular, $Q_{kl} = 0$ if $k-l$ is odd. The free parameters Q_{kk} play a role similar to multipole moments in Newtonian gravitation theory.

A "gold ring" should be earned by the person who does for complex ψ what is so simple for real ψ; namely, construct the general solution of Eq. (46). While this prize may continue to be ellusive, because of the absence of a linear superposition principle, a "silver ring" might be merited for the discovery of the general relation between the values of the constants Q_{kk}, which govern the asymptotic structure of the field, and something more directly associated with the structure of the source. What immediately comes to mind is the infinite red shift surface, upon which $|\xi_0| = 1$. One might consider the shape of the infinite red shift surface, and the phase of ξ_0 thereon, as descriptive of the structure of the source. The objective of this approach would be to relate this structure to the values of the parameters Q_{kk}. In this manner perhaps one could circumvent the necessity of having exact solutions of the vacuum field equations.

In the meanwhile we hope that more people will be encouraged to search for specific exact solutions, a task which Tomimatsu and Sato have shown is not quite impossible. The next "bronze ring" should be awarded for the discovery of an exact solution outside the Tomimatsu-Sato class.

ACKNOWLEDGMENTS

The author is very grateful to A. Tomimatsu and H. Sato for sending him preprints concerning their exciting new discovery. In addition, he remains indebted to his colleague, I. Hauser, for so much of the exterior algebraic formalism used in this research.

*Work supported in part by National Science Foundation Grant GP-8553.

[1] R. P. Kerr, Phys. Rev. Lett. **11**, 237 (1963).

[2] A. Tomimatsu and H. Sato, Phys. Rev. Lett. **29**, 1344 (1972). A. Tomimatsu and H. Sato (to be published).

[3] F. J. Ernst, Phys. Rev. **167**, 1175 (1968), Paper I. F. J. Ernst, Phys. Rev. **168**, 1415 (1968), Paper II. The latter paper contains an unusual number of sign anomalies, which should be modified to correspond to the present paper. Independently a complex potential for vacuum fields was introduced by D. Kramer and G. Neugebauer, Commun. Math. Phys. **10**, 132 (1968).

[4] F. J. Ernst, Phys. Rev. D 7, 2520 (1973).

[5] E. Newman and R. Penrose, J. Math. Phys. **7**, 863 (1966). Whereas their "thop" was defined with reference to the unit sphere, ours is defined with reference to a more general conformally flat geometry.

[6] We employ throughout this section notation developed by I. Hauser.

[7] E. Newman and R. Penrose, J. Math. Phys. **3**, 566 (1962). Our $c_{-2},\ldots, c_2$ corresponds to their $\psi_4,\ldots,\psi_0$, respectively.

[8] H. Weyl, Ann. Phys. (Leipz.) **54**, 117 (1917). T. Lewis, Proc. R. Soc. A **136**, 176 (1932). A. Papapetrou, Ann. Phys. (Leipz.) **12**, 309 (1953). A. Papapetrou, Ann. Inst. Henri Poincaré **4**, 83 (1966). W. Kundt and M. Trümper, Z. Phys. **192**, 419 (1966). B. Carter, J. Math. Phys. **10**, 70 (1969).

[9] We refer here and elsewhere to convariance under analytic transformations of the complex coordinate $\xi = \rho + iz$, coupled with null rotations about the B_0 axis; viz. $t \to te^{ia}$.

Weak gravitational fields

D. Lerner and J. R. Porter

Department of Mathematics, University of Pittsburgh, Pittsburgh, Pennsylvania 15213
(Received 12 February 1973)

We consider the set of C^k bounded tensor fields of type (r,s) on $\mathbf{R}^4$ in the topology of uniform C^k convergence. For each $k \geq 2$, the map sending a metric to its curvature tensor is shown to be analytic at the Minkowski metric. The same is true of the map sending a metric to its Einstein tensor. The well-known linearized theory of gravitation amounts to studying the directional derivatives of these maps. An iterative method for solving the full field equations along an analytic curve of Einstein tensors passing through zero is proposed.

I. INTRODUCTION

A central problem in the general theory of relativity concerns the stability of solutions to Einstein's field equations. Precisely, given a four-manifold M, a stress-energy tensor T, and an exact solution g to the field equations $E(g) = -T$, the problem is to determine all "nearby" solutions and to examine, at least qualitatively, their physical properties. ($E(g) = \{R_{ab} - \frac{1}{2}Rg_{ab}\}dx^a \otimes dx^b$ is the Einstein tensor of the metric g. The map $g \to E(g)$ is called the Einstein map.) There are essentially two approaches to the problem, depending on what one means by the word "nearby."

(a) In the first instance, one considers all metrics g' which are in some sense close to g, computes the energy–momentum tensors $-E(g')$, and examines the physical properties of the resulting space–times (M, g'). One normally requires the introduction of a topology on the set of Lorentz metrics in order to determine whether or not two metrics are close to one another.

(b) In the second instance, one perturbs the energy–momentum tensor T to a nearby T' and attempts to solve the resulting field equations $E(g') = -T'$.

In connection with (a) if one regards all Lorentz metrics on M as being on an equal (mathematical footing, it appears[1] that the only acceptable choice for a topology is the Whitney fine C^k topology. However, it frequently happens that one is *not* concerned with all such metrics, but only those g' which are in some sense close to a *fixed* metric g. In such cases, it is possible to construct a topology which is considerably more tractable than the Whitney topology and at the same time appears to provide a suitable analytic framework within which to attack problem (b).

In this paper, we examine such a topology in the particular case where $M = \mathbb{R}^4$ and the preferred metric is a fixed Minkowski metric η. Section II introduces the necessary mathematical formalism; the set of Lorentz metrics close to η is shown to be an open subset of a Banach space. In Sec. III we show that the curvature map (the map associating with each Lorentz metric its Riemann tensor) is analytic in a neighborhood of η. [The metrics themselves need only be C^k $(k \geq 2)$.] It follows immediately that the Einstein map $g \to E(g)$ is analytic at η. In Sec. IV we briefly discuss the linearized theory of gravitation, which is particularly well-posed in this formalism: The linearized Einstein tensor of the metric $\eta + h$ is simply the derivative of E at η in the direction of h. In Sec. V we discuss an iterative procedure for solving the full field equations along an analytic curve of stress–energy tensors passing through zero.

II. MATHEMATICAL PRELIMINARIES

Fix, once and for all, a global coordinate system (x^a) on $\mathbb{R}^4$ and the Minkowski metric η defined by these coordinates, $\eta_{ab} = \text{diag}\{1, -1, -1, -1\}$. Let $\mathcal{S}_k$ denote the set of C^k twice-covariant symmetric tensor fields on $\mathbb{R}^4$, and for $h \in \mathcal{S}_k$, $x \in \mathbb{R}^4$, put

$$\|h(x)\|_k = \max_{1 \leq a,b,\ldots, \leq 4} \{|h_{ab}(x)|, |h_{cd,e}(x)|, \ldots, |h_{ij,m_1,m_2\cdots m_k}(x)|\}, \tag{1}$$

and set

$$|h|_k = \sup\{\|h(x)\|_k : x \in \mathbb{R}^4\}. \tag{2}$$

Define

$$\mathcal{B}_k \equiv \{h \in \mathcal{S}_k : |h|_k < \infty\}. \tag{3}$$

The $|\cdot|_k$ norm is easily seen to be equivalent to the standard C^k norm; this particular formulation is slightly easier to calculate with. $\mathcal{B}_k$ is a Banach space. Similarly, let $\mathcal{U}_k$ denote the set of four-covariant C^k tensor fields on $\mathbb{R}^4$ having the symmetries of curvature tensors ($R_{[ab][cd]} = R_{abcd}$, $R_{a[bcd]} = 0$). For $R \in \mathcal{U}_k$, define $|R|_k$ as above and let

$$\mathcal{W}_k \equiv \{R \in \mathcal{U}_k : |R|_k < \infty\}. \tag{4}$$

$\mathcal{W}_k$ is a Banach space as well. Notice that $\eta \in \mathcal{B}_k$ and that the ball of radius 1/4 about η consists entirely of Lorentz metrics; it is these which we shall call "close" to η. Thus we are concerned with an open ball in a Banach space. {Notice that the *complete* set of Lorentz metrics contained in $\mathcal{B}_k$ is not an open set; for example, $[1/(1 + r^2)]\eta[r^2 = \sum_a (x^a)^2]$ is not an interior point. This would be a real problem if we were interested in *all* Lorentz metrics.}

III. ANALYTICITY OF THE CURVATURE MAP

Let Ω be the map sending a nondegenerate C^k metric to its C^{k-2} curvature tensor. As mentioned above, the domain of Ω contains an open ball around η in $\mathcal{B}_k$.

Theorem: For any $k \geq 2$, the map $\Omega: \mathcal{B}_k \to \mathcal{W}_{k-2}$ is analytic at η. Precisely, for any g in the ball of radius 1/4 about η, write $g = \eta + h$ where $|h|_k < 1/4$; then

$$\Omega(g) = \Omega(\eta + h) = \Omega(\eta) + \sum_{j=1}^{\infty} \frac{1}{j!} D^j\Omega(\eta)\cdot(h, \ldots, h), \qquad j \text{ times}, \tag{5}$$

where, as usual,

$$D^j\Omega(\eta)\cdot(h, \ldots, h) = \frac{d}{dt_j} \cdots \frac{d}{dt_1} \Big|_{\substack{t_1 = \cdots = t_j = 0 \\ h_1 = \cdots = h_j = h}} \{R_{abcd}(\eta + t_1 h_1 + \cdots + t_j h_j)\, dx^a \otimes dx^b \otimes dx^c \otimes dx^d\}. \tag{6}$$

The series on the right converges in norm in the space $\mathfrak{W}_{k-2}$.

Proof: We exhibit the power series for $\Omega(\eta + h)$ and show that it converges to $\Omega(\eta + h)$. It is necessary to work in components; all raising and lowering of indices is done with η and the summation convention is employed throughout. We have $g = \eta + h$, where $|h|_k = a < 1/4$. By long division, the components of the inverse matrix to g are

$$g^{cd} = \eta^{cd} - h^{cd} + \sum_{j=1}^{\infty} (-1)^{j+1} h^{ci_1} h_{i_1}{}^{i_2} \cdots h_{i_j}{}^{d}. \tag{7}$$

This is a series of real-valued functions on $\mathbb{R}^4$; we need to show uniform C^k convergence. Put $b = 4a < 1$, and differentiate the series n times ($0 \le n \le k$). One finds without difficulty that, for any $x \in \mathbb{R}^4$,

$$|(h^{ci_1} h_{i_1}{}^{i_2} h_{i_2}{}^{i_3} \cdots h_{i_j}{}^{d}), a_1 a_2 \cdots a_n(x)| < (j+1)^n b^{j+1}. \tag{8}$$

Since $\sum_{j=0}^{\infty} (j+1)^n b^{j+1} < \infty$ for $b < 1$ (ratio test), all the series for $g^{cd}, \ldots, g^{cd}, a_1 \cdots a_k$ converge uniformly and absolutely on $\mathbb{R}^4$ (Weierstrass test); and in the notation of Sec. II we have

$$|g^{-1}|_k < |\eta^{-1}|_k + \sum_{j=0}^{\infty} (j+1)^k b^{j+1} = 1 + \sum_{j=0}^{\infty} (j+1)^k b^{j+1} < \infty.$$

So g^{-1} is well defined.

Let $\Gamma^a_{bc}(g)$ be the Christoffel symbols of g with respect to (x^a). Setting $H_{dbc} = \frac{1}{2}\{h_{db,c} + h_{dc,b} - h_{bc,d}\}$, we have

$$\Gamma^a_{bc}(g) = \eta^{ad} H_{dbc} - h^{ad} H_{dbc} + h^{ai} h_i{}^d H_{dbc} - + \cdots \tag{9}$$

with absolute and uniform C^{k-1} convergence. Thus $\Omega(g) = R_{abcd}(g) dx^a \otimes dx^b \otimes dx^c \otimes dx^d$, where

$$R_{abcd}(g) = 2H_{ab[c,d]} + 2H_{as[c}\Gamma^s_{d]b}(g), \tag{10}$$

and we may expand and regroup in the following way:

$$R_{abcd}(g) = 2H_{ab[c,d]} + 2H_{as[c}H_{|e|d]b}\eta^{se} - 2H_{as[c}H_{|e|d]b}h^{se} + 2H_{as[c}H_{|e|d]b}h^{si}h_i{}^e - \cdots, \tag{11}$$

where we have convergence in the space $\mathfrak{W}_{k-2}$, with $H_{as[c}H_{|e|d]b} = \frac{1}{2}(H_{asc}H_{edb} - H_{asd}H_{ecb})$.

Remark: Because of the absolute and uniform convergence, it follows that the series for $\mathrm{Ric}(g) = R_{cabd} g^{-1\,cd} dx^a \otimes dx^b = R_{ab} dx^a \otimes dx^b$ and $R(g) = R_{ab} g^{ab}$ are also convergent. From this it follows immediately that the map E: $\mathfrak{G}_k \to \mathfrak{G}_{k-2}$ sending a Lorentz metric to its Einstein tensor is also analytic at η in the ball of radius 1/4. Similar remarks apply to the map sending a Lorentz metric to its conformal curvature tensor.

IV. THE LINEARIZED THEORY OF GRAVITATION

The best-known method for obtaining approximate solutions to the field equations is called the linearized theory (see Pirani, Ref. 2, for a fairly complete exposition and references). It has often been remarked that it is not a particularly good method, and in this section we shall see precisely why this is so. The linearized theory proceeds roughly as follows. An energy–momentum tensor T is given; instead of solving the full equations $E(g) = -T$, one replaces E by a linear operator L and considers the simpler equations $L(g) = -T$. $L(g)$ is defined simply by writing $g = \eta + h$, calculating $E(\eta + h)$, and retaining only those terms which are first order in h. The resulting linear system is then solved for h, and one obtains the approximate solution $g = \eta + h$.

Of course, if one now calculates the full Einstein tensor $E(\eta + h)$ for this metric, it will not be equal to $-T$. However, there is a fairly obvious relation between the two quantities, namely

$$DE(\eta)\cdot h = -T. \tag{12}$$

This should be evident from the remarks in the preceding section; $DE(\eta)\cdot h$ is just the first term in the power series expansion of $E(\eta + h)$. In words, the linearized Einstein tensor is the derivative of the Einstein map at η in the direction of h. Similarly, the first term $D\Omega(\eta)\cdot h$ in the series (5) or (11) is just the usual linearized curvature tensor of the metric $\eta + h$.

Once it is recast in this formalism, the shortcomings of the linearized theory are readily apparent. The relationship between $\eta + h$ and an exact solution to $E(g) = -T$ is essentially nonexistent. What we have instead is

$$E(\eta + h) + T = \sum_{k=2}^{\infty} \frac{1}{k!} D^k E(\eta)\cdot h^k; \tag{13}$$

a real solution (if it exists) to $E(g) = -T$ is well approximated by the linearized solution only in the case that the entire power series on the right can be neglected.

V. AN ITERATIVE METHOD FOR SOLVING THE FIELD EQUATIONS

Consider a curve of the form

$$g(t) = \eta + \sum_{i=1}^{\infty} \underset{(i)}{h} \frac{t^i}{i!}, \tag{14}$$

where, for the sake of definiteness, $|\underset{(i)}{h}|_k < (\frac{1}{4})^{i+1}$. Then for $t \in (-1, 1)$, this defines an analytic curve of metrics passing through η and lying in the ball of radius 1/4 about η in $\mathfrak{G}_k$. The image of this curve under the Einstein map will be an analytic curve passing through 0 in $\mathfrak{G}_{k-2}$. Setting

$$H(t) = \sum_{i=1}^{\infty} \underset{(i)}{h} \frac{t^i}{i!},$$

we have

$$E(g(t)) = DE(\eta)\cdot H(t) + (1/2!)\, D^2E(\eta)\cdot(H(t), H(t)) + (1/3!)\, D^3E(\eta)\cdot(H(t), H(t), H(t)) + \cdots. \tag{15}$$

Expanding and regrouping according to powers of t, we have

$$E(g(t)) = \{DE(\eta)\cdot(\underset{(1)}{h})\}t + \{DE(\eta)(\underset{(2)}{h}) + D^2E(\eta)\cdot(\underset{(1)}{h}, \underset{(1)}{h})\}t^2/2! + \cdots$$

I II

$$+\{DE(\eta)(\underset{(3)}{h}) + 3D^2E(\eta)(\underset{(1)}{h}, \underset{(2)}{h}) + D^3E(\eta)(\underset{(1)}{h}, \underset{(1)}{h}, \underset{(1)}{h})\}$$

$$\times t^3/3! + \cdots. \qquad (16)$$

Now conversely, suppose we are given an analytic curve

$$T(t) = \sum_{i=1}^{\infty} \underset{(i)}{T} \frac{t^i}{i!}$$

of stress–energy tensors with $T(0) = 0$. Then we can try to find a solution curve of the form (14). According to (16), the equations to be solved are then (in order)

$$\text{I: } DE(\eta)\cdot(\underset{(1)}{h}) = -\underset{(1)}{T}, \quad \text{for } \underset{(1)}{h},$$

$$\text{II: } DE(\eta)\cdot(\underset{(2)}{h}) = -\underset{(2)}{T} - D^2E(\eta)(\underset{(1)}{h}, \underset{(1)}{h}), \quad \text{for } \underset{(2)}{h},$$

$$\text{III: } DE(\eta)(\underset{(3)}{h}) = -\underset{(3)}{T} - 3D^2E(\eta)(\underset{(1)}{h}, \underset{(2)}{h}) - D^3E(\eta)(\underset{(1)}{h}, \underset{(1)}{h}, \underset{(1)}{h}), \quad \text{for } \underset{(3)}{h},$$

$$\cdots \text{etc.} \qquad (17)$$

It should be noted that at each stage of the iteration process, one has only to solve a linear equation, which is, in principle, possible.

[1] D. Lerner, Comm. Math. Phys. (to be published).
[2] F. A. E. Pirani, in *Lectures on General Relativity*, 1964 Brandeis Summer Institute in Theoretical Physics, Vol. 1 (Prentice-Hall, Englewood Cliffs, N. J., 1965).

Asymptotically simple space-time manifolds

D. Lerner and J. R. Porter

Department of Mathematics, University of Pittsburgh, Pittsburgh, Pennsylvania 15213
(Received 12 February 1973)

Asymptotic simplicity is shown to be k-stable ($k \geq 3$) at any Minkowski metric on $\mathbb{R}^4$ in both the Whitney fine C^k topology and a coarser topology (in which the C^k twice-convariant symmetric tensors form a Banach manifold whose connected components consist of tensor field asymptotic to one another at null infinity). This result, together with a sequential method for solving the field equations previously proposed by the authors, allows a fairly straightforward proof that a well-known result in the linearized theory holds in the full nonlinear theory as well: There are no nontrivial (i.e., non-Minkowskian) asymptotically simple vacuum metrics on $\mathbb{R}^4$ whose conformal curvature tensors result from prescribing zero initial data on past null infinity.

I. INTRODUCTION

The concept of asymptotically simple space-time manifolds, introduced by Penrose,[1,2] is a fruitful one in the study of asymptotic conditions in general relativity and one which Penrose has used to good advantage.[3] One would like to have more examples of asymptotically simple space-times than the single example now known, namely Minkowski space-time. In this paper it is shown that there are many asymptotically simple space-times; in fact, there is an open neighborhood of any Minkowski metric on $\mathbb{R}^4$ in the Whitney fine C^k topology ($k \geq 3$) on the set of Lorentz metrics on $\mathbb{R}^4$ all of whose elements are asymptotically simple metrics. Using this result and a formulation for weak gravitational fields developed by the authors,[4] we show that a certain linearized solution to the vacuum field equations has an exact counterpart in the full nonlinear theory. Section II deals with definitions and preliminaries; Sec. III gives a proof of the asserted result for asymptotically simple space-times; Sec. IV extends the known linear result to the full theory; and Sec. V gives some concluding remarks and conjectures.

II. MATHEMATICAL PRELIMINARIES

A space-time manifold is a pair (M, g) where M is a four-dimensional C^∞ differentiable manifold and g is a C^3 pseudo-Riemanian metric on M of signature -2 which is time-oriented and possesses no closed timelike C^1 curves. A space-time manifold (M, g) is said to be *asymptotically simple* if there exists a space-time manifold $(\tilde{M}, \tilde{g})$ with boundary whose interior is conformal to (M, g) with $\tilde{g} = \Omega^2 g$, $\Omega > 0$, which satisfies:

(1) $\tilde{M}$ is a C^4 differentiable manifold with boundary $\mathscr{I}$ and $\tilde{g}$ is a C^3 pseudo-Riemannian metric on m,

(2) Ω is C^3 on $\tilde{M}$, $\Omega = 0$ on $\mathscr{I}$, and $d\Omega \neq 0$ on $\mathscr{I}$,

(3) Every maximally extended null geodesic in the interior of M^2 intersects $\mathscr{I}$ in precisely two points.[1]

Minkowski space-time $(\mathbb{R}^4, \eta)$ is an example of an asymptotically simple space-time. The manifold $(\tilde{M}, \tilde{\eta})$ whose interior is conformal to $(\mathbb{R}^4, \eta)$ is obtained by constructing the conformally-compactified Minkowski space-time[5,6] and then slicing this manifold apart along the light cone at infinity. The result is represented pictorially in Fig. 1. The points I^+, I^0, and I^- are not points of $\mathscr{I}$.

An exposition on the Whitney topologies applied to problems in general relativity is given by Lerner,[7] and only a brief summary will be given here. Given an arbitrary Riemannian metric μ on a differentiable manifold M, a W_0 (Whitney fine C^0) neighborhood base of a

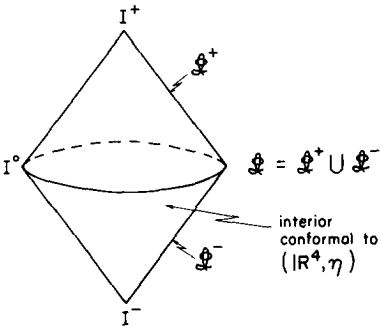


FIG. 1

C^0 tensor field $t \in \Gamma^0(T^r_s(M))$ is given by sets of the form

$$W_\mu(t, \epsilon(x)) = \{s \mid s \in \Gamma^0(T^r_s(M)) \text{ and } \|s - t\|_\mu(x) < \epsilon(x), \forall\, x \in M\}.$$

Here $\epsilon(x)$ is any positive continuous function on M and $\|\cdot\|_\mu(x)$ is the μ norm in $T^r_s M_x$. A W_k (Whitney fine C^k) neighborhood base ($k \geq 0$) of a tensors field $t \in \Gamma^k(T^r_s(M))$ is given by sets of the form

$$W^k_\mu(t, \epsilon(x)) = \{s \mid s \in \Gamma^k(T^r_s(M)) \text{ and } \|s - t\|_\mu(x) < \epsilon(x),$$

$$\|\nabla(s - t)\|_\mu(x) < \epsilon(x), \ldots, \|\nabla^k_s(s - t)\|_\mu(x) < \epsilon(x), \forall\, x \in M\}.$$

where ∇^i_s denotes the totally symmetrized ith covariant derivative with respect to the Riemannian metric μ. This gives a convenient description of the W^k topologies (the Whitney fine C^k topologies); an altogether equivalent formulation, which is manifestly independent of μ and perhaps more physically intuitive as well, is given in Ref. 7.

The set of all C^k Lorentz metrics (pseudo-Riemannian metric tensor fields of signature -2) on M will be denoted by $L^k(M)$. A property P on $\Gamma^k(T^r_s(M))$ is said to be k-stable at a tensor $t \in \Gamma^k(T^r_s(M))$ if there exists an open W^k neighborhood of t all the elements of which possess property P. For example, geodesic completeness is k-stable, $k \geq 2$, on $L^k(M)$ as is time orientability. The purpose of Sec. III is to prove that asymptotic simplicity is k-stable, $k \geq 3$, at η in L^k.

Another topology is used on $L^k(\mathbb{R}^4)$ in Sec. IV which was introduced by Lerner and Porter.[4] Given η and a Minkowski coordinate systems $\{x^i\}$ on $\mathbb{R}^4$ in which $\eta = ds^2 = (dx^0)^2 - (dx^1)^2 - (dx^2)^2 - (dx^3)^2$, for any $t \in \Gamma^k(T^r_s(\mathbb{R}^4))$ set

$$\|t(x)\|_k = \max_{\text{components}} \{|t^{\cdots}_{\cdots}(x)|, |t^{\cdots}_{\cdots,\cdot}(x)|, \ldots, |t^{\cdots}_{\cdots,\cdot\cdots\cdot_k}(x)|\}$$

where $t_{;\cdots}(x)$ stands for the components of t in $\{x^i\}$ and $,\cdots$ denotes coordinate derivatives with respect to $\{x^i\}$. Let

$$|t|_k = \sup\{\|t(x)\|_k \mid x \in \mathbb{R}^4\}.$$

Define

$$B_k^{(r,s)} = \{t \in \Gamma^k(T_s^r) \mid |t|_k < \infty\}.$$

Then $\langle B_k^{(r,s)}, |\cdot|_k\rangle$ is a Banach space; the topology is equivalent to that of uniform C^k convergence.

Given a particular Minkowski metric η on $\mathbb{R}^4$ and the conformal factor Ω mentioned in the definition of asymptotic simplicity, one can define the set of tensor fields of type (r, s) *asymptotic to zero at null infinity:*

$$A_k^{(r,s)} \equiv \{t \in B_k^{(r,s)}: \Omega^2 t \text{ extends to a } C^k \text{ tensor field on } \bar{M} \text{ with } \Omega^2 t = 0 \text{ on } \mathcal{I}\}.$$

[The conformal factor (Ω^2) used in the definition of $A_k^{(r,s)}$ for certain applications may be changed to another power of Ω. The factor Ω^2 is the correct one to use for metrics asymptotic to η in $B^{(0,2)}$.] $A_k^{(r,s)}$ is a closed subspace of $B_k^{(r,s)}$ and thus a Banach space in its own right. The set $\Gamma^k(T_s^r(\mathbb{R}^4)$ is made into a Banach manifold by taking the sets

$$\{U_\alpha \equiv \alpha + A_k^{(r,s)}: \alpha \in \Gamma^k(T_s^r(\mathbb{R}^4))\}$$

for an (analytic) atlas. Two tensor fields $\alpha, \beta \in \Gamma^k(T_s^r(\mathbb{R}^4))$ lie in the same connected component in this topology iff $\alpha - \beta$ is asymptotic to zero at null infinity. The fields α and β are then said to be asymptotic at infinity. This topology is called the A^k topology on tensor fields. It is clear that the W^k topology is finer than the A^k topology (any open set in the A^k topology is open in the W^k topology as well).

III. STABILITY AND ASYMPTOTIC SIMPLICITY

Theorem 1: Asymptotic simplicity is a k-stable property, $k \geq 3$, in $L^k(\mathbb{R}^4)$ at any Minkowski metric η on $\mathbb{R}^4$.

Proof: It is necessary to exhibit an open W^k neighborhood, $k \geq 3$, of η in $L^k(\mathbb{R}^4)$, the Lorentz metrics in which are all asymptotically simple. First, there exists an open W^k neighborhood U_1 in $\bar{M} - \mathcal{I}$ of $\tilde{\eta}$, all the Lorentz metrics in which are equal to $\tilde{\eta}$ on $\mathcal{I}$. Since $\bar{M} - \mathcal{I}$ is differomorphic to $\mathbb{R}^4$ by virtue of the conformal relatedness, an open W^k neighborhood U_1 of η in $\mathbb{R}^4$ is obtained. All the Lorentz metrics in U_1 are asymptotic to the Minkowski metric η. The Lorentz metrics in U_1 satisfy the first two defining properties of asymptotic simplicity (using the same $\bar{M}$ and Ω as for $(\mathbb{R}^4, \eta)$ and defining $\tilde{g} = \Omega^2 g$ for $g \in U_1$).

The third defining property of asymptotic simplicity states that null geodesics are complete and that, intuitively, they reach "infinity". This is certainly true for Minkowski space and an open W^k neighborhood of η must be exhibited, all the Lorentz metrics in which exhibit this feature. To this end, it is noted that the spray of η is certainly complete and that the completeness of a vector field is a stable property.[7] [The spray of a metric g on M is the map sp: $TM \to TTM$ defined locally as sending $(p, v_p) \to (p, v_p, v_p, -\Gamma(p)(v_p, v_p))$ where $\Gamma(p)$ is the connection of g at p. The spray of a metric is a second-order differential equation on M and its curves give the geodesics of g. Geodesic completeness of g means that the vector field sp(g) is complete.] Thus there exists a W^{k-1} neighborhood of the spray of η in $\Gamma^{k-1}(TTM)$ all the vector fields in which are complete. As the map sending a Lorentz metric to its spray is continuous, there exists an open neighborhood U_2 of η, all the metrics in which are geodesically complete. Let $U_3 = U_1 \cap U_2$.

If a Minkowski coordinate system $\{t = x^0, x^1, x^2, x^3\}$ is chosen for η, so that $ds^2 = dt^2 - (dx^1)^2 - (dx^2)^2 - (dx^3)^2$, then any null geodesic in $(\mathbb{R}^4, \eta)$ has the property that it crosses each of the t = const hypersurfaces. Thus the null spray of η [in these coordinates $(x^i, v^i) \to (x^i, v^i, v^i, 0)$ where $\eta_{ij}v^iv^j = 0$] is transverse to the hypersurfaces $t \circ \pi$ = const, where π is the projection associated with the tangent bundle, $T\mathbb{R}^4$. This transversal property is stable and so there exists a W^k open neighborhood U_4 of η containing only metrics whose null sprays are transverse to the hypersurfaces $t \circ \pi$ = const. Let $U = U_3 \cap U_4$. The metrics in U satisfy the properties (1) and (2) of asymptotically simplicity, and have complete null sprays transverse to each $t \circ \pi$ = constant hypersurface. The claim is that the metrics will also satisfy property (3). Null geodesics (maximally extended) for a metric g in U cross each t = const hypersurface. Since null geodesics are conformally invariant, the image of a null geodesic of $(\mathbb{R}^4, g)$ under the conformal map is a null geodesic of $(\tilde{M}, \tilde{g})$. Thus a null geodesic (maximally extended) of (M, g) can only fail to have two points on $\mathcal{I}$ if it contains I^+, I^0, or I^-. See Fig. 2. But this is impossible as the past light cone of I^+ is $\mathcal{I}^+$, the light cone of I^0 is $\mathcal{I}$, and the future light cone of I^- is $\mathcal{I}^-$. Thus, for example, the only null geodesics in $(\tilde{M}, \tilde{g})$ containing I^+ are those that lie on $\mathcal{I}^+$ and no null geodesic from $\tilde{M} - \mathcal{I}$ can contain I^+. Thus, we have an open W^k neighborhood U of η in $L(\mathbb{R}^4)$ containing only asymptotically simple space-times.

So there are many asymptotically simple space-times based in $\mathbb{R}^4$; in particular, ones which are not conformally flat.

A slightly stronger version of Theorem 1 can be proved.

Theorem 1': Asymptotic simplicity is stable at η in the A^k topology on the set B_k of C^k Lorentz metrics which are asymptotic to η.

The A^k topology is coarser than the W^k topology and the additional asymptotic condition is essential. The proof proceeds similarly to that above, with one making certain that at each step the W^k open sets can be replaced with A^k open sets all the elements in which are asymptotic to η.

IV. CURVES OF LORENTZ METRICS

Given a curve in $L^k(\mathbb{R}^4)$, $k \geq 2$, of Lorentz metrics at η, g: $t \to g(t)$ such that $g(0) = \eta$, the induced curve of Riemann tensors is denoted by Ω: $t \to \text{Riem}(g(t)) \in \Gamma^{k-2}(T_3^1\mathbb{R}^4)$, the induced curve of Ricci tensors is denoted by Ric: $t \to \text{Ricci}(g(t)) \in \Gamma^{k-2}(T_2^0\mathbb{R}^4)$, and the in-

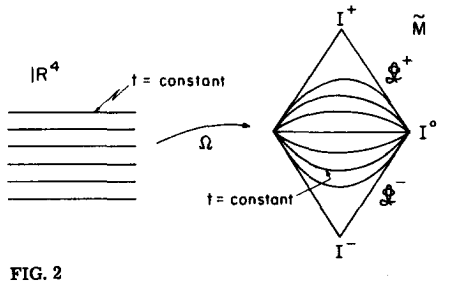


FIG. 2

duced curve of conformal tensors is denoted by Co: $t \to \text{conf.}(g(t)) \in \Gamma^{k-2}(T^0_4(\mathbb{R}^4))$. If g is in $B_k^{(0,2)}$ for η and Minkowski coordinate system $\{x^i\}$, then the three induced curves Ω, Ric, and Co have images in $B_{k-2}^{(0,4)}$, $B_{k-2}^{(0,2)}$, and $B_{k-2}^{(0,4)}$, respectively.[4] If, in addition, g is analytic at η, then the induced curves are all analytic at zero in the respective spaces.

On the other hand, if an analytic curve is given at 0 in $B_{k-2}^{(0,2)}$, the question arises as to whether there exists an analytic curve in $B_k^{(0,2)}$ of Lorentz metrics at η which induces the given curve in $B_{k-2}^{(0,2)}$ as its curve of Ricci tensors. This, in general, will not be the case; when it is true one can solve the field equations in general relativity by solving sequentially a set of linear partial differential equations. The difficulties are with respect to boundedness of the terms and convergence of the resulting sequence. An additional freedom in the resulting curve is fixed by appropriate initial conditions for g. This additional freedom is a useful adjunct in trying to find such an analytic curve of Lorentz metrics.

Theorem 2: Let $t \to g(t) \in B_k^{(0,2)}$ $(k \geq 3)$ be an analytic curve of vacuum metrics $(\text{Ric}(g(t)) = 0$, all $t)$ on $\mathbb{R}^4$ with $g(0) = \eta$. Suppose that

(a) for all $t, g(t)$ is asymptotic to η,

(b) for each t, the conformal curvature tensor of $g(t)$ results from zero initial data on $\mathcal{I}^-$ in the conformally related $(\tilde{M}, \tilde{\eta})$.

Then $g(t)$ is a curve of Minkowski metrics.

Proof: Given a curve of Lorentz metrics analytic at η, g: $t \to g(t)$, the analyticity at η requires that

$$g(t) = \eta + \sum_{i=1}^{\infty} \underset{(i)}{h} \frac{t^i}{i!}.$$

The requirement that the curve be asymptotic to η is that the curve

$$\sum_{i=1}^{\infty} \underset{(i)}{h} \frac{t^i}{i!}$$

by asymptotic to zero. Tensor fields of any given type in $B_k^{(\cdot,\cdot)}$ asymptotic to zero form a Banach space. There is an open interval about zero for which the curve g has its image in the set of asymptotically simple metrics on $\mathbb{R}^4$ as guaranteed by Theorem 1′; let t be restricted to such an interval. The solution of the equations for the Ricci tensor and conformal tensor for this curve then proceeds sequentially. Since the corresponding maps Ric and Co are analytic at η, the corresponding curves are completely determined by their derivatives for $t = 0$. The zeroth derivative gives conditions automatically satisfied since Minkowski space-time has zero Ricci tensor and zero conformal tensor. The first derivative of the curve of Ricci tensors at $t = 0$ simply gives the vanishing of the linearized Ricci tensor for $\eta + h_{(1)}$,

$$\eta^{ad} \underset{(1)}{h}_{ac,bd} - \eta^{ad} \underset{(1)}{h}_{bc,ad} - \eta^{ad} \underset{(1)}{h}_{ad,bc} + \eta^{ad} \underset{(1)}{h}_{bd,ac} = 0.$$

The first derivative of the curve of conformal tensors at $t = 0$ yields the linearized conformal tensor $C_{(1)abcd}$ of $\eta + h_{(1)}$ which must in the conformally related picture $(\tilde{M}, \tilde{\eta})$ be obtained from zero initial data on the null Cauchy hypersurface $\mathcal{I}^-$. The linearized conformal tensor, $C_{(1)abcd}$, satisfies the linearized Bianchi identities $\eta^{de} C_{(1)abcd,e} = 0$, whose spinor equivalent is $\nabla^{AA'}\psi_{ABCD} = 0$ if ψ_{ABCD} represents $C_{(1)abcd}$. By using the techniques developed by Penrose[1] for handling such zero rest-mass field equations in a conformally invariant manner, a spinor field $\phi_{ABCD} = \Omega^{-1}\psi_{ABCD}$ is obtained on $(\tilde{M}, \tilde{\eta})$ satisfying $\nabla^{AA'}\phi_{ABCD} = 0$. Initial data for ϕ_{ABCD} is given on the null Cauchy hypersurface $\mathcal{I}^-$, namely zero data, and the solution at a point $p \in \tilde{M}$ of the initial value problem for the equation $\nabla^{AA'}\phi_{ABCD} = 0$ can be obtained as a generalized Kirchoff integral over the intersection of the past light cone of p and the initial data surface $\mathcal{I}^-$.[3] For zero initial data, the resulting field ϕ_{ABCD} is zero and thus $\psi_{ABCD} = 0$ or equivalently $C_{(1)abcd} = 0$.

The fact that the linearized Ricci tensor and the linearized conformal tensor of $\eta + h_{(1)}h$ are both zero is equivalent to the vanishing of the linearized Riemann tensor. This fact is best exploited by using the Cartan structure equations in their linearized forms for the determination of $h_{(1)}$. The structure equations are

$$d\theta^a + \omega^a{}_b \wedge \theta^b = 0,$$

$$d\omega^a{}_b + \omega^a{}_c \wedge \omega^c{}_b = \tfrac{1}{2} R^a{}_b,$$

where θ^a is a basis for cotangent vector fields, $\omega^a{}_b$ is the connection form whose Riemann curvature form is $R^a{}_b$. In the frame $\theta^a = dx^a$ the metric $\eta + h_{(1)}$ yields the linearized structure equations with linearized connection $\omega^a{}_b$

$$\omega^a{}_b \wedge \theta^b = 0, \tag{1}$$

$$d\omega^a{}_b = 0. \tag{2}$$

Equation (2) says that the 1-form $\omega^a{}_b = \omega^a{}_{bc}dx^c$ is closed and since the manifold is simply connected, $\omega^a{}_b$ is exact,

$$\omega^a{}_{bc} = \mu^a{}_{b,c}. \tag{3}$$

Equation (1) then gives $\mu^a{}_{b,c}dx^b \wedge dx^c = 0$ or that the 1-form $\mu^a = \mu^a{}_b dx^b$ is closed. Again, simple connectivity says that μ^a is exact,

$$\mu^a{}_b = \underset{(1)}{\sigma}{}^a{}_{,b}.$$

and thus

$$\omega^a{}_{bc} = \underset{(1)}{\sigma}{}^a{}_{,bc}. \tag{4}$$

The tensor $h_{(1)ab}$ is obtained from the linearized equation for the vanishing of the covariant derivative of the metric, i.e., $Dg_{ab;c}(t)|_{t=0}(h_{(1)ab})$, resulting in

$$\underset{(1)}{h}{}_{ab,c} = \eta_{ad} \underset{(1)}{\sigma}{}^d{}_{,bc} + \eta_{bd} \underset{(1)}{\sigma}{}^d{}_{,ac}, \tag{5}$$

whose solution (absorbing constants of integration into $\sigma_{(1)}{}^a$) is

$$\underset{(1)}{h}{}_{ab} = \underset{(1)}{\sigma}{}_{a,b} + \underset{(1)}{\sigma}{}_{b,a}, \tag{6}$$

where

$$\underset{(1)}{\sigma}{}_a = \eta_{ad} \underset{(1)}{\sigma}{}^d.$$

The four functions $\sigma_{(1)a}$ are required to be C^4 and to be appropriately bounded in $B_4^{(0,0)}$.

A coordinate transformation $x'^a = x^a + t\sigma_{(1)}{}^a$ yields the same curve of metrics which, when expressed in the new coordinate system, has zero linear term:

$$\eta + 0\cdot t + \sum_{i=2}^{\infty} h'_{(i)} \frac{t^i}{i!}. \tag{7}$$

The next step in the sequential solution is to determine $h'_{(2)}$ from the conditions imposed on the curve of Ricci terms (namely that it be the constant curve 0) and the curve of conformal tensors. The resulting equations for $h'_{(2)}$ are exactly those previously solved for $h_{(1)}$ and the same technique that resulted in Eq. (7) yields a curve of the form

$$\eta + 0\cdot t + 0\cdot\frac{t^2}{2!} + \sum_{i=3}^{\infty} h''_{(i)} \frac{t^i}{i!}. \tag{8}$$

Continuing, it is seen that the resulting curve is a curve of Minkowski metrics as required. For each t, the transformation

$$\tilde{x}^a = x^a + \sum_{i=1}^{\infty} \sigma^a{}_{(i)} \frac{t^i}{i!}$$

exhibits $g(t)$ in a Minkowski coordinate system.

V. CONCLUSION

The following questions arise:

(1) Is asymptotic simplicity W^k stable? Theorem 1 is the proof that this is the case at any Minkowski metric on $\mathbb{R}^4$, $(\mathbb{R}^4, \eta)$. The proof utilizes some special properties of Minkowski space and no obvious generalizations of the techniques involved exist. Also note that Theorem 1 gives the stability of weakly asymptotically simple spaces[1] at any weakly asymptotically simple space whose corresponding asymptotically simple space is in the open W^k neighborhood of $(\mathbb{R}^4, \eta)$ exhibited in the theorem.

(2) Is it possible to use the techniques in Sec. IV to generate nontrivial analytic curves of solutions to prescribed field equations? Considerations of this nature are to appear in a forthcoming paper by the authors.

(3) Is it possible to obtain conditions under which the linearized solutions to the field equations actually determine the behavior of solutions in the full theory? Included in this might be a formulation of stable properties in terms of Lyapunov functionals.

[1]R. Penrose, in *Battelle Recontres*, 1967 Lectures in Mathematics and Physics (Benjamin, New York, 1968).
[2]R. Penrose, *An analysis of the Structure of Space-Time* (Princeton U. P., Princeton, N. J., 1966).
[3]E. T. Newman and R. Penrose, Proc. R. Soc. A **305**, 175 (1968).
[4]D. Lerner and J.R. Porter, J. Math. Phys. 15, 1413 (1974).
[5]N. H. Kupier, Ann. Math. **50**, 916 (1949).
[6]H. Rudberg, thesis (Uppsula, 1957).
[7]D. E. Lerner, Comm. Math. Phy. (to be published).

On the three-body linear problem with three-body interaction*

J. Wolfes†

Internationale Centre for Theoretical Physics, Trieste, Italy
(Received 8 November 1972)

A three-body potential is introduced for which Schrödinger's equation of the three-body linear problem with additional harmonic and inverse cube forces is solved exactly.

I. INTRODUCTION

The three-body linear problem has been solved exactly for only two distinct potentials[1]:

$$V_1=\sum_1^3 a_i\,\delta(x_i-x_j), \tag{1a}$$

$$V_2=\sum_1^3 [\omega^2(x_i-x_j)^2+g(x_i-x_j)^{-2}]. \tag{1b}$$

Its seems reasonable to assert that this list exhausts all possibilities of exact solution in terms of two-body interactions.

Therefore, in order to extend the application of the three-body linear model, an exact solving three-body potential might be of interest since, as far as we know, no example with this potential is available up to date. The potential to be introduced has the generic form

$$V_{ij,k}=f[(x_i-x_k)+(x_j-x_k)]^{-2}, \tag{2}$$

where f is a real parameter. It represents the interaction between the i–j couple due to the presence of the third k-body. The potential is reasonable since it depends, in a symmetric way, on the distance of each particle with respect to the third particle and, furthermore, it tends to zero when the third body approaches infinity with respect to the other two. Besides, it can also be understood as a generalization of a two-body potential, since for $k=i$ or $k=j$ Eq. (2) becomes a two-body inverse square potential.

We apply the three-body potential to two different problems for which Schrödinger's equation separates in a nontrivial way.

The main characteristic of this model is the appearance of anisotropic states for distinguishable particles. Order and statistics are the other two features of the solutions to be discussed.

In Sec. II, Schrödinger's equation of the most general problem is established and the symmetry properties of the coordinate system are discussed. In Secs. III and IV we solve the problem of three equal-mass linear bodies interacting via three-body plus harmonic and/or inverse cube forces. Section V is devoted to some general remarks and future outlook.

II. SCHRÖDINGER'S EQUATION AND COORDINATES

In this section we write Schrödinger's equation for the general problem which includes harmonic, inverse cube, and three-body forces. We start with one-dimensional cartesian coordinates x_i, and then transform the equation into center-of-mass polar coordinates (r,φ) for which separation and exact solution is possible.

We have then $(\hbar^2/2m=1)$

$$\left(-\sum_{i=1}^3\frac{\partial^2}{\partial x_i^2}+\sum_{i,j=1}^3[\tfrac{1}{8}\omega^2(x_i-x_j)^2+2g(x_i-x_j)^{-2}]\right.$$
$$\left.+\sum_{i,j,k=1}^3 6f[(x_i-x_k)+(x_j-x_k)]^{-2}-E\right)\underset{i\neq j\neq k}{\Psi}(x_1,x_2,x_3)$$
$$=0. \tag{3}$$

Introducing center-of-mass coordinates

$$\sqrt{2}\,u=x_1-x_2,$$
$$\sqrt{6}\,v=x_1+x_2-2x_3, \tag{4a}$$
$$3R=x_1+x_2+x_3,$$

so that

$$x_2-x_3=\tfrac{1}{2}\sqrt{2}\,(\sqrt{3}\,v-u),\quad x_3-x_1=-\tfrac{1}{2}\sqrt{2}\,(\sqrt{3}\,v+u), \tag{4b}$$

we obtain for (3), after eliminating the R coordinate,

$$\left[-\frac{\partial^2}{\partial u^2}-\frac{\partial^2}{\partial v^2}+\omega^2(u^2+v^2)+9g\frac{(u^2+v^2)^2}{u^2(3v^2-u^2)^2}\right.$$
$$\left.+9f\frac{(u^2+v^2)^2}{v^2(3u^2-v^2)^2}-E\right]\Psi(u,v)=0, \tag{5}$$

where E is now the energy in the c.m. frame. Finally we introduce the polar coordinates (r,φ):

$$u=r\sin\varphi,\quad v=r\cos\varphi, \tag{6}$$

so that

$$\sin 3\varphi=\sin\varphi(\sin^2\varphi-3\cos^2\varphi)=r^{-3}u(u^2-3v^2),$$
$$\cos 3\varphi=\cos\varphi(\cos^2\varphi-3\sin^2\varphi)=r^{-3}v(v^2-3u^2) \tag{7}$$

and Eq. (5) becomes

$$\left[-\frac{\partial^2}{\partial r^2}-\frac{1}{r}\frac{\partial}{\partial r}-\frac{\partial^2}{\partial\varphi^2}+\omega^2r^2+9g(r\sin3\varphi)^{-2}\right.$$
$$\left.+9g(r\cos3\varphi)^{-2}-E\right]\Psi(r,\varphi)=0. \tag{8}$$

This equation can be separated (λ^2 is the separation constant):

$$\left(-\frac{\partial^2}{\partial r^2}-\frac{1}{r}\frac{\partial}{\partial r}+\omega^2r^2+\frac{\lambda^2}{r^2}-E\right)\chi(r)=0, \tag{9a}$$

$$\left(-\frac{\partial^2}{\partial\varphi^2}+\frac{9g}{\sin^2 3\varphi}+\frac{9f}{\cos^2 3\varphi}-\lambda^2\right)\quad\phi(\varphi)=0. \tag{9b}$$

Both equations can be solved exactly. By putting $f=0$ and $g=0$, one obtains the well-known harmonic oscillator problem. The case where only $f=0$ was solved by Calogero.[1b] In this paper we discuss the solutions obtained for $g=0$ and $f\neq0$ as well as $g\neq0$ and $f\neq0$. The r equation is the same for all cases mentioned above and its solution is well known. Only for the sake of com-

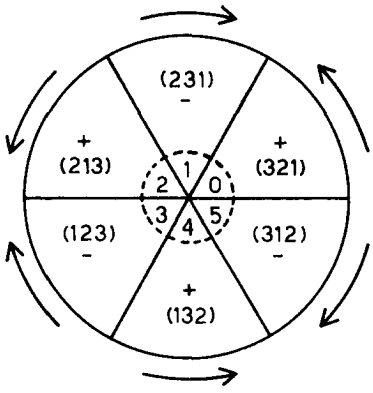


FIG. 1. p sectors of order due to inverse cube forces.

pleteness shall we solve it at this stage. To this end one makes the ansatz

$$\chi(r)=r^b\exp(cr^2)\,F(ar^2). \tag{10}$$

And replacing (10) in (9a), one obtains a confluent hypergeometric equation for F if

$$a=-2c=\omega, \quad b=\pm\lambda, \tag{11}$$

then

$$F(\alpha;\gamma;\omega r^2)=F[\tfrac{1}{2}(1\pm\lambda)-E/4\omega;\ 1\pm\lambda;\ \omega r^2].$$

The physically acceptable solution is obtained by choosing b positive and by requiring F to be a (Laguerre) polynomial, that is,

$$b=\lambda>0, \quad \tfrac{1}{2}(1+\lambda)-E/4\omega=-n,$$

so that the solution of the r equation is

$$\chi_{\lambda n}(r)=r^\lambda\exp(-\tfrac{1}{2}\omega r^2)\,L_n^\lambda(\omega r^2) \tag{12}$$

with

$$E=2\omega(2n+\lambda+1), \quad n=0,1,2,\cdots. \tag{13}$$

Next we discuss the very important symmetry properties of the coordinate system (r,φ). From (4) and (6) one obtains

$$x_1-x_2=\sqrt{2}\ r\sin\varphi, \quad x_1+x_2-2x_3=\sqrt{6}\ r\cos\varphi. \tag{14}$$

Since the three x_i coordinates are collinear, a specific value of φ gives a specific ordering of all three particles. It can readily be seen from (14) that if the complete φ circle is divided into six sectors of $\pi/3$ each, then to each sector we may attach one and only one order of the three particles (Fig. 1). A given sector will be called the p sector if all angles $\varphi=\varphi_p$ belonging to it can be written $\varphi_p=\varphi_0+p\pi/3$ for $0\leq\varphi_0\leq\pi/3$.

For identical particles, permutation operators will be of importance. We shall call P^- an odd permutation operator which interchanges any two indices of an ordered trio. A product of any two odd permutation operators gives the even permutation operator P^+. Besides, two configurations of the three particles will be called "identical configurations" if one can be obtained from the other by application of either P^+ or P^-. Then, if we call $C(\varphi)$ a specific configuration associated with a given angle φ, all six identical configurations generated by the six operators $P^\pm$ are given by

$$P^\pm C(\varphi)=C(\pm\varphi+2n/3\pi) \tag{15}$$

for $n=0$, 1, or 2.

A value of φ which will be of interest in our problem is given by $\varphi=(2p+1)/6\pi$. At these points one has $3\sqrt{2}\ r\sin[(2p+1)/6\pi]=\sqrt{6}\ r\cos[(2p+1)/6\pi]$. That is, if the order (i,j,k) corresponds to the p sector, then for $\varphi_p=(2p+1)/6\pi$, the midcoordinate of the ordered trio (i,j,k) will be zero $(x_j=0)$. Furthermore, for $\varphi_p<(2p+1)/6\pi$, the same coordinate is positive $(x_j>0)$ for even sectors, while it is negative $(x_j<0)$ for odd sectors. Now, in order to distinguish between these two configurations, we write for the pth configuration (p even):

$$(ijk)=\begin{cases} i(jk) & \text{if } \varphi_p<\frac{1}{6}(2p+1)\pi \\ (ij)k & \text{if } \varphi_1>\frac{1}{6}(2p+1)\pi \end{cases}. \tag{16}$$

Figure 3 (Sec. III A) shows the 12 sectors of $\pi/6$ each for which this particular arrangement of particles holds. Furthermore, we call, for instance, R^+ the arrangement $i(jk)$ if (ijk) is an even permutation of (321); but if (ijk) is an odd permutation of (321), then, for instance, $(ij)k=L^-$.

Ordering plays an important role in our problem since inverse cube forces do not allow for interpenetration of particles, so that a given order will be preserved under the presence of these forces. On the other hand, the three-body force (2) will be shown to allow for interchange of two particles while only the third conserves order. This particular effect will give rise to the R and L configurations.

III. HARMONIC AND THREE-BODY FORCES

In this section we solve and discuss Schrödinger's equation (3) for $g=0$. The corresponding r equation (9a) has already been solved, (12); the φ equation (9b) of this problem reads

$$\left(-\frac{\partial^2}{\partial\varphi^2}+\frac{9f}{\cos^2 3\varphi}-\lambda^2\right)\phi(\varphi)=0. \tag{17}$$

In order to solve this problem, we introduce the variable $z=\cos^2 3\varphi$. Indeed, the inverse of this transformation is not unique. But, since solutions will be found for the range $(q-1)/2\pi\leq 3\varphi\leq(q+1)/2\pi$, for each sector so defined there will be only one value of φ for each value of z. Substitution of z into (17) gives

$$\left[z(1-z)\frac{d^2}{dz^2}+\left(\frac{1}{2}-z\right)\frac{d}{dz}-\frac{1}{4}\left(\frac{f}{z}-\frac{\lambda^2}{9}\right)\right]\phi(z)=0. \tag{18}$$

If we now put $\phi(z)=z^a H(z)$ and require that

$$a=\tfrac{1}{4}[1\pm(1+4g)^{1/2}], \tag{19}$$

we obtain the following hypergeometric equation for $H(z)$:

$$\left[z(1-z)\frac{d^2}{dz^2}+\left(\frac{1}{2}+2a-z(1+2a)\right)\frac{d}{dz}+\left(\frac{\lambda}{6}\right)^2-a^2\right]H(z)=0; \tag{20}$$

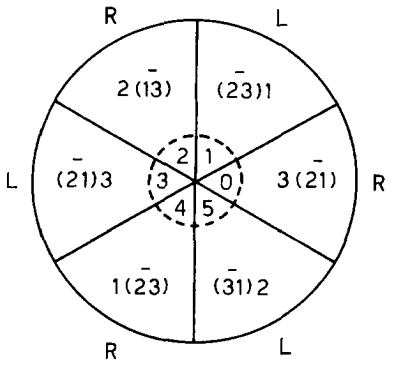


FIG. 2. q sectors of polarization due to three-body forces.

and the complete solution is:

$$H(z)=A\,H(\alpha,\beta;\gamma;z)+B\,z^{1-\gamma}\,H(\alpha+1-\gamma,\ \beta+1-\gamma;\ 2-\gamma;z), \tag{21}$$

where

$$\alpha=a-\lambda/6,\quad \beta=a+\lambda/6,\quad \gamma=\tfrac{1}{2}+2a. \tag{22}$$

If we now require both $|\phi|^2$ and $\phi\phi'$ to be continuous, zero at the boundaries ($z=0,1,$) and, in order to avoid collapsing of the system, $g>-\frac{1}{4}$ (see Ref. 1b), then since, at $z=0$, $\phi\phi'$ will contain terms like $A\,z^{2a-1}+Bz^{-2a}$, we should ask $B=0$ and $2a-1\geq 0$. Therefore, one must choose the positive sign of (19) and require $g>-3/16$. On the other hand, at $z=1$ the hypergeometric function behaves like $(1-z)^{\gamma-\alpha-\beta}=(1-z)^{1/2}$, and its derivative like $(1-z)^{-1/2}$, which should be continuous. Therefore, we have to cut the infinite series H by requiring α(or β) to be a negative integer:

$$a-\tfrac{1}{6}\lambda=-l,\quad \text{for } l=0,1,2,\cdots. \tag{23}$$

Then we obtain

$$\phi(z)=z^a\,H(-l,l+2a;\ \tfrac{1}{2}+2a;z). \tag{24}$$

And recalling the definition of Gegenbauer polynomials,

$$G^s_{2n}(x)=H(-n,\ n+s;\ s+\tfrac{1}{2};\ 1-x^2),$$
$$G^s_{2n+1}(x)=xH(-n,\ n+s+1;\ n+\tfrac{1}{2};\ 1-x^2), \tag{25}$$

we obtain as the physically acceptable solution of (17)

$$\phi_q(\varphi)=\cos^{2a}3\varphi\,G_l^{2a}(\sin 3\varphi) \tag{26}$$

for

$$\tfrac{1}{2}(q-1)\pi\leq 3\varphi\leq\tfrac{1}{2}(q+1)\pi,\quad \text{otherwise zero.} \tag{27}$$

We are ready now to write down the complete solution of this problem. But, despite the fact that the three particles are identical, we shall first treat them as if they were distinguishable, in order to get a better insight into the kinematics of our problem. By doing so we facilitate the symmetrization procedure, which follows once we treat the particles, as they are, as indistinguishable.

A. Distinguishable particles

We shall show now that, for a given energy E_{nl}, six degenerate Ψ_{nlq} states are obtained from (26) if we allow q to range from 0 to 5.

Now, each ϕ_q is different from zero only if φ_q lies in its respective q sector given by (26a). If we compare these with the p sectors defined by the coordinates (Sec. II), we observe that in each q sector only one particle conserves order while the other two may permute. We call "free particle" the first one and "bounded pair" the other two. Furthermore, for $q=0,2,$ and 4 the bounded pair stays on the right of the free particle while for $q=1,3,$ and 5 the bounded pair stays on the left of the free particle. We call these R and L states, respectively. The motivation for all these definitions stems from the fact that the boundaries of the q sectors coincide with those values of φ for which the midparticle of the trio lies either on the right (R state) or on the left (L state) of the center of mass of the other two (see Sec. II). Certainly, if the particles are distinguishable, each of these three R and three L states can be distinguished from one another (Fig. 2). That is, three distinguishable particles give rise to six distinguishable states, as it should be. But these six states cannot be generated from each other by permutation P, since R and L states do not mix under the action of P. Therefore, there are still six states missing. These, as we shall see, can be obtained by permuting the pair of bounded particles. The whole effect on the wavefunction will be a phase factor $(-1)^l$. In fact, let us consider the R state $3(\overline{21})$. From Fig. 3 we see that while φ stays in the lower sector (labelled R^-) the particles arrange according to 3(12). The identical configuration, with 1—2 permuted, belongs to the R^+ sector, φ being changed by $-\varphi$. But since $G_l^{2a}\sin(-3\varphi)=(-1)^l\,G_l^{2a}(\sin 3\varphi)$, we have $P(1,2)\phi_{nl0}=(-1)^l\,\phi_{nl0}$. Are these states physically different? That is, is it possible by means of an observation to tell whether we are dealing with 3(21) or 3(12)? Certainly not, since both configurations are equally probable for particles in the state $3(\overline{21})$. This, by the way, is reflected by the fact that both wavefunctions differ, at most, in sign. That is, although all three particles are assumed to be distinguishable, our three-body interaction allows one to distinguish only between states for which the free particle is different. This is why we have only three R and three L states.

Finally we are in a condition to write down the six energy-degenerate wavefunctions which are solutions of Eq. (3) if $g=0$ and the particles are distinguishable [see Eqs. (12) and (26)]

$$\Psi_{nlq}(r,\varphi)=N_{nlq}\,r^{6(a+l)}\exp(-\tfrac{1}{2}\omega r^2)\,L_n^{6(a+l)}(\omega r^2)$$
$$\times\cos^{2a}3\varphi_q\,G_l^{2a}(\sin 3\varphi_q), \tag{28}$$

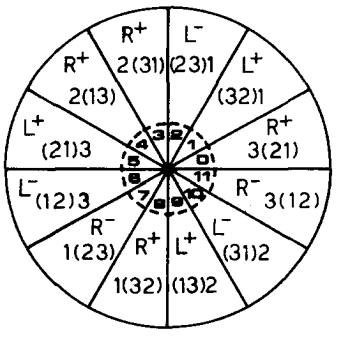


FIG. 3. Polarization and order.

where $q=0,1,\ldots,5$, $n,l=0,1,2,\ldots$, $(q-1)/6\pi \leq \varphi_q \leq (q+1)/6\pi$, $a=\frac{1}{4}[1+(1+4g)^{1/2}]$, $g \geq -\frac{3}{16}$, and N_{nlq} is a normalization factor. The corresponding energy is

$$E_{nl}=2\omega(2n+6a+2l+1). \tag{29}$$

For q even, Ψ_{nlq} is R-polarized while for q odd it is L-polarized. Bosons are described by l-even while fermions by l-odd wavefunctions.

B. Indistinguishable particles

Our next task is to symmetrize all six wavefunctions given by (28). To this end we make use of the permutation operator $P^{\pm}$, defined by (15), obtaining

$$P^{\pm}\,\phi_q(\varphi_q)=\phi_q(\pm\varphi_q+2m/3\pi), \tag{30}$$

where $m=0,1,2$. This result is equivalent to

$$P^{+}\,\phi_q(\varphi_q)=\phi_q(\varphi_q+2m), \quad P^{-}\,\phi_q(\varphi_q)=(-1)^l\,\phi_q(\varphi_q+2m). \tag{31}$$

But the state $\phi_q(\varphi_q+2m)$, unless $m=0$, has not been defined. Without changing anything in our previous calculations, we may now regard q as a mere parameter (not a "quantum number of distinguishability" as in Sec. IIIB) of the argument of only one unique function ϕ, so that $\phi_q=\phi_{q'}$ for all q and q'. This is equivalent to requiring $N_{nl0}=N_{nl2}=N_{nl4}=N_R$ and $N_{nl1}\,N_{nl3}=N_{nl5}=N_L$. But the relative phase between N_L and N_R cannot be decided on statistical grounds since permutation does not mix R and L states. These two are connected by parity so that $N_L=N_R=N$ for even parity and $N_L=-N_R=-N$ for odd parity.

We are then left with only one symmetrized state for each energy E_{nl}, that is,

$$\Psi_{nls}(\varphi_0)=N r^{6(a+l)}\exp(-\tfrac{1}{2}\omega r^2)\,L_n^{6(a+l)}(\omega r^2)$$
$$\times\sum_{q=0}^{5}(s)^q\cos^{2a}3(\varphi_0+q\pi/3)\,G_l^{2a}[\sin 3(\varphi_0+q\pi/3)], \tag{32}$$

where the energy is given by (29), $-\pi/6 \leq \varphi_0 \leq \pi/6$, and $s=+1$ for even parity and (-1) for odd parity. Bosons are described by l-even and fermions by l-odd wavefunctions.

Before closing this section we wish to comment on polarization.

If the simultaneous position of all three particles is an observable, then an R state can certainly be distinguished from an L state. Or, what amounts to the same thing, if we were able to prepare a specific polarized state, then due to the singular nature of the three-body interaction which forbids a particle to cross the center of mass of the other two, this particular state of polarization would certainly be conserved. Therefore one would have, instead of one wavefunction for each energy level (nl), two degenerate states, one for q-even (R states) and one for q-odd (L states). These states would not be parity eigenstates. But is the simultaneous position an observable?

The answer to this question is no if the particles are indistinguishable. In fact, any observable related to identical particles must be permutation invariant.[2] Therefore, for three indistinguishable particles, the following are well-defined position operators:

$$X=x_1+x_2+x_3,$$
$$Y=|x_1-x_2|+|x_1-x_3|+|x_2-x_3|,$$
$$Z=|x_1+x_2-2x_3|+|x_1+x_3-2x_2|+|x_2+x_3-2x_1|.$$

But a simultaneous measurement of X, Y, and Z does not allow one to infer unambiguously the value of all three coordinates (given a specific order). Although one may say that for R or L states one could infer precise values of x_i from the knowledge of the position observables, the converse is not true. This can be seen by computing Z for a specific order, say (321). Then $|x_1+x_3-2x_2|$ is either $x_1+x_2-2x_2$ or $2x_2-x_1-x_3$, and there is no way of deciding between these two values from the knowledge of Y and X alone.

However, the answer to this question would be yes if one accepted the possibility of polarization destroying indistinguishability.

IV. HARMONIC, INVERSE CUBIC AND THREE-BODY FORCES

This time we may go straight on since the details of the main operations have already been given in the last section. We start with Eq. (9b):

$$\left(-\frac{d^2}{d\varphi^2}+\frac{g}{\sin^2 3\varphi}+\frac{f}{\cos^2 3\varphi}-\lambda^2\right)\phi(\varphi)=0, \tag{33a}$$

where now

$$q/2\pi \leq 3\varphi \leq (q+1)/2\pi. \tag{33b}$$

By transforming according to $z=\sin^2 3\varphi$, we obtain

$$\left[z(1-z)\frac{d^2}{dx^2}+\left(\frac{1}{2}-z\right)\frac{d}{dz}+\frac{1}{4}\left(\frac{\lambda^2}{9}-\frac{g}{z}-\frac{f}{1-z}\right)\right]\phi(z)=0. \tag{34}$$

Now we try $\phi(z)=z^a(1-z)^b H(z)$ and obtain a hypergeometric equation for H:

$$\left\{z(1-z)\frac{d^2}{dz^2}+\left[\left(\frac{1}{2}+2a\right)-z(1+2a+2b)\right]\frac{d}{dz}+\left[\left(\frac{\lambda}{6}\right)^2-(a+b)^2\right]\right\}H(z)=0. \tag{35}$$

if

$$a=\tfrac{1}{4}[1+(1+4g)^{1/2}], \quad b=\tfrac{1}{4}[1+(1+4f)^{1/2}], \tag{36}$$

and $H=H(\alpha,\beta;\gamma;z)$ for

$$\alpha=a+b-\lambda/6, \quad \beta=a+b+\lambda/6, \quad \gamma=2a+\tfrac{1}{2}. \tag{37}$$

Again, for the correct physical solution we must require α (or β) to be a negative integer; then H becomes a Jacobi polynomial $P_r^{s,t}(z)$, so that for each sector defined by (33b) one obtains as a solution of (33a)

$$\phi(\varphi)=\sin^{2a}3\varphi\,\cos^{2b}3\varphi\,P_l^{2a-1/2,2b-1/2}(\cos 6\varphi). \tag{38}$$

Next we shall discuss the solution for distinguishable particles.

A. Distinguishable particles

The main new element which appears by addition of

the pair inverse square potential to the problem treated in Sec. II is that of order. This ordering factor is already present in Calogero's problem ($g\neq 0; f=0$) and stems from the fact that the potential term $(x_i-x_j)^2$ does not allow particle i to overcome particle j. We saw that the three-body potential induces polarization and as a consequence it conserves order between the free particle and the bounded pair. By combining both potentials the net effect is a superposition of both order and polarization. States like $|a(bc)\rangle$ and $|a(cb)\rangle$, which were indistinguishable even for distinguishable particles, now become different. Therefore, energy eigenstates of Boltzmann particles are now twelvefold degenerate. We may divide them into four classes:

$$R^+=\chi\ \phi_{4q'+4},\quad R^-=\chi\ \phi_{4q'+3},$$
$$L^+=\chi\ \phi_{4q'+1},\quad L^-=\chi\ \phi_{4q'+2}, \tag{39}$$

for $q'=0,1,2$, $q=4q'+s$, and $\phi_q=\phi(\varphi_q)$. The twelve complete normalized degnerate wavefunctions of energy are then

$$\Psi_{nlq}(r,\varphi)=N_{nlq}\,r^\lambda \exp(-\tfrac{1}{2}\omega r^2)\,L_n^\lambda(\omega r^2)\sin^{2a}(3\varphi_0 + q\pi/2)\cos^{2b}(3\varphi_0+q\pi/2)\,P_l^{2a-1/2,2b-1/2}[\cos(6\varphi_0+q\pi)], \tag{40}$$

where $0\leq\varphi_0\leq\pi/6$, $n,l=0,1,2,\ldots$, $\lambda=6(a+b+l)$. Different values of q give polarized even or odd states according to (39). The energy of this state is given by

$$E_{nl}=2\omega(2n+6a+6b+6l+1). \tag{41}$$

B. Indistinguishable particles

We must symmetrize all twelve functions given by (40).

Again we consider the permutation operators $P^\pm$, defined by Eq. (15). One has, as usual,

$$P^+(\varphi_0+q\pi/6)=\phi[\varphi_q+\tfrac{1}{6}(q+4m)\pi]\equiv\phi_{q+4m}(\varphi_{q+4m})\equiv P(4m)\,\phi(\varphi_q),$$
$$P^-\phi_q(\varphi_0+\tfrac{1}{6}q\pi)=\phi_q[(\tfrac{1}{6}\pi-\varphi_0)+\tfrac{1}{6}(q+4m-1)\pi]\equiv\phi_{q+4m-1}(\bar\varphi_{q+4m-1})\equiv P(4m-1)\,\phi(\varphi_q), \tag{42}$$

where again we have chosen $\phi_q=\phi_{q'}$, and, at the same time, given a definition of $\bar\varphi_p=(\pi/6-\varphi_0)+p/6\pi$. We see then that starting with any ϕ_{2n} $(n=0,1,2)$ via $P^\pm$, we generate all R states and that starting with ϕ_{2n-1}, we generate all L states in such a way that a symmetrized version of our problem becomes a linear superposition of all states given by (39).

Having defined $P(n)$ in (42) and recalling that symmetrized R states are connected to L states via parity, we may write the indistinguishable particle solution of this problem in a rather compact form, that is,

$$\Psi(r,\varphi)=\sum_{p=0}^{2}\sum_{s=0}^{1}[P(4p)\pm P(4p-1)]\,(\pm)^s\,\Psi(r,\varphi_s), \tag{43}$$

where φ_q, as usual, is different from zero only if $q/6\pi\leq\varphi_q\leq(q+1)/6\pi$; the $\pm$ sign between the P operators is due to statistics (+ for bosons and − for fermions), while $(\pm)^s$ is due to parity. In order to obtain a more explicit form, we notice that

$$P(4p)\,\phi(\varphi_q)=N\sin^{2a}[3\varphi_0+\tfrac{1}{2}(4p+q)\pi]\times\cos^{2b}[3\varphi_0+\tfrac{1}{2}(4p+q)\pi]\,P_l^{2a-1/2,2b-1/2}\cos[6\varphi_0+(4p+q)\pi]$$
$$P(4p-1)\,\phi(\varphi_q)=N(-1)^{1+2a+2b}\cos^{2a}[3\varphi_0+\tfrac{1}{2}(4p+q-1)\pi]\times\sin^{2b}[3\varphi_0+\tfrac{1}{2}(4p+q-1)\pi]\,P_l^{2a-1/2,2b-1/2}[\cos(6\varphi_0+(4p+q-1)\pi)]. \tag{44}$$

We now define

$$\phi_{p,s}(\varphi_0)=N_s\sin^\alpha[3\varphi_0+\tfrac{1}{2}(4p+s)\pi]\cos^\beta[3\varphi_0+\tfrac{1}{2}(4p+s)\pi]\,P_l^{2a-1/2,2b-1/2}[\cos(6\varphi_0+(4p+s)\pi)], \tag{45a}$$

where

$$\alpha=a[1+(-)^s]+b[1-(-)^s],$$
$$\beta=a[1-(-1)^s]+b[1+(-1)^s],$$
$$N_s=(-1)^{(l/2+a+b)[1-(-1)^s]}. \tag{45b}$$

Then we can write for the complete symmetrized solution of the three-body linear problem (3) which is an eigenstate of energy (nl) and parity (π):

$$\Psi_{nl\pi}(r,\varphi)=\chi_{nl}(r)\sum_{p=0}^{2}\left(\sum_{s=0}^{1}(-)^{\pi s}\sum_{s=2}^{3}\right)(\pm)^s\,\phi_{p,s}(\varphi_0), \tag{46}$$

where $\chi(r)$ is given by (12), ϕ_{ps} is given by (45a), $0\leq\varphi_0\leq\pi/6$. $(+)^s$ corresponds to bosons and $(-)^s$ to fermions. Parity is positive for $\pi=0$ and negative for $\pi=1$.

The corresponding energy is given by (41).

V. FINAL REMARKS

Although we realize the lack of realism of the linear model treated in this paper, the fact that an exact solution for a three-body interaction is available may give some insight into more realistic problems.

Extension of this model to two or three dimensions should not be an easy task. Besides, the specific polarization and order effect is certainly a consequence of linearity.

Our intention is to extend this model to four and, if possible, to n particles. This has been done by Calogero for the harmonic and inverse square potential.[3]

ACKNOWLEDGMENTS

It is a real pleasure to acknowledge the constant advice and critical discussions of G.C. Ghirardi. I wish to thank M. de Llano and R. Avalos for helpful comments. I also wish to thank Professor Abdus Salam and Professor P. Budini as well as the International Atomic Energy Agency and UNESCO for hospitality at the International Centre for Theoretical Physics, Trieste.

*On leave of absence from Facultad de Ciencias, Universidad de Chile, Santiago, Chile.

[1](a) For the δ-function potential see H.M. Nussenzweig, Proc. Roy Soc. (London) A 264, 408 (1961). (b) For the inverse square and harmonic potential see F. Calogero, J. Math. Phys. 10, 2191 (1969).
[2]L. Fonda and G.-C. Ghirardi, *Symmetry Principles in Quantum Physics* (Decker, New York, 1970).
[3]F. Calogero, J. Math. Phys. 12, 419 (1971).

Exact solution of a one-dimensional three-body scattering problem with two-body and/or three-body inverse-square potentials

F. Calogero

Istituto di Fisica, Università di Roma, 00185 Rome, Italy

Istituto Nazionale di Fisica Nucleare, Sezione di Roma, Rome, Italy

C. Marchioro*

Istituto di Matematica, Università di Roma, Rome, Italy

(Received 7 March 1973)

The exact solution is presented of the scattering problem of three equal particles interacting in one-dimension via two- and/or three-body inverse-square potentials. Both the classical and the quantal problems are treated. It is shown that the outcome of this scattering problem is an extremely simple relation between initial and final momenta, the latter being univocally determined by the former even in the quantal case. The solvability of the problem, and the simple results just mentioned, are peculiar to the equal particle case.

1. INTRODUCTION

Four years ago the one-dimensional quantal problem of three equal particles interacting pairwise via quadratic ("harmonical") *and* inversely quadratic ("centrifugal") pair potentials was solved, namely all its eigenfunctions were explicitly exhibited, together with the corresponding eigenvalues.[1] The spectrum of this problem, that is, of course, discrete (in the c.m. frame) since the harmonical potentials prevent the particles from escaping to infinity, turns out to be extremely simple; in fact it coincides, except for a constant shift of all energy levels, with the spectrum of the identical-particle problem with harmonical forces only.[2] If the harmonical potential is instead absent, the spectrum is continuous, and only scattering states exist. This scattering problem has also been solved and the following surprisingly simple result has been obtained: an ingoing scattering configuration, characterized by (initial) momenta p_i, $i=1,2,3$, goes over into a unique outgoing configuration, characterized by (final) momenta p_i', with $p_i'=p_{4-i}$.[3] This has been proved in the quantal case, the proof being indeed simpler in this case than in the classical case. The same outcome obtains, of course, in the classical case, independently of the initial position of the incoming particles; this has been explicitly proved, and a simple relation has also been obtained between the next-to-leading terms in the asymptotic expressions for the positions of the particles.[3]

These remarkably simple results are a peculiarity of the case with equal particles (i.e., equal masses, and equal strengths of all pair potentials).[4] One conjectures them to originate from an underlying group-theoretical structure, but this has not yet been fully brought to light.[5] Such a conjecture is supported by the observation that the simple results described thus far for the three body problem are also valid (with obvious extensions) in the N-body case.[2]

Quite recently Wolfes has shown that a similar, but more general, one-dimensional 3-body model, characterized by quadratic ("harmonical") pair potentials, and by inversely quadratic ("centrifugal") pair potentials *and/or* by inversely quadratic three-body potentials (whose exact structure is specified below), is also amenable to exact solution.[6] The presence of the harmonical potential guarantees that the spectrum is purely discrete; and again this spectrum turns out to be simply related to the spectrum of the problem with only harmonical forces, this being certainly again a peculiarity of the case with equal particles. If the harmonical potential is missing the spectrum becomes continuous and only scattering states exist. The quantal scattering problem can be solved exactly, exploiting the eigenfunctions explicitly given by Wolfes[6]; and again surprisingly simple results obtain, the property being preserved, that any ingoing scattering configuration, characterized by (initial) momenta p_i, $i=1,2,3$, goes over into a unique outgoing configuration, characterized by momenta p_i'. If both two- and three-body potentials are present, the rule relating p_i' to p_i is simply $p_i'=-p_i$; if only the three-body potential is present, the rule reads instead $p_1'=-p_2$, $p_2'=-p_1$, $p_3'=-p_3$, the 3 particles being labeled so that initially (namely, in the asymptotic past) particles 1 and 2 are closer to each other than to particle 3 (this property coincides with the requirement that in the c.m. system, the initial momenta of particles 1 and 2, p_1 and p_2, have the same sign, while the momentum p_3 of particle 3 has the opposite sign, so that the CM condition $p_1+p_2+p_3=0$ implies $|p_3|=|p_1|+|p_2|=|p_1+p_2|$). These results are proved below, first in the quantal case, and then in the classical case (when simple relations are also obtained between the next-to-leading terms in the asymptotic expressions for the positions of the particles). They are again peculiar to the equal particle problem, although an explicit proof of this will not be given here.

2. THE SCATTERING PROBLEM IN THE QUANTAL CASE

The quantal problem is characterized by the Hamiltonian[7]

$$H=-\frac{\hbar^2}{(2m)}\sum_{i=1}^{3}\frac{\partial^2}{\partial x_i^2}+\sum_{i=1}^{3}[g(x_i-x_{i+1})^{-2}+3f(x_i+x_{i+1}-2x_{i+2})^{-2}]. \tag{2.1}$$

The coordinate x_i indicates the position of the ith particle, with the cyclic convention

$$x_{i+3}\equiv x_i. \tag{2.2}$$

We always assume validity of the inequalities

$$g > -\hbar^2/(4m), \quad f > -\hbar^2/(4m) \tag{2.3}$$

that are required to prevent collapse.[1,6]

We shall work throughout in the c.m. frame, taking moreover, for simplicity, the origin of the x axis to coincide with the position of the center of mass of the three-body system, so that

$$x_1 + x_2 + x_3 = 0. \tag{2.4}$$

Note that in this reference frame the "three-body" potential in the Hamiltonian becomes simply $\frac{1}{3} f \sum_{i=1}^{3} x_i^{-2}$.

It is convenient to introduce the "polar" coordinates r and φ setting

$$x_1 - x_2 = \sqrt{2} r \sin\varphi, \tag{2.5a}$$

$$x_1 + x_2 - 2x_3 = \sqrt{6} r \cos\varphi. \tag{2.5b}$$

These equations, together with (2.4), imply the relations

$$x_{i+1} - x_{i+2} = \sqrt{2} r \sin[\varphi + i(2\pi/3)], \quad i = 1, 2, 3, \tag{2.6a}$$

$$x_{i+1} + x_{i+2} - 2x_{i+3} = \sqrt{6} r \cos[\varphi + i(2\pi/3)], \quad i = 1, 2, 3, \tag{2.6b}$$

$$x_i = -\sqrt{2/3}\, r \cos[\varphi + i(2\pi/3)], \quad i = 1, 2, 3. \tag{2.7}$$

In the polar coordinates the Hamiltonian reads (after elimination of the CM part)[1,6]

$$H = -[\hbar^2/(2m)]\left(\frac{\partial^2}{\partial r^2} + r^{-1}\frac{\partial}{\partial r}\right) + r^{-2} M \tag{2.8}$$

with

$$M = -\frac{\hbar^2}{(2m)}\frac{\partial^2}{\partial\varphi^2} + \tfrac{9}{2}[g(\sin 3\varphi)^{-2} + f(\cos 3\varphi)^{-2}]. \tag{2.9}$$

The singular nature of the interactions disconnects the wavefunctions (apart from a symmetry requirement in the case of identical particles) in different sectors of configuration space, corresponding to different intervals of values of the "angular" variable φ.[1,6] To discuss the scattering process, we assume for simplicity the particles to be distinguishable; accordingly we consider wavefunctions that differ from zero only in one sector. If $g \neq 0$ and $f \neq 0$, it is sufficient to restrict attention to the sector $0 < \varphi < \pi/6$; this may be replaced by any one of the other 11 sectors, $n\pi/6 < \varphi < (n+1)\pi/6$, $n = 1, 2, \ldots, 11$, by inverting the orientation of the x axis and/or by permuting the particles.[1,6] If instead $g = 0$, $f \neq 0$, attention can be restricted to the sector $-\pi/6 < \varphi < \pi/6$, the other 5 sectors, $(2n-1)\pi/6 < \varphi < (2n+1)\pi/6$, $n = 1, 2, \ldots, 5$, being obtainable in a similar fashion. Finally, if $g \neq 0$, $f = 0$, one considers the sector $0 < \varphi < \pi/3$; although this case has already been solved,[3] we shall report here, for completeness, also the results appropriate to this case.

It is important to note that Eqs. (2.6) imply that the sector $0 < \varphi < \pi/3$ is characterized by the property

$$x_1 > x_2 > x_3, \tag{2.10c}$$

the sector $-\pi/6 < \varphi < \pi/6$ is characterized by

$$x_1 > x_3, \quad x_2 > x_3, \quad |x_1 - x_2| < x_1 - x_3, \quad |x_1 - x_2| < x_2 - x_3, \tag{2.10b}$$

and the sector $0 < \varphi < \pi/6$ is characterized by the simultaneous validity of Eqs. (2.10b) and (2.10c), implying

$$x_1 > x_2 > x_3, \quad x_1 - x_2 < x_2 - x_3. \tag{2.10a}$$

Thus the problem with $g \neq 0$, $f \neq 0$ is investigated assuming that the middle particle is closer to that on the right than to that on the left, and assigning to the particles the labels 1, 2, 3 from right to left; the singular nature of the interactions guarantees that both properties are preserved throughout the motion. Note that the second property corresponds to the requirement that particle 2 stay to the right of the c.m. of the system; in the c.m. system defined above, this corresponds simply to $x_2 < 0$. As for the problem with $g = 0$ and $f \neq 0$, it is investigated in the case where there are two particles (labeled 1 and 2) to the right of the center of mass of the system, and one (labeled 3) to the left (in the c.m. sector defined above, this corresponds simply to the sector $x_1 > 0$, $x_2 > 0$, $x_3 < 0$). Finally, the problem with $g \neq 0$, $f = 0$ is investigated labeling the particles in increasing order from right to left. These properties are again preserved throughout the motion.

The scattering problem is treated in the time-independent framework. The eigensolutions of the stationary Schrödinger equation,

$$H\psi = E\psi, \tag{2.11}$$

can be written in the separated form

$$\psi = R(r)\Phi(\varphi), \tag{2.12}$$

and it is easily proved that[1,6]

$$R(r) = J_\nu(pr), \tag{2.13}$$

where

$$E = \hbar^2 p^2/(2m) \tag{2.14}$$

and

$$M\Phi(\varphi) = [\hbar^2/(2m)]\nu^2\Phi(\varphi). \tag{2.15}$$

The explicit expressions of ν are[1,6]

$$\nu = 3(2l + a + b + 1), \quad l = 0, 1, 2, \cdots, \quad \text{if } g \neq 0,\ f \neq 0, \tag{2.16a}$$

$$\nu = 3(l + b + \tfrac{1}{2}), \quad l = 0, 1, 2, \cdots, \quad \text{if } g = 0,\ f \neq 0, \tag{2.16b}$$

$$\nu = 3(l + a + \tfrac{1}{2}), \quad l = 0, 1, 2, \cdots, \quad \text{if } g \neq 0,\ f = 0. \tag{2.16c}$$

The corresponding eigenfunctions of M are

$$\Phi_l(\varphi) = (\sin 3\varphi)^{a+1/2}(\cos 3\varphi)^{b+1/2} \times P_l^{(a,b)}(\cos 6\varphi)\theta(\varphi)\theta[(\pi/6) - \varphi] \quad (g \neq 0,\ f \neq 0), \tag{2.17a}$$

$$\Phi_l(\varphi) = (\cos 3\varphi)^{b+1/2} C_l^{(b+1/2)}(\sin 3\varphi) \times \theta[\varphi - (\pi/6)]\theta[(\pi/6) - \varphi] \quad (g = 0,\ f \neq 0), \tag{2.17b}$$

$$\Phi_l(\varphi)=(\sin 3\varphi)^{a+1/2}C_l^{(a+1/2)}(\cos 3\varphi)\times\theta(\varphi)\theta[(\pi/3)-\varphi] \quad (g\neq 0,\ f=0). \tag{2.17c}$$

In these equations

$$a=\tfrac{1}{2}(1+g4m\hbar^{-2})^{1/2}, \tag{2.18a}$$

$$b=\tfrac{1}{4}(1+f4m\hbar^{-2})^{1/2}, \tag{2.18b}$$

$P_l^{(a,b)}$ is a Jacobi polynomial, $C_l^{(c)}$ is a Gegenbauer polynomial, and θ is the usual step function, $\theta(x)=\frac{1}{2}(1+x/|x|)$.

The eigenvalues ν and eigenfunctions Φ of the problem with $g\neq 0$, $f\neq 0$, go over, as $g\to 0$ resp. $f\to 0$, into those eigenvalues and eigenfunctions of the problems with $g=0$ resp. $f=0$ that satisfy the additional condition $\Phi(\bar{\varphi})=0$, with $\bar{\varphi}=0$ resp. $\bar{\varphi}=\pi/6$.[1,6] This corresponds to the known relationship[8]

$$\cos\alpha P_l^{(a,1/2)}(\cos 2\alpha)=C_{2l+1}^{(a+1/2)}(\cos\alpha)\times\{(2l+1)!!\,\Gamma(a+\tfrac{1}{2})/[2^l\Gamma(\alpha+l+\tfrac{3}{2})]\}, \tag{2.19a}$$

or, equivalently,[9]

$$\sin\alpha P_l^{(a+1/2)}(\cos 2\alpha)=C_{2l+1}^{(a+1/2)}(\sin\alpha)\times\{(-)^l(2l+1)!!\,\Gamma(a+\tfrac{1}{2})/[2^l\Gamma(\alpha+l+\tfrac{3}{2})]\}. \tag{2.19b}$$

Let us also recall the explicit expressions of $\sin 3\varphi$ and $\cos 3\varphi$ in terms of the original coordinates:

$$\sin 3\varphi=-\sqrt{2}\,(x_1-x_2)(x_2-x_3)(x_3-x_1)/r^3, \tag{2.20a}$$

$$\cos 3\varphi=(2/27)^{1/2}(x_1+x_2-2x_3)(x_2-x_3-2x_1)\times(x_3-x_1-2x_2)/r^3. \tag{2.20b}$$

Here r is the "radial" coordinate of Eqs. (2.5), that is also expressible, in terms of the original variables, in the manifestly symmetrical form

$$r^2=\tfrac{1}{3}[(x_1-x_2)^2+(x_2-x_3)^2+(x_3-x_1)^2]. \tag{2.21}$$

It is now convenient to introduce the two symmetry operations defined by

$$T^{(2)}x_1=x_2, \quad T^{(2)}x_2=x_1, \quad T^{(2)}x_3=x_3, \tag{2.22a}$$

$$T^{(3)}x_i=-x_{4-i}, \quad i=1,2,3. \tag{2.22b}$$

Clearly they leave invariant the r variable,

$$T^{(n)}r=r, \quad n=2,3, \tag{2.23}$$

while they act as follows on the φ variable:

$$T^{(2)}\varphi=-\varphi, \tag{2.24a}$$

$$T^{(3)}\varphi=(\pi/3)-\varphi, \tag{2.24b}$$

so that

$$T^{(2)}\sin 3\varphi=-\sin 3\varphi, \quad T^{(2)}\cos 3\varphi=\cos 3\varphi, \tag{2.25a}$$

$$T^{(3)}\sin 3\varphi=\sin 3\varphi, \quad T^{(3)}\cos 3\varphi=-\cos 3\varphi. \tag{2.25b}$$

Moreover, and most important, $T^{(2)}$ transforms the interval $-\pi/6<\varphi<\pi/6$ into itself, while $T^{(3)}$ transforms into itself the interval $0<\varphi<\pi/3$. Therefore $T^{(2)}$ can be applied to the angular eigenfunctions (2.17b) of the problem with $g=0$, $f\neq 0$, yielding

$$T^{(2)}\Phi_l=(-)^l\Phi_l, \tag{2.26}$$

while $T^{(3)}$ can be applied to the angular eigenfunctions (2.17c) of the problem with $g\neq 0$, $f=0$, yielding

$$T^{(3)}\Phi_l=(-)^l\Phi_l. \tag{2.27}$$

The scattering problem is now easily treated, in complete analogy to the already known case.[10] The most general eigenfunction of the Hamiltonian (2.1) (in the c.m. frame) is written in the form

$$\psi=\sum_{l=0}^{\infty}c_l J_\nu(pr)\Phi_l(\varphi), \tag{2.28}$$

where the coefficients c_l are complex constants, and p, ν and Φ_l are given by Eqs. (2.14), (2.15) and (2.17). To discuss scattering, only the asymptotic behavior of this wavefunction when all particles are far apart from each other is needed. Then

$$\psi\sim\psi_{\rm in}+\psi_{\rm out}, \tag{2.29}$$

where

$$\psi_{\rm in}=(-2\pi pr)^{-1/2}\sum_{l=0}^{\infty}c_l\exp[-ipr+i\tfrac{1}{2}\pi(\nu+\tfrac{1}{2})]\Phi_l, \tag{2.30}$$

and

$$\psi_{\rm out}=(-2\pi pr)^{-1/2}\sum_{l=0}^{\infty}c_l\exp[ipr-i\tfrac{1}{2}\pi(\nu+\tfrac{1}{2})]\Phi_l \tag{2.31a}$$

$$=(-2\pi\bar{p}r)\sum_{l=0}^{\infty}c_l\exp[-i\bar{p}r+i\tfrac{1}{2}\pi(\nu+\tfrac{1}{2})]\Phi_l\times\exp[-i\pi\nu]. \tag{2.31b}$$

Equation (2.31b) differs only notationally from Eq. (2.31a), with

$$\bar{p}=-p. \tag{2.32}$$

The wavy symbol ~ in Eq. (2.29) and below indicates asymptotic equality, i.e., equality up to corrections of order r^{-2}.

The stationary eigenfunction describing, in the CM frame, the scattering situation is characterized by the asymptotic condition

$$\psi_{\rm in}\sim c\exp\left(i\sum_{j=1}^{3}p_jx_j\right), \tag{2.33}$$

with

$$p^2=\sum_{j=1}^{3}p_j^2 \tag{2.34a}$$

and

$$\sum_{j=1}^{3}p_j=0. \tag{2.34b}$$

The restrictions (2.34) correspond to energy conservation [see Eqs. (2.11) and (2.14)] and to momentum conservation (in the c.m. frame). Additional restrictions must be placed on the initial momenta p_j, in order that the plane wave (2.33) describes an *incoming* scattering state in the sector under consideration for each problem. For completeness we now report these conditions in detail, even though they play no explicit rôle in the proof. We leave it to reader the verify that, in each case, the final momenta p'_j (see below) satisfy the conditions characteristic of an *outgoing* scattering configuration in the same section of configuration space.

An incoming scattering configuration is characterized

by the condition that the particles move freely (because they are still far apart from each other) and approach each other (so that, going backward in time, they do not collide). This requirement, together with the c.m. condition (2.34b) and the conditions (10) characterizing, for each problem, the sector under consideration, imply the following restrictions for the initial momenta: in the case with $g\neq 0$ and $f\neq 0$,

$$p_3 > 0 \geq p_2 > p_1; \tag{2.35a}$$

in the case with $g=0$, $f\neq 0$,

$$p_3 > 0, \quad p_2 \leq 0, \quad p_1 < 0; \tag{2.35b}$$

in the case with $g\neq 0$, $f=0$,

$$p_3 > p_2 > p_1. \tag{2.35c}$$

It is now easy to prove that, if the constants c_l of Eq. (2.28) are chosen so that Eq. (2.30) yields (2.33), then from Eqs. (2.31) there also follows

$$\psi_{\text{out}} = \exp(-i\pi A) c \exp\, i\sum_{j=1}^{3} p'_j x_j \ , \tag{2.36}$$

the constant A and the final momenta p'_j being given by the following prescriptions: in the case $g\neq 0$, $f\neq 0$,

$$A = 3(a+b+1), \quad p'_j = -p_j, \ j=1,2,3\,; \tag{2.37a}$$

in the case $g=0$, $f\neq 0$,

$$A = 3(b+\tfrac{1}{2}), \quad p'_1 = -p_2, \quad p'_2 = -p_1, \quad p'_3 = -p_3; \tag{2.37b}$$

in the case $g\neq 0$, $f=0$,

$$A = 3(a+\tfrac{1}{2}), \quad p'_j = p_{4-j}, \quad j=1,2,3. \tag{2.37c}$$

The proof obtains[10] inserting the explicit expression of ν, Eqs. (2.16), in Eq. (2.31b), applying the transformations $T^{(2)}$ resp. $T^{(3)}$ in the cases $g=0$, $f\neq 0$ resp. $g\neq 0$, $f=0$ [to get rid of the factors $\exp(3i\pi l)=(-)^l$; see Eqs. (2.26) and (2.27) and recall that $T^{(n)}$ do not act on r, Eq. (2.23)], comparing the resulting expression with Eqs. (2.30) and (2.33), and finally using Eqs. (2.22) (unless both $g\neq 0$ and $f\neq 0$) and the definition (2.32).

The "initial" plane wave (2.33) describes, in the sector under consideration in each case, the (free) motion of particle i with momentum p_i, the conditions (2.35) insuring in each case that this motion corresponds to an initial scattering configuration, i.e., one where each particle gets less close to every other particle if time runs backward. The "final" wavefunction (2.36) describes (in each case, in the same sector of configuration space), the (free) motion of particle i with momentum p'_i. The result just proven implies that the stationary eigenfunction of the Hamiltonian H, Eq. (2.1), that is identified by the requirement that its incoming part coincide with Eq. (2.33), contains only the outgoing plane wave (2.36). Thus an initial scattering state, characterized by particle 1 having momentum p_1, particle 2 having momentum p_2 and particle 3 having momentum p_3, can go only into the final state characterized by particle 1 having momentum p'_1, particle 2 having momentum p'_2, particle 3 having momentum p'_3, the values of these momenta being related to those of the initial momenta by the simple rules (2.37). The final momenta satisfy the equations (2.34), implied by energy and momentum conservation; note however that these 2 equations would not be sufficient to determine the 3 momenta p'_i. Indeed, if the 3 particles under consideration were not equal, to an initial ingoing scattering state characterized by given momenta p_i, there would generally correspond a continuum of ∞^1 possible outgoing final states, a (continuous) function giving the probability density that any one of them be the outcome of the scattering process.

3. THE SCATTERING PROBLEM IN THE CLASSICAL CASE

The classical problem is characterized by the Hamiltonian

$$H = (2m)^{-1}\sum_{i=1}^{3} p_i^2 + \sum_{i=1}^{3}\left[\, g(x_i - x_{i+1})^{-2} + 3f(x_i + x_{i+1} - 2x_{i+2})^{-2}\right]. \tag{3.1}$$

To exclude collapse we must now assume the conditions

$$g \geq 0, \quad f \geq 0. \tag{3.2}$$

We shall again investigate the 3 problems characterized by $g>0$, $f>0$, by $g=0$, $f>0$, and by $g>0$, $f=0$, restricting attention in each case to the appropriate sectors of configuration space, introduced in the previous section.

We work again the the CM frame introduced above, and use the "polar" coordinates r and φ, related to the particle coordinates by Eqs. (2.5)–(2.7). The Hamiltonian (3.1) is therefore written in the separated form

$$H = E = p_r^2/(2m) + B^2/r^2, \tag{3.3}$$

$$B^2 = p_\varphi^2/(2m) + \tfrac{9}{2}\left[\, g(\sin 3\varphi)^{-2} + f(\cos 3\varphi)^{-2}\right], \tag{3.4}$$

E and B being two constants of motion.

From these equations, and the explicit forms of p_r and p_φ,

$$p_r = m\frac{dr}{dt}\,, \tag{3.5a}$$

$$p_\varphi = mr^2\frac{d\varphi}{dt}\,, \tag{3.5b}$$

there follows[11]

$$r(t) = \left[(2E/m)(t-t_0)^2 + B^2/E\right]^{1/2} \tag{3.6}$$

and

$$\varphi(t) = \tfrac{1}{3}\left(\arcsin\left\{\left[\Phi^{(1)}(t)\right]^{1/2}\right\}\right), \tag{3.7}$$

with

$$\Phi^{(1)}(t) = \alpha + \beta\sin\left\{\gamma^{(1)} + 6\arctan\left[(t-t_0)/T\right]\right\}, \tag{3.8}$$

where

$$\alpha = \tfrac{1}{2}\left[1 + 9(g-f)/(2B^2)\right], \tag{3.9}$$

$$\beta = \left[\alpha^2 - 9g/(2B^2)\right]^{1/2} \tag{3.10a}$$

$$= \tfrac{1}{2}\left\{\left[1 - 9(g+f)/(2B^2)\right]^2 - gf(9/B^2)^2\right\}, \tag{3.10b}$$

$$T = (m/2)^{1/2}(B/E). \tag{3.11}$$

Note that the definition of B^2, Eq. (3.4), implies that α is positive and β is real (by convention, positive).

In Eq. (3.6) t_0 is the time when $r(t)$ assumes its minimal value $r_0 = BE^{-1/2}$, while the quantity $\gamma^{(1)}$ in Eq.

(3.8) is related to the values of φ at given times by any one of the following relations:

$$\sin 3\varphi(t_0) = [\alpha + \beta \sin\gamma^{(1)}]^{1/2}, \tag{3.12a}$$

$$\sin 3\varphi(\pm\infty) = [\alpha - \beta \sin\gamma^{(1)}]^{1/2}, \tag{3.12b}$$

that incidentally imply $\varphi(+\infty) = \varphi(-\infty)$.

These equations refer to the case with $g > 0$, $f > 0$; in Eq. (3.7), the positive determination of the square root, and the principal determination of the arcsin function, are intended. It is easy to convince oneself that the positivity of both g and f guarantees $\Phi^{(1)}(t)$ to be always positive and less then unity, implying that $\varphi(t)$ is a continuous smooth function of time satisfying the restriction $0 < \varphi(t) < \pi/6$. However, in the limiting form that these equations take for $g = 0$, $f > 0$,

$$\Phi^{(1)}(t) = \tfrac{1}{2}[1 - 9f/(2B^2)](1 + \sin\{\gamma^{(1)} + 6\arctan[(t - t_0)/T]\}) \tag{3.13}$$

is still less than unity, but does reach the value zero; therefore the coordinate $\varphi(t)$, Eq. (3.7), has a discontinuous time derivative at the time $\bar{t}$ defined by

$$\sin\{\gamma^{(1)} + 6\arctan[(\bar{t} - t_0)/T]\} = -1, \tag{3.14}$$

namely at the time $\bar{t}$ such that $\varphi(\bar{t}) = 0$ or, equivalently, $x_1(\bar{t}) = x_2(\bar{t})$. Indeed, this solution describes the motion that obtains if the two-body potential in the Hamiltonian (3.1) is replaced by a zero-range infinitely repulsive pair potential,[12] whose only effect is to exchange the momenta of two particles whenever they encounter (elastic collision), thereby preventing them from overtaking each other.

The motion in the case $g = 0$, $f > 0$ can also be evinced from Eqs. (3.7) and (3.13); it is sufficient to take the appropriate determination of the square root in Eq. (3.7), so that, at the time $\bar{t}$, $\varphi(t)$ changes sign. This prescription is automatically taken care of by replacing Eqs. (3.7) and (3.8) by

$$\varphi(t) = \tfrac{1}{3}\arcsin[\Phi^{(2)}(t)] \tag{3.15}$$

$$\Phi^{(2)}(t) = [1 - 9f/(2B^2)]^{1/2} \times \sin\{\gamma^{(2)} + 3\arctan[(t - t_0)/T]\}. \tag{3.16}$$

Here the principal determination is understood for the arcsin and arctan functions, T is always defined by Eq. (3.11), while $\gamma^{(2)}$ is now related to the values of φ at given times by any one of the following relations:

$$\sin 3\varphi(t_0) = [1 - 9f/(2B^2)]^{1/2} \sin\gamma^{(2)}, \tag{3.17a}$$

$$\sin 3\varphi(\pm\infty) = \mp[1 - 9f/(2B^2)]^{1/2} \cos\gamma^{(2)}, \tag{3.17b}$$

which incidentally imply $\varphi(+\infty) = -\varphi(-\infty)$. Note that now $\varphi(t)$ is again a continuous smooth function of time, but it is restricted by $-\pi/6 < \varphi(t) < \pi/6$. The function $r(t)$ is always given by the same formula, Eq. (3.6).

An analogous discussion can be made for the limiting case $g > 0$, $f = 0$. In this case the function $\Phi^{(1)}(t)$ is always positive, but can reach the value 1. The question that now arises refers to the determination of the arcsin function in Eq. (3.7), rather than the square root (that must always be taken positive). If the principal determination is always maintained, $\varphi(t)$ gets reflected back at the time $\tilde{t}$ defined by

$$\sin\{\gamma^{(1)} + 6\arctan[(t - t_0)/T]\} = 1,$$

namely at the time $\tilde{t}$ such that $\varphi(\tilde{t}) = \pi/6$, or, equivalently, $x_2(\tilde{t}) = \tfrac{1}{2}[x_1(\tilde{t}) + x_3(\tilde{t})] = 0$. On the other hand, the formulas relevant for the motion in the $g > 0$, $f = 0$ case can be obtained by changing appropriately the determination of the arcsin function in Eq. (3.7), as t crosses the value $\tilde{t}$. Alternatively, and more straightforwardly, one can use the equivalent formula[3]

$$\varphi = \tfrac{1}{3}\arccos[\Phi^{(3)}(t)], \tag{3.18}$$

with

$$\Phi^{(3)}(t) = [1 - 9g/(2B^2)]^{1/2} \times \sin\{\gamma^{(3)} - 3\arctan[(t - t_0)/T]\}, \tag{3.19}$$

using now the principal determination for both the arccos and arctan functions. Here T is always defined by Eq. (3.11), but $\gamma^{(3)}$ is related to the values of φ at given times by

$$\cos 3\varphi(t_0) = [1 - 9g/(2B^2)]^{1/2} \sin\gamma^{(3)}, \tag{3.20a}$$

$$\cos 3\varphi(\pm\infty) = \pm[1 - 9g/(2B^2)]^{1/2} \cos\gamma^{(3)}, \tag{3.20b}$$

implying $\varphi(+\infty) = (\pi/3) - \varphi(-\infty)$ [note that, in this case, $0 < \varphi(t) < \pi/3$].

These equations, together with Eqs. (2.7), provide the explicit determination of the motion of each particle, once the constants E, B, t_0, and $\gamma^{(n)}$, are given. These constants, on the other hand, are easily determined if the position and speed of the particles at any one time is given. In a scattering process the "initial" conditions are assigned in the asymptotic past, setting

$$x_i(t) \underset{t\to-\infty}{\to} (p_i/m)t + a_i + O(t^{-1}), \quad i = 1, 2, 3, \tag{3.21}$$

with

$$\sum_{i=1}^{3} p_i = \sum_{i=1}^{3} a_i = 0, \tag{3.22}$$

the last two conditions corresponding to our choice of reference frame, with the center of mass sitting at the origin of the coordinates. Then the constants E, B, and t_0 are given by the explicit formulas

$$E = (2m)^{-1} \sum_{i=1}^{3} p_i^2 = (p_1^2 + p_2^2 + p_1 p_2)/m, \tag{3.23}$$

$$B^2 = p_\varphi^2(-\infty)/(2m) + \tfrac{9}{2}\{g[\sin 3\varphi(-\infty)]^{-2} + f[\cos 3\varphi(-\infty)]^{-2}\}, \tag{3.24}$$

$$t_0 = -[a_1(2p_1 + p_2) + a_2(2p_2 + p_1)]/(2E), \tag{3.25}$$

with

$$\varphi(-\infty) = \arctan[\sqrt{3}(p_1 - p_2)/(p_1 + p_2 - 2p_3)] \tag{3.26}$$

and

$$p_\varphi(-\infty) = \sqrt{3}\, m(p_1 a_2 - p_2 a_1), \tag{3.27}$$

while the constant $\gamma^{(n)}$ is given by Eqs. (3.12b), (3.17b), or (3.20b), depending upon which one of the 3 problems one is considering. There formulas obtain in a straightforward manner from the definitions (3.3), (3.4), (3.5), and (2.5), and provide, together with Eqs. (2.7), (3.6), and (3.7) or (3.15) or (3.18), the complete explicit solutions of our problems. It should, of course, be remembered that the momenta are also assumed to satisfy

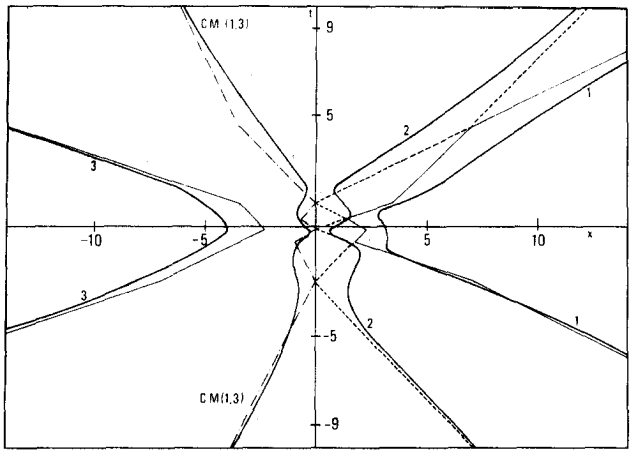


FIG. 1. The classical motion for the three-body problem discussed in this paper. The continuous lines indicate the position of the three particles, and of the center of mass of the two external ones, as a function of time, for $g=f=5$ (with $m=1$, and initial ($t\to-\infty$) conditions $p_1=-2$, $p_2=-1$, $p_3=3$, $a_1=2$, $a_2=-2.5$, $a_3=0.5$). As $g\to 0^+$, $f\to 0^+$, the motion, with the same initial conditions, degenerates into the shown sequence of straight segments, yielding finally the same asymptotic outcome.

the restrictions discussed in the previous section, Eqs. (2.35), that are required in each case in order that in each case the initial configuration belong to the appropriate sector of configuration space, and describe in that sector an incoming scattering state, with all the particles approaching each other.

In the asymptotic future the particles move again freely (but now away from each other):

$$x_i(t) \underset{t\to+\infty}{\to} (p_i'/m)t + a_i' + O(t^{-1}), \quad i=1,2,3. \tag{3.28}$$

It is easy to verify that the quantities p_i' and a_i' are given by the following simple rules: if $g>0$ and $f>0$,

$$p_i' = -p_i, \quad i=1,2,3, \tag{3.29a}$$

$$a_i' = -a_i, \quad i=1,2,3; \tag{3.29b}$$

if $g=0$, $f>0$,

$$p'_1=-p_2, \quad p_2'=-p_1, \quad p_3'=-p_3, \tag{3.30a}$$

$$a'_1=-a_2, \quad a_2'=-a_1, \quad a_3'=-a_3; \tag{3.30b}$$

if $g>0$, $f=0$,

$$p_i' = p_{4-i}, \quad i=1,2,3, \tag{3.31a}$$

$$a_i' = a_{4-i}, \quad i=1,2,3. \tag{3.31b}$$

In fact, these follow rather directly from the equation of motions, even without using the explicit expressions (3.23)—(3.27) of the constants that appear in them.

The staggering simplicity of this result is clearly a peculiarity of the equal particle case; if the masses of the particles were unequal, or the coupling constants of the potentials for different pairs and triplets of particles were different, then both p_i' and a_i' would depend on all the p_j's and the a_j's, and moreover on the values of the masses and of the coupling constants. Of course, even in the equal-particle case, the actual motion does depend on these parameters; it is only the asymptotic free motion, Eq. (3.28), that is independent of (almost all) the parameters, being completely described by the simple rules of Eqs. (3.29)—(3.31). This point is illustrated in Fig. 1, that displays the actual motion in one specific case, for $g>0$ and $f>0$. The motion that would result for different values of g and f (but with the same initial, and therefore final, asymptotics), can be easily inferred from the case displayed, noting that an increase in g and f, implying more repulsion between the particles, has the effect of separating the different trajectories, while on the contrary for very small (positive) values of g and f the distance between adjacent particles, and between each particle and the center of mass of the other two, can become quite small. As g and f tend to zero (through positive values), the motion degenerates into a sequence of free trajectories, separated by sharp collisions between adjacent particles (those on the rhs, in the sector on which we have focused in this paper; see Fig. 1) and between the middle particle and the center of mass of the other two (or, equivalently, the center of mass of the whole system). In the first type of collision, the two particles involved exchange their momenta; in the second type of collision, the middle particle reverses its momentum, and the other two exchange and reverse their momenta. The corresponding motion is also reported in Fig. 1. Note that, if the initial momenta of the particles are p_1, p_2, and p_3, in the limiting motion corresponding to $g\to 0^+$, $f\to 0^+$, no other values of the momenta appear besides these 3 and the three values $-p_1$, $-p_2$, and $-p_3$.[13]

*Research partially supported by the Consiglio Nazionale delle Ricerche (G.N.F.M.).

[1]F. Calogero, J. Math. Phys. 10, 2191 (1969).
[2]F. Calogero, J. Math. Phys. 12, 419 (1971).
[3]C. Marchioro, J. Math. Phys. 11, 2193 (1970).
[4]For an explicit proof, see Appendices D and E of Ref. 2.
[5]Progress in this direction has been made, in the case with the harmonical potential present (discrete spectrum), by A.M. Perelomov, Teor. Mat. Fis. 6, 364 (1971) [Sov. J. Teor. Mat. Fis. 6, 263 (1971)].
[6]J. Wolfes, J. Math. Phys. 15, 1420 (1974).
[7]We use a slightly different notation from that of Wolfes, Ref. 6.
[8]M. Abramowitz and I.A. Stegun, *Handbook of Mathematical Functions* (U.S. Department of Commerce, Washington, D.C., 1965), Eq. 22. 5.21.
[9]The equivalence between Eqs. (2.19a) and (2.19b) obtains from the substitution $\alpha\to\alpha+\pi/2$, using the well-known property of Jacobi polynomials, $P_l^{(a,b)}(-x)=(-)^l P_l^{(a,b)}(x)$.
[10]See Ref. 3 and, more specifically, Sec. 4 of Ref. 2.
[11]The derivation of these results is straightforward; see Ref. 3 and Appendix D of Ref. 2.
[12]Actually, in the sector under consideration, this potential gets a chance to act only between particles 1 and 2.
[13]This remark suggests the possibility to solve the 3-body (and maybe even the N-body) problem with a zero-range interaction as that considered by J.B. McGuire, J. Math. Phys. 5, 622 (1964), but acting not only between pairs.

Wave propagation in a random medium: A complete set of the moment equations with different wavenumbers

L. C. Lee

Downs Laboratory of Physics, California Institute of Technology, Pasadena, California 91109
(Received 7 March 1974)

Propagation of waves in a random medium is studied under the "quasioptics" and the "Markov random process" approximations. Under these assumptions, a Fokker–Planck equation satisfied by the characteristic functional of the random wave field is derived. A complete set of the moment equations with different transverse coordinates and *different* wavenumbers is then obtained from the Fokker–Planck equation of the characteristic functional. The applications of our results to the pulse smearing of the pulsar signal and the frequency correlation function of the wave intensity in interstellar scintillation are briefly discussed.

I. INTRODUCTION

Phenomena such as the twinkling of starlight and the ionospheric, interplanetary, and interstellar radio wave scintillations involve the propagation of an electromagnetic wave in a random medium. A complete statistical description of the wave field requires the solution of all moments of the wave field with different positions and different wavenumbers.

A complete set of the moment equations of the wave field with different transverse coordinates but the *same* wavenumbers has been derived under the "quasioptics" and the "Markov random process" approximations,[1,2] which can be applied to both weak and strong scatterings. However, such a set of the moment equations with the *same* wavenumbers is not sufficient to describe all the statistical properties of the random wave field. Some observed quantities in interstellar scintillations, such as the pulse smearing and the correlation function of the intensity fluctuation with different wavenumbers,[3–5] need the solution of the moment equations with different wavenumbers. It is the purpose of this paper to derive a complete set of the moment equations with different transverse positions and *different* wavenumbers under the quasioptics and the Markov random process approximations. The results reduce to those of Tatarskii[1,2] in the case of the *same* wavenumbers. It is noted that the method of the derivation used here is new, and simpler than that by Tatarskii.[1,2]

It is the idea of Hopf[6] to introduce the "characteristic functional" as an alternative way to describe the complete statistical properties of a random field. In Sec. II, we will derive a Fokker—Planck equation for the characteristic functional of the random electromagnetic field. In Sec. III, a complete set of the moment equations will be derived from the Fokker—Planck equation satisfied by the characteristic functional. Some applications of the results will be briefly discussed in Sec. IV.

II. FOKKER–PLANCK EQUATION FOR THE CHARACTERISTIC FUNCTIONAL OF THE WAVE FIELD

We consider the propagation of a monochromatic wave $E_\omega(\mathbf{r}, t)$ obeying the scalar wave equation

$$\nabla^2\Phi_\omega(\mathbf{r}) + (\omega^2/c^2)\epsilon_\omega(\mathbf{r})\Phi_\omega(\mathbf{r}) = 0, \tag{1}$$

where

$$E_\omega(\mathbf{r}, t) = \Phi_\omega(\mathbf{r})e^{-i\omega t}. \tag{2}$$

$\Phi_\omega(\mathbf{r})$ may be regarded as a Fourier component in time of a general wavefunction. Here $(\omega/2\pi)$ is the frequency of the monochromatic wave, c is the speed of light, and $\epsilon_\omega(\mathbf{r})$ is the refractive index of the medium in which the wave propagates.

The refractive index $\epsilon_\omega(\mathbf{r})$ is a random function and depends on both the position $\mathbf{r}$ and the wave frequency ω. As an example, we will consider in this paper the propagation of the high frequency waves with $\omega \gg \omega_p$, the plasma frequency of the medium, in the plasma medium. This applies to the propagation of the radio waves in the ionosphere, the interplanetary space, or the interstellar medium. If N_e is the electron density, then we have

$$\epsilon_\omega(\mathbf{r}) = 1 - \omega_p^2/\omega^2 \tag{3}$$

and

$$\omega_p^2 = 4\pi N_e e^2/m, \tag{4}$$

where m is the mass and e is the charge of an electron.

Now N_e and $\epsilon_\omega(\mathbf{r})$ fluctuate irregularly. Let $\langle\ \rangle$ denote an average over an ensemble of propagation volumes. Then define

$$\langle\epsilon_\omega(\mathbf{r})\rangle = \epsilon_{\omega 0}(\mathbf{r}),$$
$$N_e(\mathbf{r}) = \langle N_e(\mathbf{r})\rangle + \delta N_e(\mathbf{r}),$$
$$\beta(\mathbf{r}) = -4\pi e^2\delta N_e(\mathbf{r})/mc^2.$$

We have

$$\nabla^2\Phi_\omega(\mathbf{r}) + k^2[1 + \beta(\mathbf{r})/k^2]\Phi_\omega(\mathbf{r}) = 0, \tag{5}$$

where now $\beta(\mathbf{r})$ is a wave-frequency independent random variable with zero mean and where the wavenumber $k = (\omega/c)\sqrt{\epsilon_{\omega 0}}$.

It is useful to define

$$\Phi_\omega(\mathbf{r}) = u(k, \mathbf{r})e^{ikz}, \tag{6}$$

from which we obtain

$$2ik\frac{\partial u(k,\mathbf{r})}{\partial z} + \left(\frac{\partial^2}{\partial z^2} + \frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2}\right)u(k,\mathbf{r}) + \beta(\mathbf{r})u(k,\mathbf{r}) = 0. \tag{7}$$

Let

$$\mathbf{r} = (z, \rho), \quad \rho = (x, y), \quad \text{and} \quad s = (\rho, k).$$

In order to proceed further, we will make two assumptions about the wave equation and the properties of the medium.

First, we assume that the term $\partial^2 u/\partial z^2$ in Eq. (7) can be neglected. This is called the "quasioptics" approximation or "parabolic" approximation. Physically this assumption is equivalent to neglect the reflected wave since the equation has been reduced to one with a first-order derivative in z from the one with a second-order derivative. Thus we have

$$\frac{\partial}{\partial z}u(z,\rho,k)+\frac{1}{2ik}\nabla_\rho^2 u(z,\rho,k)+\frac{1}{2ik}\beta(z,\rho)u(z,\rho,k)=0, \quad (8)$$

where

$$\nabla_\rho^2=\partial^2/\partial x^2+\partial^2/\partial y^2.$$

Second, we assume that $\beta(z,\rho)$ is delta-correlated in z direction. This is called the Markov random process approximation. As we can see later, this is equivalent to assume that the correlation scale of $\beta(z,\rho)$ in z direction is much less than the correlation scale of the wave field u in z direction. We then have

$$\langle\beta(z,\rho)\beta(z',\rho')\rangle=2\delta(z-z')A(\rho-\rho') \quad (9a)$$

and

$$A(\rho-\rho')=\int_{-\infty}^{z}\langle\beta(z,\rho)\beta(z',\rho')\rangle\,dz'. \quad (9b)$$

Note that the z dependence of $A(\rho)$ is not explicitly expressed for convenience.

The validity of the above two assumptions has been discussed.[2,7] We will only note that the "quasioptics" approximation and the "Markov" approximation can be applied in the strong scattering cases.

It is known that the probability distribution function at time t of a random variable $x(t)$ that satisfies a differential equation of the first order in time with a delta-correlated external random force satisfies the Fokker–Planck equation. In our case, z plays the role of time. However, for a fixed value of z, the random field $u(z,\rho,k)$ does not have just a discrete value but has an infinite number of values and is a function of ρ and k. It is the idea of Hopf[6] to introduce a characteristic functional Ψ to describe the statistical properties of a random field. One defines the characteristic functional as

$$\Psi(z,\nu,\nu^*)=\langle\exp(iR_z)\rangle$$
$$=\langle\exp\{i\int\int[u(z,\rho,k)\nu(\rho,k)+u^*(z,\rho,k)\nu^*(\rho,k)]\,d^2\rho\,dk\}\rangle, \quad (10)$$

where * denotes complex conjugate and the range of integration is over all the allowed values of ρ and k. Here ν and ν^* are treated as independent functions of ρ and k.

It is the purpose of this section to derive a Fokker–Planck equation for the characteristic functional Ψ defined above. Tatarskii[1] derived an equation for the characteristic functional with constant wavenumber k. It is noted that we treat in Eq. (10) the wavenumber k as a variable.

Using $s=(\rho,k)$, we write Eq. (10) as

$$\Psi(z,\nu,\nu^*)=\langle\exp\{i\int[u(z,s)\nu(s)+u^*(z,s)\nu^*(s)]\,ds\}\rangle. \quad (10')$$

We differentiate Eq. (10) with respect to z and obtain by Eq. (8)

$$\frac{\partial}{\partial z}\Psi(z,\nu,\nu^*)=\left\langle\exp(iR_z)i\int\left[\left(\frac{-1}{2ik}\right)[\nabla_\rho^2u(z,s)+\beta(z,\rho)u(z,s)]\nu(s)+\left(\frac{1}{2ik}\right)[\nabla_\rho^2u^*(z,s)+\beta(z,\rho)u^*(z,s)]\nu^*(s)\right]ds\right\rangle. \quad (11)$$

First we calculate the terms $\langle\exp(iR_z)\nabla_\rho^2u(z,s)\rangle$ and $\langle\exp(iR_z)\nabla_\rho^2u^*(z,s)\rangle$ in Eq. (11). From Eq. (10), we have

$$\frac{\delta\Psi(z,\nu,\nu^*)}{d\nu(s)}=i\langle u(z,s)\exp(iR_z)\rangle \quad (12a)$$

and

$$\frac{\delta\Psi(z,\nu,\nu^*)}{\delta\nu^*(s)}=i\langle u^*(z,s)\exp(iR_z)\rangle. \quad (12b)$$

The operators $\delta/\delta\nu(s)$ and $\delta/\delta\nu^*(s)$ denote functional derivatives.[6,8] Operating ∇_ρ^2 on Eqs. (12a) and (12b), we have respectively

$$\langle\nabla_\rho^2u(z,s)\exp(iR_z)\rangle=\frac{1}{i}\nabla_\rho^2\frac{\delta\Psi(z,\nu,\nu^*)}{\delta\nu(s)} \quad (13a)$$

and

$$\langle\nabla_\rho^2u^*(z,s)\exp(iR_z)\rangle=\frac{1}{i}\nabla_\rho^2\frac{\delta\Psi(z,\nu,\nu^*)}{\delta\nu^*(s)}. \quad (13b)$$

Next we consider the other terms in Eq. (11), namely, $\langle\exp(iR_z)\beta(z,\rho)u(z,s)\rangle$ and $\langle\exp(iR_z)\beta(z,\rho)u^*(z,s)\rangle$. We define

$$g(\nu,\nu^*,z,s)=\langle\exp(iR_z)\beta(z,\rho)\rangle. \quad (14)$$

Expand $\exp(iR_z)$ in power series as follows:

$$\exp(iR_z)=\sum_{m=0}^{\infty}\frac{1}{m!}\{i\int[u(z,s)\nu(s)+u^*(z,s)\nu^*(s)]\,ds\}^m. \quad (15)$$

Then we have

$$g(\nu,\nu^*,z,s)=\sum_{m=0}^{\infty}\frac{i^m}{m!}\langle[\int(u_1\nu_1+u_1^*\nu_1^*)\,ds_1]\times[\int(u_2\nu_2+u_2^*\nu_2^*)\,ds_2]\cdots[\int(u_m\nu_m+u_m^*\nu_m^*)\,ds_m]\beta(z,\rho)\rangle, \quad (16)$$

where we define $s_i=(\rho_i,k_i)$, $\nu_i=\nu(s_i)$, $u_i=u(z,s_i)$, and etc. for $i=1,2,3,\cdots$. In the expansion of Eq. (16), the existence of moments of all orders is assumed.

Consider now the term in Eq. (16) like $\langle u_1^{\alpha_1}u_2^{\alpha_2}\cdots u_m^{\alpha_m}\beta\rangle$, where $u_i^{\alpha_i}$ denotes either u_i or u_i^*. From Eq. (8), we may write $u(z,s)$ as

$$u(z,s)=u(0,s)+\frac{i}{2}\int_0^z\frac{1}{k}\times[\nabla_\rho^2u(z',s)+\beta(z',\rho)u(z',s)]\,dz'. \quad (17)$$

Note that $u(z,s)$ does not depend on $\beta(z',s)$ for $z'>z$. Let Δz be an increment in z, which is larger than the correlation scale of $\beta(z,\rho)$ in z direction, and write

$$u(z,s)=u(z-\Delta z,s)+\frac{i}{2}\int_{z-\Delta z}^{z}\frac{i}{k}\times[\nabla_\rho^2u(z',s)+\beta(z',\rho)u(z',s)]\,dz', \quad (18)$$

where $u(z-\Delta z, s)$ has no correlation with $\beta(z,\rho)$. Suppose Δz is small, and expand $u(z,s)$ as

$$u(z,s)=u(z-\Delta z,s)+\frac{i}{2}\left(\frac{\Delta z}{k}\right)\nabla_\rho^2 u(z-\Delta z,s)$$
$$+\frac{i}{2}\frac{u(z-\Delta z,s)}{k}\int_{z-\Delta z}^{z}\beta(z',\rho)\,dz'+O(\Delta^2 z). \tag{19}$$

Under the Markov approximation, the correlation scale of $\beta(z,\rho)$ in z direction is zero. Therefore, we let $\Delta z \to 0$. We note that

$$\lim_{\Delta z\to 0} u(z-\Delta z,s)=u(z,s) \tag{20}$$

and

$$\langle\beta(z,\rho')\int_{z-\Delta z}^{z}\beta(z',\rho)\,dz'\rangle=A(\rho-\rho'). \tag{21a}$$

For higher moments such as

$$T_i=\langle\beta(z,\rho)\int_{z-\Delta z}^{z}\beta(z_1,\rho_1)\,dz_1\cdots$$
$$\times\int_{z-\Delta z}^{z}\beta(z_i,\rho_i)\,dz_i\rangle,\quad i\geq 2,$$

we will assume as in the derivation of ordinary Fokker–Planck equation[9]

$$\lim_{\Delta z\to 0} T_i=0,\quad i\geq 2. \tag{21b}$$

This assumption can be satisfied if the random function $\beta(z,\rho)$ has a Gaussian, or normal statistics. However, the assumption made in (21b) is more general and does not require the Gaussian statistics of $\beta(z,\rho)$ in general.

It follows directly from Eqs. (16), (19), (20), (21a), and (21b) that, as $\Delta z\to 0$,

$$\langle u(z,s_1)\beta(z,\rho)\rangle=(i/2k_1)\langle u(z,s_1)\rangle A(\rho-\rho_1)$$

and, in general,

$$\langle(u_1\nu_1+u_1^*\nu_1^*)\cdots(u_m\nu_m+u_m^*\nu_m^*)\beta(z,\rho)\rangle$$
$$=\sum_{j=1}^{m}A(\rho-\rho_j)\left(\frac{i}{2k_j}\right)\langle(u_1\nu_1+u_1^*\nu_1^*)\cdots$$
$$\times(u_{j-1}\nu_{j-1}+u_{j-1}^*\nu_{j-1}^*)(u_j\nu_j-u_j^*\nu_j^*)$$
$$\times(u_{j+1}\nu_{j+1}+u_{j+1}^*\nu_{j+1}^*)\cdots(u_m\nu_m+u_m^*\nu_m^*)\rangle \tag{22}$$

by noting that $\langle u(z-\Delta z,s)\beta(z,\rho')\rangle=0$. Other than the assumption made in (21b), Eq. (22) is exact under the delta-correlation assumption. But we see that really we only require the existence of an intermediate scale Δz which is larger than the coherence scale of $\beta(z,\rho)$ but smaller than the scale of variation of u such that $u(z-\Delta z,s)\approx u(z,s)$. The existence of the intermediate scale and Eq. (21) are the essence of the Fokker-Planck equation.

Substituting Eq. (22) into (16) and noting that all the s_i's are dummy variables, we then have

$$g(\nu,\nu^*,z,s)$$
$$=\sum_{m=1}^{\infty}\left(\frac{i^m}{m!}\right)\left(\frac{i}{2k_m}\right)(m)\int\cdots\int A(\rho-\rho_m)\langle(u_1\nu_1+u_1^*\nu_1^*)\cdots$$
$$\times(u_{m-1}\nu_{m-1}+u_{m-1}^*\nu_{m-1}^*)(u_m\nu_m-u_m^*\nu_m^*)\rangle\,ds_1\cdots ds_m. \tag{23}$$

We can also write Eq. (23) as

$$g(\nu,\nu^*,z,s)$$
$$=\sum_{m=1}^{\infty}\frac{i^{m-1}}{(m-1)!}\left(-\frac{1}{2}\right)\int\cdots\int\frac{1}{k'}A(\rho-\rho')\langle(u'\nu'-u^{*\prime}\nu^{*\prime})$$
$$\times(u_1\nu_1+u_1^*\nu_1^*)\cdots(u_{m-1}\nu_{m-1}+u_{m-1}^*\nu_{m-1}^*)\rangle\,ds'ds_1\cdots ds_{m-1}. \tag{24}$$

where $s'=(\rho',k')$, $u'=u(z,s')$, and $\nu'=\nu(s')$.

Setting $m-1=n$, we have

$$g(\nu,\nu^*,z,s)$$
$$=\sum_{n=0}^{\infty}\frac{(i)^n}{n!}\left(\frac{-1}{2}\right)\int\cdots\int\frac{i}{k'}A(\rho-\rho')\langle(u'\nu'-u^{*\prime}\nu^{*\prime})$$
$$\times(u_1\nu_1+u_1^*\nu_1^*)\cdots(u_n\nu_n+u_n^*\nu_n^*)\rangle\,ds'ds_1\cdots ds_n. \tag{25}$$

From Eq. (25) it is easy to show

$$g(\nu,\nu^*,z,s)=\left(\frac{-1}{2}\right)\int\frac{ds'}{k'}A(\rho-\rho')[\nu(s')\langle u(z,s')\exp(iR_z)\rangle$$
$$-\nu^*(s')\langle u^*(z,s')\exp(iR_z)\rangle]. \tag{25'}$$

By Eqs. (12a) and (12b), we write g as

$$g(\nu,\nu^*,z,s)$$
$$=\left(\frac{i}{2}\right)\int\frac{ds'}{k'}A(\rho-\rho')\left(\nu(s')\frac{\delta\Psi}{\delta\nu(s')}-\nu^*(s')\frac{\delta\Psi}{\delta\nu^*(s')}\right). \tag{26}$$

Define the operator $\hat{M}(s)$ as

$$\hat{M}(s)=\nu(s)\frac{\delta}{\delta\nu(s)}-\nu^*(s)\frac{\delta}{\delta\nu^*(s)}. \tag{27}$$

We then have

$$g(\nu,\nu^*,z,s)=\left(\frac{i}{2}\right)\int\frac{ds'}{k'}A(\rho-\rho')\hat{M}(s')\Psi(z,\nu,\nu^*). \tag{28}$$

We also note that

$$\langle\beta(z,\rho)u(z,s)\exp(iR_z)\rangle=\frac{1}{i}\frac{\delta g(\nu,\nu^*,z,s)}{\delta\nu(s)} \tag{29a}$$

and

$$\langle\beta(z,\rho)u^*(z,s)\exp(iR_z)\rangle=\frac{1}{i}\frac{\delta g(\nu,\nu^*,z,s)}{\delta\nu^*(s)}. \tag{29b}$$

By Eqs. (11), (13a), (13b), (28), (29a), and (29b), we obtain

$$\frac{\delta\Psi(z,\nu,\nu^*)}{\delta z}=\left(\frac{i}{2}\right)\int\frac{ds}{k}\left(\nu(s)\nabla_\rho^2\frac{\delta\Psi}{\delta\nu(s)}-\nu^*(s)\nabla_\rho^2\frac{\delta\Psi}{\delta\nu^*(s)}\right)$$
$$-\frac{1}{4}\iint\frac{ds\,ds'}{kk'}A(\rho-\rho')\hat{M}(s)\hat{M}(s')\Psi. \tag{30}$$

This is the Fokker–Planck equation for the characteristic functional Ψ of the random electromagnetic field $u(z,\rho,k)$. Since the characteristic functional is the Fourier transform of the probability functional, Eq. (30) is in fact the Fourier transform of the Fokker-Planck equation. Our technique used here can also be applied to the derivation of the Fokker-Planck equation for the ordinary characteristic function of a random function $x(t)$.

III. MOMENT EQUATIONS

We want to derive a complete set of moment equations in this section. First, we expand $\Psi(z,\nu,\nu^*)$ as a power series

$$\Psi(z,\nu,\nu^*)=\sum_{m=0}^{\infty}\sum_{n=0}^{\infty}\frac{i^{m+n}}{m!n!}\left(\int u(z,s)\nu(z,s)\,ds\right)^m \times\left(\int u^*(z,s')\nu^*(z,s')\,ds'\right)^n$$

$$=\sum_{m=0}^{\infty}\sum_{n=0}^{\infty}\frac{i^{m+n}}{m!n!}K_{m,n}(z,\nu,\nu^*), \tag{31}$$

where

$$K_{m,n}(z,\nu,\nu^*)=\int\cdots\int\Gamma_{m,n}(z,s_1,\cdots,s_m;s_1',\cdots,s_n') \times\nu_1\cdots\nu_m\nu_1^{*\prime}\cdots\nu_n^{*\prime}\,ds_1\cdots ds_m ds_1'\cdots ds_n'$$

and

$$\Gamma_{m,n}(z,s_1,\cdots,s_m;s_1',\cdots,s_n')=\langle u_1u_2\cdots u_m u_1^{*\prime}\cdots u_n^{*\prime}\rangle. \tag{32}$$

$\Gamma_{m,n}$ is the m-nth moment of the random field $u(z,s)$. The object of this section is to derive a differential equation satisfied by $\Gamma_{m,n}$.

We note that, for any function $f(s)$ of s, we have

$$\int f(s)\nu(s)\frac{\delta}{\delta\nu(s)}K_{m,n}(z,\nu,\nu^*)\,ds =\int\cdots\int\sum_{i=1}^{m}\Gamma_{m,n}(z,s,\cdots,s_m;s_1',\cdots,s_n')f(s_i) \times\nu_1\cdots\nu_m\nu_1^{*\prime}\cdots\nu_n^{*\prime}\,ds_1\cdots ds_m ds_1'\cdots ds_n' \tag{33a}$$

and

$$\int f(s)\nu^*(s)\frac{\delta}{\delta\nu^*(s)}K_{m,n}(z,\nu,\nu^*)\,ds =\int\cdots\int\sum_{i=1}^{n}\Gamma_{m,n}(z,s_1\cdots s_m;s_1'\cdots s_n')f(s_i') \times\nu_1\cdots\nu_m\nu_1^{*\prime}\cdots\nu_n^{*\prime}\,ds_1\cdots ds_m ds_1'\cdots ds_n'. \tag{33b}$$

From (33a) and (33b) we obtain

$$-\frac{1}{4}\int\int\frac{ds\,ds'}{kk'}A(\rho-\rho')\hat{M}(s)\hat{M}(s')K_{m,n}(z,\nu,\nu^*) =-\frac{1}{4}\int\cdots\int\Bigg(\sum_{i=1}^{m}\sum_{j=1}^{n}\frac{A(\rho_i-\rho_j)}{k_ik_j} -\sum_{i=1}^{m}\sum_{j=1}^{n}\frac{[A(\rho_i-\rho_i')+A(\rho_i'-\rho_i)]}{k_ik_i'} +\sum_{i=1}^{n}\sum_{j=1}^{n}\frac{A(\rho_i'-\rho_j')}{k_i'k_j'}\Bigg)\Gamma_{m,n}\nu_1\cdots\nu_m\nu_1^{*\prime}\cdots\nu_n^{*\prime} \times ds_1\cdots ds_m ds_1'\cdots ds_n'. \tag{34}$$

We also note that

$$\int\nu(s)\nabla_\rho^2\frac{\delta K_{m,n}}{\delta\nu(s)}\,ds =\int\cdots\int(\nabla_1^2+\nabla_2^2+\cdots+\nabla_m^2)\Gamma_{m,n}\nu_1\cdots\nu_m\nu_1^{*\prime}\cdots\nu_n^{*\prime} \times ds_1\cdots ds_m ds_1'\cdots ds_n' \tag{35a}$$

and

$$\int\nu^*(s)\nabla_\rho^2\frac{\delta K_{m,n}}{\delta\nu^*(s)}\,ds =\int\cdots\int(\nabla_1'^2+\nabla_2'^2+\cdots+\nabla_n'^2)\Gamma_{m,n}\nu_1\cdots\nu_m\nu_1^{*\prime}\cdots\nu_n^{*\prime} \times ds_1\cdots ds_m ds_1'\cdots ds_n', \tag{35b}$$

where $\nabla_j^2=\nabla_{\rho_j}^2$ and $\nabla_j'^2=\nabla_{\rho_j'}^2$.

By Eqs. (31), (34), (35a), and (35b), we can write Eq. (30) as

$$\sum_{m=0}^{\infty}\sum_{n=0}^{\infty}\frac{i^{m+n}}{m!n!}\int\cdots\int\Bigg[\frac{\partial\Gamma_{m,n}}{\partial z}-\frac{i}{2}\left(\frac{\nabla_1^2}{k_1}+\cdots+\frac{\nabla_m^2}{k_m}-\frac{\nabla_1'^2}{k_1'}-\cdots-\frac{\nabla_n'^2}{k_n'}\right)\Gamma_{m,n}+\frac{1}{4}\Bigg(\sum_{i=1}^{m}\sum_{j=1}^{m}\frac{A(\rho_i-\rho_j)}{k_ik_j} -\sum_{i=1}^{m}\sum_{j=1}^{n}\frac{(A(\rho_i-\rho_j')+A(\rho_j'-\rho_i))}{k_ik_j'}+\sum_{i=1}^{n}\sum_{j=1}^{n}\frac{A(\rho_i'-\rho_j')}{k_i'k_j'}\Bigg)\Gamma_{m,n}\Bigg]\nu_1\cdots\nu_m\nu_1^{*\prime}\cdots\nu_n^{*\prime}ds_1\cdots ds_m ds_1'\cdots ds_n'=0. \tag{36}$$

Since $\nu(s)$ and $\nu^*(s)$ are arbitrarily defined, the quantity inside the bracket in Eq. (36) must be zero. We have then the following differential equation for the moment function $\Gamma_{m,n}$:

$$\frac{\partial\Gamma_{m,n}}{\partial z}(z,s_1\cdots s_m,s_1'\cdots s_n') =\frac{i}{2}\left[\frac{\nabla_1^2}{k_1}+\cdots+\frac{\nabla_m^2}{k_m}-\frac{\nabla_1'^2}{k_1'}-\cdots-\frac{\nabla_n'^2}{k_n'}\right]\Gamma_{m,n} -\frac{1}{4}\Bigg(\sum_{i=1}^{m}\sum_{j=1}^{m}\frac{A(\rho_i-\rho_j)}{k_ik_j}-\sum_{i=1}^{m}\sum_{j=1}^{n}\frac{[A(\rho_i-\rho_j')+A(\rho_j'-\rho_i)]}{k_ik_j'} +\sum_{i=1}^{n}\sum_{j=1}^{n}\frac{A(\rho_i'-\rho_j')}{k_i'k_j'}\Bigg)\Gamma_{m,n}. \tag{37}$$

It is noted that we can also derive the moment equation (37) directly from the wave equation (8), using the same technique in obtaining Eq. (22). Equation (37) thus gives us a complete set of the moment equations of the random wave field with different transverse coordinates and *different* wavenumbers.

IV. APPLICATIONS

First we note that we have derived a complete set of the moment equations with different transverse coordinates and *different* wavenumbers for the high-frequency waves propagating in a plasma medium. However, we can easily extend the argument to the other cases when the index of refraction $\epsilon_\omega(\mathbf{r})$ has a different frequency dependence.

Next we consider some applications.

A. Identical wavenumbers

When all the wavenumbers are identical, Eq. (37) becomes

$$\frac{\partial \Gamma_{m,n}}{\partial z}(z, \rho_1 \cdots \rho_m, \rho_1' \cdots \rho_n')$$

$$= \frac{i}{2k}(\nabla_1^2 + \cdots + \nabla_m^2 - \nabla_1'^2 - \cdots - \nabla_n'^2)\Gamma_{m,n} - \frac{1}{4k^2} \times \Bigg(\sum_{i=1}^{m}\sum_{j=1}^{n} A(\rho_i - \rho_j) - \sum_{i=1}^{m}\sum_{j=1}^{n}[A(\rho_i - \rho_j') + A(\rho_j' - \rho_i)] + \sum_{i=1}^{n}\sum_{j=1}^{n} A(\rho_i' - \rho_j') \Bigg) \Gamma_{m,n}, \tag{38}$$

which is identical to that obtained by Tatarskii.[2] However, the derivation by Tatarskii requires that the refraction index fluctuations possess Gaussian statistics while we do not require the assumption of Gaussian statistics in our derivation in general.

B. $\Gamma_{1,1}(z, s_1, s_2)$

When $m=1$, and $n=1$, Eq. (37) gives

$$\frac{\partial \Gamma_{1,1}}{\partial z}(z, \rho_1, k_1, \rho_2, k_2) = \frac{i}{2}\left(\frac{\nabla_1^2}{k_1} - \frac{\nabla_2^2}{k_2}\right)\Gamma_{1,1} - \frac{1}{4}\left[\left(\frac{1}{k_1^2} + \frac{1}{k_2^2}\right)A(0) - \frac{2A(\rho_1 - \rho_2)}{k_1 k_2}\right]\Gamma_{1,1}, \tag{39}$$

where $\Gamma_{1,1}(z, \rho_1, k_1, \rho_2, k_2) = \langle u(z, \rho_1, k_1)u^*(z, \rho_2, k_2)\rangle$.

Equation (39) can be used to calculate the mean intensity profile $\langle I(\mathbf{r}, t)\rangle$ at position r. Consider the random wave observed by a detector with a bandwidth function $f_B(k)$. Then we have the total observed wave amplitude $h(z, \rho, t)$ at position z, ρ and time t

$$h(z, \rho, t) = \int_{-\infty}^{\infty} u(z, \rho, k) f_B(k) \exp\{i[kz - \omega(k)t]\}\, dk. \tag{40}$$

The average intensity profile is then

$$\langle I(\mathbf{r}, t)\rangle = \langle h(z, \rho, t)h^*(z, \rho, t)\rangle = \int\int_{-\infty}^{\infty} \langle u(z, \rho, k_1)u^*(z, \rho, k_2)\rangle f_B(k_1) f_B(k_2) \times \exp\{i[k_1 z - \omega(k_1)t]\} \exp\{-i[k_2 z - \omega(k_2)t]\}\, dk_1 dk_2. \tag{41}$$

Thus $\Gamma_{1,1}$ is related to the average intensity profile $\langle I(\mathbf{r}, t)\rangle$ by Eq. (41). Equations (39) and (41) have been applied to calculate the pulse profile of pulsar in interstellar scintillation. The details will be given in a later paper.

C. $\Gamma_{2,2}$

When $m=2$, and $n=2$, Eq. (39) becomes

$$\frac{\partial}{\partial z}\Gamma_{2,2}(z, s_1, s_2, s_3, s_4) = \frac{i}{2}\left(\frac{\nabla_1^2}{k_1} + \frac{\nabla_2^2}{k_2} - \frac{\nabla_3^2}{k_3} - \frac{\nabla_4^2}{k_4}\right)\Gamma_{2,2} - \frac{1}{4}\Bigg[\left(\frac{1}{k_1^2} + \frac{1}{k_2^2} + \frac{1}{k_3^2} + \frac{1}{k_4^2}\right)A(0) + 2\frac{A(\rho_1 - \rho_2)}{k_1 k_2} + 2\frac{A(\rho_3 - \rho_4)}{k_3 k_4} - 2\frac{A(\rho_1 - \rho_3)}{k_1 k_3} - 2\frac{A(\rho_1 - \rho_4)}{k_1 k_4} - 2\frac{A(\rho_2 - \rho_3)}{k_2 k_3} - 2\frac{A(\rho_2 - \rho_4)}{k_2 k_4}\Bigg]\Gamma_{2,2}, \tag{42}$$

where

$$\Gamma_{2,2}(z, s_1, s_2, s_3, s_4) = \langle u(z, s_1)u(z, s_2)u^*(z, s_3)u^*(z, s_4)\rangle. \tag{43}$$

If one sets $s_3 = s_1$, $s_4 = s_2$, and $\rho_1 = \rho_2$, then

$$\Gamma_{2,2}(z, s_1, s_2, s_3, s_4) = \langle |u(z, \rho_1, k_1)|^2 |u(z, \rho_1, k_2)|^2\rangle = \langle I(z, \rho_1, k_1), I(z, \rho_1, k_2)\rangle = P_I(k_1 - k_2). \tag{44}$$

Here I is the intensity and P_I is the correlation function of intensity at different frequencies. Thus $\Gamma_{2,2}$ gives in this special case the intensity correlation function $P_I(k_1 - k_2)$ at a given observation point with different wavenumbers. The intensity correlation function has been measured in interstellar scintillations,[3-5,10] and Eq. (42) provides a theoretical base of interpretation.

ACKNOWLEDGMENTS

I would like to thank Professor J.R. Jokipii for helpful suggestions and discussions, and Professor J. Mathews for reading the manuscript and for his suggestions. This work has been supported, in part, by NASA under Grant NGR-05-002-160 and by the National Science Foundation under Grant GP-395-07.

[1] V.I. Tatarskiĭ, Zh. Eksp. Teor. Fiz. 56, 2106 (1969) [Sov. Phys. JETP 29, 1133 (1969)].
[2] V.I. Tatarskii, "The Effects of the Turbulent Atmosphere on Wave Propagation," Israel Program for Scientific Translation (1971).
[3] P.A.G. Scheuer, Nature 218, 920 (1968).
[4] E.E. Salpeter, Nature 221, 31 (1969).
[5] K.R. Lang, Ap. J. 166, 1401 (1971).
[6] E. Hopf, J. Ratl. Mech. Anal. 1, 87 (1952).
[7] L.C. Lee and J.R. Jokipii, to be published (1974).
[8] V. Volterra, *Theory of Functionals and of Integral and Integro-differential Equations* (Blackie, London and Glasgow, 1930).
[9] Ming Chen Wang and G.E. Uhlenbeck, in *Selected Papers on Noise and Stochastic Processes* edited by N. Wax (Dover, New York, 1954).
[10] B.F. Rickett, Monthly Not. Roy. Astron. Soc. 150, 67 (1970).

On the automorphisms of real Lie algebras

G. Burdet, M. Perrin, and P. Sorba

Centre de Physique Théorique-C.N.R.S., 31 chemin J. Aiguier, 13274 Marseille Cedex 2, France
(Received 26 January 1973; revised manuscript received 10 April 1973)

We establish some properties of automorphisms of real Lie algebras, which in particular allow us to construct the derivation algebra of a Lie algebra from the derivations of its radical. We apply this construction to some familiar kinematical algebras.

I. INTRODUCTION

Up to now the notion of a group of automorphisms of a finite dimensional Lie algebra has practically been used mainly in the simple case of Abelian Lie algebras for which the group of automorphisms is $Gl(n,\mathbb{R})$, where n is the dimension of the real algebra. In the non-Abelian case, two physically interesting examples are well known with the groups of automorphisms of the Poincaré[1] and the Galilei[2] algebras.

Let us add that the knowledge of the Lie algebra of automorphisms, also called the derivation algebra,[3] has been useful to classify, up to a conjugation, the subalgebras of a given Lie algebra,[4] and had led to build a theorem on the derivations of the semidirect sum of two Lie algebras $\mathfrak{A} \square \mathcal{L}$, where $\mathfrak{A}$ is Abelian and $\mathcal{L}$ semi-simple, provided that the decomposition of $\mathfrak{A}$ into invariant subspaces under $\mathcal{L}$ is unique.

In this paper it is proposed to extend the above theorem to the case of the semidirect sum $\mathcal{A}=\mathcal{R} \square \mathcal{L}$, where $\mathcal{R}$ is the radical, i.e., the maximal solvable Lie algebra in $\mathcal{A}$, and $\mathcal{L}$ plays the role of a Levi's factor. From Levi's theorem[3] any Lie algebra admits such a decomposition and it will be shown how it is possible to construct the derivations of $\mathcal{R} \square \mathcal{L}$ using some well-defined derivations of the radical, the action of these derivations being extended to the whole algebra.

To find out the group of derivations of a Lie algebra, we could have used a pure cohomological approach based on the structural theorem between cohomology groups of G. Hochschild and J.P. Serre (Theorem 13 in Ref. 5). In particular this theorem shows that the first cohomology group $H^1(\mathcal{A},A)$, namely the quotient group of derivations by inner derivations, is essentially made of some linear maps defined in the radical into the whole algebra $\mathcal{A}$ considered as a vector space A. But these maps cannot in general be easily found. Moreover it is clear that the knowledge of the first cohomology group does not give immediately neither the structure of the derivation algebra, nor the action of this algebra on $\mathcal{A}$.

This paper is organized as follows:

Section II is devoted to some definitions and useful lemmas we need for the following sections.

The two following sections deal with some properties of $\mathcal{A}$ in connection with the automorphisms. In Sec. III it is shown that the radical $\mathcal{R}$ is stable for all automorphisms of the whole algebra $\mathcal{A}$. This result is essential and leads to the study of the derivations of $\mathcal{A}$ through the derivations of $\mathcal{R}$. The Levi's factor does not possess a so strong property. However $\mathcal{L}$ decomposes as $\mathcal{L}_{\mathrm{I}} \oplus \mathcal{L}_{\mathrm{II}}$ with $\mathcal{L}_{\mathrm{I}}$ acting "effectively" on $\mathcal{R}$ and $\mathcal{L}_{\mathrm{II}}$ commuting with $\mathcal{R}$, and it can be proved (Sec. IV) that the algebras $\mathcal{A}_{\mathrm{I}}=\mathcal{R} \square \mathcal{L}_{\mathrm{I}}$ and $\mathcal{L}_{\mathrm{II}}$ are both stable for the connected part $\mathrm{Aut}^{\circ}(\mathcal{A})$ of the automorphisms of $\mathcal{A}$.

Section V emphasizes the role played by a particular subgroup R_0 of the group of the inner automorphisms of $\mathcal{A}_{\mathrm{I}}$. It follows a fundamental relationship between $\mathrm{Aut}^{\circ}(\mathcal{A}_{\mathrm{I}})$ and a subgroup G of $\mathrm{Aut}^{\circ}(\mathcal{R})$. In Sec. VI this subgroup G is fully characterized which allows to carry out the construction of the derivation algebra.

To illustrate the practical applications of this theorem, some relevant physical examples involving kinematical algebras are studied in Sec. VII.

In another paper[6] it has been shown how the notion of derivation algebra can be used to increase an explicitly time-dependent invariance algebra of a given quantum mechanical system.

II. PRELIMINARY PROPERTIES

All along this paper we are concerned with real Lie algebras.

(A) *Definition*: Let $\mathcal{A}$ be a finite dimensional Lie algebra and a an element of $\mathcal{A}$; we call ideal relative to a the subalgebra $\mathcal{J}_{\mathcal{A}}(a)$ constructed as follows: let us consider the subalgebra $[\Lambda_a^1]$ obtained by closing under the Lie product the vector subspace Λ_a^1 defined as:

$$\Lambda_a^1=\{a_0 \mid \exists\, a' \in \mathcal{A} \quad \text{such that } [a',a]=a_0\}=[\mathcal{A},a] \quad \text{if } a\in[\mathcal{A},a],$$

$$\Lambda_a^1=\{[\mathcal{A},a]\}\oplus\{a\} \quad \text{with } \{a\}=\{\lambda a \mid \lambda\in\mathbb{R}\} \quad \text{if } a\notin[\mathcal{A},a].$$

Then let us form $[\mathcal{A},[\Lambda_a^1]]$; two cases may appear:

(i) $[\mathcal{A},[\Lambda_a^1]]\subset[\Lambda_a^1]$ and $[\Lambda_a^1]$ furnishes the ideal relative to a.

(ii) $[\mathcal{A},[\Lambda_a^1]]\not\subset[\Lambda_a^1]$ then the subalgebra $[\Lambda_a^k]$ can be constructed on the vector subspace Λ_a^k deduced from the set:

$$\{[\Lambda_a^{k-1}]\}\cup\{[\mathcal{A},[\Lambda_a^{k-1}]]\}$$

and the iterative procedure will be stopped at the first stage n for which

$$[\mathcal{A},[\Lambda_a^n]]\subset[\Lambda_a^n],$$

then $\mathcal{J}_{\mathcal{A}}(a)$ must be identified with $[\Lambda_a^n]$.

In the case of a simple algebra the following property can be deduced.

Lemma: Let $\mathcal{S}$ be a simple algebra; then $\mathcal{J}(s)=\mathcal{S}$ for any $s\in\mathcal{S}$.

Indeed the ideal $\mathcal{J}_{\mathcal{S}}(s)$ being different from zero by construction, can only be, in a simple algebra, the whole algebra.

(B) *Lemma*: Let $\mathcal{L}$ be a semisimple algebra; then $[\mathcal{L}, l] \neq 0$ for any $l \in \mathcal{L}$.

The proof of this property is obvious for a simple algebra $\mathcal{S}$, since the elements $s \in \mathcal{S}$ such that $[s, \mathcal{S}] = 0$ would form an ideal in $\mathcal{S}$. We can deduce easily the same property for $\mathcal{L}$ semisimple owing to the decomposition of any semisimple algebra into a direct sum of simple subalgebras.

(C) Let $\mathcal{A}$ be a Lie algebra defined as a semidirect sum $\mathcal{R} \square \mathcal{L}$ and $\mathcal{B}$ a proper subalgebra of $\mathcal{A}$ ($\mathcal{B} \neq 0$). We can define[4]:

$$\mathcal{L}_{\mathcal{B}} = \mathcal{L} \cap \mathcal{B} \quad \text{and} \quad \mathcal{R}_{\mathcal{B}} = \mathcal{R} \cap \mathcal{B}.$$

It can be easily seen that $\mathcal{L}_{\mathcal{B}}$ and $\mathcal{R}_{\mathcal{B}}$ are subalgebras of $\mathcal{B}$:

$$[\mathcal{L}_{\mathcal{B}}, \mathcal{L}_{\mathcal{B}}] \subseteq \mathcal{L}_{\mathcal{B}}, \quad [\mathcal{R}_{\mathcal{B}}, \mathcal{R}_{\mathcal{B}}] \subseteq \mathcal{R}_{\mathcal{B}}$$

and moreover

$$[\mathcal{L}_{\mathcal{B}}, \mathcal{R}_{\mathcal{B}}] \subseteq \mathcal{R}_{\mathcal{B}}$$

which prove that

$$\mathcal{R}_{\mathcal{B}} \square \mathcal{L}_{\mathcal{B}} \subseteq \mathcal{B}.$$

If $\mathcal{R}_{\mathcal{B}} \square \mathcal{L}_{\mathcal{B}} \subset \mathcal{B}$, we denote by $\mathcal{M}_{\mathcal{B}}$ a complementary subspace of $\mathcal{R}_{\mathcal{B}} + \mathcal{L}_{\mathcal{B}}$ in $\mathcal{B}$:

$$\mathcal{B} = \mathcal{R}_{\mathcal{B}} + \mathcal{L}_{\mathcal{B}} + \mathcal{M}_{\mathcal{B}}.$$

One can be easily convinced that any nonzero element m_j of $\mathcal{M}_{\mathcal{B}}$ can be written in an unique way as:

$$m_j = r_j + l_j,$$

with

$$r_j \in \mathcal{R}, \quad r_j \notin \mathcal{R}_{\mathcal{B}}, \quad r_j \neq 0,$$

$$l_j \in \mathcal{L}, \quad l_j \notin \mathcal{L}_{\mathcal{B}}, \quad l_j \neq 0.$$

Let us denote by $\mathcal{R}'_{\mathcal{B}}$ (respectively, $\mathcal{L}'_{\mathcal{B}}$) the subspace spanned by the r_j's (resp. l_j's), then the following property can be shown:

$$[\mathcal{L}_{\mathcal{B}} + \mathcal{L}'_{\mathcal{B}}, \mathcal{L}_{\mathcal{B}} + \mathcal{L}'_{\mathcal{B}}] \subseteq \mathcal{L}_{\mathcal{B}} + \mathcal{L}'_{\mathcal{B}},$$

i.e., that $\mathcal{L}_{\mathcal{B}} + \mathcal{L}'_{\mathcal{B}}$ is a subalgebra of $\mathcal{B}$.

(D) In the algebra $\mathcal{A}$ written as $\mathcal{R} \square \mathcal{L}$, the Levi's factor $\mathcal{L}$ being semisimple can be decomposed into a direct sum of simple subalgebras:

$$\mathcal{L} = \bigoplus_{i=1}^{n} \mathcal{L}_i.$$

It is useful to remark that this sum can be settled as:

$$\mathcal{L} = \left(\bigoplus_{i=1}^{k} \mathcal{L}_i \right) \oplus \left(\bigoplus_{j=k+1}^{n} \mathcal{L}_j \right) = \mathcal{L}_{\mathrm{I}} \oplus \mathcal{L}_{\mathrm{II}},$$

with $\mathcal{L}_i$, $i = 1, \ldots, k$ acting effectively on $\mathcal{R}$, i.e., such that $[\mathcal{L}_i, \mathcal{R}] \neq 0$ and $\mathcal{L}_j$, $j = k+1, \ldots, n$ commuting with $\mathcal{R}$, i.e., such that $[\mathcal{L}_j, \mathcal{R}] = 0$.

It is worth noticing the following property:

Lemma: If $\mathcal{L}$ simple acts effectively on $\mathcal{R}$, then $[l, \mathcal{R}] \neq 0$ for each $l \in \mathcal{L}$.

Proof: Let us consider the set

$$\mathcal{E} = \{ l_0 \in \mathcal{L} \mid [l_0, \mathcal{R}] = 0 \}$$

this set is an ideal in $\mathcal{L}$ since for any $l \in \mathcal{L}$, $l_0 \in \mathcal{E}$, the Jacobi's identity implies $[[l_0, l], r] = 0$ for each $r \in \mathcal{R}$. By hypothesis $\mathcal{E}$ is different from $\mathcal{L}$ and thus can only be zero. QED

(E) We recall here a property given in Ref. 7.

Lemma: Let σ be a representation of a semisimple Lie algebra $\mathcal{L}$ in a finite dimensional vector space V. If W is an invariant subspace of V such that $\dim V = \dim W + 1$ and if $\sigma(l) V \subseteq W$ for all $l \in \mathcal{L}$, then there exists a vector $v \in V$, $v \notin W$ such that $\sigma(l) v = 0$ for all $l \in \mathcal{L}$.

III. STABILITY OF THE RADICAL

Let $\mathcal{A}$ be a real finite dimensional Lie algebra which can be written following the Levi's decomposition as the semidirect sum $\mathcal{R} \square \mathcal{L}$ corresponding to a (nontrivial) homomorphism from $\mathcal{L}$ into $D(\mathcal{R})$ the derivation algebra of $\mathcal{R}$, where $\mathcal{R}$ is the radical, i.e., the maximal solvable ideal and $\mathcal{L}$ a Levi's factor (semisimple subalgebra by definition). Then:

Theorem 1: $\mathcal{R}$ is stable for each automorphism of $\mathcal{A}$, i.e., is a characteristic ideal of $\mathcal{A}$.

Assume that there exists an automorphism σ of $\mathcal{A}$ and an element r of $\mathcal{R}$ such that

$$\sigma(r) = l + r' \quad \text{with } l \in \mathcal{L},\ l \neq 0, \text{ and } r' \in \mathcal{R}, \tag{1}$$

$\mathcal{R}$ being an ideal of $\mathcal{A}$ $[\mathcal{A}, \mathcal{R}] \subseteq \mathcal{R}$ and we have $[\sigma(\mathcal{A}), \sigma(\mathcal{R})] \subseteq \sigma(\mathcal{R})$. But $\sigma(\mathcal{A}) = \mathcal{A}$, by definition, and it follows that

$$[\mathcal{L}, \sigma(\mathcal{R})] \subseteq \sigma(\mathcal{R}),$$

in particular,

$$[\mathcal{L}, l + r'] \subseteq \sigma(\mathcal{R}). \tag{2}$$

Obviously, $\mathcal{L}$ being semisimple can be decomposed into a direct sum of simple algebras $\mathcal{L}_i$:

$$\mathcal{L} = \bigoplus_{i=1}^{n} \mathcal{L}_i.$$

Then the element l used in (1) can be decomposed as

$$l = \sum_{i=1}^{n} l_i \quad \text{with } l_i \in \mathcal{L}_i.$$

Let us choose in this sum $l_j \in \mathcal{L}_j$, $l_j \neq 0$, which is possible since $l \neq 0$.

The relation (2) ensures that

$$[\mathcal{L}_j, l + r'] \subseteq \sigma(\mathcal{R}).$$

Let us remark that

$$[\mathcal{L}_j, l + r'] = [\mathcal{L}_j, l_j + r'].$$

Then, using the lemmas A and B, it is not difficult to deduce that $\mathcal{J}_{\mathcal{A}}(l + r')$, which is included in $\sigma(\mathcal{R})$, contains a subalgebra $\mathcal{B}$

$$\mathcal{B} = \mathcal{L}_{\mathcal{B}} + \mathcal{M}_{\mathcal{B}} + \mathcal{R}_{\mathcal{B}}$$

such that

$$\mathcal{L}_\beta + \mathcal{L}'_\beta = \mathcal{J}_{\mathcal{L}_j}(l_j) = \mathcal{L}_j.$$

Therefore $\mathcal{B}$ is not solvable. But $\mathcal{R}$ solvable implies $\sigma(\mathcal{R})$ solvable and then $\sigma(\mathcal{R})$ contains only solvable subalgebras, which contradicts the above result. So an element $l \neq 0$ cannot appear in (1) and for any automorphism σ of $\mathcal{A}$ we have $\sigma(\mathcal{R}) \subseteq \mathcal{R}$. Since $\sigma(\mathcal{R})$ and $\mathcal{R}$ have the same dimension, it follows that $\mathcal{R}$ is stable under all the automorphisms of $\mathcal{A}$, hence is a characteristic subalgebra of $\mathcal{A}$.[1]

IV. PROPERTIES OF THE LEVI'S FACTOR

Lemma: Let $\pi_{\mathcal{L}}$ be the projection operator mapping $\mathcal{A}$ on $\mathcal{L}$, considered as vector spaces, then $\pi_{\mathcal{L}} \circ \sigma$ is an automorphism of $\mathcal{L}$ for any σ automorphism of $\mathcal{A}$. In particular $\pi_{\mathcal{L}} \circ \sigma$ is an inner automorphism of $\mathcal{L}$ if σ belongs to the connected part $\mathrm{Aut}^\circ(\mathcal{A})$ of the automorphisms of $\mathcal{A}$.

Let us define

$$\sigma(l) = \lambda(l) + \rho(l), \tag{3}$$

with $\lambda(l) \in \mathcal{L}$ and $\rho(l) \in \mathcal{R}$ whenever $l \in \mathcal{L}$ and $\sigma \in \mathrm{Aut}(\mathcal{A})$. By definition

$$[\sigma(l_1), \sigma(l_2)] = \sigma([l_1, l_2]) \text{ for each pair } l_1, l_2 \in \mathcal{L}.$$

This relation can be decomposed into the two following ones:

$$[\lambda(l_1), \lambda(l_2)] = \lambda([l_1, l_2]), \tag{4}$$

$$[\lambda(l_1), \rho(l_2)] + [\rho(l_1), \lambda(l_2)] + [\rho(l_1), \rho(l_2)] = \rho([l_1, l_2]). \tag{5}$$

Equation (4) together with the linearity property of λ ensure us that $\lambda = \pi_{\mathcal{L}} \circ \sigma$ is an homomorphism of $\mathcal{L}$. From Theorem 1 we can deduce that λ is a one-to-one mapping, so that λ is an automorphism of $\mathcal{L}$.

In particular, if σ belongs to the connected part of the group of automorphisms of $\mathcal{A}$, it is easy to show, by using the corresponding derivation, that λ is a connected automorphism of $\mathcal{L}$, and thus an inner automorphism of $\mathcal{L}$, since any connected automorphism of a semisimple Lie algebra is an inner one.

Let us now decompose $\mathcal{A}$ as $\mathcal{A} = (\mathcal{R} \square \mathcal{L}_{\mathrm{I}}) \oplus \mathcal{L}_{\mathrm{II}}$, with $\mathcal{L}_{\mathrm{I}}$ acting effectively on $\mathcal{R}$; we have then:

Lemma 2: The algebras $\mathcal{R} \square \mathcal{L}_{\mathrm{I}} = \mathcal{A}_{\mathrm{I}}$ and $\mathcal{L}_{\mathrm{II}}$ are stable for the connected part of the automorphisms of $\mathcal{A}$. This property remains valid if $\mathcal{L}_{\mathrm{I}} = 0$.

Let us first show that $\sigma(\mathcal{L}_{\mathrm{II}}) = \mathcal{L}_{\mathrm{II}}$, for any $\sigma \in \mathrm{Aut}^\circ(\mathcal{A})$. As $[\mathcal{L}_{\mathrm{II}}, \mathcal{R}] = 0$, and using also the result of the Theorem 1: $\sigma(\mathcal{R}) = \mathcal{R}$ we have

$$[\sigma(\mathcal{L}_{\mathrm{II}}), \sigma(\mathcal{R})] = [\sigma(\mathcal{L}_{\mathrm{II}}), \mathcal{R}] = 0. \tag{6}$$

Let us consider $l_2 \in \mathcal{L}_{\mathrm{II}}$ and form

$$\sigma(l_2) = \lambda(l_2) + \rho(l_2).$$

Equation (6) allows us to write

$$[\lambda(l_2) + \rho(l_2), \mathcal{R}] = 0. \tag{7}$$

Moreover we know, from Lemma 1, that if σ is a connected automorphism of $\mathcal{A}$, λ is an inner automorphism of $\mathcal{L}$, so that[8] $\lambda(l_2) \in \mathcal{L}_{\mathrm{II}}$.

Then it follows from (7) that $\rho(l_2) \in \mathcal{C}(\mathcal{R})$. Thus we can deduce

$$[\sigma(\mathcal{L}_{\mathrm{II}}), \sigma(\mathcal{L}_{\mathrm{II}})] \subseteq \mathcal{L}_{\mathrm{II}}. \tag{8}$$

But $\mathcal{L}_{\mathrm{II}}$ being semisimple, $[\mathcal{L}_{\mathrm{II}}, \mathcal{L}_{\mathrm{II}}] = \mathcal{L}_{\mathrm{II}}$ and also

$$[\sigma(\mathcal{L}_{\mathrm{II}}), \sigma(\mathcal{L}_{\mathrm{II}})] = \sigma(\mathcal{L}_{\mathrm{II}}). \tag{9}$$

Together with (8), Eq. (9) implies $\sigma(\mathcal{L}_{\mathrm{II}}) = \mathcal{L}_{\mathrm{II}}$, since $\sigma(\mathcal{L}_{\mathrm{II}})$ and $\mathcal{L}_{\mathrm{II}}$ have the same dimension.

Consider now the action of σ on $\mathcal{A}_{\mathrm{I}} = \mathcal{R} \square \mathcal{L}_{\mathrm{I}}$. We know from Theorem 1 that $\sigma(\mathcal{R}) = \mathcal{R}$. Moreover for any $l_1 \in \mathcal{L}_{\mathrm{I}}$ we have

$$\sigma(l_1) = \lambda(l_1) + \rho(l_1),$$

with $\lambda(l_1) \in \mathcal{L}_{\mathrm{I}}$ for the reason already mentioned.[8] Then one gets $\sigma(\mathcal{R} \square \mathcal{L}_{\mathrm{I}}) \subseteq \mathcal{R} \square \mathcal{L}_{\mathrm{I}}$ and $\sigma(\mathcal{R} \square \mathcal{L}_{\mathrm{I}}) = \mathcal{R} \square \mathcal{L}_{\mathrm{I}}$.

Let us mention that this proof is valid for $\mathcal{L}_{\mathrm{I}} = 0$. In this case $\mathcal{A} = \mathcal{R} \oplus \mathcal{L}$, $\sigma(\mathcal{R}) = \mathcal{R}$, and $\sigma(\mathcal{L}) = \mathcal{L}$ for any $\sigma \in \mathrm{Aut}^\circ(\mathcal{A})$.

V. ROLE OF THE ALGEBRA $\mathcal{R}_0 \cong \mathcal{C}(\mathcal{R})/\mathcal{C}(\mathcal{A}_{\mathrm{I}})$

Let $\mathcal{R}_0$ be the Abelian subalgebra of the inner derivation algebra $\mathcal{ID}(\mathcal{A}_{\mathrm{I}})$ of $\mathcal{A}_{\mathrm{I}}$ generated from elements of the center $\mathcal{C}(\mathcal{R})$ of $\mathcal{R}$, namely $\mathcal{R}_0 \cong \mathcal{C}(\mathcal{R})/\mathcal{C}(\mathcal{A}_{\mathrm{I}})$.

Lemma 3: The Lie subgroup R_0 of the inner automorphisms of $\mathcal{A}$ generated by the elements of $\mathcal{R}_0$ is an Abelian invariant subgroup in the group $\mathrm{Aut}(\mathcal{A})$ of the automorphisms of $\mathcal{A}$.

Let w_{r_0} be an inner automorphism of $\mathcal{A}$ generated by $r_0 \in \mathcal{R}_0$; its action on $\mathcal{R} \square \mathcal{L}$ is given by

$$w_{r_0}(l) = l + [r_0, l] \quad \forall\, l \in \mathcal{L},$$

$$w_{r_0}(r) = r \qquad \forall\, r \in \mathcal{R}.$$

Consider now an automorphism σ of $\mathcal{A}$ and form the product $\sigma^{-1} \circ w_{r_0} \circ \sigma$. One sees immediately that

$$(\sigma^{-1} \circ w_{r_0} \circ \sigma)(r) = r \quad \forall\, r \in \mathcal{R}. \tag{10}$$

In order to study the action of this product on $\mathcal{L}$ we use the decomposition (3) for σ. It follows for σ^{-1}:

$$\sigma^{-1}(l) = \lambda^{-1}(l) + \rho'(l), \tag{11}$$

such that

$$\rho'(l) + \sigma^{-1}(\rho(l)) = 0 \quad \forall\, l \in \mathcal{L}. \tag{12}$$

A simple calculation gives

$$\begin{aligned}(\sigma^{-1} \circ w_{r_0} \circ \sigma)(l) &= l + [\sigma^{-1}(r_0), l + \rho'(\lambda(l))] \\ &= l + [\sigma^{-1}(r_0), l]\end{aligned} \tag{13}$$

since $\sigma^{-1}(r_0)$ belongs to $\mathcal{C}(\mathcal{R})$ [$\mathcal{C}(\mathcal{R})$ characteristic subalgebra]. Let us notice that, if $\sigma^{-1}(r_0) \in \mathcal{C}(\mathcal{A})$, we obtain the identity automorphism.

From Eqs. (10) and (13) we deduce that

$$\sigma^{-1} \circ w_{r_0} \circ \sigma = w_{\sigma^{-1}(r_0)} \tag{14}$$

which expresses that the w_{r_0}'s form an invariant subgroup we shall denote $\mathcal{R}_0$, in $\mathrm{Aut}(\mathcal{A})$.

Lemma 4: (a) Every nontrivial connected automorphism of $\mathcal{A}_{\mathrm{I}} = \mathcal{R} \square \mathcal{L}_{\mathrm{I}}$ which acts as the identity on $\mathcal{R}$ is an inner automorphism w_{r_0} of $\mathcal{A}_{\mathrm{I}}$, generated from an element r_0 of $\mathcal{R}_0$.

(b) Two connected automorphisms of $\mathcal{A}_I$, σ and σ' whose restrictions on $\mathcal{R}$ are identical, are $\mathcal{R}_0$ equivalent, i.e., there exists an automorphism w_{r_0} generated by an element r_0 of $\mathcal{R}_0$ such that

$$\sigma = \sigma' \circ w_{r_0}.$$

Let w be a connected automorphism of $\mathcal{A}_I$, $w \neq \mathbb{1}_{\mathcal{A}_I}$ such that its restriction to $\mathcal{R}$ is the identity $\mathbb{1}_{\mathcal{R}}$ on $\mathcal{R}$. For any pair $l \in \mathcal{L}_I$, $r \in \mathcal{R}$, the relation

$$[w(l),\ w(r)] = w([l,r])$$

becomes

$$[w(l), r] = [l, r]. \tag{15}$$

As previously, we consider the decomposition

$$w(l) = \lambda(l) + \rho(l), \tag{16}$$

where $\lambda(l) \in \mathcal{L}$ and $\rho(l) \in \mathcal{R}$.

In a first step, let us show that λ must be the identity on $\mathcal{L}_I$. Assume that $\lambda \neq \mathbb{1}_{\mathcal{L}_I}$; therefore $l_I \in \mathcal{L}_I$ exists such that

$$\lambda(l_I) - l_I \neq 0.$$

Since

$$\mathcal{L}_I = \bigoplus_{i=1}^{k} \mathcal{L}_i$$

with $\mathcal{L}_i$ simple, l_I can be decomposed as

$$l_I = \sum_{i=1}^{k} \alpha^i l_i \quad \text{with } l_i \in \mathcal{L}_i.$$

Then using Lemma 1 and Ref. 8, we can write

$$\lambda(l_I) = \sum_{i=1}^{k} \alpha^i \lambda(l_i) \quad \text{with } \lambda(l_i) \in \mathcal{L}_i.$$

In consequence

$$\lambda(l_I) - l_I = \sum_{i=1}^{k} \alpha^i (\lambda(l_i) - l_i) \neq 0,$$

ensures us that there is at least one $l_i \in \mathcal{L}_i$ $(l_i \neq 0)$ such that

$$\lambda(l_i) - l_i \neq 0.$$

Equations (15) and (16) written for l_i lead to

$$[\lambda(l_i) - l_i, r] = [-\rho(l_i), r] \quad \text{for any } r \in \mathcal{R}.$$

Owing to Lemma D of Sec. II,

$$[\lambda(l_i) - l_i, \mathcal{R}] \neq 0.$$

Therefore $\rho(l_I) \neq 0$ and $\rho(l_i) \notin \mathcal{C}(\mathcal{R})$.

Let us then consider the set

$$\mathcal{L}_0 = \{l_0 \in \mathcal{L}_i \mid \exists\ r_0 \in \mathcal{R} \text{ s.t. } [l_0, r] = [r_0, r]\ \forall r \in \mathcal{R}\}.$$

It is easy to prove by repeated use of the Jacobi's identity that $\mathcal{L}_0$ is an ideal in $\mathcal{L}_i$, and that, if r_0 corresponds to $l_0 \in \mathcal{L}_0$, then $[l, r_0]$ corresponds to $[l, l_0]$ for any $l \in \mathcal{L}_i$. Moreover one can verify, owing to the Lemma D, that $\mathcal{L}_0$ is isomorphic to a subalgebra of $\mathcal{R}/\mathcal{C}(\mathcal{R})$. But $\mathcal{L}_i$ simple cannot contain solvable ideal other than zero; hence $\mathcal{L}_0$ is zero and λ cannot be different from $\mathbb{1}_{\mathcal{L}_I}$. Thus Eq. (15) becomes

$$[l + \rho(l), r] = [l, r] \quad \text{for any } r \in \mathcal{R},$$

which implies $\rho(l) \in \mathcal{C}(\mathcal{R})$.

From Theorem 1 we know that the restriction of w to $\mathcal{R}$, denoted $w/\mathcal{R}$, belongs to $\mathrm{Aut}^\circ(\mathcal{R})$; but $\mathcal{C}(\mathcal{R})$ is a characteristic subalgebra of $\mathcal{R}$,[1] thus $w/\mathcal{C}(\mathcal{R})$ belongs to $\mathrm{Aut}^\circ(\mathcal{C}(\mathcal{R}))$. Consequently we can consider the semidirect sum $\mathcal{C}(\mathcal{R}) \square \mathcal{L}_I$ and deduce:

$$w/\mathcal{C}(\mathcal{R}) \square \mathcal{L}_I \in \mathrm{Aut}^\circ(\mathcal{C}(\mathcal{R}) \square \mathcal{L}_I).$$

Consider now

$$w([l, l']) = [w(l), w(l')] \quad \text{for any pair } l, l' \in \mathcal{L}_I. \tag{17}$$

Using the specific form of w, Eq. (17) becomes

$$\rho([l, l']) = [l, \rho(l')] + [\rho(l), l'].$$

This relation is a 1-cocycle equation and also a 1-coboundary since $\mathcal{L}_I$ is semisimple. So that, there exists $r_w \in \mathcal{C}(\mathcal{R})$ such that

$$\rho(l) = [r_w, l], \quad l \in \mathcal{L}_I,$$

and

$$w(l) = l + [r_w, l].$$

It follows that every connected automorphism of $\mathcal{A}_I$ which reduce to the identity on $\mathcal{R}$ is an inner automorphism of $\mathcal{A}_I$ generated by an element of $\mathcal{R}_0$.

To prove the part (b) let us first note that, if σ_1 and σ_2 are two automorphisms of $\mathcal{A}$, then:

$$\sigma_1 | \mathcal{R} \circ \sigma_2 | \mathcal{R} = \sigma_1 \circ \sigma_2 / \mathcal{R}. \tag{18}$$

Consider now the two connected automorphisms σ and σ' whose restrictions on $\mathcal{R}$ are identical. It is easy to go back to (a) by setting

$$\mathfrak{A} = \sigma'^{-1} \circ \sigma$$

which verifies, following (18), the relation:

$$\mathfrak{A}/\mathcal{R} = \sigma'^{-1}/\mathcal{R} \circ \sigma/\mathcal{R} = \mathbb{1}_{\mathcal{R}}.$$

Therefore there exists w_{r_0}, $r_0 \in \mathcal{R}_0$, such that

$$\sigma'^{-1} \circ \sigma = w_{r_0}. \tag{19}$$

Since, from Lemma 3, R_0 is an invariant subgroup of $\mathrm{Aut}(\mathcal{A})$, the relation (19) defines the R_0 equivalence of σ and σ'.

From these two lemmas we deduce:

Theorem 2: The quotient group $\mathrm{Aut}^\circ(\mathcal{A}_I)/R_0$ is isomorphic to a subgroup G of $\mathrm{Aut}^\circ(\mathcal{R})$.

The property given in Lemma 3 allows to consider the quotient group $\mathrm{Aut}^\circ(\mathcal{A}_I)/R_0$ which is, according to Lemma 4, in a one-to-one correspondence φ with a subset of $\mathrm{Aut}^\circ(\mathcal{R})$. Note that φ maps any class $\bar{\sigma} \in \mathrm{Aut}^\circ(\mathcal{A}_I)/R_0$ on the automorphism $\sigma/\mathcal{R}$, restriction to $\mathcal{R}$ of any element of $\bar{\sigma}$. Using the simple property given in Eq. (18) we can deduce that φ is an isomorphism of $\mathrm{Aut}^\circ(\mathcal{A}_I)/R_0$ onto a subgroup G of $\mathrm{Aut}^\circ(\mathcal{R})$.

VI. CONSTRUCTION OF THE DERIVATION ALGEBRA[9]

Let $\mathcal{G}$ be the Lie algebra of the group G defined in Theorem 2, we have:

Theorem 3: The algebra $\mathcal{G}$ can be decomposed into two parts:

(i) The subalgebra $\hat{\mathcal{L}}_I$ which corresponds to the action of $\mathcal{L}_I$ on $\mathcal{R}$ in $\mathcal{A}_I$.

(ii) The subalgebra $\hat{D}$ which contains all the elements $\hat{d}$ of $D(R)$ such that $[\hat{d},\hat{L}_I]\subseteq\hat{R}$, where $\hat{R}$, the algebra of the inner derivations of R ($\hat{R}\cong R/C(R)$) is an ideal in $\hat{D}$. Moreover a basis in $\hat{D}$ can be chosen such that any element of this basis is either an inner derivation, or an outer derivation $\hat{d}'$ satisfying $[\hat{d}',\hat{L}_I]=0$. Hence $\mathcal{G}$ can be written as the semidirect sum $\mathcal{G}=\hat{D}\square\hat{L}_I$ where $\hat{D}$ is not in general a solvable Lie algebra.

It is easily seen first, that $\hat{L}_I$ can be identified with L_I and also that the algebra $\mathcal{J}D(\mathcal{A}_I)$ of inner derivations of $\mathcal{A}_I$ isomorphic to the quotient $R\square L_I/C(R\square L_I)$ exists in $\mathcal{G}$ modulo the Lie algebra R_0. This result can be written: $\mathcal{J}D(\mathcal{A}_I)/R_0\cong\hat{R}\square\hat{L}_I$.

Consider now $\hat{d}\in\mathcal{G}$, $\hat{d}$ belonging to a complementary subspace of $\mathcal{A}_I$ in the vector space $\mathcal{G}$. We shall show that $[\hat{d},L_I]\subseteq\hat{R}$. Indeed to the element $\hat{d}$ corresponds the class $\underline{d}$ in $D(\mathcal{A}_I)/R_0$: $\underline{d}=d+R_0$, $d\in D(\mathcal{A}_I)$. We know that $[d,\mathcal{J}D(\mathcal{A}_I)]\subseteq\mathcal{J}D(\mathcal{A}_I)$. But, more precisely, let d_l be the element of $\mathcal{J}D(\mathcal{A}_I)$ corresponding to the element $l\in L_I$. From the choice of $\hat{d}$ in $\mathcal{G}$, and using Lemma 1 it can be deduced that $d(l')\in R$ for any $l'\in L_I$ Then remembering that the inner derivation d_a acts on a' as:

$$d_a(a')=[a,a']$$

and using the elementary properties of the derivations:

$$[d,d_l](a')=d(d_l(a'))-d_l(d(a))$$

one deduces the action of the derivation $[d,d_l]$ on any element $a'\in\mathcal{A}_I$, one gets

$$\begin{aligned}[d,d_l](a')&=d([l,a'])-[l,d_l(a')]\\&=[d(l),a']\end{aligned}$$

and since $d(l)\in R$, $[d,d_l]$ can therefore be identified with an inner derivation corresponding to an element of R. Going back to $\mathcal{G}$, we easily deduce that

$$[\hat{d},\hat{L}_I]\subseteq\hat{R}. \tag{20}$$

Conversely it is possible to prove that any outer derivation $\hat{d}$ of R verifying (20) is an element of $\mathcal{G}$.

Let us show that to any such outer derivation $\hat{d}$ of R, such that $[\hat{d},\hat{L}]\neq0$, can be associated another outer derivation $\hat{d}'$ which differs from $\hat{d}$ by an inner derivation $\hat{d}_r$ and satisfying

$$[\hat{d}',\hat{L}_I]=0. \tag{21}$$

Indeed, respectively, to each $\hat{d}$ defined above, one can associate the subalgebra $D_{\hat{d}}$ of the derivation algebra $D(R)$ of R, which is generated by $\hat{d}$ and all the inner derivations of $\hat{R}$. Noting that $\hat{R}$ form an ideal in $D(R)$, one can write

$$[\hat{d},\hat{R}]\subseteq\hat{R}. \tag{22}$$

Equations (20) and (22) allow us to consider the following semidirect sum, subalgebra of $D(R)$: $D_{\hat{d}}\square\hat{L}_I$. It is then interesting to notice that, from the definition of a semidirect sum, $D_{\hat{d}}$—considered as vector space—is the representation space for the representation σ of L_I associated with the semidirect sum $D_{\hat{d}}\square L_I$. We are then in the conditions of Lemma E ($D_{\hat{d}}$ playing the role of V and $\hat{R}$ the role of W). In consequence, there exists a vector $\hat{d}'\in D_{\hat{d}}$, $\hat{d}'\notin\hat{R}$ such that

$$\hat{d}'=\hat{d}+\hat{d}_r \text{ with } \hat{d}_r\in\hat{R} \tag{23}$$

and

$$(\sigma(L_I))(\hat{d}')=[L_I,d']=0.$$

Now let us show that any such $\hat{d}'$ belongs to $\mathcal{G}$. Indeed, one can extend the action of $\hat{d}'$ on the whole algebra $\mathcal{A}_I$ by setting

$$\begin{aligned}&d'(r)=\hat{d}'(r) \quad \text{for any } r\in R,\\&d'(l)=0 \qquad \text{for any } l\in L_I.\end{aligned} \tag{24}$$

Then, d' so defined, is a derivation of $\mathcal{A}_I$; from its definition (24) we have just to verify its action on a Lie bracket $[l,r]$, where $l\in L_I$, $r\in R$,

$$d'([l,r])=d'\circ d_l(r)=\hat{d}'\circ d_l(r)=d_l\circ d'(r)=d_l\circ d'(r),$$

and also

$$d'([l,r])=[l,d'(r)].$$

Taking into account the action of d' on L_I, this last relation ensures that d' is a derivation of $\mathcal{A}_I$.

Moreover it is easy to see that d' belongs to the class of $D(\mathcal{A}_I)/R_0$ which corresponds to $\hat{d}'$ by the isomorphism φ defined in Sec. IV. Therefore $\hat{d}'$ is an element of $\mathcal{G}$ and this holds for any $\hat{d}$ satisfying (20) since in the relation (23) and $\hat{d}_r$ is in $\mathcal{G}$.

The algebra $D(\mathcal{A})$

It is easy now to determine $D(\mathcal{A}_I)$ from the knowledge of $\mathcal{G}$. Indeed the subalgebra of inner derivations $\mathcal{J}D(\mathcal{A}_I)$ is *a priori* known $\mathcal{J}D(\mathcal{A}_I)\cong\mathcal{A}_I/C(\mathcal{A}_I)$ and we have just to study the Lie brackets involving outer derivations.

This can be achieved by remembering (proof of Theorem 3) that any $\hat{d}'\in\mathcal{G}$ such that $[\hat{d}',\hat{L}]=0$ can be extended into a derivation d' on $\mathcal{A}_I$, the action of which is given by Eq. (24). Thus, the action of all the elements of a basis in $D(\mathcal{A}_I)$ is known, hence the action of the general element in $D(\mathcal{A}_I)$ is also known, which fully characterizes the derivation algebra since, by definition, for any pair of derivations d_1 and d_2 we have

$$[d_1,d_2](a)=d_1\circ d_2(a)-d_2\circ d_1(a),\quad a\in\mathcal{A}_I.$$

Finally from the property stated in Lemma 2, we deduce the whole algebra

$$D(\mathcal{A})=D(\mathcal{A}_I)\oplus D(L_{II}).$$

We recall that $D(L_{II})\approx L_{II}$, L_{II} being semisimple.

VII. APPLICATIONS TO LIE ALGEBRAS OF PHYSICAL INTEREST

In order to illustrate the above results we study the derivation algebras of some physically relevant Lie algebras. More precisely, we shall pick up examples among kinematical invariance Lie algebras which are endomorphisms of space—time or phase space.

In practice, the knowledge of the derivations of the radical R is not necessary. Indeed, $D(R)$ is a subalgebra of $\mathcal{G}l(n,\mathbb{R})$ ($n=\dim R$) and it is sufficient first to determine the generators of $\mathcal{G}l(n,\mathbb{R})$ which commute with those corresponding to the action of the semisimple

part $\mathcal{L}$ on $\mathcal{R}$, and then to select among these elements those which are derivations of $\mathcal{R}$. This can be achieved by using a nxn matricial representation of $\mathcal{G}l(n,\mathbb{R})$.

A. Algebras of the three-dimensional Euclidian group $E(3)$ and of the Poincaré group P

These two examples can be treated by the theorem of the Appendix B of the Ref. 4 already mentioned in the introduction. Indeed, these two algebras can be written $\mathcal{E}(3)=\mathfrak{A}(3)\square SO(3)$ and $\mathcal{P}=\mathfrak{A}(4)\square SO(3,1)$, respectively, and $\mathfrak{A}(3)$ [respectively, $\mathfrak{A}(4)$] is an irreducible vector space under $SO(3)$ [resp. $SO(3,1)$].

To use our method, first for the Euclidean algebra $\mathcal{E}(3)$, we note that the radical $\mathfrak{A}(3)$ is Abelian and therefore its derivation algebra, which contain only outer derivations is isomorphic to $\mathcal{G}l(3,\mathbb{R})$. Hence we have to select in the algebra of $GL(3,\mathbb{R})$ the generators which commute with those of the algebra $\widehat{SO(3)}$ corresponding by construction to the action on $\mathfrak{A}(3)$ in the semidirect sum $\mathfrak{A}(3)\square SO(3)$. After calculations we find there exists only one such generator which acts as a dilatation on $\mathfrak{A}(3)$. In consequence the derivation algebra can be written

$$D(\mathcal{E}(3))=\mathfrak{A}_D(3)\square(SO_J(3)\oplus\mathcal{R}_D).$$

In the same way, it is easy to see there exists only one generator in $\mathcal{G}l(4,\mathbb{R})$ which commutes with the generators of $SO(3,1)$ and that this generator still acts as a dilatation on $\mathfrak{A}(4)$. So we obtain the well-known result[1]

$$D(\mathcal{P})=\mathfrak{A}(4)\square(SO(3.1)\oplus\mathcal{R}_D).$$

B. Algebra of the isochronous Galilei group G'

This algebra which is the derived Galilei group algebra, can be written

$$\mathcal{G}'=(\mathfrak{A}_P(3)\oplus\mathfrak{A}_K(3))\square SO_J(3).$$

Although the solvable part $\mathfrak{A}$ is Abelian, the theorem of Ref. 4 cannot be applied, the decomposition of $\mathfrak{A}$ into irreducible subspaces under $SO(3)$ being not unique (more precisely, the three-dimensional vector representation appears with the multiplicity two). But from the point of view of our technic we are in the same situation as in Sec. VII A. We have to exhibit in the derivation algebra of $\mathfrak{A}$ isomorphic to $\mathcal{G}l(6,\mathbb{R})$, the maximal subalgebra of the form $\hat{D}\oplus\widehat{SO_J(3)}$, where $\widehat{SO_J(3)}$ still corresponds to the action on $\mathfrak{A}_P(3)\oplus\mathfrak{A}_K(3)$ in the semidirect sum $(\mathfrak{A}_P(3)\oplus\mathfrak{A}_K(3))\square SO_J(3)$. A simple calculation permits to show that $\hat{D}$ is isomorphic to $\mathcal{G}l(2,\mathbb{R})$. Hence the derivation algebra of $\mathcal{G}'$ is

$$D(\mathcal{G}')=(\mathfrak{A}_P(3)\oplus\mathfrak{A}_K(3))\square(SO_J(3)\oplus\mathcal{G}l(2,\mathbb{R})).$$

We shall give more details about this algebra in the Sec. VII D.

C. Algebra of the Galilei group G

$$\mathcal{G}=(\mathfrak{A}_P(3)\oplus\mathfrak{A}_K(3)\square\mathfrak{A}_H(1))\square SO_J(3).$$

The solvable part is no more Abelian. Its derivation algebra is a subalgebra of $\mathcal{G}l(7,\mathbb{R})$ and in this subalgebra there exist two generators, denoted D_1 and D_2, which commute with $SO(3)$ and which correspond to dilatations on the generators $\mathbf{P},\mathbf{K},H$. Thus $D(\mathcal{G})$ can be decomposed as[2]:

$$D(\mathcal{G})=(\mathfrak{A}_P(3)\oplus\mathfrak{A}_K(3)\square\mathfrak{A}_H(1))\square(SO_J(3)\oplus\mathcal{R}_{D_1}\oplus\mathcal{R}_{D_2})$$

and we have the following commutation relations:

$$\begin{aligned}&[D_1,H]=2H, \quad [D_2,H]=0,\\ &[D_1,K_j]=-K_j, \quad [D_2,K_j]=K_j, \quad (j=1,2,3)\\ &[D_1,P_j]=P_j, \quad [D_2,P_j]=P_j,\\ &[D_1,D_2]=0.\end{aligned} \tag{25}$$

D. Algebra of the extended isochronous Galilei group $\tilde{G}'$

To apply our method we consider this algebra written under the form

$$\tilde{\mathcal{G}}'=(\mathfrak{A}_M(1)\oplus\mathfrak{A}_P(3)\square\mathfrak{A}_K(3))\square SO_J(3).$$

The solvable part is the Heisenberg algebra $\mathcal{H}$ of which the derivations are well known.

$$D(\mathcal{H})=(\mathfrak{A}_P(3)\oplus\mathfrak{A}_K(3))\square(\mathrm{Sp}(6,\mathbb{R})\oplus\mathcal{R}_D).$$

The $SO(3)$ is contained into the symplectic algebra $\mathrm{Sp}(6,\mathbb{R})$ and the generators of $D(\mathcal{H})$ which commute with the $\widehat{SO(3)}$ form a subalgebra isomorphic to $\mathcal{G}l(2,\mathbb{R})$. This algebra was already encountered in Sec. VII B where its action was limited on $\mathfrak{A}_P(3)\oplus\mathfrak{A}_K(3)$. We can the write:

$$D(\tilde{\mathcal{G}}')=(\mathfrak{A}_P(3)\oplus\mathfrak{A}_K(3))\square(SO_J(3)\oplus\mathcal{G}l(2,\mathbb{R})).$$

It is interesting to decompose $\mathcal{G}l(2,\mathbb{R})$ into the direct sum:

$$\mathcal{G}l(2,\mathbb{R})\cong SU(1,1)\oplus\mathcal{R}_{D_2}$$

which makes appear the Schrödinger algebra $\mathcal{S}$[10] defined as the largest algebra which leaves invariant the free Schrödinger equation:

$$\mathcal{S}=\mathfrak{A}_P(3)\oplus\mathfrak{A}_K(3)\square SO_J(3)\oplus SU(1,1).$$

This algebra contains the Galilei algebra and the generator H also belongs to the $SU(1,1)$ algebra of which the two other generators are the dilatation D_1 and the "expansion" C. The new generators satisfy the following commutation relations:

$$\begin{aligned}&[C,P_j]=K_j, \quad [C,K_j]=0,\\ &[C,J_j]=0, \quad [C,H]=D_1,\\ &[C,D_1]=2C, \quad [C,D_2]=0.\end{aligned}$$

The other commutation relations are identical to the relations (25). So, we have established that the Schrödinger algebra is the derivation algebra of the extended Galilei group, up to a dilatation which acts on the mass. It may be interesting to notice that:

(i) The $SU(1,1)$ algebra which appears above is in fact the $\mathrm{Sp}(2,\mathbb{R})$ algebra which subsists in $D(\mathcal{H})$ if we consider only a one-dimensional space.

(ii) It is easy to convince himself that following our technics and by its construction the Schrödinger algebra is complete, i.e., it has no center and no outer derivations.

E. Algebra of the extended Galilei group $\tilde{G}$

$$\tilde{\mathcal{G}} = (\mathfrak{A}_H(1) \oplus \mathfrak{A}_M(1) \oplus \mathfrak{A}_P(3) \square \mathfrak{A}_K(3)) \square SO_J(3).$$

This case differs from the case of the Galilei algebra by the appearance of a supplementary generator z, the automorphisms $\sigma_{\alpha z}$ of which act on $\tilde{\mathcal{G}}$ as

$$\sigma_{\alpha z}(H) = H + \alpha M,$$

$$\sigma_{\alpha z}(x) = x, \quad \text{for any } x \in \tilde{\mathcal{G}}'.$$

z commutes with all inner derivations, but with the two dilatations we have the following commutation relations:

$$[D_1, z] = -2z, \quad [D_2, z] = 2z.$$

Hence the derivation algebra of $\tilde{\mathcal{G}}$ can be written

$$\mathcal{D}(\tilde{\mathcal{G}}) = (\mathcal{I}\mathcal{D}(\mathcal{G}) \oplus \mathcal{R}_z) \square (\mathcal{R}_{D_1} \oplus \mathcal{R}_{D_2}).$$

ACKNOWLEDGMENTS

We thank Professor L. Michel for fruitful comments and Ph. Combe, J.L. Richard, and M. Sirugue for many helpful discussions.

[1] L. Michel, *Cargese Lectures in Theoretical Physics*, edited by F. Lurçat (Gordon and Breach, New York, 1967).
[2] J.M. Levy Leblond, *Group Theory and Its Applications*, edited by E.M. Loebl (Academic, New York, 1971), Vol. 2.
[3] N. Jacobson, *Lie Algebras* (Interscience, London, 1965).
[4] H. Bacry, Ph. Combe, and P. Sorba, Preprint Marseille 72, p. 449 (1972).
[5] G. Hochschild and J.P. Serre, Ann. Math. 57, 591 (1953).
[6] G. Burdet, M. Perrin, and P. Sorba, "On the Explicit Time-dependent Invariance Properties of Quantum Mechanical Systems", Preprint 72, p. 499 (1972).
[7] M. Hausner and J. Schwartz, *Lie Groups; Lie Algebras* (Gordon and Breach, New York, 1968), pp. 143–144.
[8] λ being an inner automorphism of $\mathcal{L} = \oplus_{i=1}^{n} \mathcal{L}_i$ there exists $l_0 = \sum_{i=1}^{n} \alpha^i l_i$ with $l_i \in \mathcal{L}_i$ such that $\lambda(l) = \exp(l_0) l \exp(-l_0)$ for any $l \in \mathcal{L}$. In consequence, for any $l'_j \in \mathcal{L}_j$, $\lambda(l'_j) = \exp(\sum \alpha^i l_i) l'_j \times \exp(-\sum \alpha^k l_k) = \exp(\alpha^j l_j) l'_j \exp(-\alpha^j l_j)$, which proves that $\lambda(l'_j) \in \mathcal{L}_j$.
[9] Let $\mathcal{A}$ be a finite-dimensional algebra over the field of real numbers: We denote $\mathcal{D}(\mathcal{A})$ the Lie algebra of the group of automorphisms of $\mathcal{A}$ also called the Lie algebra of derivations or derivation algebra of $\mathcal{A}$. We recall that a derivation of $\mathcal{A}$ is a linear map $d: \mathcal{A} \to \mathcal{A}$ such that

$$d[a, a'] = [d(a), a'] + [a, d(a')], \quad \text{for } a, a' \in \mathcal{A},$$

and that in $\mathcal{D}(\mathcal{A})$ the Lie bracket is defined by

$$[d_1, d_2](a) = d_1(d_2(a)) - d_2(d_1(a)), \quad \text{for } d_1, d_2 \in \mathcal{D}(\mathcal{A}) \text{ and } a \in \mathcal{A}.$$
[10] U. Niederer, Helv. Phys. Acta 45, 5, 802 (1972); G. Burdet and M. Perrin, Nuovo Cimento Lett. 4, 13, 651 (1972).

Spectral representation of the pentagon diagram amplitude*

J. S. Frederiksen

Institute for Theoretical Physics, University of Groningen, The Netherlands
(Received 19 January 1973)

A method developed in two previous papers is used to derive a double spectral representation with Mandelstam boundary for the pentagon diagram amplitude for the production process $A\,B \to C\,D\,N$. Restrictions on the masses and kinematic invariants for which this representation is valid are found and it is discussed how a representation can be obtained for wider ranges of these variables. Finally, a comparison is made with the results of other authors.

1. INTRODUCTION

Different aspects of the properties of the pentagon diagram amplitude or five-point function have been discussed by a number of authors. Cutkosky used the Landau—Cutkosky rules[1,2] to show that, unlike the leading singularities of the triangle and box diagram amplitudes, the leading singularity of the pentagon diagram amplitude is not a branch point. The discontinuity associated with this singularity, as calculated by the Cutkosky rules, is a delta function.[3] Cook and Tarski[4] made a detailed study of the leading Landau curve of the pentagon diagram amplitude and determined the singular points of this amplitude for several specific processes. A reduction formula expressing the pentagon diagram amplitude in terms of five box daigram amplitudes was obtained by Halpern.[5]

The pentagon diagram amplitude has also been studied with a view to writing it as a double spectral representation, for a restricted range of masses and kinematic invariants, by Zav'yalov and Pavlov.[6] Their analysis however contains a number of errors. In particular, the double spectral representation obtained by them [Eq. (23) of Ref. 6] is divergent, that is, infinity is obtained when the integration is carried out. Further, the properties of the roots of the quadratic equation yielding the leading Landau curve of the pentagon diagram amplitude are more complicated than indicated in Ref. 6. The roots can under certain circumstances becomes complex and this is another reason why their spectral representation is incorrect.

In this paper we extend a method used in two previous papers, Ref. 7 (referred to as VF) and Ref. 8 (referred to as I), to obtain a double spectral representation for the pentagon diagram amplitude, for a restricted range of masses and kinematic invariants. (Equations from I will be denoted by placing an I- in front of the equation number).

In Sec. 2, the pentagon diagram amplitude associated with the pentagon diagram in Fig. 1 is transformed from its Feynman parametrized form into a more convenient form and the restrictions made on the values of the masses and kinematic invariants are discussed. The boundary of the region of integration in the quadruple integral obtained in Sec. 2 is studied in Sec. 3 and in Sec. 4 we obtain some results necessary for reversing the order of integration.

The order of integration is reversed in Sec. 5 and a triple integral representation is obtained. In Sec. 6 the boundary of the region of integration in the triple integral is studied and in Sec. 7 results necessary for reversing the order of integration are obtained. Finally, in Sec. 8 the order of integration is reversed and an integration is carried out to obtain a double spectral representation in s and t for the pentagon diagram amplitude. We also note in Sec. 8 that one of the integrations can be carried out to obtain a single dispersion relation in s and in principle the method of Ref. 9 (referred to as II) can be used to obtain a representation for the pentagon diagram amplitude for general physical invariants. Using this method it should be possible to determine directly how and when complex triangle, box, and pentagon singularities occur, resulting in a breakdown of even a single dispersion integral over a real domain.

2. TRANSFORMATION OF THE PENTAGON DIAGRAM AMPLITUDE

With plane wave states normalized so that $\langle \mathbf{p}' | \mathbf{p} \rangle = \delta^{(3)}(\mathbf{p}' - \mathbf{p})$ we define the scalar invariant production amplitude $P(s_1, s_2, s_3, s_4, s_5)$ for the process $AB \to CDN$ in terms of the S-operator by

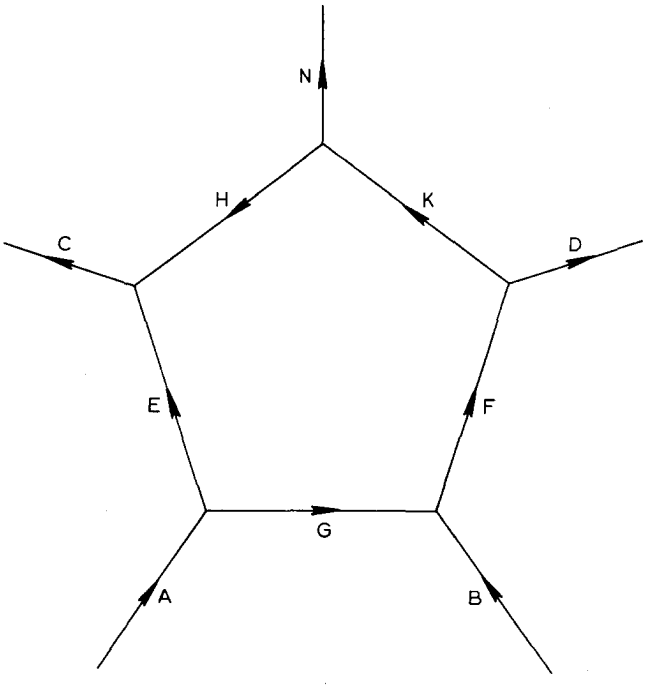


FIG. 1. Pentagon amplitude for the production process $AB \to CDN$.

$$\langle \mathbf{p}_C \mathbf{p}_D \mathbf{p}_N | S | \mathbf{p}_A \mathbf{p}_B \rangle = -i(2\pi)^4 \delta^{(4)}(p_C + p_D + p_N - p_A - p_B)(2\pi)^{-15/2} \times (2E_A)^{-1/2}(2E_B)^{-1/2}(2E_C)^{-1/2}(2E_D)^{-1/2}(2E_N)^{-1/2} \times P(s_1, s_2, s_3, s_4, s_5), \quad (1)$$

where $s_1=(p_A+p_B)^2$, $s_2=(p_A-p_C)^2$, $s_3=(p_B-p_D)^2$, $s_4=(p_D+p_N)^2$, $s_5=(p_C+p_N)^2$ are five independent kinematic invariants. (The notation has been chosen so that the results of I can be applied without the need to relabel the variables.) Then, using standard Feynman rules[10] and the Feynman identity, we find that the amplitude arising from the pentagon diagram of Fig. 1 takes the form

$$P_{\text{pent.}}(s_1, s_2, s_3, s_4, s_5) = -(g/64\pi^2 EFGH) I(x_1, x_2, x_3, x_4, x_5), \quad (2)$$

where, writing $I(x_i)$ for $I(x_1, x_2, x_3, x_4, x_5)$,

$$I(x_i) = -4EFGH \int_0^1 d\alpha \int_0^{1-\alpha} d\beta \int_0^{1-\alpha-\beta} d\gamma \int_0^{1-\alpha-\beta-\gamma} d\delta \times [E^2\alpha + F^2\beta + G^2(1-\alpha-\beta-\gamma-\delta) + H^2\gamma + K^2\delta - A^2(1-\alpha-\beta-\gamma-\delta)\alpha - B^2(1-\alpha-\beta-\gamma-\delta)\beta - C^2\alpha\gamma - D^2\beta\delta - N^2\gamma\delta - s_1\alpha\beta - s_2(1-\alpha-\beta-\gamma-\delta)\gamma - s_3(1-\alpha-\beta-\gamma-\delta)\delta - s_4\beta\gamma - s_5\alpha\delta]^{-3}. \quad (3)$$

In Eq. (3), the s_i are to be expressed in terms of the new variables x_i or X_i defined by

$$x_1 = -X_1 = (2EF)^{-1}(s_1 - E^2 - F^2),$$
$$x_2 = -X_2 = (2GH)^{-1}(s_2 - G^2 - H^2),$$
$$x_3 = -X_3 = (2GK)^{-1}(s_3 - G^2 - K^2),$$
$$x_4 = -X_4 = (2FH)^{-1}(s_4 - F^2 - H^2),$$
$$x_5 = -X_5 = (2EK)^{-1}(s_5 - E^2 - K^2). \quad (4)$$

We shall find it convenient to use both the quantities x_i and X_i $(i=1,\ldots,5)$ in the following. The factor g in Eq. (2) is given by $g = g_{AEG}g_{BFG}g_{CEH}g_{DFK}g_{NHK}$, where $g_{AEG}, \ldots, g_{NHK}$ are the usual rationalized coupling constants.

We begin by generalizing the transformation used in I. The change of variables is

$$\lambda = (\alpha+\beta)^{-1}(1-\alpha-\beta-\gamma-\delta), \quad \mu = (\alpha+\beta)^{-1}\gamma,$$
$$\xi = (\alpha+\beta)^{-1}\delta, \quad \nu = \beta^{-1}(\alpha+\beta),$$

with the inverse

$$\alpha = \nu^{-1}(\nu-1)(1+\lambda+\mu+\xi)^{-1}, \quad \beta = \nu^{-1}(1+\lambda+\mu+\xi)^{-1},$$
$$\gamma = \mu(1+\lambda+\mu+\xi)^{-1}, \quad \delta = \xi(1+\lambda+\mu+\xi)^{-1}.$$

The Jacobian of the transformation is given by

$$|\partial(\alpha,\beta,\gamma,\delta)/\partial(\lambda,\mu,\nu,\xi)| = (1+\lambda+\mu+\xi)^{-5}\nu^{-2}$$

and we find that

$$I(x_i) = -4EFGH \int_0^\infty d\xi \int_0^\infty d\mu \int_0^\infty d\lambda \int_0^\infty d\nu (1+\lambda+\mu+\xi)\nu^{-2} \times [E^2\nu^{-1}(\nu-1)(1+\lambda+\mu+\xi) + F^2\nu^{-1}(1+\lambda+\mu+\xi) + G^2\lambda(1+\lambda+\mu+\xi) + H^2\mu(1+\lambda+\mu+\xi) + K^2\xi(1+\lambda+\mu+\xi) - A^2\lambda\nu^{-1}(\nu-1) - B^2\lambda\nu^{-1} - C^2\mu\nu^{-1}(\nu-1) - D^2\xi\nu^{-1} - N^2\mu\xi - s_1\nu^{-2}(\nu-1) - s_2\lambda\mu - s_3\lambda\xi - s_4\nu^{-1}\mu - s_5\nu^{-1}(\nu-1)\xi]^{-3}$$

$$= 2EFGH \int_0^\infty \frac{d\xi}{\xi} \int_0^\infty d\mu \int_0^\infty d\lambda \int_1^\infty d\nu. \times \frac{\partial}{\partial K^2}[(\nu-1)\phi(\lambda,\mu,\xi) + \psi(\lambda,\mu,\xi) - \nu^{-1}(\nu-1)v(x_1)]^{-2}, \quad (5)$$

where

$$\phi(\lambda,\mu,\xi) = G^2\lambda^2 + H^2\mu^2 + K^2\xi^2 + 2GHX_2\lambda\mu + 2GKX_3\lambda\xi + 2HKe\mu\xi + 2EGa\lambda + 2EHc\mu + 2EKX_5\xi + E^2, \quad (6)$$

$$\psi(\lambda,\mu,\xi) = G^2\lambda^2 + H^2\mu^2 + K^2\xi^2 + 2GHX_2\lambda\mu + 2GKX_3\lambda\xi + 2HKe\mu\xi + 2FGb\lambda + 2FHX_4\mu + 2FKd\xi + F^2, \quad (7)$$

$$v(x_1) = 2EFx_1 + E^2 + F^2. \quad (8)$$

The constants A^2, B^2, C^2, D^2, N^2 have been expressed in terms of a, b, c, d, e defined by

$$2EGa = E^2 + G^2 - A^2, \quad 2FGb = F^2 + G^2 - B^2,$$
$$2EHc = E^2 + H^2 - C^2, \quad 2FKd = F^2 + K^2 - D^2,$$
$$2HKe = H^2 + K^2 - N^2, \quad (9)$$

and we have also used Eq. (4).

To simplify the proof of a spectral representation we restrict the quantities defined in Eqs. (4) and (9) as follows:

$$a, b, c, d, e > 0, \quad X_i > 0 \quad (i=1,\ldots,5). \quad (10)$$

Equation (10) ensures that $\phi(\lambda,\mu,\xi) > 0$, $\psi(\lambda,\mu,\xi) > 0$ for $\lambda \geq 0$, $\mu \geq 0$, $\xi \geq 0$; in fact the term in square brackets in Eq. (5) is always positive and $I(x_i)$ is well defined. While Eq. (10) can be satisfied with physical invariants by choosing the internal masses sufficiently large, the restrictions on X_i mean that the amplitude we are considering does not in general correspond to a physical process since for the physical amplitude associated with the pentagon diagram in Fig. 1 X_1, X_4, and X_5 would in general be negative. However, from the form of $I(x_i)$ we see that a spectral representation cannot in general be proved for negative X_1, X_4, and X_5, using real analysis only. One way of obtaining the physical amplitude would be to start with the spectral representation for the unphysical amplitude [Eq. (65)] and do an analytic continuation in X_1, X_4, and X_5 using, for example, a generalization of the method used in II. We discuss this problem further in Sec. 8.

The argument leading to Eqs. (I-19) and (I-20) can now be used to show that

$$I(x_i) = \lim_{\epsilon \downarrow 0} \frac{\partial}{\partial K^2} \int_\epsilon^\infty \frac{d\xi}{\xi} \int_0^\infty \frac{d\mu}{\mu} \times \lim_{\delta \downarrow 0} \frac{\partial}{\partial x_2} \int_\delta^\infty \frac{d\lambda}{\lambda} \int_{h(\lambda,\mu,\xi)}^\infty \frac{d\xi}{(\xi - x_1)[U(\xi,\lambda,\mu,\xi)]^{1/2}}, \quad (11)$$

where

$$U(\xi,\lambda,\mu,\xi) = (\xi - h(\lambda,\mu,\xi))(\xi - k(\lambda,\mu,\xi)), \quad (12)$$

$$\left.\begin{matrix} h(\lambda,\mu,\xi) \\ k(\lambda,\mu,\xi) \end{matrix}\right\} = (2EF)^{-1}\{[\sqrt{\phi(\lambda,\mu,\xi)} \pm \sqrt{\psi(\lambda,\mu,\xi)}]^2 - E^2 - F^2\}, \quad (13)$$

and $\phi(\lambda, \mu, \zeta)$, $\psi(\lambda, \mu, \zeta)$ are given by Eqs. (6), (7).

3. STUDY OF $h(\lambda,\mu,\zeta)$

To reverse the order of integration in Eq. (11) we need to examine the function $h(\lambda, \mu, \zeta)$ for $\lambda \geq 0$, $\mu \geq 0$, $\zeta \geq 0$. As in Sec. 4 of I (or of VF) we write

$$\phi(\lambda, \mu, \zeta) = p_1\lambda^2 + 2q_1(\mu, \zeta)\lambda + r_1(\mu, \zeta),$$
$$\psi(\lambda, \mu, \zeta) = p_1\lambda^2 + 2q_1'(\mu, \zeta)\lambda + r_1'(\mu, \zeta), \tag{14}$$

where

$$p_1 = G^2,$$
$$q_1(\mu, \zeta) = G(HX_2\lambda + KX_3\zeta + Ea),$$
$$r_1(\mu, \zeta) = H^2\mu^2 + K^2\zeta^2 + 2HKe\mu\zeta + 2EHc\mu + 2EKX_5\zeta + E^2,$$
$$q_1'(\mu, \zeta) = G(HX_2\lambda + KX_3\zeta + Fb),$$
$$r_1'(\mu, \zeta) = H^2\mu^2 + K^2\zeta^2 + 2HKe\mu\zeta + 2FHX_4\mu + 2FKd\zeta + F^2. \tag{15}$$

Then the argument of Sec. 4 of I (or of VF) shows that for fixed $\mu \geq 0$, $\zeta \geq 0$, $h(\lambda, \mu, \zeta)$ increases strictly from $h(0, \mu, \zeta)$ to $+\infty$ as λ increases from 0 to $+\infty$, whenever $h_\lambda(0, \mu, \zeta) \geq 0$. Now

$$h_\lambda(0, \mu, \zeta) = (EF)^{-1}\left(\sqrt{r_1(\mu, \zeta)} + \sqrt{r_1'(\mu, \zeta)}\right) l_1(\mu, \zeta), \tag{16}$$

where

$$l_1(\mu, \zeta) = [q_1(\mu, \zeta)/\sqrt{r_1(\mu, \zeta)}] + [q_1'(\mu, \zeta)/\sqrt{r_1'(\mu, \zeta)}]. \tag{17}$$

Thus, when Eq. (10) holds, it follows from Eqs. (15), (17), and (16) that for fixed $\mu \geq 0$, $\zeta \geq 0$, $h(\lambda, \mu, \zeta)$ increases strictly from $h(0, \mu, \zeta)$ to $+\infty$ as λ increases from 0 to $+\infty$. Similarly, for fixed $\lambda \geq 0$, $\zeta \geq 0$, $h(\lambda, \mu, \zeta)$ increases strictly from $h(\lambda, 0, \zeta)$ to $+\infty$ as μ increases from 0 to $+\infty$ and for fixed $\mu \geq 0$, $\zeta \geq 0$, $h(\lambda, \mu, \zeta)$ increases strictly from $h(\lambda, \mu, 0)$ to $+\infty$ as ζ increases from 0 to $+\infty$.

4. SOLUTIONS OF $U(\xi,\lambda,\mu,\zeta)= 0$

In this section we study the behavior of the zeros of $U(\xi, \lambda, \mu, \zeta)$ first when ξ, μ and ζ are held fixed, then when ξ, λ, and ζ are held fixed and finally when ξ, λ, and μ are held fixed. From Eqs. (12), (13), (6), and (7) we have

$$\begin{aligned} 4E^2F^2\,U(\xi, \lambda, \mu, \zeta) &= a_1(\xi)\lambda^2 + 2b_1(\xi, \mu, \zeta)\lambda + c_1(\xi, \mu, \zeta) \\ &= a_2(\xi)\mu^2 + 2b_2(\xi, \lambda, \zeta)\mu + c_2(\xi, \lambda, \zeta) \\ &= a_3(\xi)\zeta^2 + 2b_3(\xi, \lambda, \mu)\zeta + c_3(\xi, \lambda, \mu), \end{aligned} \tag{18}$$

where

$$a_1(\xi) = 4G^2[(Ea - Fb)^2 - v(\xi)],$$
$$a_2(\xi) = 4H^2[(Ec - FX_4)^2 - v(\xi)],$$
$$a_3(\xi) = 4K^2[(EX_5 - Fd)^2 - v(\xi)],$$
$$b_1(\xi, \mu, \zeta) = \beta(\xi, -X_2)\mu + \gamma(\xi, -X_3)\zeta + b_1,$$
$$b_2(\xi, \lambda, \zeta) = \beta(\xi, -X_2)\lambda + \delta(\xi, -e)\zeta + b_2,$$
$$b_3(\xi, \lambda, \mu) = \gamma(\xi, -X_3)\lambda + \delta(\xi, -e)\zeta + b_3,$$
$$\beta(\xi, -X_2) = 4GH[(Ea - Fb)(Ec - FX_4) - X_2\,v(\xi)],$$
$$\gamma(\xi, -X_3) = 4GK[(Ea - Fb)(EX_5 - Fd) - X_3\,v(\xi)],$$
$$\delta(\xi, -e) = 4HK[(Ec - FX_4)(EX_5 - Fd) - e\,v(\xi)],$$
$$b_1 \equiv b_1(\xi, 0, 0) = 2G[(Ea - Fb)(E^2 - F^2) - (Ea + Fb)v(\xi)],$$
$$b_2 \equiv b_2(\xi, 0, 0) = 2H[(Ec - FX_4)(E^2 - F^2) - (Ec + FX_4)v(\xi)],$$
$$b_3 \equiv b_3(\xi, 0, 0) = 2K[(EX_5 - Fd)(E^2 - F^2) - (EX_5 + Fd)v(\xi)], \tag{19}$$

and $v(\xi)$ is given in Eq. (8).

The quantities $c_1(\xi, \mu, \zeta)$, $c_2(\xi, \lambda, \zeta)$, $c_3(\xi, \lambda, \mu)$ are determined from Eq. (18) by putting λ, μ, ζ, respectively, equal to zero and using in addition Eq. (19) and the fact that

$$c_1(\xi, 0, 0) = c_2(\xi, 0, 0) = c_3(\xi, 0, 0) = 4E^2F^2(\xi^2 - 1). \tag{20}$$

The argument of Sec. 5 of VF (see also Sec. 5 of I) shows that for each $\xi \geq h(0, \mu, \zeta)$, where μ and ζ are fixed and ≥ 0, the quadratic equation in λ

$$U(\xi, \lambda, \mu, \zeta) = 0$$

has two real roots given by

$$\lambda_\pm(\xi, \mu, \zeta) = [a_1(\xi)]^{-1}\{-b_1(\xi, \mu, \zeta) \mp \{[b_1(\xi, \mu, \zeta)]^2 - a_1(\xi)c_1(\xi, \mu, \zeta)\}^{1/2}\}. \tag{21}$$

From Eqs. (19), (13), (14), and (15) we see that

$$b_1(h(0, \mu, \zeta), \mu, \zeta) = -4[\sqrt{r_1(\mu, \zeta)} + \sqrt{r_1'(\mu, \zeta)}]\sqrt{r_1(\mu, \zeta)}\sqrt{r_1'(\mu, \zeta)}\,l_1(\mu, \zeta), \tag{22}$$

where $l_1(\mu, \zeta)$ is given in Eq. (17). Since $l_1(\mu, \zeta) > 0$ when Eq. (10) holds it follows that $\lambda_+(h(0, \mu, \zeta), \mu, \zeta) = 0 \neq \lambda_-(h(0, \mu, \zeta), \mu, \zeta)$ and in fact $\lambda_+(\xi, \mu, \zeta)$ is the inverse of the strictly increasing function $h(\lambda, \mu, \zeta)$ on $0 \leq \lambda < \infty$. Thus $\lambda_+(\xi, \mu, \zeta)$ increases strictly from 0 to $+\infty$ as ξ increases from $h(0, \mu, \zeta)$ to $+\infty$. Similarly for each $\xi \geq h(\lambda, 0, \zeta)$, where λ and ζ are fixed and ≥ 0, the quadratic equation in μ

$$U(\xi, \lambda, \mu, \zeta) = 0$$

has two real roots, $\mu_\pm(\xi, \lambda, \zeta)$ given by the right-hand side of Eq. (21) with $\mu \to \lambda$, $1 \to 2$. The root $\mu_+(\xi, \lambda, \zeta)$ is the inverse of the strictly increasing function $h(\lambda, \mu, \zeta)$ on $0 \leq \mu < \infty$. Further, for each $\xi \geq h(\lambda, \mu, 0)$, where λ and μ are fixed and ≥ 0, the quadratic equation in ζ

$$U(\xi, \lambda, \mu, \zeta) = 0$$

has two real roots, $\zeta_\pm(\xi, \lambda, \mu)$ given by the right-hand side of Eq. (21) with $\mu \to \lambda$, $\zeta \to \mu$, $1 \to 3$. Again $\zeta_+(\xi, \lambda, \mu)$ is the inverse of the strictly increasing function $h(\lambda, \mu, \zeta)$ on $0 \leq \zeta < \infty$.

5. REVERSAL OF ORDER OF INTEGRATION

Since Eq. (10) holds, we showed in Sec. 3 that $h_\lambda(0, \mu, \zeta) > 0$ for all $\mu \geq 0$, $\zeta \geq 0$ and so from Sec. 4, $\lambda_+(\xi, \mu, \zeta)$ is the inverse of the strictly increasing function $h(\lambda, \mu, \zeta)$ on $0 \leq \lambda < \infty$ for each $\mu \geq 0$, $\zeta \geq 0$. Thus Eq. (11) can be written

$$I(x_i) = 2EF \lim_{\epsilon \downarrow 0} \frac{\partial}{\partial K^2} \int_\epsilon^\infty \frac{d\zeta}{\zeta} \int_0^\infty \frac{d\mu}{\mu} \times \lim_{\delta \downarrow 0} \frac{\partial}{\partial x_2} \int_{h(\delta, \mu, \zeta)}^\infty \frac{d\xi}{\xi - x_1}\, \Lambda(\xi, \delta, \mu, \zeta), \tag{23}$$

where

$$\Lambda(\xi,\delta,\mu,\zeta) = \int_{\delta}^{\lambda_+(\xi,\mu,\zeta)} \frac{d\lambda}{\lambda[a_1(\xi)\lambda^2+2b_1(\xi,\mu,\zeta)\lambda+c_1(\xi,\mu,\zeta)]^{1/2}}. \tag{24}$$

Note that $h(\delta,\mu,\zeta)$, $b_1(\xi,\mu,\zeta)$, $\lambda_+(\xi,\mu,\zeta)$, and $\Lambda(\xi,\delta,\mu,\zeta)$ depend on x_2.

Now since for fixed $\zeta \geq 0$, $\mu_+(\xi,0,\zeta)$ is the inverse of the strictly increasing function $h(0,\mu,\zeta)$ on $0 \leq \mu < \infty$, the argument of Sec. 6 of I can be used to show that

$$I(x_i) = \lim_{\epsilon\downarrow 0} \frac{\partial}{\partial K^2} \int_{\epsilon}^{\infty} \frac{d\zeta}{\zeta} \int_{h(0,0,\zeta)}^{\infty} \frac{d\xi}{\xi - x_1} X(\xi,\zeta), \tag{25}$$

where

$$X(\xi,\zeta) = \frac{16EFGH\,v(\xi)}{[c_2(\xi,0,\zeta)]^{1/2}} \int_0^1 \frac{du(lu^2-m)}{(lu^2-m)^2+4nu^2}. \tag{26}$$

Here

$$\begin{aligned} l &= [\mu_-(\xi,0,\zeta)]^{-1} b_1(\xi,\mu_-(\xi,0,\zeta),\zeta) \\ m &= [\mu_+(\xi,0,\zeta)]^{-1} b_1(\xi,\mu_+(\xi,0,\zeta),\zeta) \\ n &= -a_1(\xi)[c_2(\xi,0,\zeta)]^{-1}\{[b_2(\xi,0,\zeta)]^2 - a_2(\xi)c_2(\xi,0,\zeta)\}. \end{aligned} \tag{27}$$

Since from Eq. (18) $c_1(\xi,0,\zeta) = c_2(\xi,0,\zeta)$, we have

$$\begin{aligned} lm-n &= [c_2(\xi,0,\zeta)]^{-2}([\bar{G}(\xi,x_2,\zeta)]^2 \\ &\quad - \{[b_1(\xi,0,\zeta)]^2 - a_1(\xi)c_1(\xi,0,\zeta)\} \\ &\quad \times \{[b_2(\xi,0,\zeta)]^2 - a_2(\xi)c_2(\xi,0,\zeta)\}) \\ &= [c_2(\xi,0,\zeta)]^{-1}\bar{F}(\xi,x_2,\zeta), \end{aligned} \tag{28}$$

where

$$\bar{G}(\xi,x_2,\zeta) = -[\beta(\xi,x_2)c_2(\xi,0,\zeta) - b_1(\xi,0,\zeta)b_2(\xi,0,\zeta)], \tag{29}$$

and

$$\begin{aligned} \bar{F}(\xi,x_2,\zeta) &= [\beta(\xi,x_2)]^2 c_2(\xi,0,\zeta) - 2\beta(\xi,x_2)b_1(\xi,0,\zeta)b_2(\xi,0,\zeta) \\ &\quad + [b_1(\xi,0,\zeta)]^2 a_2(\xi) + [b_2(\xi,0,\zeta)]^2 a_1(\xi) - a_1(\xi)a_2(\xi)c_2(\xi,0,\zeta) \\ &= 16G^2H^2[v(\xi)]^2 c_2(\xi,0,\zeta)[x_2 - f_+(\xi,\zeta)][x_2 - f_-(\xi,\zeta)]. \end{aligned} \tag{30}$$

Here

$$\begin{aligned} f_\pm(\xi,\zeta) &= [4GH\,v(\xi)c_2(\xi,0,\zeta)]^{-1}(-4GH\,c_2(\xi,0,\zeta)(Ea-Fb) \\ &\quad \times (Ec - FX_4) + b_1(\xi,0,\zeta)b_2(\xi,0,\zeta) \pm \{[b_1(\xi,0,\zeta)]^2 \\ &\quad - a_1(\xi)c_1(\xi,0,\zeta)\}^{1/2} \\ &\quad \times \{[b_2(\xi,0,\zeta)]^2 - a_2(\xi)c_2(\xi,0,\zeta)\}^{1/2}) \end{aligned} \tag{31}$$

and the argument of Sec. 5 of I (or of VF) shows that

$$\begin{aligned} &[b_1(\xi,0,\zeta)]^2 - a_1(\xi)c_1(\xi,0,\zeta) > 0, \\ &[b_2(\xi,0,\zeta)]^2 - a_2(\xi)c_2(\xi,0,\zeta) > 0, \end{aligned} \tag{32}$$

for $\xi \geq h(0,0,\zeta)$. Further since $c_2(\xi,0,\zeta) > 0$ for $\xi > h(0,0,\zeta)$ it follows from Sec. 6 of I that

$$X(\xi,\zeta) = 8EFGH\,v(\xi) \int_{f_+(\xi,\zeta)}^{\infty} \frac{d\eta}{(\eta - x_2)[\bar{F}(\xi,\eta,\zeta)]^{1/2}}. \tag{33}$$

Note that

$$\bar{F}(\xi,x_2,0) = 64E^2F^2G^2H^2[v(\xi)]^2 F(\xi,x_2), \tag{34}$$

where $F(\xi,x_2)$ is given in Eq. (A1) (and in Eq. (I-12) with $\eta \to x_2$, $d \to X_4$).

Now since, as shown in Sec. 4, $\zeta_+(\xi,0,0)$ is the inverse of the strictly increasing function $h(0,0,\zeta)$ on $0 \leq \zeta < \infty$ we find on inserting Eq. (33) into Eq. (25) and reversing the order of the ζ and ξ integrations that

$$\begin{aligned} I(x_i) &= \lim_{\epsilon\downarrow 0} \frac{\partial}{\partial K^2} \int_{h(0,0,\epsilon)}^{\infty} \frac{d\xi}{\xi - x_1} \int_{\epsilon}^{\zeta_+(\xi,0,0)} \frac{d\zeta}{\zeta} \\ &\quad \times \int_{f_+(\xi,\zeta)}^{\infty} \frac{d\eta}{\eta - x_2} \frac{8EFGH\,v(\xi)}{[\bar{F}(\xi,\eta,\zeta)]^{1/2}}. \end{aligned} \tag{35}$$

Note that $h(0,0,\epsilon)$, $\zeta_+(\xi,0,0)$, and $\bar{F}(\xi,\eta,\zeta)$ depend on K^2 through Eqs. (4) and (9).

6. STUDY OF $f_+(\xi,\zeta)$

To reverse the order of the ζ and η integrations in Eq. (35), we need to examine the function $f_+(\xi,\zeta)$ for $\xi \geq 1$, $0 \leq \zeta \leq \zeta_+(\xi,0,0)$. First we examine the behavior of $f_+(\xi,\zeta)$ as $\zeta \uparrow \zeta_+(\xi,0,0)$ with ξ fixed and ≥ 1. We showed in Sec. 3 that when Eq. (10) holds, $l_1(0,\zeta) > 0$ and thus from Eq. (22)

$$b_1(\xi,0,\zeta_+(\xi,0,0)) < 0. \tag{36}$$

Similarly

$$b_2(\xi,0,\zeta_+(\xi,0,0)) < 0 \tag{37}$$

and hence from Eqs. (18) and (31) and the fact that

$$v(\xi) > 0 \tag{38}$$

for $\xi \geq 1$ it follows that $f_+(\xi,\zeta) \to +\infty$ as $\zeta \uparrow \zeta_+(\xi,0,0)$.

Next, from Eq. (31) we see that the derivative of $f_+(\xi,\zeta)$ with respect to ζ is

$$\begin{aligned} f_{+\zeta}(\xi,\zeta) &= [4GH\,v(\xi)]^{-1}[c_2(\xi,0,\zeta)]^{-2} \\ &\quad \times (-b_1(\xi,0,\zeta)\{[b_2(\xi,0,\zeta)]^2 - a_2(\xi)c_2(\xi,0,\zeta)\}^{1/2} \\ &\quad - b_2(\xi,0,\zeta)\{[b_1(\xi,0,\zeta)]^2 - a_1(\xi)c_1(\xi,0,\zeta)\}^{1/2}) \\ &\quad \times L(\xi,\zeta), \end{aligned} \tag{39}$$

where

$$L(\xi,\zeta) = [Q(\xi,\zeta)/\sqrt{R(\xi,\zeta)}] + [Q'(\xi,\zeta)/\sqrt{R'(\xi,\zeta)}], \tag{40}$$

$$\begin{aligned} Q(\xi,\zeta) &= \zeta(a_3b_1 - \gamma b_3) + (b_3b_1 - \gamma c_3), \\ Q'(\xi,\zeta) &= \zeta(a_3b_2 - \delta b_3) + (b_3b_2 - \delta c_3), \\ R(\xi,\zeta) &= \zeta^2(\gamma^2 - a_3a_1) + 2\zeta(\gamma b_1 - a_1b_3) + b_1^2 - a_1c_1, \\ R'(\xi,\zeta) &= \zeta^2(\delta^2 - a_3a_2) + 2\zeta(\delta b_2 - a_2b_3) + b_2^2 - a_2c_2. \end{aligned} \tag{41}$$

In Eq. (41) a_i has been written for $a_i(\xi)$, c_i for $c_i(\xi,0,0)$, γ for $\gamma(\xi,-X_3)$ etc. where these quantities are defined in Eqs. (19) and (20). From Eq. (36) together with the facts that $b_1 \equiv b_1(\xi,0,0) < 0$ when Eq. (10) holds and $b_1(\xi,0,\zeta)$ is linear in ζ it follows that

$$b_1(\xi,0,\zeta) < 0 \tag{42}$$

for all $0 \leq \zeta \leq \zeta_+(\xi,0,0)$. Similarly

$$b_2(\xi,0,\zeta) < 0, \tag{43}$$

for all $0 \leq \zeta \leq \zeta_+(\xi, 0, 0)$. Thus the term in boldface parentheses in Eq. (39) is always positive and $f_{+\zeta}(\xi, \zeta)$ vanishes if and only if $L(\xi, \zeta)$ vanishes.

Now

$$(b_3^2 - a_3c_3)R(\xi, \zeta) = Q^2(\xi, \zeta) - c_2(\xi, 0, \zeta)C_2(\xi, -X_3),$$
$$(b_3^2 - a_3c_3)R'(\xi, \zeta) = Q'^2(\xi, \zeta) - c_2(\xi, 0, \zeta)C_3(\xi, -e), \tag{44}$$

where $(b_3^2 - a_3c_3) > 0$ for $\xi \geq 1$,

$$C_2(\xi, -X_3) = c_3(\gamma^2 - a_1a_3) + b_1^2a_3 + b_3^2a_1 - 2\gamma b_1b_3,$$
$$C_3(\xi, -e) = c_3(\delta^2 - a_2a_3) + b_2^2a_3 + b_3^2a_2 - 2\delta b_2b_3, \tag{45}$$

and $c_2(\xi, 0, \zeta) > 0$ for $\xi > 1$, $0 \leq \zeta < \zeta_+(\xi, 0, 0)$. In Eqs. (44) and (45) the abbreviations described after Eq. (41) have again been used. The argument of Sec. 4 of VF then shows that $f_{+\zeta}(\xi, \zeta_0(\xi)) = 0$ with $0 \leq \zeta_0(\xi) < \zeta_+(\xi, 0, 0)$ if and only if

$$\text{(i) } C_2(\xi, -X_3) < 0, \quad C_3(\xi, -e) < 0 \tag{46}$$

and

$$\text{(ii) } Q(\xi, \zeta_0(\xi))/\sqrt{-C_2(\xi, -X_3)} = -Q'(\xi, \zeta_0(\xi))/\sqrt{C_3(\xi, -e)} \tag{47}$$

solves to give $0 \leq \zeta_0(\xi) < \zeta_+(\xi, 0, 0)$. Thus, for fixed $\xi \geq 1$ $f_+(\xi, \zeta)$ is strictly increasing on $0 \leq \zeta \leq \zeta_+(\xi, 0, 0)$ if and only if $L(\xi, 0) > 0$ or

$$\frac{G_2(\xi, -X_3)}{[\xi^2 + 2ab\xi + a^2 + b^2 - 1]^{1/2}} + \frac{G_3(\xi, -e)}{[\xi^2 + 2cX_4\xi + c^2 + X_4^2 - 1]^{1/2}} > 0, \tag{48}$$

where $G_2(\xi, -X_3)$ and $G_3(\xi, -e)$ are given by Eqs. (A13) and (A14). When Eq. (10) holds we see that each of the two terms in Eq. (48) is positive and hence $f_+(\xi, \zeta)$ increases strictly on $0 \leq \zeta \leq \zeta_+(\xi, 0, 0)$ for fixed $\xi \geq 1$.

7. SOLUTIONS OF $\bar{F}(\xi, \eta, \zeta) = 0$

Next we study the behavior of the zeros of $\bar{F}(\xi, \eta, \zeta)$ when ξ and η are held fixed. From Eqs. (30) and (19) we find that

$$\bar{F}(\xi, \eta, \zeta) = A(\xi, \eta; K^2)\zeta^2 + 2B(\xi, \eta; K^2)\zeta + C(\xi, \eta), \tag{49}$$

where

$$A(\xi, \eta; K^2) = a_3([\beta(\xi, \eta)]^2 - a_1a_2) + \gamma^2a_2 + \delta^2a_1 - 2\beta(\xi, \eta)\gamma\delta,$$
$$B(\xi, \eta; K^2) = b_3([\beta(\xi, \eta)]^2 - a_1a_2) + b_1\gamma a_2 + b_2\delta a_1 - \beta(\xi, \eta)b_1\delta - \beta(\xi, \eta)b_2\gamma,$$
$$C(\xi, \eta) = c_3([\beta(\xi, \eta)]^2 - a_1a_2) + b_1^2a_2 + b_2^2a_1 - 2\beta(\xi, \eta)b_1b_2. \tag{50}$$

The abbreviations described after Eq. (41) have again been used except for $\beta(\xi, \eta)$, which is the only term that depends on η.

The discriminant of the quadratic function of ζ in Eq. (49) is

$$[B(\xi, \eta; K^2)]^2 - A(\xi, \eta; K^2)C(\xi, \eta) = (\{\beta(\xi, \eta)\}^2 - a_1a_2) \times ((b_3^2 - a_3c_3)\{[\beta(\xi, \eta)]^2 - a_1a_2\} + 2[-b_3(b_1\delta + b_2\gamma) + a_3b_1b_2 + c_3\gamma\delta]\beta(\xi, \eta) + (b_1\delta - b_2\gamma)^2 + 2b_3(b_1\gamma a_2 + b_2\delta a_1) - c_3(\gamma^2a_2 + \delta^2a_1) - a_3(b_1^2a_2 + b_2^2a_1)) \tag{51}$$

and the term in boldface parentheses vanishes when

$$\beta(\xi, \eta) = (b_3^2 - a_3c_3)^{-1}\{b_3(b_1\delta + b_2\gamma) - a_3b_1b_2 - c_3\gamma\delta \pm [C_2(\xi, -X_3)C_3(\xi, -e)]^{1/2}\} \tag{52}$$

giving $\eta = p_\pm(\xi; K^2)$, with $p_\pm(\xi; K^2)$ defined in Eq. (A6). The terms $C_2(\xi, -X_3)$ and $C_3(\xi, -e)$ are defined in Eq. (45). Thus

$$[B(\xi, \eta; K^2)]^2 - A(\xi, \eta; K^2)C(\xi, \eta) = \{[\beta(\xi, \eta)]^2 - a_1a_2\}(16)^2E^2F^2G^2H^2K^2[v(\xi)]^3\,P(\xi, \eta; K^2), \tag{53}$$

where

$$P(\xi, \eta; K^2) = (\xi^2 + 2dX_5\xi + d^2 + X_5^2 - 1)(\eta - p_+(\xi; K^2)) \times (\eta - p_-(\xi; K^2)). \tag{54}$$

The discriminant in Eq. (53) is always nonnegative since the inverse of $f_+(\xi, \zeta)$ is real. To show that it is in fact positive we note first that when Eq. (10) holds

$$\{[\beta(\xi, \eta)]^2 - a_1(\xi)a_2(\xi)\} > 0 \tag{55}$$

for all $\zeta \geq 1$, $\eta \geq f_+(\xi, 0)$ [or equivalently for $\eta \geq 1$, $\xi \geq g_+(\eta)$ where $g_+(\eta)$, defined in Eq. (A3) and in Eq. (I-36) with $d \to X_4$, is the inverse of $f_+(\xi, 0)$]. Establishing Eq. (55) is straightforward but tedious. Secondly in Appendix A we show that either $p_\pm(\xi; K^2)$ are complex conjugates or

$$p_-(\xi; K^2) \leq p_+(\xi; K^2) < f_+(\xi, 0), \tag{56}$$

for $\xi \geq 1$. Thus, since the first factor in Eq. (54) is positive when Eq. (10) holds and Eq. (38) is satisfied it follows that $P(\xi, \eta; K^2)$ and hence the right-hand side of Eq. (53) is in fact positive for $\xi \geq 1$, $\eta \geq f_+(\xi, 0)$.

The two real solutions of

$$\bar{F}(\xi, \eta, \zeta) = 0 \tag{57}$$

are

$$\zeta_{\substack{a \\ b}}(\xi, \eta; K^2) = [A(\xi, \eta; K^2)]^{-1}(-B(\xi, \eta; K^2) \mp \{[B(\xi, \eta; K^2)]^2 - A(\xi, \eta; K^2)C(\xi, \eta)\}^{1/2}). \tag{58}$$

Now from Eqs. (30) and (49) it follows that

$$C(\xi, f_+(\xi, 0)) = 0 \tag{59}$$

and from Eqs. (50), (40), (41), (42), and (43)

$$B(\xi, f_+(\xi, 0); K^2) = [c_3]^{-2}\{b_1(R'(\xi, 0))^{-1/2} + b_2(R(\xi, 0))^{-1/2}\}L(\xi, 0) < 0. \tag{60}$$

Thus

$$\zeta_a(\xi, f_+(\xi, 0); K^2) = 0 \neq \zeta_b(\xi, f_+(\xi, 0); K^2)$$

and also as $\eta \to +\infty$

$$\zeta_{\substack{a \\ b}}(\xi, f_+(\xi, 0); K^2) \sim \zeta_\pm(\xi, 0, 0),$$

where $\zeta_\pm(\xi, 0, 0)$ are defined by the right-hand side of Eq. (21) with $1 \to 3$, $\mu \to 0$, $\zeta \to 0$. It now follows that $\zeta_a(\xi, \eta; K^2)$ is the inverse of the strictly increasing function $f_+(\xi, \zeta)$ on $0 \leq \zeta \leq \zeta_+(\xi, 0, 0)$. Hence $\zeta_a(\xi, \eta, K^2)$ increases from 0 to $\zeta_+(\xi, 0, 0)$ as η increases from $f_+(\xi, 0)$ to $+\infty$.

8. SPECTRAL REPRESENTATION OF THE PENTAGON DIAGRAM AMPLITUDE

Since for fixed $\xi \geq 1$ $f_+(\xi, \zeta)$ is strictly increasing on $0 \leq \zeta \leq \zeta_+(\xi, 0, 0)$ and $\zeta_a(\xi, \eta; K^2)$ is the inverse of $f_+(\xi, \zeta)$ on $0 \leq \zeta \leq \zeta_+(\xi, 0, 0)$, Eq. (35) can be written

$$I(x_i) = \lim_{\epsilon \downarrow 0} \frac{\partial}{\partial K^2} \int_{h(0,0,\epsilon;K^2)}^{\infty} \frac{8EFGH\, v(\xi)\, d\xi}{\xi - x_1} \int_{f_+(\xi,\epsilon;K^2)}^{\infty} \frac{d\eta}{\eta - x_2} \times \Gamma(\xi, \eta, \epsilon; K^2), \tag{61}$$

where

$$\Gamma(\xi, \eta, \epsilon; K^2) = \int_{\epsilon}^{\zeta_a(\xi,\eta;K^2)} \frac{d\zeta}{\zeta[A(\xi,\eta;K^2)\zeta^2 + 2B(\xi,\eta;K^2)\zeta + C(\xi,\eta)]^{1/2}}$$

and $A(\xi, \eta; K^2)$, $B(\xi, \eta; K^2)$, and $C(\xi, \eta)$ are given in Eq. (50). Note that in Eq. (61) $h(0, 0, \epsilon)$ and $f_+(\xi, \epsilon)$ depend on K^2. From Eqs. (30) and (49) and the fact that $f_+(\xi, \zeta)$ is strictly increasing on $0 \leq \zeta \leq \zeta_+(\xi, 0, 0)$ we have $C(\xi, \eta) > 0$ for $\xi > 1$, $\eta > f_+(\xi, 0)$. The integration in Eq. (61) can then be performed (c.f. Sec. 5 of VF) to give

$$\Gamma(\xi, \eta, \epsilon; K^2) = [C(\xi, \eta)]^{-1/2} \times \ln \frac{C(\xi,\eta) + \epsilon B(\xi,\eta;K^2) + [C(\xi,\eta)]^{1/2}[\bar{F}(\xi,\eta,\epsilon;K^2)]^{1/2}}{\epsilon\{[B(\xi,\eta;K^2)]^2 - A(\xi,\eta;K^2)C(\xi,\eta)\}^{1/2}}. \tag{63}$$

The method of differentiating with respect to K^2 and taking the limit $\epsilon \downarrow 0$ is now very similar to that given in Sec. 6 (and 7) of I and in Ref. 11. We find that

$$I(x_i) = \int_1^{\infty} \frac{d\xi}{\xi - x_1} \int_{f_+(\xi,0)}^{\infty} \frac{d\eta}{\eta - x_2} \frac{8EFGH\, v(\xi)}{[C(\xi,\eta)]^{1/2}} \times \frac{(-\frac{1}{2})(\partial/\partial K^2)\{[B(\xi,\eta;K^2)] - A(\xi,\eta;K^2)C(\xi,\eta)\}}{\{[B(\xi,\eta;K^2)]^2 - A(\xi,\eta;K^2)C(\xi,\eta)\}}. \tag{64}$$

From Eqs. (8) and (19) it follows that the factor $\{[\beta(\xi,\eta)]^2 - a_1 a_2\}[v(\xi)]^3$ in Eq. (53) does not depend on K^2. Thus

$$I(x_i) = -\tfrac{1}{2} \int_1^{\infty} \frac{d\xi}{\xi - x_1} \int_{f_+(\xi)}^{\infty} \frac{d\eta}{\eta - x_2} \frac{1}{\sqrt{F(\xi,\eta)}} \times \frac{(\partial/\partial K^2)[K^2 P(\xi,\eta;K^2)]}{K^2 P(\xi,\eta;K^2)}$$

$$= -\tfrac{1}{2} \int_1^{\infty} \frac{d\xi}{\xi - x_1} \int_{f_+(\xi)}^{\infty} \frac{d\eta}{\eta - x_2} \frac{1}{\sqrt{F(\xi,\eta)}} \times \frac{(\partial/\partial K^2)\tilde{P}(\xi,\eta;K^2)}{\tilde{P}(\xi,\eta;K^2)}, \tag{65}$$

where

$$\tilde{P}(\xi,\eta;K^2) = 32E^2F^2G^2H^2K^2\, P(\xi,\eta;K^2) \tag{66}$$

and $P(\xi, \eta; K^2)$ is given in Eq. (54). The functions $f_+(\xi)$ ($\equiv f_+(\xi, 0)$) and $F(\xi, \eta)$ are defined in Eqs. (A2) and (A1) [and in Eqs. (I-11) and (I-12) with $d \to X_4$]. Their properties are studied in detail in Sec. 8 of I. Note that the relationship between $\tilde{P}(\xi, \eta; K^2)$ and $P(\xi, \eta; K^2)$ is similar to that between $\tilde{F}(v, w)$ and $F(\xi, \eta)$ in Eq. (I-12); that is, $\tilde{P}(\xi, \eta; K^2)$ would be the function we would choose to describe the leading Landau curve of the pentagon diagram amplitude had we been working directly in the masses and kinematic invariants rather than in the related quantities in Eqs. (4) and (9).

As discussed in Sec. 2, while Eq. (10) can be satisfied with physical invariants for sufficiently large internal masses, the spectral representation in Eq. (65) does not in general correspond to the physical amplitude since for the physical amplitude associated with the pentagon diagram in Fig. 1 X_1, X_4, and X_5 would in general be negative. To obtain the physical amplitude one might then start with Eq. (65) and do an analytic continuation in X_1, X_4, and X_5. Continuation in $x_1 (= -X_1)$ (and also in x_2) is straightforward since x_1 occurs only in the Cauchy kernel. The continuation in X_4 and X_5 is much more difficult since $F(\xi, \eta)$ depends on X_4 and $P(\xi, \eta; K^2)$ depends on both X_4 and X_5. The inner integration in Eq. (65) can, of course, be carried out, for example by using real and, if $p_\pm(\xi; K^2)$ are complex, complex partial fractions, to give a single integral representation of $I(x_i)$. Thus in principle it should be possible to generalize the method of analytic continuation used in II to apply to the pentagon diagram amplitude. In this way it should be possible to determine directly how and when complex triangle, box and pentagon singularities occur, resulting in a breakdown of even a single dispersion integral over a real domain. For the general mass case that we have been considering in this paper this would be a very difficult problem because of the increased number of singularities and their more complicated behavior. However, it is likely that this program can be carried out for some specific processes of physical interest. The method used in II could in principle also be generalized to obtain $I(x_i)$ for the case when the stability conditions $a, b, c, d, e > -1$, rather than just $a, b, c, d, e > 0$, hold.

Finally we compare our spectral representation in Eq. (65) with that given in Eq. (23) of Ref. 6. First note that $I(x_i)$ given in Eq. (65) is real and well defined since $f_+(\xi)$ is real when Eq. (10) holds and, from Eqs. (A1) and (A10), $F(\xi, \eta) > 0$ for $\xi > 1$, $\eta > f_+(\xi)$ and $P(\xi, \eta; K^2) > 0$ for $\xi \geq 1$, $\eta \geq f_+(\xi)$. Further for fixed ξ both $F(\xi, \eta)$ and $P(\xi, \eta; K^2)$ are quadratic functions of η and for fixed η they are quadratic functions of ξ. In comparison, in Eq. (23) of Ref. 6 it is assumed that $p_\pm(\xi; K^2)$ are always real whereas we show in Appendix A that they can in fact be complex for the case considered there. More important, the spectral representation in Eq. (23) of Ref. 6 is divergent, that is, infinity is obtained when the integration is carried out.

APPENDIX A

We collect here a number of results involving the various functions needed in the main body of the paper. It is assumed throughout that Eq. (10) holds. From Eqs. (34) and (30),

$$\begin{aligned} F(\xi, x_2) &\equiv F(\xi, x_2; a, b, c, X_4), \\ &= (\xi^2 - 1)(x_2^2 - 1) - 2(\xi - 1)(x_2 - 1)(aX_4 + bc) \\ &\quad - 2(\xi - 1)(a + c)(b + X_4) - 2(x_2 - 1)(a + b)(c + X_4) \\ &\quad + (aX_4 - bc)^2 - (a + b + c + X_4)^2, \\ &= (\xi^2 - 1)[x_2 - f_+(\xi)][x_2 - f_-(\xi)], \\ &= (x_2^2 - 1)[\xi - g_+(x_2)][\xi - g_-(x_2)], \end{aligned} \tag{A1}$$

where from Eqs. (31) and (19)

$$f_\pm(\xi)\equiv f_\pm(\xi,0)$$
$$=(\xi^2-1)^{-1}[(\xi-1)(aX_4+bc)+(a+b)(c+X_4)$$
$$\pm(\xi^2+2ab\xi+a^2+b^2-1)^{1/2}(\xi^2+2cX_4\xi+c^2+X_4^2-1)^{1/2}] \tag{A2}$$

and

$$g_\pm(x_2)=(x_2^2-1)^{-1}[(x_2-1)(aX_4+bc)+(a+c)(b+X_4)$$
$$\pm(x_2^2+2acx_2+a^2+c^2-1)^{1/2}(x_2^2+2bX_4x_2+b^2+X_4^2-1)^{1/2}]. \tag{A3}$$

Note also that

$$(\xi^2-1)F(\xi,x_2)=[G(\xi,x_2)]^2-\{\xi^2+2ab\xi+a^2+b^2-1\}$$
$$\times\{\xi^2+2cX_4\xi+c^2+X_4^2-1\}, \tag{A4}$$

where

$$G(\xi,x_2)\equiv G(\xi,x_2;a,b,c,X_4)$$
$$=-x_2(\xi^2-1)+(\xi-1)(aX_4+bc)+(a+b)(c+X_4). \tag{A5}$$

The above functions (with $X_4\to d$, $x_2\to y$) were also defined in Eqs. (I-12), (I-11), (I-36), and (I-A5) and their properties were discussed in detail in Sec. 8 and Appendix A of I and in Sec. 4 of II.

From Eqs. (52) and (45) we find that

$$p_\pm(\xi;K^2)=(\xi^2+2dX_5\xi+d^2+X_5^2-1)^{-1}$$
$$\times[-E(\xi,X_3,e)\pm\{F_2(\xi,-X_3)F_3(\xi,-e)\}^{1/2}], \tag{A6}$$

where

$$E(\xi,X_3,e)$$
$$=(\xi^2-1)eX_3+(\xi-1)[e(ad+bX_5)+X_3(cd+X_4X_5)-aX_4-bc]$$
$$+e(a+b)(X_5+d)+X_3(c+X_4)(X_5+d)+d^2ac+X_5^2bX_4$$
$$-X_5d(bc+aX_4), \tag{A7}$$

$$F_2(\xi,-X_3)\equiv F(\xi,-X_3;a,b,X_5,d), \tag{A8}$$

$$F_3(\xi,-e)\equiv F(\xi,-e;c,X_4,X_5,d), \tag{A9}$$

with $F(\xi,x_2;a,b,c,X_4)$ given in Eq. (A1).

Since

$$(\xi^2+2dX_5\xi+d^2+X_5^2-1)>0$$

the inequality

$$P(\xi,\eta;K^2)>0 \tag{A10}$$

will hold for $\xi\geq 1$, $\eta\geq f_+(\xi)$ if either $p_\pm(\xi;K^2)$ are complex conjugates or if Eq. (56) holds. That it is possible, when Eq. (10) holds, for $p_\pm(\xi;K^2)$ to be either real or complex conjugates depending on the value of ξ, where $\xi\geq 1$, can be seen as follows. Consider first the case when

$$0<a,b,X_5,d<1,\quad X_3>0. \tag{A11}$$

Then one of the four possible configurations of the curve Γ defined by $F_2(\xi,-X_3)=0$ is as shown in Fig. 1 of II with $\eta\to -X_3$. We see that $F_2(\xi,-X_3)$ may be positive, zero or negative depending on the values of ξ and X_3. When a,b,X_5,d are no longer restricted to be less than 1, then there are more different configurations of Γ.

Examples of the possible configurations of Γ are sketched in Ref. 12. Again $F_2(\xi,-X_3)$, and also $F_3(\xi,-e)$, may be positive, zero or negative. This statement is still true if the zeros on the right-hand sides of the inequalities in Eq. (10) are replaced by ones, the case initially considered in Ref. 6.

We now have the following cases to consider

(i) $F_2(\xi,-X_3)F_3(\xi,-e)<0$. Then $p_\pm(\xi;K^2)$ are complex conjugates and $P(\xi,\eta;K^2)>0$;

(ii) $F_2(\xi,-X_3)\geq 0$, $F_3(\xi,-e)\geq 0$. From Eq. (A7) it follows that

$$(\xi^2-1)E(\xi,X_3,e)+(\xi^2+2dX_5\xi+d^2+X_5^2-1)$$
$$\times[(\xi-1)(aX_4+bc)+(a+b)(c+X_4)]$$
$$=G_2(\xi,-X_3)G_3(\xi,-e), \tag{A12}$$

where

$$G_2(\xi,-X_3)\equiv G(\xi,-X_3;a,b,X_5,d)>0, \tag{A13}$$

$$G_3(\xi,-e)\equiv G(\xi,-e;c,X_4,X_5,d)>0, \tag{A14}$$

and $G(\xi,x_2;a,b,c,X_4)$ is given in Eq. (A5). Then using two equations similar to Eq. (A4), relating $F_2(\xi,-X_3)$ and $G_2(\xi,-X_3)$ and relating $F_3(\xi,-e)$ and $G_3(\xi,-e)$, and defining

$$\cosh\kappa_1=\frac{G_2(\xi,-X_3)}{(\xi^2+2dX_5\xi+d^2+X_5^2-1)^{1/2}(\xi^2+2ab\xi+a^2+b^2-1)^{1/2}}, \tag{A15}$$

$$\cosh\kappa_2=\frac{G_3(\xi,-e)}{(\xi^2+2dX_5\xi+d^2+X_5^2-1)^{1/2}(\xi^2+2cX_4\xi+c^2+X_4^2-1)^{1/2}} \tag{A16}$$

we find that

$$p_\pm(\xi;K^2)-f_+(\xi)=(\xi^2-1)^{-1}(\xi^2+2ab\xi+a^2+b^2-1)^{1/2}$$
$$\times(\xi^2+2cX_4\xi+c^2+X_4^2-1)^{1/2}$$
$$\times[-1-\cosh\kappa_1\cosh\kappa_2\pm\sinh\kappa_1\sinh\kappa_2]$$
$$=(\xi^2-1)^{-1}(\xi^2+2ab\xi+a^2+b^2-1)^{1/2}$$
$$\times(\xi^2+2cX_4\xi+c^2+X_4^2-1)^{1/2}$$
$$\times[-1-\cosh(\kappa_1\mp\kappa_2)]<0. \tag{A17}$$

Thus Eqs. (56) and (A10) hold.

(iii) $F_2(\xi,-X_3)<0$, $F_3(\xi,-e)<0$. In this case we define $\cos\phi_1$ by the right-hand side of Eq. (A15) and $\cos\phi_2$ by the right-hand side of Eq. (A16). Then

$$p_\pm(\xi;K^2)-f_+(\xi)$$
$$=(\xi^2-1)^{-1}(\xi^2+2ab\xi+a^2+b^2-1)^{1/2}(\xi^2+2cX_4\xi+c^2$$
$$+X_4^2-1)^{1/2}[-1-\cos(\phi_1\pm\phi_2)]<0 \tag{A18}$$

since the inequalities in Eqs. (A13) and (A14) hold. Again Eqs. (56) and (A10) are valid.

ACKNOWLEDGMENTS

It is my pleasure to thank Dr. W. S. Woolcock for discussing several points of this paper with me and Professor D. Atkinson for reading the manuscript and making several helpful suggestions.

*Research supported by Stichting voor Fundamenteel Onderzoek der Materie. Part of this work was carried out while the author was at the Australian National University, Canberra.

[1]R. E. Cutkosky, Phys. Rev. Lett. 4, 624 (1960); J. Math. Phys. 1, 294 (1960).
[2]L. D. Landau, Nucl. Phys. 13, 181 (1959); Zh. Eksp. Teor. Fiz. 37, 62 (1959) [Sov. Phys. JETP 10, 45 (1960)].
[3]R. J. Eden, P. V. Landshoff, D. I. Olive, and J. C. Polkinghorne, *The Analytic S-Matrix* (Cambridge U. P., Cambridge, 1966), p. 115.
[4]L. F. Cook and J. Tarski, Phys. Rev. Lett. 5, 585 (1960); J. Math. Phys. 3, 1 (1962).
[5]F. R. Halpern, Phys. Rev. Lett. 10, 310 (1963).
[6]O. I. Zav'yalov and V. P. Pavlov, Zh. Eksp. Teor. Fiz. 44, 1500 (1963) [Sov. Phys. JETP 17, 1009 (1963)].
[7]J. S. Frederiksen and W. S. Woolcock, Nucl. Phys. B 28, 605 (1971).
[8]J. S. Frederiksen and W. S. Woolcock, Ann. Phys. (N. Y.) 75, 503 (1973).
[9]J. S. Frederiksen and W. S. Woolcock, Ann. Phys. (N. Y.) 80, 86 (1973).
[10]S. S. Schweber, *An Introduction to Relativistic Quantum Field Theory* (Harper and Row, New York, 1961).
[11]J. S. Frederiksen, Ph. D. thesis, Australian National University (unpublished); Bull. Austral. Math. Soc. 7, 455 (A) (1972).
[12]M. Fowler, P. V. Landshoff, and R. W. Lardner, Nuovo Cimento 17, 956 (1960).

Schrödinger equation with inverse fourth-power potential, a differential equation with two irregular singular points

Wolfgang Bühring

Physikalisches Institut der Universität Heidelberg, Heidelberg, Germany
(Received 29 January 1974)

The Schrödinger radial equation with inverse fourth-power potential is treated analytically. Solutions in the form of integral representations of the generalized Laplace type are considered. Standard solutions are defined relative to each of the two irregular singular points of the differential equation. The coefficients in the linear relations persisting between any three of the standard solutions are obtained. The expressions for the coefficients, which contain some Taylor and Laurent series and finite determinants, are suitable for electronic computation. From the coefficients the S matrix and the scattering phase shifts may be obtained immediately.

1. INTRODUCTION

The inverse fourth-power potential is one of the few singular potentials for which the Schrödinger radial equation can be treated analytically. Several authors[1–4] have utilized the fact that, by suitable changes of the variables, the Schrödinger equation with this potential can be transformed into the Mathieu equation the properties of which, although complicated, are rather well known.[5–9] In this way it was possible to derive analytical expressions for the S matrix, the partial wave amplitudes, and the scattering phase shifts. The results have been reviewed recently by Frank, Land, and Spector.[10]

In the present paper the Schrödinger equation with inverse fourth-power potential is considered from a different point of view. Since the Mathieu equation is in some sense a more complicated differential equation than the Schrödinger equation is, we prefer to treat the Schrödinger equation directly. Then we are able to derive in Sec. 2 some of the already known results in a more transparent way than before.[10] While in Sec. 2 we implicitly use some results from the theory of the Mathieu equation, a more general treatment is presented in Sec. 3 which constitutes the main part of this work. The emphasis is on the fact that we are concerned with a linear differential equation with two irregular singular points of rank one. Accordingly, we consider solutions in the form of integral representations of the generalized Laplace type, thereby modifying and extending the work of Erdélyi.[11] We find exact expressions for the coefficients in the linear relation between the two fundamental sets of solutions defined relative to the two irregular singular points. These expressions, which contain some Taylor and Laurent series and *finite* determinants, are shown to be suitable for electronic computation under certain conditions. They may be used to compute the S matrix and the scattering phase shifts.

There is a different method of treating the Schrödinger equation by Fubini and Stroffolini,[12] which yields the coefficients and the S matrix in terms of *infinite* determinants.

2. MODIFIED DERIVATION OF EARLIER RESULTS

The Schrödinger radial equation with inverse fourth-power potential can be written

$$y'' + 2r^{-1}y' + [k^2 - l(l+1)r^{-2} - \beta^2 r^{-4}]y(r) = 0. \tag{1}$$

Here the potential parameter β and the momentum k are real in case of the scattering from a repulsive potential, and positive integer values of the angular momentum l are particularly important. Equation (1) has two irregular singular points of the same species at zero and infinity. In order to obtain an equation which is more symmetric with respect to interchange of zero and infinity, it is advantageous to extract an appropriate power of r by

$$y(r) = r^{-1/2} f(r). \tag{2}$$

Then $f(r)$ is a solution of the equation

$$r^2 f'' + rf' + [k^2 r^2 - (l+\tfrac{1}{2})^2 - \beta^2 r^{-2}] f(r) = 0. \tag{3}$$

This equation has power series solutions

$$f_\mu(r) = r^\mu \sum_{n=-\infty}^{\infty} c_{2n}^\mu r^{2n}, \tag{4}$$

where the coefficients obey the three-term recurrence relation

$$[(\mu+2n)^2 - (l+\tfrac{1}{2})^2] c_{2n}^\mu + k^2 c_{2n-2}^\mu - \beta^2 c_{2n+2}^\mu = 0 \tag{5}$$

with $n = \cdots, -2, -1, 0, 1, 2, \cdots$. The value of the characteristic index μ is determined by the requirement that this infinite system of linear equations have a nontrivial solution, i.e., its determinant be zero. If this condition is satisfied by $\mu = \nu$, then it is satisfied by $\mu = -\nu$ too. Consequently, there are two solutions,

$$f_{+\nu}(r) \quad \text{and} \quad f_{-\nu}(r), \tag{6}$$

which are linearly independent provided that ν is not an integer. Furthermore, the condition is satisfied if $\mu = \nu + 2m$ with $m = \cdots, -2, -1, 0, 1, 2, \cdots$, but this simply corresponds to a different labeling of the coefficients c_{2n}^μ with respect to the index n and does not give anything new. The coefficient c_0^μ may be chosen arbitrarily, the other coefficients are then defined uniquely. It is sometimes convenient to use the coefficients

$$C_{2n}^\mu = (i\beta/k)^n c_{2n}^\mu \tag{7}$$

which obey the more symmetric recurrence relation

$$[(\mu+2n)^2 - (l+\tfrac{1}{2})^2] C_{2n}^\mu + i\beta k C_{2n-2}^\mu + i\beta k C_{2n+2}^\mu = 0. \tag{8}$$

To obtain expressions for these coefficients and the characteristic index μ is a problem which has been investigated extensively in the context of the Mathieu equation.[5–9]

The representation of the solutions (6) by Eq. (4) is not useful in so far as we cannot infer their behavior for small and large values of the argument r. By analogy with the Mathieu equation, we therefore con-

sider four other solutions which are expansions in terms of products of Bessel functions:

$$f^{\mathrm{I}}_{+\nu}(r) = (C_0^{+\nu})^{-1} \sum_{n=-\infty}^{\infty} (-1)^n C_{2n}^{+\nu} J_n(kr) J_{+\nu+n}\left(\frac{i\beta}{r}\right), \tag{9a}$$

$$f^{\mathrm{I}}_{-\nu}(r) = (C_0^{-\nu})^{-1} \sum_{n=-\infty}^{\infty} (-1)^n C_{2n}^{-\nu} J_n(kr) J_{-\nu+n}\left(\frac{i\beta}{r}\right), \tag{9b}$$

$$f^{\mathrm{II}}_{+\nu}(r) = (C_0^{+\nu})^{-1} \sum_{n=-\infty}^{\infty} (-1)^n C_{2n}^{+\nu} J_n\left(\frac{i\beta}{r}\right) J_{+\nu+n}(kr), \tag{9c}$$

$$f^{\mathrm{II}}_{-\nu}(r) = (C_0^{-\nu})^{-1} \sum_{n=-\infty}^{\infty} (-1)^n C_{2n}^{-\nu} J_n\left(\frac{i\beta}{r}\right) J_{-\nu+n}(kr). \tag{9d}$$

That these four functions are solutions of Eq. (3) and the coefficients are the same as in Eq. (8) can be seen if any of Eqs. (9) is inserted into Eq. (3) and the resulting derivatives of the Bessel functions eliminated by means of the Bessel differential equation and the relation

$$2xyJ_n'(x)J_{\mu+n}'(y) = -2n(\mu+n)J_n(x)J_{\mu+n}(y) + xy[J_{n-1}(x)J_{\mu+n-1}(y) + J_{n+1}(x)J_{\mu+n+1}(y)]. \tag{10}$$

From the properties of the Bessel functions the asymptotic behavior of the solutions (9) can easily be seen when $r \to 0$ or $r \to \infty$, respectively:

$$f^{\mathrm{I}}_{+\nu}(r) \to \left(\frac{2r}{\pi i\beta}\right)^{1/2} \cos\left(\frac{i\beta}{r} - \frac{\pi}{2}\nu - \frac{\pi}{4}\right), \tag{11a}$$

$$r \to 0, \quad -\pi/2 < \arg r - \arg\beta < 3\pi/2,$$

$$f^{\mathrm{I}}_{-\nu}(r) \to \left(\frac{2r}{\pi i\beta}\right)^{1/2} \cos\left(\frac{i\beta}{r} + \frac{\pi}{2}\nu - \frac{\pi}{4}\right), \tag{11b}$$

$$r \to 0, \quad -\pi/2 < \arg r - \arg\beta < 3\pi/2,$$

$$f^{\mathrm{II}}_{+\nu}(r) \to \left(\frac{2}{\pi kr}\right)^{1/2} \cos\left(kr - \frac{\pi}{2}\nu - \frac{\pi}{4}\right), \tag{11c}$$

$$r \to \infty, \quad -\pi < \arg r + \arg k < \pi,$$

$$f^{\mathrm{II}}_{-\nu}(r) \to \left(\frac{2}{\pi kr}\right)^{1/2} \cos\left(kr + \frac{\pi}{2}\nu - \frac{\pi}{4}\right), \tag{11d}$$

$$r \to \infty, \quad -\pi < \arg r + \arg k < \pi.$$

When the power series expansions of the Bessel functions are inserted into Eqs. (9), both $f^{\mathrm{I}}_{+\nu}(r)$ and $f^{\mathrm{II}}_{-\nu}(r)$ are seen to have the form of $r^{-\nu}$ times an even Laurent expansion and therefore are proportional to one another and to the solution $f_{-\nu}(r)$ of Eq. (4). Similarly, both $f^{\mathrm{I}}_{-\nu}(r)$ and $f^{\mathrm{II}}_{+\nu}(r)$ are proportional to the solution $f_{+\nu}(r)$. Furthermore, we observe by inspection of Eqs. (9) that for $r = X$, where

$$X = \sqrt{i\beta/k}, \tag{12}$$

the arguments of the Bessel functions become equal, and we have

$$f^{\mathrm{I}}_{+\nu}(X) = f^{\mathrm{II}}_{+\nu}(X), \tag{13a}$$

$$f^{\mathrm{I}}_{-\nu}(X) = f^{\mathrm{II}}_{-\nu}(X). \tag{13b}$$

Combining these facts, we find

$$f^{\mathrm{II}}_{-\nu}(r) = R f^{\mathrm{I}}_{+\nu}(r), \tag{14a}$$

$$f^{\mathrm{I}}_{-\nu}(r) = R f^{\mathrm{II}}_{+\nu}(r), \tag{14b}$$

where, because of Eqs. (13), only one constant

$$R = f^{\mathrm{I}}_{-\nu}(X)/f^{\mathrm{II}}_{+\nu}(X) = f^{\mathrm{II}}_{-\nu}(X)/f^{\mathrm{I}}_{+\nu}(X) \tag{15}$$

appears. A factor $\exp(-i\frac{1}{2}\pi\nu)$ may be extracted by

$$R = \tilde{R}\exp(-i\tfrac{1}{2}\pi\nu) \tag{16}$$

so that $\tilde{R}$ is real (if β, k, l, and ν are), as can be seen if any of Eqs. (14) is considered for real values of r. Explicitly we have from Eq. (15)

$$R = \frac{C_0^{+\nu}\sum_{n=-\infty}^{\infty}(-1)^n C_{2n}^{-\nu} J_n(\sqrt{i\beta k}) J_{-\nu+n}(\sqrt{i\beta k})}{C_0^{-\nu}\sum_{n=-\infty}^{\infty}(-1)^n C_{2n}^{+\nu} J_n(\sqrt{i\beta k}) J_{+\nu+n}(\sqrt{i\beta k})}. \tag{17}$$

Taking suitable linear combinations of Eqs. (9a) and (9b) and returning to the original differential equation (1), we define two standard solutions

$$y^{(1)}(r) = \left(\frac{\pi\beta}{2r}\right)^{1/2} (\sin\pi\nu)^{-1} \left[\exp\left(+i\frac{\pi}{2}\nu\right) f^{\mathrm{I}}_{-\nu}(r) - \exp\left(-i\frac{\pi}{2}\nu\right) f^{\mathrm{I}}_{+\nu}(r)\right], \tag{18a}$$

$$y^{(2)}(r) = \left(\frac{\pi\beta}{2r}\right)^{1/2} i(\sin\pi\nu)^{-1} \left[\exp\left(-i\frac{\pi}{2}\nu\right) f^{\mathrm{I}}_{-\nu}(r) - \exp\left(+i\frac{\pi}{2}\nu\right) f^{\mathrm{I}}_{+\nu}(r)\right], \tag{18b}$$

which, according to Eqs. (11a) and (11b), have a simple behavior at the origin,

$$y^{(1)}(r) \sim \exp(-\beta/r), \tag{19a}$$

$$y^{(2)}(r) \sim \exp(+\beta/r). \tag{19b}$$

Using Eqs. (14) and (16) we find the alternative representation

$$y^{(1)}(r) = \left(\frac{\pi\beta}{2r}\right)^{1/2} \frac{1}{\sin\pi\nu} [\tilde{R} f^{\mathrm{II}}_{+\nu}(r) - \tilde{R}^{-1} f^{\mathrm{II}}_{-\nu}(r)], \tag{20a}$$

$$y^{(2)}(r) = \left(\frac{\pi\beta}{2r}\right)^{1/2} \frac{1}{\sin\pi\nu} i[\tilde{R}\exp(-i\pi\nu) f^{\mathrm{II}}_{+\nu}(r) - \tilde{R}^{-1}\exp(+i\pi\nu) f^{\mathrm{II}}_{-\nu}(r)], \tag{20b}$$

from which, by means of Eqs. (11c) and (11d), the behavior at infinity can be found

$$y^{(1)}(r) \sim \left(\frac{\beta}{k}\right)^{1/2} \frac{1}{2r\sin\pi\nu} \left\{\left[\tilde{R}\exp\left(-i\frac{\pi}{2}\nu\right) - \tilde{R}^{-1}\exp\left(+i\frac{\pi}{2}\nu\right)\right] \times \exp\left(-i\frac{\pi}{4}\right)\exp(+ikr) + \left[\tilde{R}\exp\left(+i\frac{\pi}{2}\nu\right) - \tilde{R}^{-1}\exp\left(-i\frac{\pi}{2}\nu\right)\right]\exp\left(+i\frac{\pi}{4}\right)\exp(-ikr)\right\}, \tag{21a}$$

$$y^{(2)}(r) \sim \left(\frac{\beta}{k}\right)^{1/2} \frac{i}{2r\sin\pi\nu} \left\{\left[\tilde{R}\exp\left(-i\frac{3\pi}{2}\nu\right) - \tilde{R}^{-1}\exp\left(+i\frac{3\pi}{2}\nu\right)\right] \times \exp\left(-i\frac{\pi}{4}\right)\exp(+ikr) + \left[\tilde{R}\exp\left(-i\frac{\pi}{2}\nu\right) - \tilde{R}^{-1}\exp\left(+i\frac{\pi}{2}\nu\right)\right]\exp\left(+i\frac{\pi}{4}\right)\exp(-ikr)\right\}. \tag{21b}$$

The S matrix and the scattering phase shift δ are defined by

$$S = \exp(2i\delta) = -(K^+/K^-)\exp(i\pi l), \tag{22}$$

where K^+ and K^- are the factors of $\exp(+ikr)$ and $\exp(-ikr)$, respectively, in Eq. (21a). Accordingly, we have

$$S = \exp(2i\delta) = \frac{\tilde{R}^2 - \exp(+i\pi\nu)}{\tilde{R}^2 - \exp(-i\pi\nu)} \exp[i\pi(l+\tfrac{1}{2}-\nu)] \tag{23}$$

or, in view of Eq. (16),

$$S = \frac{R^2-1}{R^2 - \exp(-2\pi i\nu)} \exp[i\pi(l+\tfrac{1}{2}-\nu)], \tag{24}$$

which is a formula appearing in the review article,[10] apart from the factor $\exp(-i\pi\nu)$ which evidently has been lost there. (This factor has been lost in a trivial way during preparation of the review article, for it is present in the formulas of both the original papers[3,4] quoted.)

3. GENERAL TREATMENT OF THE DIFFERENTIAL EQUATION

A. Symmetry of the differential equation

We want to consider solutions of the differential equation (1) and are particularly interested in their behavior when $r \to 0$ and when $r \to \infty$. Both these cases can be dealt with simultaneously if we take advantage of the symmetry of the equation with respect to zero and infinity. For we observe that if we introduce a new independent variable x by

$$x = kr \quad (\text{or } x = -kr) \tag{25a}$$

or by

$$x = i\beta/r \quad (\text{or } x = -i\beta/r), \tag{25b}$$

then

$$g(x) = f(r) = r^{1/2}y(r) \tag{26}$$

is a solution of the same differential equation in either case. It therefore suffices to investigate the behavior when $x \to \infty$ of the solutions of the x equation

$$x^2g'' + xg' + [x^2 - (l+\tfrac{1}{2})^2 - \beta^2k^2x^{-2}]g(x) = 0 \tag{27}.$$

B. Solutions of the x-equation

In a similar way as in Sec. 2 we introduce two standard solutions $g_{+\nu}(x)$ and $g_{-\nu}(x)$ by

$$g_\mu(x) = \sum_{n=-\infty}^{\infty} d_{2n}^\mu x^{\mu+2n}, \quad \mu = -\nu, +\nu, \tag{28}$$

where ν is the characteristic index as before and the coefficients d_{2n}^μ obey the recurrence relation

$$[(\mu+2n)^2 - (l+\tfrac{1}{2})^2]d_{2n}^\mu + d_{2n-2}^\mu - \beta^2k^2d_{2n+2}^\mu = 0. \tag{29}$$

We take for the constants of integration

$$d_0^{+\nu} = d_0^{-\nu} = 1. \tag{30}$$

The coefficients then satisfy the symmetry relation

$$d_{-2n}^{-\nu} = (-\beta^2k^2)^n d_{+2n}^{+\nu}. \tag{31}$$

In order to investigate the behavior of the solutions (28) near infinity, we first extract a factor x^λ in view of greater flexibility and then consider integral representations of the generalized Laplace type

$$g(x) = \frac{1}{2\pi i} x^\lambda \int_C \exp(xt)v(t)\,dt. \tag{32}$$

Here $v(t)$ is a solution of the t equation

$$(t^2+1)v^{IV} + (7-2\lambda)tv''' - (l+\tfrac{7}{2}-\lambda)(l-\tfrac{5}{2}+\lambda)v'' - \beta^2k^2v(t) = 0, \tag{33}$$

and the contour C has to be chosen such that the bilinear concomitant[13]

$$\begin{aligned}\exp(xt)[&(t^2+1)(x^3v - x^2v' + xv'' - v''') - 2t(x^2v - 2xv' + 3v'') \\ &+ 2(xv - 3v') + (2\lambda+1)(tx^2v - txv' + tv'' - xv + 2v') \\ &- (l+\tfrac{1}{2}+\lambda)(l+\tfrac{1}{2}-\lambda)(xv - v')]\end{aligned} \tag{34}$$

has the same value (identically in x) at both the termini of the contour. The parameter λ, which is quite arbitrary, will be specified later as circumstances demand.

C. Construction of an appropriate solution of the t-equation

1. Solutions outside the unit circle

The t equation (33) has two regular singular points at $t = i$ and $t = -i$ and one irregular singular point at infinity. Outside the unit circle, a solution may be represented by an appropriate power times an even Laurent expansion

$$v_\mu(t) = t^{\lambda-\mu-1}\phi_\mu(t^2), \tag{35}$$

$$\phi_\mu(z) = \sum_{n=-\infty}^{\infty} b_{2n}^\mu z^{-n}, \tag{36}$$

where the coefficients obey the recurrence relation

$$\begin{aligned}&[(\mu+2n)^2 - (l+\tfrac{1}{2})^2](\mu-\lambda+2n+1)(\mu-\lambda+2n+2)b_{2n}^\mu \\ &+ (\mu-\lambda+2n-1)(\mu-\lambda+2n)(\mu-\lambda+2n+1)(\mu-\lambda+2n+2)b_{2n-2}^\mu \\ &- \beta^2k^2b_{2n+2}^\mu = 0.\end{aligned} \tag{37}$$

This recurrence relation is satisfied if

$$b_{2m}^\mu = \frac{\Gamma(\mu-\lambda+1+2m)}{\Gamma(\mu-\lambda+1)} d_{2m}^\mu, \tag{38}$$

as may be seen by comparison with Eq. (29). Possible values of μ are therefore $\mu = +\nu$ and $\mu = -\nu$, where ν is the characteristic index as before. But since $v_\mu(t)$ is a solution of a fourth order equation, there are two further significantly different values. In fact, the recurrence relation is also satisfied if all the coefficients with positive indices > 0 vanish and $\mu = \lambda-1, \lambda-2, \lambda-3$, $\lambda-4$ is one of the roots of the corresponding fourth order indicial equation. Consequently, there are two further solutions:

$$v_{\lambda-1}(t) = \sum_{n=0}^{\infty} b_{-2n}^{\lambda-1}t^{2n}, \tag{39a}$$

$$v_{\lambda-2}(t) = \sum_{n=0}^{\infty} b_{-2n}^{\lambda-2}t^{2n+1}. \tag{39b}$$

While here the coefficients b_0^μ are arbitrary constants of integration, the coefficients b_{-2}^μ, which also are not determined by the recurrence relation, have to be chosen such that the series (39) converge outside the unit circle. But then they converge also on and inside the unit circle and represent entire functions of t. Accordingly, there are two solutions of the t equation, one even and one odd function of t, which are analytic at both the regular singular points $t = -i, +i$. These solutions are not of interest since they do not contribute to the con-

tour integrals we will consider. Their existence is relevant in so far as, at a later stage of our investigation, it will suffice to consider two linearly independent solutions of the t equation instead of four.

2. Solutions relative to the regular singular points

The exponents of the t equation relative to any of the two regular singular points $t=-i,+i$ are $0,1,2,\lambda-\frac{1}{2}$. Provided that λ is not half an integer, the solution does not contain logarithmic terms, even though all the exponents apart one are integers. The solutions can therefore be written

$$v^+(t)=F(1+it), \quad |t-i|<2, \tag{40}$$

$$u_j^+(t)=G_j(1+it), \quad |t-i|<2, \quad j=0,1,2, \tag{41}$$

and

$$v^-(t)=F(1-it), \quad |t+i|<2, \tag{42}$$

$$u_j^-(t)=G_j(1-it), \quad |t+i|<2, \quad j=0,1,2, \tag{43}$$

where

$$G_j(z)=\sum_{n=0}^{\infty}A_n(j)z^{j+n}, \quad |z|<2, \tag{44}$$

$$F(z)=z^{\lambda-1/2}H(z), \tag{45}$$

$$H(z)=\sum_{n=0}^{\infty}A_n(\lambda-\tfrac{1}{2})z^n, \quad |z|<2. \tag{46}$$

Here we may choose the initial coefficients arbitrarily as

$$A_0(q)=1 \quad \text{for } q=0,1,2,\lambda-\tfrac{1}{2}, \tag{47}$$

$$A_1(0)=0, \quad A_2(0)=0, \quad A_1(1)=0. \tag{48}$$

The other coefficients then are determined by the recurrence relation

$$A_n(q)=\frac{(q+n-\lambda+l+\frac{1}{2})(q+n-\lambda-l-\frac{1}{2})}{2(q+n-\lambda+\frac{1}{2})(q+n)}A_{n-1}(q) + \frac{\beta^2k^2}{2(q+n-\lambda+\frac{1}{2})(q+n)(q+n-1)(q+n-2)}A_{n-3}(q), \tag{49}$$

$[n>0$ if $q=2,\lambda-\frac{1}{2}$; $n>1$ if $q=1$; $n>2$ if $q=0$; $A_{-1}(q)=A_{-2}(q)=0]$.

By Eqs. (40), (41) and (42), (43) we have two fundamental sets of solutions, valid in different but overlapping domains of the t plane. Any solution of one set may therefore be expressed as a linear combination of the solutions of the other set, in particular,

$$v^+(t)=Ev^-(t)+\sum_{j=0}^{2}B_ju_j^-(t). \tag{50}$$

The coefficients E and B_j may be determined by evaluating this equation and its first three derivatives at $t=0$. They therefore appear as the solution of the system of linear equations

$$\begin{pmatrix} F(1) & G_0(1) & G_1(1) & G_2(1) \\ F'(1) & G_0'(1) & G_1'(1) & G_2'(1) \\ F''(1) & G_0''(1) & G_1''(1) & G_2''(1) \\ F'''(1) & G_0'''(1) & G_1'''(1) & G_2'''(1) \end{pmatrix} \begin{pmatrix} E \\ B_0 \\ B_1 \\ B_2 \end{pmatrix} = \begin{pmatrix} F(1) \\ -F'(1) \\ F''(1) \\ -F'''(1) \end{pmatrix} \tag{51}$$

and will be considered as known numbers. Introducing

$$G(z)=\sum_{j=0}^{2}B_jG_j(z), \tag{52}$$

$$u^+(t)=G(1+it), \tag{53}$$

$$u^-(t)=G(1-it), \tag{54}$$

we have

$$v^+(t)=Ev^-(t)+u^-(t). \tag{55}$$

There is also a relation

$$u^+(t)=(1-E^2)v^-(t)-Eu^-(t), \tag{56}$$

as may be seen by evaluating Eqs. (55), (56) and their first derivatives at $t=0$. Similarly, we have

$$v^-(t)=Ev^+(t)+u^+(t), \tag{57}$$

$$u^-(t)=(1-E^2)v^+(t)-Eu^+(t). \tag{58}$$

3. Multiplicative solutions and analytical continuation

Let us consider a path in the t plane, in the form of a loop enclosing the two regular singular points $t=-i,+i$, such that it lies inside the region where $v^+(t)$ or $v^-(t)$ converges. If we start at some point P of this path with any of the appropriate solutions, say $v^+(t)$, and continue it analytically along the path until we arrive at the point P again, we generally do not reproduce $v^+(t)$ but obtain some linear combination of $v^+(t)$ and $u^+(t)$. By taking appropriate linear combinations of $v^+(t)$ and $u^+(t)$, we now want to construct multiplicative solutions $w(t)$ such that $w(t)$ is reproduced, apart from a constant factor, after the loop has been described. Considering the special loop shown in Fig. 1, we define at the point

$$P_2: \quad w(t)=\alpha v^+(t)+\gamma u^+(t). \tag{59a}$$

According to Eqs. (55), (56) we also have at

$$P_2: w(t)=[\alpha E+\gamma(1-E^2)]v^-(t)+[\alpha-\gamma E]u^-(t). \tag{59b}$$

Following the loop from P_2 in the negative direction, the description of the circle around $-i$ increases $\arg(1-it)$ by $-2\pi i$, so that we obtain at

$$P_1: \quad w(t)=[\alpha E+\gamma(1-E^2)]\exp[-2\pi i(\lambda-\tfrac{1}{2})]v^-(t)+[\alpha-\gamma E]u^-(t) = w^{(1)}(t). \tag{60}$$

Following the loop from P_2 in the positive direction, the description of the circle around $+i$ increases $\arg(1+it)$ by $+2\pi i$ so that we obtain at

$$P_3: \quad w(t)=\alpha\exp[+2\pi i(\lambda-\tfrac{1}{2})]v^+(t)+\gamma u^+(t). \tag{61a}$$

According to Eqs. (55), (56) we also have at

$$P_3: \quad w(t)=\{\alpha E\exp[+2\pi i(\lambda-\tfrac{1}{2})]+\gamma(1-E^2)\}v^-(t) +\{\alpha\exp[+2\pi i(\lambda-\tfrac{1}{2})]-\gamma E\}u^-(t) = w^{(3)}(t). \tag{61b}$$

The constants α and γ have to be chosen such that the solutions at P_1 and P_3 are proportional to each other, say

$$w^{(3)}(t)=p\exp(2\pi i\lambda)w^{(1)}(t), \tag{62}$$

where the constant of proportionality has been denoted by $p\exp(2\pi i\lambda)$ for later convenience. Then α and γ are a solution of the homogeneous system of linear equations

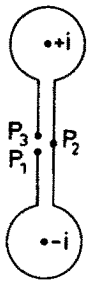


FIG. 1. t plane: path along which analytical continuation is considered. The points P_1 and P_3 are to coincide but have been drawn separately to indicate the termini of the path. Furthermore, the straight lines may be assumed to coincide with the imaginary axis.

$$E[p-\exp(2\pi i\lambda)]\alpha+(1-E^2)(p+1)\gamma=0, \tag{63a}$$

$$-(p+1)\alpha+E[p-\exp(-2\pi i\lambda)]\gamma=0. \tag{63b}$$

The requirement that the determinant be zero leads to

$$p^2+2[1-2(E\cos\pi\lambda)^2]p+1=0. \tag{64}$$

The roots p_1 and p_2 of this equation obey the relations

$$p_1+p_2=-2[1-2(E\cos\pi\lambda)^2], \tag{65}$$

$$p_1p_2=1. \tag{66}$$

Because of the last equation, the two roots may conveniently be represented by means of one (not necessarily real) parameter ν in the form

$$p_1=\exp(-2\pi i\nu),\quad p_2=\exp(+2\pi i\nu). \tag{67}$$

Then we have

$$p_1+p_2=-2(1-2\cos^2\pi\nu) \tag{68}$$

and, by comparison with Eq. (65),

$$\cos^2\pi\nu=(E\cos\pi\lambda)^2. \tag{69}$$

The further discussion may be restricted to one choice of p if all the quantities depending on p carry an index μ corresponding to $p=\exp(-2\pi i\mu)$. The name ν for the new parameter of Eqs. (67) is appropriate in view of the fact that it is the characteristic index. For we are constructing two functions $w_\mu(t)$ which, according to Eqs. (62) and (67), obey the circuit relations

$$w_\mu[\exp(2\pi i)t]=\exp[2\pi i(\lambda-\mu)]w_\mu(t) \tag{70}$$

with $\mu=\nu$ or $\mu=-\nu$, respectively. But as shown in Sec. 3C1, there are just two solutions which are not single-valued and obey the same circuit relations with ν equal to the characteristic index. It should be noted, however, that the quantity ν here is defined by Eqs. (67) or (69) mod 1 only, while the characteristic index has been defined mod 2. This fact is reflected in the sign ambiguity arising if we want to take the square root of Eq. (69). This sign has to be chosen such that the functions $w_\mu(t)$ obey also the appropriate half-circuit relations suggested by Eq. (35), namely

$$w_\mu[\exp(i\pi)t]=\exp[i\pi(\lambda-\mu-1)]w_\mu(t). \tag{71}$$

Equation (59b) may be simplified by means of Eqs. (63) so that we have at P_2

$$w_\mu(t)=\begin{cases}\alpha_\mu v^+(t)+\gamma_\mu u^+(t) & \text{if } |t-i|<2,\\ \dfrac{E\cos\pi\lambda}{\cos\pi\mu}\exp[i\pi(\lambda-\mu-1)]\{\alpha_\mu\exp[2\pi i(\lambda-\frac12)]v^-(t)+\gamma_\mu u^-(t)\} & \text{if } |t+i|<2.\end{cases} \tag{72}$$

In order to check the half-circuit relation (71) we evaluate Eq. (72) for $t=2i$ and $t=-2i$ assuming that $\arg(1+it)=\arg(1-it)=0$ on the imaginary axis near P_2. By means of Eqs. (40), (42), (45), (53), (54) we find

$$w_\mu(2i)=\alpha_\mu\exp[i\pi(\lambda-\tfrac12)]H(-1)+\gamma_\mu G(-1), \tag{73a}$$

$$w_\mu(-2i)=\frac{E\cos\pi\lambda}{\cos\pi\mu}\exp[-i\pi(\lambda-\mu-1)]\times\{\alpha_\mu\exp[i\pi(\lambda-\tfrac12)]H(-1)+\gamma_\mu G(-1)\}. \tag{73b}$$

By comparison with the half-circuit relation (71) we conclude that the equation determining the characteristic index is

$$\cos\pi\nu=E\cos\pi\lambda \tag{74}$$

rather than Eq. (69). Using Eq. (74) the ratio of the coefficients α_μ and γ_μ may be found from any of Eqs. (63) to be

$$\alpha_\mu/\gamma_\mu=\exp(-i\pi[\lambda-\tfrac12])\sin\pi(\lambda-\mu)/\cos\pi\lambda. \tag{75}$$

If we choose arbitrarily the normalization

$$\gamma_\mu=\exp(i\tfrac12\pi[\lambda-\mu-1])\Gamma(\tfrac12-\lambda)\cos\pi\lambda/\sin\pi(\lambda-\mu) \tag{76}$$

in view of later convenience, all the quantities in Eq. (72) are known. The analytical continuation of $w_\mu(t)$ for larger values of $|t|$ is, according to the discussion following Eq. (70), proportional to the function $v_\mu(t)$ given by Eq. (35). We then obtain finally

$$w_\mu(t)=\begin{cases}\Gamma(\frac12-\lambda)\exp[+i(\pi/2)(\lambda-\mu-1)]\left\{\exp[-i\pi(\lambda-\frac12)]v^+(t)+\dfrac{\cos\pi\lambda}{\sin\pi(\lambda-\mu)}u^+(t)\right\} & \text{if } |t-i|<2,\\ \Gamma(\frac12-\lambda)\exp[-i(\pi/2)(\lambda-\mu-1)]\left\{\exp[+i\pi(\lambda-\frac12)]v^-(t)+\dfrac{\cos\pi\lambda}{\sin\pi(\lambda-\mu)}u^-(t)\right\} & \text{if } |t+i|<2,\\ \Delta_\mu\Gamma(\mu-\lambda+1)v_\mu(t) & \text{if } |t|>1.\end{cases} \tag{77}$$

Here the constant of proportionality

$$\Delta_\mu=\frac{\Gamma(\frac12-\lambda)\Gamma(\lambda-\mu)[\sin\pi(\lambda-\mu)\,H(-1)+\cos\pi\lambda\,G(-1)]}{\pi 2^{\lambda-\mu-1}\phi_\mu(-4)} \tag{78}$$

may be checked by comparing the first and last line of Eq. (77) for $t=2i$. The powers are defined such that near the point P_2 of Fig. 1 we have $\arg(1-it)=\arg(1+it)=0$ if t is on the imaginary axis while $|t|<1$, and $\arg t=0$ if t is real and positive.

D. Contour integral solutions of the x-equation

1. Suitable contours

The t equation has one irregular singular point at infinity. It can be shown that there are four linearly in-

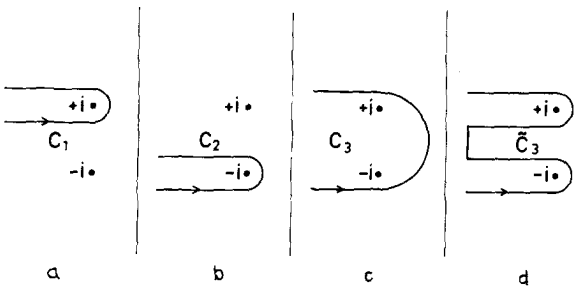


FIG. 2. t-plane: contours suitable for the integral representation.

dependent solutions which, when $|t| \to \infty$, behave asymptotically as $t^{-1+(1/2)\lambda}\exp(2\epsilon\sqrt{\beta k t})$ with $\epsilon = -i, 1, i, -1$, respectively. This behavior is dominated by the kernel $\exp(xt)$ of the integral representation (32) so that the bilinear concomitant (34) tends to zero when $|t| \to \infty$ in appropriate sectors of the t plane depending on $\arg x$. Consequently, there are contours, suitable for the integral representation, which start somewhere at infinity, enclose at least one of the regular singular points, and return to the starting point at infinity. Some examples of contours appropriate for our purpose are shown in Fig. 2.

2. Solutions relative to the irregular singular point at infinity

We are going to define two standard solutions of the x equation relative to the irregular singular point at infinity. This can be done in analogy to the definition of Hankel functions, since for $\beta^2 k^2 \to 0$ the differential equation reduces to the Bessel equation of the order $l+\frac{1}{2}$. For our purpose, however, it is more convenient to use a different normalization. Our standard solutions are

$$g^{(1)}(x) = \frac{1}{2\pi i} x^{\lambda} \int_{C_1} \exp(xt) w^{+}(t)\, dt, \quad -\pi/2 < \arg x < \pi/2, \tag{79a}$$

$$g^{(2)}(x) = \frac{1}{2\pi i} x^{\lambda} \int_{C_2} \exp(xt) w^{-}(t)\, dt, \quad -\pi/2 < \arg x < \pi/2, \tag{79b}$$

where

$$\left.\begin{aligned} w^{+}(t) &= \Gamma(\tfrac{1}{2}-\lambda)\exp[-i(\pi/2)\lambda] v^{+}(t) \\ &= \Gamma(\tfrac{1}{2}-\lambda)\exp[-i(\pi/2)\lambda](1+it)^{\lambda-1/2} H(1+it) \end{aligned}\right\} \quad \text{if } |t-i|<2, \tag{80a}$$

$$\left.\begin{aligned} w^{-}(t) &= \Gamma(\tfrac{1}{2}-\lambda)\exp[+i(\pi/2)\lambda] v^{-}(t) \\ &= \Gamma(\tfrac{1}{2}-\lambda)\exp[+i(\pi/2)\lambda](1-it)^{\lambda-1/2} H(1-it) \end{aligned}\right\} \quad \text{if } |t+i|<2. \tag{80b}$$

Here the powers are defined such that $\arg(1-it) = \arg(1+it) = 0$ when t is on the imaginary axis while $|t|<1$.

The solutions (79) may be continued analytically into larger sectors of the x plane by rotation of the contour. In each case the possible angles of rotation are limited by the presence of the other regular singular point so that $g^{(1)}(x)$ may be defined in the sector $-\pi < \arg x < 2\pi$ and $g^{(2)}(x)$ in the sector $-2\pi < \arg x < \pi$. If the Taylor series are inserted for $H(1+it)$ and $H(1-it)$ and the integrals performed term by term, then the asymptotic expansions of the standard solutions for $|x| \to \infty$ are obtained in the usual way:

$$g^{(1)}(x) \sim \exp(-i\pi/4) x^{-1/2} \exp(+ix) \sum_{n=0}^{\infty} a_n (ix)^{-n}, \quad -\pi < \arg x < 2\pi, \tag{81a}$$

$$g^{(2)}(x) \sim \exp(+i\pi/4) x^{-1/2} \exp(-ix) \sum_{n=0}^{\infty} a_n (-ix)^{-n}, \quad -2\pi < \arg x < \pi. \tag{81b}$$

Here the coefficients

$$a_n = A_n(\lambda - \tfrac{1}{2})\Gamma(\lambda + \tfrac{1}{2} + n)/\Gamma(\lambda + \tfrac{1}{2}), \tag{82}$$

which do not depend on λ, can be obtained from the recurrence relation

$$a_n = \frac{(n+l)(n-l-1)}{2n} a_{n-1} + \frac{\beta^2 k^2}{2n} a_{n-3} \tag{83}$$

with $a_0 = 1$, $a_{-1} = a_{-2} = 0$.

3. Laurent series solutions

Let us consider the integral representation (32) with the contour C_3 and $v(t) = \Gamma(\mu - \lambda + 1) v_{\mu}(t)$, where $\arg t = 0$ if t is real and positive. Inserting the Laurent series for $v_{\mu}(t)$ according to Eqs. (35), (36) and integrating term by term, we find that the result is just the function $g_{\mu}(x)$ defined by Eq. (28). The standard solutions (28) therefore have the integral representation

$$g_{\mu}(x) = \frac{1}{2\pi i} x^{\lambda} \Gamma(\mu - \lambda + 1) \int_{C_3} \exp(xt) v_{\mu}(t)\, dt, \quad -\pi/2 < \arg x < \pi/2, \tag{84}$$

for $\mu = \nu, -\nu$. The analytical continuation for other sectors may be obtained by rotation of the contour by an appropriate angle. Therefore Eq. (28) defines $g_{\mu}(x)$ for arbitrary values of $\arg x$.

4. Linear relations between the solutions

Since we are concerned with a second order differential equation, the four standard solutions $g^{(1)}(x)$, $g^{(2)}(x)$, $g_{\nu}(x)$, $g_{-\nu}(x)$ we have introduced so far are not independent, but linear relations persist between any three of them. In order to determine the coefficients we consider the integral representation (32) with the contour C_3 and the function $w_{\mu}(t)$ of Eq. (77) inserted for $v(t)$. Then, according to Eq. (84), the integral is equal to $\Delta_{\mu} g_{\mu}(x)$. Now the contour C_3 may be deformed into the contour $\tilde{C}_3$ which is just the sum of the contours C_1 and C_2, apart from a straight line, connecting them at infinity, which does not contribute to the integral (Fig. 2). According to Eqs. (79), (80), the integrals along the contours C_1 and C_2 give $g^{(1)}(x)$ and $g^{(2)}(x)$, respectively, multiplied with the appropriate factors following by comparison of Eqs. (80) with the singular parts of Eq. (77). We therefore obtain

$$\Delta_{\mu} g_{\mu}(x) = \exp[-i(\pi/2)\mu] g^{(1)}(x) + \exp[+i(\pi/2)\mu] g^{(2)}(x), \quad \mu = \nu, -\nu. \tag{85}$$

Then, provided that the characteristic index ν is not an integer, we also have

$$g^{(1)}(x) = \frac{1}{2i\sin\pi\nu}\{\exp[+i(\pi/2)\nu]\Delta_{-\nu}g_{-\nu}(x) - \exp[-i(\pi/2)\nu]\,\Delta_{+\nu}g_{+\nu}(x)\}, \tag{86a}$$

$$g^{(2)}(x) = \frac{-1}{2i\sin\pi\nu}\{\exp[-i(\pi/2)\nu)]\Delta_{-\nu}g_{-\nu}(x) - \exp[+i(\pi/2)\nu]\Delta_{+\nu}g_{+\nu}(x)\}. \tag{86b}$$

5. Circuit relations

From Eq. (28) the circuit relations for $g_\nu(x)$ and $g_{-\nu}(x)$ are known to be

$$g_\mu[\exp(i\pi m)x] = \exp(i\pi m\mu)g_\mu(x), \tag{87}$$

where m is any integer and $\mu = \nu, -\nu$. Then, by means of Eqs. (85)—(87), the circuit relations for $g^{(1)}(x)$ and $g^{(2)}(x)$ may be obtained:

$$g^{(1)}[\exp(m\pi i)x] = \frac{\sin(1-m)\pi\nu}{\sin\pi\nu}\,g^{(1)}(x) - \frac{\sin m\pi\nu}{\sin\pi\nu}g^{(2)}(x), \tag{88a}$$

$$g^{(2)}[\exp(m\pi i)x] = \frac{\sin m\pi\nu}{\sin\pi\nu}\,g^{(1)}(x) + \frac{\sin(1+m)\pi\nu}{\sin\pi\nu}g^{(2)}(x). \tag{88b}$$

These equations extend the definitions of the functions $g^{(1)}(x)$ and $g^{(2)}(x)$ and allow one to obtain their asymptotic expansions for values of arg$x$ other than those considered in Sec. 3D2.

6. The case when the characteristic index is an integer

It may happen for particular values of β^2k^2 that the characteristic index ν is an integer. Then $g_{+\nu}(x)$ and $g_{-\nu}(x)$ are no longer linearly independent, so that Eqs. (86) become meaningless. But the linear relation (85) and the coefficient Δ_μ (where $\mu = \nu, -\nu$) remain relevant. The Eq. (78) defining Δ_μ, however, breaks down whenever λ is half an integer or $\lambda - \mu$ is an integer. Here we see one advantage of having introduced the arbitrary parameter λ by extracting the factor x^λ in front of the Laplace integral. For otherwise, i.e., if $\lambda = 0$, Eq. (78) would break down whenever μ is an integer. But Δ_μ is in fact independent of λ, so we may choose any convenient value, e.g., $\lambda = \frac{1}{4}$. Then Δ_μ remains well-defined by Eq. (78) even if μ is an integer.

We do not want to explain the other aspects of the case when μ is an integer, which is quite analogeous to the case of Bessel functions of integer order.[11]

E. Solutions of the r-equation

1. Solutions relative to the singular points

The solutions $g^{(1)}(x)$ and $g^{(2)}(x)$ may now be used, according to Sec. 3A, to define two fundamental sets of solutions of the r equation (1) relative to the irregular singular points at the origin and infinity, respectively,

$$y^{(1)}(r) = i\left(\frac{\beta}{r}\right)^{1/2} g^{(1)}\left(\frac{\beta}{r}\exp(i\pi/2)\right), \tag{89a}$$

$$y^{(2)}(r) = \left(\frac{\beta}{r}\right)^{1/2} g^{(2)}\left(\frac{\beta}{r}\exp(i\pi/2)\right), \tag{89b}$$

$$y^{(3)}(r) = \exp(+i\pi/4)\left(\frac{k}{r}\right)^{1/2} g^{(1)}(kr), \tag{90a}$$

$$y^{(4)}(r) = \exp(-i\pi/4)\left(\frac{k}{r}\right)^{1/2} g^{(2)}(kr). \tag{90b}$$

They have the asymptotic expansions

$$y^{(1)}(r) \sim \exp(-\beta/r)\sum_{n=0}^{\infty} a_n\left(-\frac{r}{\beta}\right)^n, \quad -3\pi/2 < \arg r/\beta < 3\pi/2, \quad r \to 0, \tag{91a}$$

$$y^{(2)}(r) \sim \exp(+\beta/r)\sum_{n=0}^{\infty} a_n\left(\frac{r}{\beta}\right)^n, \quad -\pi/2 < \arg r/\beta < 5\pi/2, \quad r \to 0, \tag{91b}$$

$$y^{(3)}(r) \sim \frac{1}{r}\exp(+ikr)\sum_{n=0}^{\infty} a_n(ikr)^{-n}, \quad -\pi < \arg kr < 2\pi, \quad r \to \infty, \tag{92a}$$

$$y^{(4)}(r) \sim \frac{1}{r}\exp(-ikr)\sum_{n=0}^{\infty} a_n(-ikr)^{-n}, \quad -2\pi < \arg kr < \pi, \quad r \to \infty, \tag{92b}$$

with the coefficients a_n from Eq. (83).

2. Linear relations between the solutions

The linear relations persisting between any three of the solutions (89), (90) may be obtained by means of the functions $g_\mu(x)$, which obey the symmetry relation

$$g_\mu\left(\frac{\beta}{r}\exp(i\pi/2)\right) = \exp[i(\pi/2)\mu](\beta k)^\mu g_{-\mu}(kr), \quad \mu = \nu, -\nu. \tag{93}$$

Using Eqs. (85), (86), (89), (90), and (93) the coefficients of the linear relations

$$y^{(1)}(r) = D_{13}y^{(3)}(r) + D_{14}y^{(4)}(r), \tag{94a}$$

$$y^{(2)}(r) = D_{23}y^{(3)}(r) + D_{24}y^{(4)}(r) \tag{94b}$$

are found to be

$$D_{13} = \frac{1}{2\sin\pi\nu}\left(R - \frac{1}{R}\right)\left(\frac{\beta}{k}\right)^{1/2}\exp(-i\pi/4), \tag{95a}$$

$$D_{14} = \frac{1}{2\sin\pi\nu}\left(R\exp(+i\pi\nu) - \frac{1}{R}\exp(-i\pi\nu)\right)\left(\frac{\beta}{k}\right)^{1/2}\exp(+i\pi/4), \tag{95b}$$

$$D_{23} = \frac{1}{2\sin\pi\nu}\left(R\exp(-i\pi\nu) - \frac{1}{R}\exp(+i\pi\nu)\right)\left(\frac{\beta}{k}\right)^{1/2}\exp(+i\pi/4), \tag{95c}$$

$$D_{24} = -D_{13}, \tag{95d}$$

where

$$R = [\beta k\exp(i\pi/2)]^{-\nu}\Delta_{-\nu}/\Delta_{+\nu} = \exp[-i(\pi/2)\nu]\tilde{R}. \tag{96}$$

This quantity is identical to the quantity R of Eq. (17), as may be seen by comparing Eqs. (94)—(96) with Eqs. (21). But we have obtained here an entirely different analytical expression for it.

The case when ν happens to be equal to any integer or zero requires special attention. Then we have $R=1$ or $R=-1$ so that the expressions (95) are undefined and have to be replaced by their limiting values.

3. Integer values of the characteristic index

If $\nu=0$ we have $R=1$ evidently by Eq. (96). It will now be shown that $R=1$ or $R=-1$ if ν is equal to any integer $\neq 0$. For this purpose it is necessary to consider the way of generating the coefficients of the Laurent series by means of continued fractions, a method well-known from the Mathieu equation. Let us introduce the "right" and "left" ratios of the coefficients

$$R_m = d^{\mu}_{2m}/d^{\mu}_{2m-2} \tag{97}$$

and

$$L_m = d^{\mu}_{-2m-2}/d^{\mu}_{-2m}. \tag{98}$$

With the abbreviation

$$V_m = (\mu+2m)^2 - (l+\tfrac{1}{2})^2 \tag{99}$$

it then follows from the recurrence relation (29) of the coefficients that

$$R_m = \frac{-1}{V_m - \beta^2k^2R_{m+1}} = \frac{-1}{V_m+}\,\frac{\beta^2k^2}{V_{m+1}+}\,\frac{\beta^2k^2}{V_{m+2}+\cdots}, \tag{100}$$

$$L_m = \frac{\beta^2k^2}{V_{-m-1}+L_{m+1}} = \frac{\beta^2k^2}{V_{-m-1}+}\,\frac{\beta^2k^2}{V_{-m-2}+}\,\frac{\beta^2k^2}{V_{-m-3}+\cdots}. \tag{101}$$

According to Eqs. (97), (98) we have

$$R_0L_0 = 1. \tag{102}$$

This condition of consistency is satisfied by the ratios R_0 and L_0 computed from the continued fractions (100), (101) only if μ has the value of a characteristic index ($\mu = \nu, -\nu \bmod 2$). Let us first assume that this value $\mu = 2M$ is an even integer. Then we have the symmetry relation

$$V_{-M+n} = V_{-M-n} \tag{103a}$$

and therefore

$$L_{+M+n} = -\beta^2k^2R_{-M+n+1}. \tag{104a}$$

It then follows that, with $\epsilon = +1$ or -1,

$$d^{2M}_{-2M-2n} = \epsilon(i\beta k)^{2n}d^{2M}_{-2M+2n} \tag{105a}$$

or, by means of Eq. (31),

$$d^{2M}_{-2M-2n} = \epsilon(i\beta k)^{2M}d^{-2M}_{+2M-2n}. \tag{106a}$$

If, on the other hand, $\mu = 2M+1$ is an odd integer, Eqs. (103a)–(106a) have to be replaced by

$$V_{-M+n} = V_{-M-n-1}, \tag{103b}$$

$$L_{+M+n} = -\beta^2k^2R_{-M+n}, \tag{104b}$$

$$d^{2M+1}_{-2M-2n} = \epsilon(i\beta k)^{2n-1}d^{2M+1}_{-2M+2n-2}, \tag{105b}$$

$$d^{2M+1}_{-2M-2n} = \epsilon(i\beta k)^{2M+1}d^{-2M-1}_{+2M-2n+2}, \tag{106b}$$

respectively. By means of Eqs. (106) it is not difficult to show that $R=\epsilon(-1)^{\nu}$ whenever the characteristic index ν is an integer $\neq 0$. Whether $\epsilon=1$ or $\epsilon=-1$ depends on the values of the parameters l and β^2k^2. There are these two possibilities, in analogy to the existence of even and odd periodic solutions of the Mathieu equation.

In order to avoid some obscurity which might arise by inspection of Eqs. (95), (96), we finally want to show that, if βk and l are real parameters, then odd integer values of the characteristic index cannot occur. For if we assume that ν be an odd integer, say $\nu=1$, we have $L_0 = -\beta^2k^2R_0$ from Eq. (104b), where the continued fractions L_0 and R_0 are real in this case according to Eqs. (99)–(101). It then follows that $R_0L_0 = -\beta^2k^2R_0^2$ is negative real. But this result violates the consistency condition (102), and therefore our assumption that the characteristic index might be an odd integer is wrong. Nevertheless, values of ν very close to the odd integers may occur. In fact, numerical computations show that the characteristic index, considered as a function of l with $\beta^2k^2 > 0$ fixed, has maxima and minima some of which are extremely close to an odd integer. Even integer values of ν, on the other hand, do occur. In their neighborhood ν may become complex, its real part then being equal to that integer.

F. Computational aspects

The formulas we have obtained for ν, Δ_{μ}, and R are relatively convenient for electronic computation. We have to evaluate twenty Taylor series, the coefficients of which can be computed recursively, for a value of the variable equal to half the convergence radius. Sixteen of these series represent the coefficients of the system of linear equations (51) which has to be solved, four additional series are needed in the evaluation of $H(-1)$ and $G(-1)$. Finally, two Laurent series $\phi_{\mu}(-4)$, $\mu=\nu,-\nu$, have to be computed for a value of the variable equal to four times the convergence radius of the principal part. The coefficients of these Laurent series or the related coefficients d^{μ}_{2m} obey a three-term recurrence relation and may be evaluated by means of continued fractions: The ratios of coefficients R_m and L_m are evaluated for the maximum value of m required by means of the continued fractions appearing on the right-hand side of Eqs. (100), (101). The ratios for the lower values of m then are obtained recursively using the left-hand part of Eqs. (100), (101). Finally the coefficients themselves are obtained recursively by

$$d^{\mu}_{2m} = R_m d^{\mu}_{2m-2}, \quad d^{\mu}_{-2m} = L_{m-1}d^{\mu}_{-2m+2}, \tag{107}$$

starting with $m=1$ and $d^{\mu}_0 = 1$.

If the system of linear equations (51) is solved by Cramer's rule and use is made of the fact that the determinant D of the system is, apart from the sign, a Wronskian of the t equation equal to

$$-D = (\lambda-\tfrac{1}{2})(\lambda-\tfrac{3}{2})(\lambda-\tfrac{5}{2})2^{-\lambda+9/2}, \tag{108}$$

then an explicit expression for the characteristic index ν is obtained in the form

$$\cos\pi\nu = \frac{-D_1\cos\pi\lambda}{(\lambda-\frac{1}{2})(\lambda-\frac{3}{2})(\lambda-\frac{5}{2})}\,2^{\lambda-9/2}, \tag{109}$$

where

$$D_1 = \begin{vmatrix} F(1) & G_0(1) & G_1(1) & G_2(1) \\ -F'(1) & G_0'(1) & G_1'(1) & G_2'(1) \\ F''(1) & G_0''(1) & G_1''(1) & G_2''(1) \\ -F'''(1) & G_0'''(1) & G_1'''(1) & G_2'''(1) \end{vmatrix}. \tag{110}$$

Here the value of the finite determinant D_1 enters, the elements of which are Taylor series. This method of generating the characteristic index therefore seems to

be advantageous as compared to one of the usual methods which requires the computation of an infinite determinant.

A simple but sensitive check of the equations themselves and of the accuracy which can be achieved if they are evaluated numerically is provided by the parameter λ. This parameter, which is quite arbitrary except that it must not be equal mod 1 neither to $\frac{1}{2}$ nor to the characteristic index, influences the computation significantly, but the final results for ν, Δ_μ, and $\tilde{R}$ do not depend on it. Therefore computations with different values of λ may serve for checking.

By inspection of the relevant equations and on the basis of several numerical computations (performed with real values of the parameters $l > 0$ and $\beta^2 k^2$) we arrive at the following conclusions concerning the choice of λ and of ν [which is defined by Eq. (109) mod 2 only]. In order to obtain an accurate value of ν even if l is not small (i.e., larger than something like 3), an integer value near $-|l+\frac{1}{2}|$ should be chosen for the parameter λ. This is important also in view of the fact that the accuracy of R computed by any method depends on the accuracy of ν. The coefficients of the Laurent series can be computed accurately only if ν is chosen such that its real part is as close to $l+\frac{1}{2}$ as is possible. For computing the Δ_μ the parameter λ should be near $-|l+\frac{1}{2}|$, but sufficiently different from $+\nu$ mod 2 and from half-integer values.

If these points are observed, accurate results may be obtained. If l is not too large, both the methods of computing R give equally accurate values, although the convergence of the series is faster in the conventional formula (17). But if ν happens to be very close to an integer, the values of R computed by the new method are often more accurate, since then the conventional formula is very sensitive to errors of ν (via the Bessel functions of nearly integer negative index). If l becomes larger, however, the conventional formula is superior since the terms with $H(-1)$ and $G(-1)$ in Δ_μ may interfere destructively so that significant figures are lost. Nevertheless the new method still gives accurate results if the scattering phase shifts is all what is wanted. For as the numerical accuracy of $|R|$ decreases, it becomes larger and larger compared with 1 so that, according to Eq. (24), the phase shift becomes more and more independent of R.

Note added in proof: Using a quite different and modern method by Naundorf,[14] we find

$$\Delta_\mu = 2\sum_{n=0}^{\infty} a_n/\Gamma(\mu+\tfrac{3}{2}+n),$$

with the coefficients a_n from Eq. (83). By means of Eq. (96) we then may obtain another expression for the important quantity R, which is remarkable because of its simplicity and because of the fact that the coefficients of the Laurent series do not enter.

[1]E. Vogt and G. Wannier, Phys. Rev. **95**, 1190 (1954).
[2]Th. F. O'Malley, L. Spruch, and L. Rosenberg, J. Math. Phys. **2**, 491 (1961).
[3]R. Spector, J. Math. Phys. **5**, 1185 (1964).
[4]H. H. Aly and H. J. W. Müller, J. Math. Phys. **7**, 1 (1966).
[5]J. Meixner and I. W. Schäfke, *Mathieusche Funktionen und Sphäroid-Funktionen* (Springer, Berlin, 1954).
[6]N. W. McLachlan, *Theory and Applications of Mathieu Functions* (Oxford U. P., Oxford, 1947).
[7]R. B. Dingle and H. J. W. Müller, J. Reine Angew. Math. **211**, 11 (1962).
[8]*Handbook of Mathematical Functions*, edited by M. Abramowitz and I. A. Stegun (Dover, New York, 1965).
[9]A. Erdélyi *et al.*, *Higher Transcendental Functions* (McGraw-Hill, New York, 1955), Vol. 3.
[10]W. M. Frank, D. J. Land, and R. M. Spector, Rev. Mod. Phys. **43**, 36 (1971).
[11]A. Erdélyi, Math. Z. **41**, 653 (1936).
[12]S. Fubini and R. Stroffolini, Nuovo Cimento **37**, 1812 (1965).
[13]E. L. Ince, *Ordinary Differential Equations* (Dover, New York, 1956).
[14]F. Naundorf, "Globale Lösungen von gewöhnlichen linearen Differentialgleichungen mit zwei stark singulären Stellen," Thesis, Universität Heidelberg, 1974.

Dynamical symmetries and constants of the motion for classical particle systems

Gerald H. Katzin* and Jack Levine†

Departments of Physics and Mathematics†, North Carolina State University, Raleigh, North Carolina*
(Received 11 February 1974)

By formulating the conditions for dynamical symmetry mappings directly at the level of the dynamical equations (which are taken in the form of Newton's equations, Lagrange's equations, Hamilton's equations, or Hamilton–Jacobi equation), we derive new expressions for dynamical symmetries and associated constants of the motion for classical particle dynamical systems. All dynamical symmetry mappings we consider are based upon infinitesimal point transformations of the form (a) $\bar{x}^i = x^i + \delta x^i$ $[\delta x^i \equiv \xi^i(x)\delta a]$ with associated changes in the independent variable t (path parameter) defined by (b) $\delta t \equiv \{\int 2\phi[x(t)]\,dt + c\}\,\delta a$. A generalized form of the related integral theorem (a method for obtaining constants of the motion based upon deformations of a known constant of the motion under dynamical symmetry mappings) is obtained. We take the "Newtonian form" of the dynamical equations to have a coordinate-covariant structure with forces defined by a general polynomial in the velocities and obtain dynamical symmetry conditions for all such systems. For the special case of conservative systems the related integral theorem is applied. Based upon Lagrange's equations with $L = L(x^i, \dot{x}^i)$ we find the conditions for dynamical symmetry mappings may be expressed in the form

(c) $(\partial/\partial x^j)[\delta L + L(d/dt)(\delta t)] - (d/dt)(\partial/\partial \dot{x}^j)[\delta L + L\,(d/dt)\,(\delta t)] = -2\phi_{,j}[(\partial L/\partial \dot{x}^i)\dot{x}^i - L]\,\delta a.$

From this form we obtain a new formula for concomitant constants of the motion: (d) $[\partial(\delta L)/\partial \dot{x}^j]\dot{x}^j - \delta L = k$. By use of the related integral theorem such constants of the motion can be expressed as deformations of the energy integral under the dynamical symmetry mappings defined by (c). A short derivation of the Noether identity is given which is independent of the integration processes of Hamilton's variational principle. For mappings of the type (a), (b) "Noether type" symmetries and associated constants of the motion are formulated. For a conservative dynamical system with $L \equiv (1/2)g_{ij}\dot{x}^i\dot{x}^j - V(x)$ we find such Noether symmetries are basically conformal motions, while those derived from (c) are basically projective collineations. For such systems the constants of the motion (d) are evaluated and shown in general to differ from those obtained from the Noether method. We show for conservative dynamical systems that the formulation of dynamical symmetry mappings directly at the level of the Hamilton–Jacobi equation leads to the Noether symmetry conditions. Dynamical symmetry conditions are formulated for Hamilton's equation in phase space and shown to be more general than canonical transformations. The formulation of the related integral theorem in phase space is found to be a generalization of Poisson's theorem. For systems with $H(x^A)$, $A = 1,\ldots,2n$, it is an immediate observation that δH induced by a symmetry mapping is a constant of the motion. Application to the isotropic harmonic oscillator shows both symmetric tensor and angular momenta constants of the motion are obtained in this manner. An additional constant of the motion $\partial_A \xi^A - 2\phi(x^A)$ is shown in general to be a concomitant of a phase space symmetry transformation.

1. INTRODUCTION

We consider in this paper dynamical symmetries and associated constants of the motion for classical particle dynamical systems. For such systems a dynamical symmetry is a transformation which maps the set of all dynamical paths into itself.[1] It is well known from the work of Noether[2,3] how conditions for dynamical symmetries may be derived by use of the formalism of Hamilton's variational principle and that certain constants of the motion are concomitant with the existence of such dynamical symmetries. It is also well known from the work of Poisson that certain types of canonical transformations associated with the existence of constants of the motion may be interpreted as dynamical symmetry mappings at the level of Hamilton's equations.[4]

In several recent papers[5–9] the present authors developed an additional symmetry-based method for obtaining constants of the motion. This method, which was published in various forms as "related integral theorems," provides a means for deriving additional constants of the motion based upon the deformations of a known constant of the motion under dynamical symmetry mappings.

It is a main purpose of this paper to give a unified proof of the above mentioned related integral theorems which will include all theorems of this type previously proven for specific classes of dynamical equations in both configuration space and phase space, and where applicable to compare various aspects of this method with the more familiar Noether and Poisson methods for obtaining constants of the motion.

We shall also show that by formulating the conditions for dynamical symmetries directly at the level of the dynamical equations (which we take in the form of Newton's equation, Hamilton's equations, Lagrange's equations, or the Hamilton–Jacobi equation) that we obtain certain new and interesting relationships between dynamical symmetries and constants of the motion.

Throughout this paper we shall base our dynamical symmetries upon the existence of infinitesimal point transformations (3.1). Any accompanying transformation (3.3) in the independent variable (dynamical path parameter) will be treated as a transformation which is correlated to the point mapping being considered. For the most part we shall also limit our discussions to con-

stants of the motion which have no explicit dependence upon the independent variable. We recognize that such mappings and/or constants of the motion are not the most general allowed in the traditional Noether and Poisson methods; however, they suffice for the purposes of explaining and comparing the various approaches to obtaining dynamical symmetries and constants of the motion. The time-dependent theory as well as field theory analogs will be considered in a later paper.

In Sec. 2 we first define the class of dynamical systems to be considered in terms of a system of general second order ordinary differential equations (2.1).

In Sec. 3 we mathematically formulate the conditions (3.13) for the existence of a dynamical symmetry mapping directly at the level of the above-mentioned differential equations.

In Sec. 4 we prove a generalized related integral theorem for the class of dynamical systems considered. We base this derivation directly upon the use of infinitesimal point mappings (with associated changes in path parameter) thereby eliminating the added conceptual complications associated with the Lie derivative approach used in previous formulations.

In Sec. 5 we specialize the general form of dynamical equations (2.1) to be of "Newtonian form" (5.1). These equations, which are taken to have a manifestly coordinate—covariant structure, embrace many important types of dynamical systems in that they allow for the inclusion of rather general velocity dependent forces. Based upon these equations we obtain a general form for the dynamical symmetry conditions as a direct application of the method explained in Sec. 3. As a detailed illustration we further specialize the Newtonian form of the dynamical equations to those of a conservative system. The specific symmetry conditions obtained [(5.15), (5.16)] indicate the dynamical symmetry mappings are basically projective collineations subject to additional restrictions dependent upon the form of the potential energy. By application of the related integral theorem to such systems it is found that the deformations of the energy integral with respect to the symmetry mappings generate additional quadratic constants of the motion. (For the Kepler problem and the three-dimensional isotropic simple harmonic oscillator the well-known Runge—Lenz vector and symmetric tensor constants of the motion respectively have recently been obtained by this method.[7])

In Sec. 6 we specialize the basic dynamical equations (2.1) to the form of Hamilton's equations in phase space.[10] The accompanying specialization in the dynamical symmetry conditions (3.13) results in symmetry conditions (6.4) which in general lead to noncanonical dynamical symmetry mappings.[11] From the form of these symmetry conditions two methods for formulating associated constants of the motion are immediately obvious. The constants of the motion (6.6) resulting from one method are shown to be a direct consequence of the invariance of the divergence-free character of the integral curves of a Hamiltonian system under dynamical symmetry mappings. The constants of the motion (6.13) given by the second method result from the deformations of the Hamiltonian of the system under the dynamical symmetry mappings. The constants of the motion associated with this latter approach are also shown to be obtainable from the phase space formulation of the related integral theorem.

It is also shown that the phase space form of the related integral theorem is a generalization of the well-known Poisson's theorem on constants of the motion.

A simple application of this (phase space form of) related integral theorem to the three-dimensional isotropic harmonic oscillator shows that the well-known symmetric tensor and angular momenta constants of the motion result from symmetry-induced deformations of the Hamiltonian.

In Sec. 7 we specialize the general dynamical equations (2.1) to the form of Lagrange's equations. The accompanying specialization in the conditions (3.13) for dynamical symmetry mappings results in symmetry equations (formulated directly at the level of Lagrange's equations) which may be reexpressed in a new form (7.6) which displays a structure similar to Lagrange's equations.

From this form of the symmetry equations it is a simple matter to obtain a new formula (7.10) for associated constants of the motion. An investigation into the nature of these constants of the motion shows they may be reexpressed as the deformation of the energy integral under the above-mentioned dynamical symmetry mappings. Thus for this case we find the essential mechanism of the related integral theorem is actually contained within the equations which determine the conditions for a dynamical symmetry mapping. Two specific Lagrangians are chosen to illustrate consistency with results obtained when the dynamical equations were taken in Newtonian form and in Hamiltonian form.

In Sec. 8 we first give a short derivation of Noether's identity (8.5) which is independent of the integration processes of Hamilton's variational principle.[3] We then make a comparison between the "Noether symmetry conditions" (8.6) based upon this well-known identity and the dynamical symmetry conditions (7.6) which were formulated directly at the level of Lagrange's equations. We find that even for the relatively simple case of a conservative dynamical system these two approaches lead to considerable differences in symmetries. For such systems we find that the Noether symmetries are fundamentally conformal mappings whereas the dynamical symmetries formulated directly at the level of Lagrange's equations are basically projective. Therefore the intersection of these two symmetry types is limited to homothetic mappings.

A comparison of the constants of the motion (7.10), (8.12) associated with these two approaches shows that, in general, they differ. For example, in the case of the three-dimensional isotropic harmonic oscillator we find the angular momenta first integrals are given by the Noether formula (8.12) and the symmetric tensor quadratic first integrals are given by the new formula (7.10). Hence the two methods of obtaining constants of the motion complement each other in this application.

In Sec. 9 we consider the problem of proving that the

Noether symmetry mapping conditions (as formulated from the Noether identity) can also be formulated directly at the level of the dynamical equations. We show for a conservative dynamical system that the demand that the set of solutions of the Hamilton–Jacobi equation maps into itself leads to the Noether symmetry conditions. Because the Hamilton–Jacobi equation is a representation of the dynamical equations in the transformation theory of mechanics we may thereby interpret the Noether mappings as dynamical symmetry mappings which are also demonstrable at the level of the dynamical equations. It then also follows that the constants of the motion associated with the Noether symmetries can be considered as concomitants of dynamical symmetries based upon the Hamilton–Jacobi equation.

2. DYNAMICAL SYSTEMS

Consider a system of n second order differential equations[12]

$$E^j(x^1,\ldots,x^n;\dot{x}^1,\ldots,\dot{x}^n;\ddot{x}^1,\ldots,\ddot{x}^n)\equiv E^j(x,\dot{x},\ddot{x})=0,$$
$$j=1,\ldots,n, \tag{2.1}$$

where

$$\dot{x}^i\equiv\frac{dx^i}{dt},\quad \ddot{x}^i\equiv\frac{d^2x^i}{dt^2}. \tag{2.2}$$

We observe that Newton's equations, Lagrange's equations, and Hamilton's equations (mentioned in the introduction) are of the form (2.1).[13]

It is assumed that equations (2.1) satisfy the conditions for solvability of $\ddot{x}^i$ in terms of x^j's and $\dot{x}^j$'s.

We represent the solutions of (2.1) in the form

$$x^i=f^i(a^1,\ldots,a^n;b^1,\ldots,b^n;t)\equiv f^i(a,b,t), \tag{2.3}$$

where the a^i and b^i are $2n$ constants of integration. For specific values of these $2n$ constants equation (2.3) will define a curve γ. The totality of curves so obtained will be denoted by Γ. The coordinates x^i of a particular curve γ are denoted by

$$x^i_\gamma=f^i(a_\gamma,b_\gamma,t). \tag{2.4}$$

In this paper we consider first integrals of the differential equations (2.1) which are of the form[14]

$$I(x,\dot{x})=\text{const}, \tag{2.5}$$

where it is understood that the left side of (2.5) when evaluated along any solution curve $\gamma\in\Gamma$ reduces to a constant h_γ. This is indicated by writing

$$I(x_\gamma,\dot{x}_\gamma)=h_\gamma. \tag{2.6}$$

It is important to note that the value of the constant h_γ will in general vary from one solution curve to another.

3. DYNAMICAL SYMMETRIES

We wish to obtain mappings of the family Γ of integral curves of (2.1) such that the family Γ is mapped into itself in that $\Gamma\to\bar{\Gamma}=\Gamma$. Such mappings will be called dynamical symmetries. For purposes of this paper we shall limit these mappings to infinitesimal point transformations of the form

$$\bar{x}^i=x^i+\delta x^i,\quad \delta x^i\equiv\xi^i(x)\delta a\quad(\delta a\equiv\text{infinitesimal}), \tag{3.1}$$

with associated change in curve parameter t based on a scalar $\phi(x)$ such that[7]

$$d\bar{t}=dt\{1+2\phi[x(t)]\delta a\}. \tag{3.2}$$

The notation $\phi[x(t)]$ indicates the function ϕ is to be evaluated along a solution curve. Hence we obtain from (3.2)

$$\bar{t}=t+\delta t,\quad \delta t\equiv\xi^0(t)\delta a\equiv\{\int 2\phi[x(t)]\,dt+c\}\delta a. \tag{3.3}$$

If (3.1), (3.2) are to satisfy the symmetry mapping requirement

$$\gamma\in\Gamma\to\bar{\gamma}\in\bar{\Gamma},\quad \Gamma=\bar{\Gamma}, \tag{3.4}$$

then we must have the coordinates of the mapped curves $\bar{\gamma}$ satisfy an equation of the form (2.4), namely

$$\bar{x}^i_{\bar{\gamma}}=f^i(a_{\bar{\gamma}},b_{\bar{\gamma}},\bar{t}). \tag{3.5}$$

It should be noted from the form of (2.4) that the parameter t is taken to be the same for all solution curves $\gamma\in\Gamma$. However, in (3.4) the parameter $\bar{t}$ will in general differ for each $\bar{\gamma}\in\bar{\Gamma}$. This follows as a result of the form of (3.3) from which it is apparent that in general each curve $x^i_\gamma(t)$ determines a specific parameter $\bar{t}$ for its mapped image curve $\bar{x}^i_{\bar{\gamma}}(\bar{t})$.

The symmetry mapping condition (3.4) [or (3.5)] is expressed in terms of (2.1) by the requirement that

$$E^j(\bar{x},\bar{\dot{x}},\bar{\ddot{x}})=0, \tag{3.6}$$

where

$$\bar{\dot{x}}^i\equiv d\bar{x}^i/d\bar{t},\quad \bar{\ddot{x}}^i\equiv d^2\bar{x}^i/d\bar{t}^2. \tag{3.7}$$

The conditions on $\xi^i(x)$ and $\phi(x)$ so that (3.1), (3.2) define a symmetry mapping will be obtained from (3.6). We first derive some basic formulas in terms of the δ-derivative defined below.

For any function $F(x,\dot{x},\ddot{x})$ we define[15]

$$\delta F(x,\dot{x},\ddot{x})\equiv\frac{\partial F}{\partial x^i}\delta x^i+\frac{\partial F}{\partial\dot{x}^i}\delta\dot{x}^i+\frac{\partial F}{\partial\ddot{x}^i}\delta\ddot{x}^i, \tag{3.8}$$

where from (3.1) and (3.2) we obtain (to first order in δa)

$$\delta\dot{x}^i\equiv\bar{\dot{x}}^i-\dot{x}^i=\frac{d}{dt}(x^i+\xi^i\delta a)\frac{dt}{d\bar{t}}-\dot{x}^i=(-2\phi\dot{x}^i+\xi^i_{,m}\dot{x}^m)\delta a, \tag{3.9}$$

$$\delta\ddot{x}^i\equiv\bar{\ddot{x}}^i-\ddot{x}^i=\frac{d}{dt}[\dot{x}^i+(-2\phi\dot{x}^i+\xi^i_{,m}\dot{x}^m)\delta a]\frac{dt}{d\bar{t}}-\ddot{x}^i$$
$$=(-4\phi\ddot{x}^i-2\phi_{,m}\dot{x}^m\dot{x}^i+\xi^i_{,mk}\dot{x}^m\dot{x}^k+\xi^i_{,m}\ddot{x}^m)\delta a, \tag{3.10}$$

where we have made use of the relation [obtained from (3.2)]

$$\frac{dt}{d\bar{t}}=(1+2\phi\delta a)^{-1}=1-2\phi\delta a. \tag{3.11}$$

We now expand the left side of (3.6) in the form

$$E^j(\bar{x},\bar{\dot{x}},\bar{\ddot{x}})=E^j(x+\delta x,\dot{x}+\delta\dot{x},\ddot{x}+\delta\ddot{x})=E^j(x,\dot{x},\ddot{x})+\delta E^j(x,\dot{x},\ddot{x}). \tag{3.12}$$

From (2.1) and (3.6) we obtain as a condition for a dynamical symmetry based upon the infinitesimal mapping (3.1), (3.2) the equation

$$\delta E^j = \frac{\partial E^j}{\partial x^i}\delta x^i + \frac{\partial E^j}{\partial \dot{x}^i}\delta \dot{x}^i + \frac{\partial E^j}{\partial \ddot{x}^i}\delta \ddot{x}^i = 0. \tag{3.13}$$

To obtain the above mentioned conditions on ξ^i and ϕ such that (3.1), (3.2) define symmetry mapping we first replace in (3.13) the δx^i, $\delta \dot{x}^i$, $\delta \ddot{x}^i$ by use of the respective formulas (3.1), (3.9), (3.10). Then all $\ddot{x}^i$ terms which appear in the resulting equations are to be replaced by use of (2.1). The equations so obtained will in general be of the form

$$F^j(x^i, \dot{x}^i, \phi, \phi_{,m}, \xi^i, \xi^i_{,m}, \xi^i_{,mk}) = 0. \tag{3.14}$$

Equations (3.14) are considered as identically zero in the $\dot{x}^i$ variables since otherwise they would impose constraints on the dynamical system. The equations resulting from the consideration of (3.14) as such identities will give the symmetry mapping conditions on the $\xi^i(x)$ and $\phi(x)$.[16] This procedure will be illustrated in later sections.

4. RELATED INTEGRAL THEOREM

With reference to the general dynamical systems (2.1) we prove a related integral theorem which is based upon the use of the δ-derivative. Previous such theorems were obtained for specific dynamical equations and were formulated in terms of Lie derivatives.[5-9]

We now assume the dynamical system (2.1) admits a first integral (constant of the motion) of the form $I(x, \dot{x})$ so that along each integral curve $\gamma \in \Gamma$ equation (2.6) is applicable. In addition, we assume that (3.1), (3.2) define a dynamical symmetry mapping (3.4). As a result we have along each integral curve $\bar{\gamma} \in \bar{\Gamma}$

$$I(\bar{x}_{\bar{\gamma}}, \bar{\dot{x}}_{\bar{\gamma}}) = h_{\bar{\gamma}} \equiv h_\gamma + \delta h_\gamma. \tag{4.1}$$

Expansion of (4.1) gives

$$I(x_\gamma, \dot{x}_\gamma) + \delta I(x_\gamma, \dot{x}_\gamma) = h_\gamma + \delta h_\gamma. \tag{4.2}$$

If in (4.2) we further expand δI by means of a formula similar to (3.8) and make use of (2.6) we obtain

$$\delta I(x_\gamma, \dot{x}_\gamma) = \left(\frac{\partial I}{\partial x^i}\xi^i + \frac{\partial I}{\partial \dot{x}^i}(-2\phi\dot{x}^i + \xi^i_{,j}\dot{x}^j)\right)_\gamma \delta a = c_\gamma \delta a,$$

$$c_\gamma \delta a \equiv \delta h_\gamma. \tag{4.3}$$

We can thus state

Theorem 4.1 (*related integral theorem*): If a dynamical system (2.1) admits a constant of the motion (2.6), and if the system admits an infinitesimal dynamical symmetry as defined in Sec. 3, then in general there will exist an additional constant of the motion defined by

$$\frac{\partial I}{\partial x^i}\xi^i + \frac{\partial I}{\partial \dot{x}^i}(-2\phi\dot{x}^i + \xi^i_{,j}\dot{x}^j) = c. \tag{4.4}$$

Remark: This theorem includes previous statements of related integral theorems.[5-8]

5. APPLICATION 1: NEWTONIAN FORM OF THE DYNAMICAL EQUATIONS

As a first application of the methods discussed in Sec. 3 (for obtaining conditions for dynamical symmetries) and Sec. 4 (for obtaining constants of the motion) we consider those systems whose dynamical equations may be expressed in what we will refer to as the "Newtonian form"

$$E^i \equiv \frac{D^2x^i}{dt^2} + \sum_{m=0}^{N} P^i_{j_1\cdots j_m}\dot{x}^{j_1}\cdots\dot{x}^{j_m} = 0 \quad \text{(Ref. 17)}. \tag{5.1}$$

In (5.1) the tensor coefficients $P^i_{j_1\cdots j_m}$ (which are completely symmetric in the lower indices) are functions of the coordinates only, and

$$\frac{D^2x^i}{dt^2} \equiv \ddot{x}^i + \Gamma^i_{jk}(x)\dot{x}^j\dot{x}^k, \tag{5.2}$$

where Γ^i_{jk} is the Christoffel symbol based upon the metric tensor g_{ij} of the configuration space. This form of E^i embraces many important types of dynamical systems in both relativistic and nonrelativistic mechanics.

To obtain the conditions for a dynamical symmetry [as indicated by (3.13)] based on (5.1) we first evaluate

$$\delta\frac{D^2x^i}{dt^2} = \delta(\ddot{x}^i + \Gamma^i_{jk}\dot{x}^j\dot{x}^k) = \Big(-4\phi\frac{D^2x^i}{dt^2} + \xi^i_{,m}\ddot{x}^m + (\Gamma^i_{jk,m}\xi^m + 2\Gamma^i_{mk}\xi^m_{,j} + \xi^i_{,jk} - 2\phi_{,j}\delta^i_k)\dot{x}^j\dot{x}^k\Big)\delta a, \tag{5.3}$$

where use has been made of (3.1), (3.8), (3.9), (3.10). Next we obtain

$$\delta\sum_{m=0}^{N} P^i_{j_1\cdots j_m}\dot{x}^{j_1}\cdots\dot{x}^{j_m} = \sum_{m=0}^{N}[P^i_{j_1\cdots j_m,k}\xi^k + m(-2\phi\delta^k_{j_m} + \xi^k_{,j_m})P^i_{j_1\cdots j_{m-1}k}]\dot{x}^{j_1}\cdots\dot{x}^{j_{m-1}}\dot{x}^{j_m}\delta a. \tag{5.4}$$

If now we substitute from (5.3) and (5.4) into

$$\delta E^i = \delta\frac{D^2x^i}{dt^2} + \delta\sum_{m=0}^{N} P^i_{j_1\cdots j_m}\dot{x}^{j_1}\cdots\dot{x}^{j_m} = 0 \tag{5.5}$$

and eliminate the $\ddot{x}^i$ terms by means of (5.1), we obtain

$$\sum_{m=0}^{N}{}' [\mathcal{L}_\xi P^i_{j_1\cdots j_m} - 2(m-2)\phi P^i_{j_1\cdots j_m}]\dot{x}^{j_1}\cdots\dot{x}^{j_m} + \mathcal{L}_\xi\Gamma^i_{jk} - \delta^i_j\phi_{,k} - \delta^i_k\phi_{,j} + \mathcal{L}_\xi P^i_{jk})\dot{x}^j\dot{x}^k = 0, \tag{5.6}$$

where ($\mathcal{L}_\xi$ indicates Lie derivative with respect to the vector ξ^i)[18]

$$\mathcal{L}_\xi\Gamma^i_{jk} = \xi^i_{,kj} + \xi^m_{,j}\Gamma^i_{mk} + \xi^m_{,k}\Gamma^i_{mj} - \xi^i_{,m}\Gamma^m_{jk} + \xi^m\Gamma^i_{jk,m}, \tag{5.7}$$

$$\mathcal{L}_\xi P^i_{j_1\cdots j_m} = P^i_{j_1\cdots j_m,k}\xi^k + P^i_{kj_2\cdots j_m}\xi^k_{,j_1} + P^i_{j_1kj_3\cdots j_m}\xi^k_{,j_2} + \cdots + P^i_{j_1\cdots j_{m-1}k}\xi^k_{,j_m} - P^k_{j_1\cdots j_m}\xi^i_{,k}, \tag{5.8}$$

and where $\sum'$ means omit $m=2$ in the sum.

As discussed in the comment following (3.14) we set the coefficients of the $\dot{x}^{j_1}\cdots\dot{x}^{j_m}$ terms equal to zero and obtain as conditions for a dynamical symmetry of (5.1)

$$\mathcal{L}_\xi P^i_{j_1\cdots j_m} - 2(m-2)\phi P^i_{j_1\cdots j_m} = 0, \quad m \neq 2, \tag{5.9}$$

$$\mathcal{L}_\xi \Gamma^i_{jk} - \delta^i_j \phi_{,k} - \delta^i_k \phi_{,j} + \mathcal{L}_\xi P^i_{jk} = 0. \tag{5.10}$$

The first three cases of (5.9) [$m=0,1,3$] are as follows:

$$m=0: \mathcal{L}_\xi P^i + 4\phi P^i = 0, \tag{5.11}$$

$$m=1: \mathcal{L}_\xi P^i_j + 2\phi P^i_j = 0, \tag{5.12}$$

$$m=3: \mathcal{L}_\xi P^i_{hjk} - 2\phi P^i_{hjk} = 0. \tag{5.13}$$

To specialize (5.1) to the case of a conservative dynamical system we take $N=0$ and $P^i \equiv g^{ij}V_{,j}$ [where $V(x)$ is the potential energy] which reduces (5.1) to

$$\frac{D^2x^i}{dt^2} + g^{ij}V_{,j} = 0. \tag{5.14}$$

The symmetry conditions (5.9), (5.10) [the appropriate form of (5.9) is (5.11) for this case] become[19]

$$4\phi g^{ij}V_{,j} + g^{ij}V_{;jk}\xi^k - g^{jk}V_{,j}\xi^i_{;k} = 0 \quad \text{(Ref. 20)}, \tag{5.15}$$

$$\mathcal{L}_\xi \Gamma^i_{jk} = \delta^i_j \phi_{,k} + \delta^i_k \phi_{,j}. \tag{5.16}$$

As an application of Theorem 4.1 (related integral theorem) for this case, we consider the deformation of the energy integral

$$I(x,\mathring{x}) \equiv \tfrac{1}{2} g_{ij}\mathring{x}^i\mathring{x}^j + V(x) = \text{const} \tag{5.17}$$

with respect to the symmetry vector ξ^i [and associated scalar $\phi(x)$] which satisfies (5.15), (5.16). We obtain from (4.4) the derived constant of the motion[21]

$$\delta I \equiv \tfrac{1}{2}(\mathcal{L}_\xi g_{ij} - 4\phi g_{ij})\mathring{x}^i\mathring{x}^j + V_{,k}\xi^k = \text{const}. \tag{5.18}$$

It should be noted that if (5.15) is satisfied by a motion vector ξ^i, i.e., if

$$\mathcal{L}_\xi g_{ij} \equiv \xi_{i;j} + \xi_{j;i} = 0, \tag{5.19}$$

then it follows that $\phi = 0$. It can be shown that (5.15) reduces to [if use be made of (5.19)]

$$(V_{,t}\xi^t)_{,j} = 0, \tag{5.20}$$

which implies

$$V_{,t}\xi^t = k, \tag{5.21}$$

where k is constant (throughout the space).

6. APPLICATION 2: HAMILTON'S EQUATIONS[10]

In this section we choose E^i of (2.1) to be all first order equations which have the form of Hamilton's equations[22]

$$E^A \equiv \mathring{x}^A - \eta^{AB}H_{,B} = 0, \quad A,B = 1,\ldots,2n, \tag{6.1}$$

where $H(x)$ the Hamiltonian, expressed in terms of the $2n$ coordinates x^A of phase space, which are defined in terms of the generalized coordinates q^i and their conjugate momenta p_i by the relation

$$(x^1,\ldots,x^n;x^{n+1},\ldots,x^{2n}) \equiv (q^1,\ldots,q^n;p_1,\ldots,p_n). \tag{6.2}$$

The symplectic matrix $[\eta^{AB}]$ is defined by[23]

$$[\eta^{AB}] \equiv \left[\begin{array}{c|c} 0_n & I_n \\ \hline -I_n & 0_n \end{array}\right]. \tag{6.3}$$

Based upon the dynamical equation (6.1) we shall utilize the method of Sec. 3 (as applied to first order equations) to formulate conditions which the mapping vector $\xi^A(x^1,\ldots,x^{2n})$ and associated scalar $\phi(x^1,\ldots,x^{2n})$ must satisfy such that (3.1), (3.2) (with $x^i \to x^A$) define a dynamical symmetry mapping within the phase space. It will then be shown that the existence of such symmetries leads in general to associated constants of the motion. In addition, we shall show that the related integral theorem in phase space may be considered as a generalization of the well-known Poisson theorem on constants of the motion.

We first substitute from (6.1) into (3.13) and make use of (3.8), (3.1), (3.9), and (6.1) (to eliminate the $\mathring{x}^A$ terms) to obtain as the condition for dynamical symmetry the equation[24]

$$-2\phi\eta^{AB}H_{,B} + \eta^{CB}\xi^A_{,C}H_{,B} - \eta^{AB}H_{,BC}\xi^C = 0. \tag{6.4}$$

By taking the divergence of (6.4) we immediately obtain

$$\eta^{AB}(\xi^C_{,C} - 2\phi)_{,A}H_{,B} = 0, \tag{6.5}$$

which implies that $\xi^C_{,C} - 2\phi$ is a constant of the motion.[25] Hence we may state

Theorem 6.1: If $\xi^A(x)$ and associated $\phi(x)$ satisfy the dynamical symmetry equation (6.4), then

$$\xi^A_{,A} - 2\phi = k \tag{6.6}$$

is a constant of the motion of the dynamical system (6.1).

We now give an alternative approach which shows that (6.6) follows as a consequence of the invariance under dynamical symmetry mapping of the divergence-free character of the integral curve congruence defined by (6.1). To do this we first rewrite (3.9) in the form

$$\bar{\mathring{x}}^A = \mathring{x}^A + (-2\phi\mathring{x}^A + \xi^A_{,M}\mathring{x}^M)\delta a. \tag{6.7}$$

We form the divergence of (6.7) to obtain[26]

$$\bar{\partial}_A\bar{\mathring{x}}^A = \partial_B[\mathring{x}^A + (-2\phi\mathring{x}^A + \xi^A_{,M}\mathring{x}^M)\delta a]\frac{\partial x^B}{\partial \bar{x}^A}. \tag{6.8}$$

From (6.8), by use of (3.1) we find

$$\bar{\partial}_A\bar{\mathring{x}}^A = \partial_A\mathring{x}^A + [(\xi^A_{,A} - 2\phi)_{,B}\mathring{x}^B - 2\phi\partial_A\mathring{x}^A]\delta a. \tag{6.9}$$

It follows immediately from (6.1) that $\partial_A\mathring{x}^A = \eta^{AB}H_{,AB} = 0$; hence (6.9) reduces to

$$\bar{\partial}_A\bar{\mathring{x}}^A = \frac{d}{dt}(\xi^A_{,A} - 2\phi)\delta a. \tag{6.10}$$

Since the demand for dynamical symmetry [refer to (3.6)] requires the left side of (6.10) be zero, if follows that (6.6) holds. Alternatively, from Theorem 6.1 it follows that the right side of (6.10) vanishes which shows the left side is zero; this in turn verifies the invariance of the divergence-free character of the integral curves

of (6.1) under dynamical symmetry mapping.

We may utilize (6.6) to evaluate the change in differential path parameter which is associated with the symmetry mapping defined by the vector ξ^A. By use of (6.6) and (3.2) we thus obtain

$$d\bar{t} = dt[1 + (\xi^A_{,A} - k)\delta a]. \tag{6.11}$$

We give now a second method for obtaining constants of the motion concomitant with the existence of dynamical symmetry mappings of (6.1). If (6.4) is contracted with $H_{,A}$, then the resulting equation may be written in the form

$$\eta^{BA}(\delta H)_{,B}H_{,A} = 0, \quad \delta H \equiv H_{,C}\xi^C \delta a, \tag{6.12}$$

which implies $\delta H/\delta a$ is a constant of the motion. This allows us to state

Theorem 6.2: If $\xi^A(x)$ satisfies the dynamical symmetry equation (6.4), then in general the dynamical system (6.1) will admit the concomitant constant of the motion

$$\frac{\delta H}{\delta a} \equiv H(x)_{,A}\xi^A. \tag{6.13}$$

We next consider the related integral theorem 4.1 as applied to phase space. Assume then that the dynamical system (6.1) has a constant of the motion of the form $I(x)$. Theorem 4.1 then states that

$$\frac{\delta I}{\delta a} = I_{,A}\xi^A \tag{6.14}$$

is also a constant of the motion, provided ξ^A is a solution of (6.4).

Since the dynamical equations (6.1) were based upon a Hamiltonian which was assumed to have no explicit time dependence it immediately follows that $H(x)$ is a constant of the motion. We may therefore take $I(x) \equiv H(x)$ in (6.14). This gives a second derivation of (6.13).

If the dynamical system (6.1) admits an r-parameter dynamical symmetry group defined by the vectors ξ^A_α, $\alpha = 1, \ldots, r$, then the first and second derived integrals I_α, $I_{\alpha\beta}$ satisfy the relation[8]

$$I_{\beta\alpha} - I_{\alpha\beta} = C^\gamma_{\beta\alpha} I_\gamma, \tag{6.15}$$

where $I_\alpha \equiv \delta_\alpha I/\delta a_\alpha$, $I_{\beta\alpha} \equiv \delta_\beta I_\alpha/\delta a_\beta$, and $C^\gamma_{\beta\alpha}$ are the structure constants of the group.[27]

It can easily be shown that the (Poisson) vector

$$\xi^A_{(P)} \equiv \eta^{AB}M_{,B}, \tag{6.16}$$

where $M(x)$ is a constant of the motion, is always a solution of (6.4) for the choice $\phi = 0$. For such a symmetry vector (6.14) takes the form

$$\frac{\delta I}{\delta a} = \eta^{AB}I_{,A}M_{,B} \tag{6.17}$$

which is Poisson's theorem on constants of the motion.[4] Hence we may consider the phase space formulation of the related integral theorem as a generalization of Poisson's theorem.

As a simple illustration of Theorem 6.2 we consider an n-dimensional isotropic harmonic oscillator in an n-dimensional Euclidean configuration space referred to rectangular coordinates. For this case the Hamiltonian $H(x)$ takes the form

$$H = \tfrac{1}{2}\delta_{AB}x^A x^B, \quad A, B = 1, \ldots, 2n. \tag{6.18}$$

It can be shown for such a Hamiltonian that a solution to (6.4) is[28]

$$\xi^A = B^A_M x^M, \quad \phi = 0, \tag{6.19}$$

where the $2n \times 2n$ matrix $[B^A_M]$ is given by

$$[B^A_M] \equiv \left[\begin{array}{c|c} a^i_j & b^i_j \\ \hline -b^i_j & a^i_j \end{array}\right], \tag{6.20}$$

where the arbitrary constants a^i_j and b^i_j are the elements of $n \times n$ matrices. It can be shown that (6.19) defines a $2n^2$-parameter group of dynamical symmetries.

It now follows from (6.13) by use of (6.18), (6.19) that this dynamical system will have the constants of the motion

$$\frac{\delta H}{\delta a} = H_{,A}\xi^A = B^A_C x^A x^C = \text{const}. \tag{6.21}$$

By use of (6.20) and (6.2) Eq. (6.21) can be written in the form

$$\frac{\delta H}{\delta a} = a^i_j(q^i q^j + p_i p_j) + b^i_j(q^i p_j - q^j p_i) = \text{const}. \tag{6.22}$$

We recognize by inspection of (6.22) the well-known symmetric tensor and angular momenta constants of the motion associated with the oscillator problem.[7] It is of particular interest that these constants of the motion may be considered as deformations in the Hamiltonian as a result of dynamical symmetry mappings.

For the particular symmetry vector (6.19), (6.20) the corresponding constant of the motion (6.6) predicted by Theorem 6.1 is trivial. However, the general symmetry vector solution of (6.4) for this problem leads to nontrivial constants of the motion (6.6).[28]

7. APPLICATION 3: LAGRANGE'S EQUATIONS

Next we shall base our formulation of dynamical symmetry conditions on dynamical equations which we assume to have the form of Lagrange's equation

$$E_j \equiv \frac{\partial L}{\partial x^j} - \frac{d}{dt}\frac{\partial L}{\partial \dot{x}^j} = \frac{\partial L}{\partial x^j} - \frac{\partial^2 L}{\partial \dot{x}^j \partial x^k}\dot{x}^k - \frac{\partial^2 L}{\partial x^j \partial x^k}\ddot{x}^k = 0, \tag{7.1}$$

where $L \equiv L(x^i, \dot{x}^i)$. For such Lagrangians it is well known that the dynamical equation (7.1) admits the constant of the motion

$$E(x, \dot{x}) \equiv \frac{\partial L}{\partial \dot{x}^i}\dot{x}^i - L. \tag{7.2}$$

We shall show that when based upon Lagrange's equation (7.1) the basic condition for dynamical symmetry (3.13) may be expressed in an interesting alternative form. This new form immediately reveals the existence of constants of the motion concomitant with these dynamical symmetries and in addition allows a direct comparison with the Noether approach to the formulation of dynamical symmetries and associated constants of the motion.

We first formulate the conditions (as outlined in Sec. 3) that the infinitesimal mapping (3.1), (3.3) define a dynamical symmetry of the system (7.1). Substitution of (7.1) into (3.13) with use of (3.1), (3.9), (3.10) gives

$$\delta E_j = \left(\frac{\partial^2 L}{\partial x^i \partial x^j} - \frac{\partial^3 L}{\partial x^i \partial \dot{x}^j \partial x^k}\dot{x}^k - \frac{\partial^3 L}{\partial x^i \partial \dot{x}^j \partial \dot{x}^k}\ddot{x}^k\right)\xi^i \delta a$$
$$+\left(\frac{\partial^2 L}{\partial \dot{x}^i \partial x^j} - \frac{\partial^3 L}{\partial \dot{x}^i \partial \dot{x}^j \partial x^k}\dot{x}^k - \frac{\partial^2 L}{\partial \dot{x}^j \partial x^i} - \frac{\partial^3 L}{\partial \dot{x}^i \partial \dot{x}^j \partial \dot{x}^k}\ddot{x}^k\right)$$
$$\times\ (-2\phi\dot{x}^i - \xi^i_{,m}\dot{x}^m)\delta a$$
$$+\frac{\partial^2 L}{\partial \dot{x}^j \partial \dot{x}^i}[4\phi\ddot{x}^i + 2\phi_{,m}\dot{x}^m\dot{x}^i$$
$$- \xi^i_{,mk}\dot{x}^m\dot{x}^k - \xi^i_{,m}\ddot{x}^m]\delta a = 0. \tag{7.3}$$

We now show that (7.3) may be expressed in a form which allows further insight into the nature of this dynamical symmetry condition. By use of (3.1), (3.8), and (3.9) we first obtain

$$\delta L = \left(\frac{\partial L}{\partial x^i}\xi^i + \frac{\partial L}{\partial \dot{x}^i}(-2\phi\dot{x}^i + \xi^i_{,m}\dot{x}^m)\right)\delta a, \tag{7.4}$$

and define the function N (which we refer to as the Noether function) by

$$N(x,\dot{x})\delta a \equiv \delta L + 2\phi L \delta a. \tag{7.5}$$

It now follows by a straightforward calculation that (7.3) may be expressed in the above-mentioned alternative form [by use of (7.1)][29]

$$\delta E_j = \frac{\partial N}{\partial x^j} - \frac{d}{dt}\frac{\partial N}{\partial \dot{x}^j} + 2\phi_{,j}E = 0, \tag{7.6}$$

where $E(x,\dot{x})$ is defined by (7.2).

The above may be summarized by stating

Theorem 7.1: A necessary and sufficient condition that an infinitesimal transformation (3.1), (3.3) define a dynamical symmetry of the Lagrangian dynamical system (7.1) is that (7.6) hold for all solutions of (7.1).

We next assume that Lagrange's equation (7.1) admits a dynamical symmetry mapping as described by Theorem 7.1. It now follows from the contraction of (7.6) with $\dot{x}^j$ that

$$\frac{\partial N}{\partial x^j}\dot{x}^j - \frac{d}{dt}\left(\frac{\partial N}{\partial \dot{x}^j}\right)\dot{x}^j + 2\frac{d\phi}{dt}E = 0. \tag{7.7}$$

By use of (7.2) and the relation

$$\frac{dN}{dt} \equiv \frac{\partial N}{\partial x^j}\dot{x}^j + \frac{\partial N}{\partial \dot{x}^j}\ddot{x}^j, \tag{7.8}$$

we may express (7.7) in the form

$$\frac{d}{dt}\left(N - \frac{\partial N}{\partial \dot{x}^j}\dot{x}^j + 2\phi E\right) = 0. \tag{7.9}$$

Expansion of the term in parentheses in (7.9) by use of (7.2) and (7.5) gives the constant of the motion

$$\frac{\partial(\delta L)}{\partial \dot{x}^j}\dot{x}^j - \delta L. \tag{7.10}$$

Hence we may state

Therem 7.2: If a Lagrangian dynamical system (7.1) based upon $L(x,\dot{x})$ admits a dynamical symmetry as described in Theorem 7.1, then in general the system will admit a constant of the motion of the form (7.10).

To further our understanding of the meaning of the constant of the motion (7.10) we note that if (7.4) is used in (7.10) the resulting expression may be expanded and regrouped to give

$$\frac{\partial(\delta L)}{\partial \dot{x}^j}\dot{x}^j - \delta L = \delta\left(\frac{\partial L}{\partial \dot{x}^j}\dot{x}^j - L\right) = \text{const.} \tag{7.11}$$

We thus find that the constant of the motion (7.10) is the deformation of the energy integral under the dynamical symmetry mapping described by Theorem 7.1. It is now apparent that (7.10) could also be obtained by the related integral theorem 4.1.

We illustrate the Lagrangian formulation of the dynamical symmetry equation (7.6) by choosing a Lagrangian which characterizes a conservative dynamical system[30]:

$$L(x,\dot{x}) \equiv \tfrac{1}{2}g_{ij}\dot{x}^i\dot{x}^j - V(x). \tag{7.12}$$

By use of formulas (3.1), (3.8), and (3.9) we find that δL based upon (7.12) is of the form

$$\delta L = [\tfrac{1}{2}(\pounds_\xi g_{jk} - 4\phi g_{jk})\dot{x}^j\dot{x}^k - V_{,j}\xi^j]\delta a, \tag{7.13}$$

and [refer to (7.5)] the Noether function is given by

$$N\delta a = [\tfrac{1}{2}(\pounds_\xi g_{jk} - 2\phi g_{jk})\dot{x}^j\dot{x}^k - (2\phi V + V_{,j}\xi^j)]\delta a. \tag{7.14}$$

From (7.2) and (7.12) we have

$$E = \tfrac{1}{2}g_{jk}\dot{x}^j\dot{x}^k + V, \tag{7.15}$$

and it then follows by use of (7.14), (7.15), and (5.14) (which is used to eliminate $\ddot{x}^k$)[31] that the expansion of the symmetry condition (7.6) leads to

$$(\pounds_\xi \Gamma^i_{jk} - \delta^i_j\phi_{,k} - \delta^i_k\phi_{,j})\dot{x}^j\dot{x}^k + (4\phi g^{ij}V_{,j} + g^{ij}V_{;jk}\xi^k$$
$$- g^{jk}V_{,j}\xi^i_{;k}) = 0. \tag{7.16}$$

Since (7.16) must hold as an identity in the $\dot{x}^k$'s, we again obtain (5.15) and (5.16) as the conditions on ξ^i and ϕ for dynamical symmetry.

It now follows with $L(x,\dot{x})$ given by (7.12) that the concomitant constant of the motion (7.10) associated with the existence of the dynamical symmetry (7.6) is the derived constant of the motion (5.18), as would be expected.

We give another illustration of the Lagrangian formulation of dynamical symmetry conditions and associated constants of the motion by taking[32]

$$L(x,\dot{x}) \equiv \tfrac{1}{2}\eta_{AB}\dot{x}^A\dot{x}^B - H(x^A),\ A, B = 1, \ldots, 2n, \tag{7.17}$$

where the matrix $[\eta_{AB}]$ is the inverse to the matrix $[\eta^{AB}]$ [defined by (6.3)], $H(x^A)$ is the Hamiltonian, and the x^A are defined by (6.2).[33] Based upon (7.17), the formal expansion of Lagrange's equation

$$E_C \equiv \frac{\partial L}{\partial x^C} - \frac{d}{dt}\frac{\partial L}{\partial \dot{x}^C} = 0, \tag{7.18}$$

gives Hamilton's equation (6.1).

The condition (7.6) for a dynamical symmetry of Lagrange's equation when based upon L gives by (7.17) requires that

$$-2\phi\eta_{CB}\dot{x}^B-\eta_{AC}\xi^A_{,B}\dot{x}^B-\eta_{BA}\xi^A_{,C}\dot{x}^B-H_{,AC}\xi^A-H_{,A}\xi^A_{,C}=0. \tag{7.19}$$

To obtain the final form of the symmetry condition we use (6.1) to eliminate $\ddot{x}^B$ terms in (7.19). The resulting equation is easily shown to be equivalent to (6.4) (if use be made of the relation $\eta_{AB}\eta^{BC}=\delta^C_A$).

Application of Theorem 7.2 based upon (7.17) and a symmetry vector ξ^A which is a solution of the dynamical symmetry condition (6.4) gives [from (7.10)] the constant of the motion

$$H_{,A}\xi^A. \tag{7.20}$$

It should be noted that the left side of (7.20) is precisely $\delta H/\delta a$. This is to be expected since $H(x^A)$ is a known constant of the motion and the related integral theorem in the form (6.14) with $I=H$ gives (7.20).

8. NOETHER SYMMETRIES AND THEIR RELATIONSHIP TO DYNAMICAL SYMMETRIES FORMULATED AT THE LEVEL OF LAGRANGE'S EQUATIONS

In this section we make a comparison between the dynamical symmetry conditions and concomitant constants of the motion formulated directly at the level of Lagrange's equation (as described in Sec. 7) and those which are obtained from the well-known Noether identity.[3] We shall show that even for the relatively simple case of a conservative dynamical system these two approaches lead to considerable differences in symmetries and associated constants of the motion.

We give first a short derivation of Noether's identity which is independent of the integration processes of Hamilton's variational principle. For purposes of making the above-mentioned comparisons we shall base this derivation upon mappings of the form (3.1), (3.3).

We take then the Lagrangian in the form $L(x^i,\dot{x}^i)$ and define

$$L_i\equiv\frac{\partial L}{\partial x^i}-\frac{d}{dt}\frac{\partial L}{\partial\dot{x}^i}. \tag{8.1}$$

If we consider

$$\frac{dL}{dt}=\frac{\partial L}{\partial x^i}\dot{x}^i+\frac{\partial L}{\partial\dot{x}^i}\ddot{x}^i, \tag{8.2}$$

then from (8.1) and (8.2) we obtain

$$\frac{d}{dt}\left(\frac{\partial L}{\partial\dot{x}^i}\dot{x}^i-L\right)+L_i\dot{x}^i=0. \tag{8.3}$$

If (8.3) is multiplied by δt [as defined by (3.3)] and the resulting expression is subtracted from the right side of (7.4), we obtain

$$\delta L=-\left(\frac{d}{dt}(E\xi^0)-2\phi L+\frac{\partial L}{\partial\dot{x}^i}\xi^i+\frac{\partial L}{\partial\dot{x}^i}\xi^i_{,m}\dot{x}^m-L_i\dot{x}^i\xi^0\right)\delta a, \tag{8.4}$$

where E is defined by (7.2). Equation (8.4) may be rewritten in the form of Noether's identity[3,4]

$$N=\frac{d}{dt}\left(\frac{\partial L}{\partial\dot{x}^i}\xi^i-E\xi^0\right)+L_i(\xi^i-\dot{x}^i\xi^0), \tag{8.5}$$

where N is defined by (7.5).

Since along a dynamical path $L_i=0$, it follows that if $N(x,\dot{x})$ can be expressed in the form[3]

$$N\delta a=-\frac{d}{dt}(\delta\Omega),\quad \delta\Omega(x)\equiv\Omega(x)_{,i}\xi^i(x)\delta a, \tag{8.6}$$

then from (8.5) we obtain

$$\frac{d}{dt}\left(\frac{\partial L}{\partial\dot{x}^i}\xi^i\delta a-E\xi^0\delta a+\delta\Omega\right)=0. \tag{8.7}$$

We refer to (8.6) as the Noether symmetry condition and to (8.7) as the concomitant Noether constant of the motion.

We next consider the problem of obtaining specific conditions on the Noether symmetry vector ξ^i defined by (8.6) for the case of a conservative dynamical system with Lagrangian defined by (7.12). From (7.14) and (8.6) we immediately obtain

$$\tfrac{1}{2}(\pounds_\xi g_{ij}-2\phi g_{ij})\dot{x}^i\dot{x}^j+(\Omega_{,i}\xi^i)_{,j}\dot{x}^j-(2\phi V+V_{,i}\xi^i)=0. \tag{8.8}$$

Since (8.8) is not to act as a constraint on the dynamical system we must have

$$\pounds_\xi g_{ij}-2\phi g_{ij}=0, \tag{8.9}$$

$$2\phi V+V_{,i}\xi^i=0, \tag{8.10}$$

$$(\Omega_{,i}\xi^i)_{,j}=0. \tag{8.11}$$

From (8.11) and (8.6) we see that $d(\delta\Omega)/dt=0$ and hence from (8.6) we must have $N=0$. From the form of (8.11) we observe that any solution ξ^i of (8.9), (8.10) will lead to a differential equation which in general admits a solution for Ω. Conversely (8.9), (8.10), (8.11) imply (8.8) and (8.6).

We summarize the above results by stating

Theorem 8.1: A conservative dynamical system defined by a Lagrangian (7.12) will admit a Noether symmetry [defined by (8.6) and based upon infinitesimal mappings (3.1), (3.3)] if any only if (8.9), (8.10), and (8.11) are satisfied. These conditions imply $N=0$ and $\delta\Omega(x)=$ const (throughout space). From (8.11) and (8.7) we may also state the following:

Corollary 8.1: If a conservative dynamical system with Lagrangian (7.12) admits a Noether symmetry as described in Theorem 8.1, then it admits a constant of the motion

$$\frac{\partial L}{\partial\dot{x}^i}\xi^i-E\xi^0, \tag{8.12}$$

where E is defined by (7.2) and ξ^0 by (3.3).

We now investigate the relationship between the Noether symmetry condition (8.6) and the condition (7.6) for a dynamical symmetry based directly at the level of Lagrange's equation. We first obtain the condition that a Noether symmetry be a solution of the dynamical symmetry equation (7.6) for the general case in which $L=L(x^i,\dot{x}^i)$. It immediately follows from (7.6)

by use of (8.6) (since $\delta\Omega$ is a function of the coordinates only) that

$$\phi_{,j}E=0. \tag{8.13}$$

The dynamical symmetry condition (7.6) was formulated for an unconstrained dynamical system which implies that E is not restricted to any prescribed value. Hence (8.13) requires $\phi_{,j}=0$, i.e., $\phi=\text{const}$. Conversely with $\phi=\text{const}$ (7.6) is obviously satisfied by a Noether symmetry (8.6).

The above may be summarized by stating

Theorem 8.2: For an (unconstrained) dynamical system characterized by a Lagrangian $L(x,\dot{x})$ a necessary and sufficient condition that a Noether symmetry [defined by (8.6) and based upon infinitesimal mappings (3.1), (3.3)] satisfy (7.6), the condition for a dynamical symmetry b: d directly at the level of Lagrange's equations, is that $\phi=\text{const}$ in (3.2) [and (3.3)].

If in Theorem 8.2 we now specialize $L(x,\dot{x})$ to the form (7.12) which is characteristic of a conservative dynamical system, it then follows by use of Theorem 8.1 that we may state

Corollary 8.2: For a conservative dynamical system characterized by Lagrangian (7.12) necessary and sufficient conditions that a Noether symmetry vector, as described in Theorem 8.1, satisfy the (unconstrained) condition (7.6) for a dynamical symmetry based directly at the level of Lagrange's equations are that (8.9), (8.10) be satisfied with $\phi=\text{const}$. Such a symmetry is at most a homothetic motion.

It is of interest to note that conditions (8.9), (8.10) (with ϕ not necessarily constant) are the conditions that the mapping (3.1), (3.2) define a natural trajectory collineation of the zero energy trajectories.[9]

Next we turn our attention to the comparison of the constants of the motion which are concomitant to the two symmetry approaches considered above.

Based upon the mappings (3.1), (3.3) it is seen [we assume here the general case in which $L=L(x,\dot{x})$] that the Noether constants of the motion resulting from (8.7) will in general be time-dependent (as a result of the time-dependence of ξ^0) whereas the constants of the motion (7.10) concomitant with the dynamical symmetry condition (7.6) are independent of time.

We now wish to compare time-independent constants of the motion which may be obtained from the Noether formalism with those given by (7.10). To obtain such (Noether time-independent) constants of the motion from the Noether formula (8.7) we must limit the associated dynamical mappings (3.1), (3.3) to those for which $\phi=0$.

To further assess the implications of this restriction and to simplify the above-mentioned comparison we now specialize the dynamical system being considered to a conservative one with Lagrangian given by (7.12). The resulting constant of the motion thus obtained from the Noether approach is found from (8.12) to be

$$g_{ij}\xi^i\dot{x}^j - c(\tfrac{1}{2}g_{ij}\dot{x}^i\dot{x}^j + V),\ c\equiv\text{const} \quad \text{(Ref. 35)}, \tag{8.14}$$

where from (8.9), (8.10) it follows that the associated Noether symmetry vector ξ^i must satisfy Killing's equation

$$\mathcal{L}_\xi g_{ij} \equiv \xi_{i;j} + \xi_{j;i} = 0 \tag{8.15}$$

and

$$V_{,i}\xi^i = 0. \tag{8.16}$$

The second term in (8.14) is the total energy, a constant of the motion.[36] It therefore follows that the first term is also a constant of the motion which may be expressed in the form

$$\xi_i\dot{x}^i = \text{const}. \tag{8.17}$$

Thus for a conservative dynamical system based upon Lagrangian (7.12) we find the time-independent Noether constants of the motion associated with the Noether symmetry vectors ξ^i to be the well-known linear first integrals (momenta integrals) (8.17).

In contrast to these results it will be recalled that for the same conservative dynamical system we found (refer to Sec. 7) from the formulation of the symmetry conditions directly at the level of Lagrange's equations that the symmetry vector ξ^i must satisfy (5.15), (5.16) and it then followed that the concomitant constants of the motion were given by (5.18). We note that (5.18) is in general a quadratic integral and in general is not the total energy.

As a specific illustration of these contrasting results we consider the case of a three-dimensional isotropic harmonic oscillator. For such a dynamical system it can be shown that the vector ξ^i which is determined by (5.15), (5.16) defines a 9-parameter group of affine collineations AC_9.[7] This AC_9 contains three rotations and six proper affine collineations. Based upon the proper affine collineation symmetry vectors we find the concomitant constants of the motion defined by (7.10) [or equivalently by (5.18)] to be the well-known symmetric tensor constant of the motion [refer to (6.22)], whereas the time-independent Noether symmetry condition (8.15) is not satisfied by these proper affine collineation symmetry vectors and hence (8.17) is inapplicable. On the other hand, the rotation part of the dynamical symmetry which satisfies both (5.15), (5.16), and (8.13) accounts for the constant angular momenta by use of (8.17) and leads to a trivial result in (7.10).[37]

It is now apparent for the purpose of determining time-independent constants of the motion of a conservative dynamical system [with Lagrangian (7.12)] that those dynamical symmetry mappings [of the type (3.1), (3.3)] based directly at the level of Lagrange's equations are more general than those based upon the appropriate Noether symmetry mappings. As a result of this added generality in the dynamical symmetry mappings we obtain [through (7.10)] time-independent constants of the motion not associated with the Noether identity.

9. DYNAMICAL SYMMETRIES BASED UPON THE HAMILTON-JACOBI EQUATION AND THEIR RELATION TO NOETHER SYMMETRIES

In the preceding sections we obtained the conditions for dynamical symmetries by requiring that infinitesimal point mappings with associated changes in differen-

tial path-parameter map the set of solutions of a dynamical equation into itself. Thus far our formulation of symmetry conditions has been based directly at the level of the dynamical equations, which we have taken to be of the form of Newton's, Hamilton's, or Lagrange's equations.

It is well-known that the Noether symmetry condition (8.6) may be interpreted as defining a dynamical symmetry mapping at the level of Hamilton's variational principle, even though we found such a mapping in general did not satisfy the conditions for a dynamical symmetry which were formulated at the level of Lagrange's equations. As a means of interpreting the Noether mappings as dynamical symmetries directly at the level of the dynamical equations we now turn our attention to the transformation theory of mechanics—namely the Hamilton–Jacobi theory. For simplicity we shall limit our considerations to conservative dynamical systems characterized by Hamiltonians of the type[38]

$$H(x^i, p_i) = \tfrac{1}{2} g^{ij}(x) p_i p_j + V(x) \tag{9.1}$$

in which case the Hamilton–Jacobi equation takes the form

$$\tfrac{1}{2} g^{ij}(x) \frac{\partial S(x,t)}{\partial x^i} \frac{\partial S(x,t)}{\partial x^j} + V(x) + \frac{\partial S(x,t)}{\partial t} = 0, \tag{9.2}$$

where $S(x,t)$ is Hamilton's principal function.

It will be shown that the demand that the set of solutions of the Hamilton–Jacobi equation maps into itself[39] leads to the Neother symmetry conditions. This (not entirely unexpected) result will allow us to interpret the Noether symmetry mappings as dynamical symmetry mappings demonstrable at the level of the dynamical equations. By use of the Noether identity it then follows that the Noether constant of the motion (8.12) may be considered as a concomitant of dynamical symmetries based upon the Hamilton–Jacobi equation.

As a first step to determine infinitesimal point mappings of the form (3.1), (3.3) which map the set of solutions of (9.2) into itself, we define

$$J[S_{,t}, S_{,t}, x] \equiv \tfrac{1}{2} g^{ij}(x) \frac{\partial S(x,t)}{\partial x^i} \frac{\partial S(x,t)}{\partial x^j} + V(x) + \frac{\partial S(x,t)}{\partial t}, \tag{9.3}$$

and note that as a result of a mapping defined by (3.1), (3.3) Eq. (9.3) takes the form

$$J[S_{,\bar{t}}, S_{,\bar{t}}, \bar{x}] \equiv \tfrac{1}{2} g^{ij}(\bar{x}) \frac{\partial S(\bar{x},\bar{t})}{\partial \bar{x}^i} \frac{\partial S(\bar{x},\bar{t})}{\partial \bar{x}^j} + V(\bar{x}) + \frac{\partial S(\bar{x},\bar{t})}{\partial \bar{t}}. \tag{9.4}$$

In (9.4) we next express $\bar{x}$, $\bar{t}$ in terms of x, t by means of (3.1), (3.3) and obtain

$$\begin{aligned} J[S_{,\bar{t}}, S_{,\bar{t}}, \bar{x}] = {} & \tfrac{1}{2} g^{ij}(x+\delta x) \left(\frac{\partial S(x+\delta x, t+\delta t)}{\partial x^k} (\delta^k_i - \xi^k_{,i} \delta a) \right) \\ & \times \left(\frac{\partial S(x+\delta x, t+\delta t)}{\partial x^m} (\delta^m_j - \xi^m_{,j} \delta a) \right) \\ & + V(x+\delta x) \\ & + \frac{\partial S(x+\delta x, t+\delta t)}{\partial t} (1 - 2\phi(x) \delta a). \end{aligned} \tag{9.5}$$

If (9.5) is expanded in powers of δa (to first order), we obtain

$$J(S_{,\bar{t}}, S_{,\bar{t}}, \bar{x}] = J[S_{,t}, S_{,t}, x] + \delta J[S_{,t}, S_{,t}, x], \tag{9.6}$$

where

$$\delta J \equiv \tfrac{1}{2} (\pounds_\xi g^{ij}) S_{,t} S_{,j} \delta a + g^{ij} S_{,t} (\delta S)_{,j} + (\delta S)_{,t} - 2\phi S_{,t} \delta a + V_{,k} \xi^k \delta a, \tag{9.7}$$

with

$$\delta S \equiv (S_{,i} \xi^i + S_{,t} \xi^0) \delta a \tag{9.8}$$

and

$$\pounds_\xi g^{ij} \equiv g^{ij}_{,k} \xi^k - g^{ik} \xi^j_{,k} - g^{kj} \xi^i_{,k}. \tag{9.9}$$

Consider next the expression

$$J[(S+\delta S)_{,t}, (S+\delta S)_{,t}, x] \equiv \tfrac{1}{2} g^{ij} (S+\delta S)_{,i} (S+\delta S)_{,j} + V(x) + (S+\delta S)_{,t}. \tag{9.10}$$

Expansion of (9.10) to first order in δa leads to

$$J[(S+\delta S)_{,t}, (S+\delta S)_{,t}, x] = J[S_{,t}, S_{,t}, x] + \delta^* J[S_{,t}, S_{,t}, x], \tag{9.11}$$

where

$$\delta^* J[S_{,t}, S_{,t}, x] \equiv g^{ij} S_{,i} (\delta S)_{,j} + (\delta S)_{,t}. \tag{9.12}$$

We now formulate the conditions for a dynamical symmetry mapping based upon the Hamilton–Jacobi equation by requiring that

$$J[S_{,t}, S_{,t}, x] + \delta J[S_{,t}, S_{,t}, x] = J[(S+\delta S)_{,t}, (S+\delta S)_{,t}, x], \tag{9.13}$$

hold for all functions $S(x,t)$ which satisfy (9.2).

By use of (9.11) we find that (9.13) requires $\delta J = \delta^* J$ and hence by (9.7) and (9.12), we obtain

$$\tfrac{1}{2} (\pounds_\xi g^{ij}) S_{,t} S_{,j} - 2\phi S_{,t} + V_{,k} \xi^k = 0. \tag{9.14}$$

We now eliminate the $S_{,t}$ term in (9.14) by use of (9.2) and obtain

$$\tfrac{1}{2} (\pounds_\xi g^{ij} + 2\phi g^{ij}) S_{,t} S_{,j} + 2\phi V + V_{,k} \xi^k = 0. \tag{9.15}$$

Since (9.15) must not act as a constraint on the Hamilton–Jacobi equation (9.2) we must require that (9.15) hold identically in the quantities $S_{,i}$. We thus obtain as the requirements for dynamical symmetry the conditions

$$\pounds_\xi g^{ij} + 2\phi g^{ij} = 0, \tag{9.16}$$

$$2\phi V + V_{,k} \xi^k = 0. \tag{9.17}$$

It immediately follows [by use of $\pounds_\xi g^{ij} \equiv - g^{ia} g^{jb} \pounds_\xi g_{ab}$ in (9.16)] that (9.16), (9.17) are the Noether symmetry conditions (for a conservative dynamical system) (8.9), (8.10), respectively.

Conversely, if the infinitesimal point mapping (3.1), (3.3) satisfies the Noether symmetry conditions (9.16), (9.17), then (9.13) will be satisfied.

We thus have

Theorem 9.1: For a conservative dynamical system with Hamiltonian (9.1) a necessary and sufficient condition that the infinitesimal transformation (3.1), (3.3) map the set of solutions of the Hamilton–Jacobi equation (9.2) into itself [in that (9.13) holds] is that the mapping be a Noether symmetry mapping, i.e., the transformation must satisfy (8.9), (8.10).

By use of the Noether identity we may now associate with the existence of the above described dynamical symmetry formulated of the level of the Hamilton–Jacobi equation a concomitant constant of the motion. Hence we may restate Corollary 8.1 in the following form.

Corollary 9.1: If a conservative dynamical system with Hamiltonian (9.1) admits an infinitesimal transformation (3.1), (3.3) which maps the set of solutions of the Hamilton–Jacobi equation into itself as defined in Theorem 9.1, then there will exist a concomitant constant of the motion which may be expressed in the form (8.12).

ACKNOWLEDGMENT

The authors wish to thank Dr. J.W. York, Jr. for his corrections and very helpful comments.

[1]For a general discussion of the conceptual foundations of symmetry principles in physics see, for example, R.M.F. Houtappel, H. Van Dam, and E.P. Wigner, Rev. Mod. Phys. **37**, 595 (1965).

[2]E. Noether, Nachr. Akad. Wiss. Göttingen, Math.-Phys. K1. 235 (1918).

[3]Numerous papers have been written which give detailed derivations of the so-called Noether approach to dynamical symmetries. The paper by E.L. Hill, Rev. Mod. Phys. **23**, 253 (1951) is of particular interest in that it gives a general derivation followed by a specialization to particle dynamical systems.

[4]E.T. Whittaker, *A Treatise on the Analytical Dynamics of Particles and Rigid Bodies* (Cambridge U.P. Cambridge, 1965), 4th ed.

[5]G.H. Katzin and J. Levine, J. Math. Phys. **9**, 8 (1968).

[6]G.H. Katzin and J. Levine, Colloquium Mathematicum (Wroclaw, Poland) **26**, 21 (1972).

[7]G.H. Katzin, J. Math. Phys. **14**, 1213 (1973).

[8]G.H. Katzin, Lett. Nuovo Cimento **7**, 213 (1973).

[9]J. Levine and G.H. Katzin, J. Math. Phys. **14**, 1886 (1973).

[10]In this section we expand and elaborate on basic results given in Ref. 8.

[11]A. Komar, Phys. Rev. D **8**, 1028 (1973), has also considered noncanonical symmetry mappings and concomitant constants of the motion.

[12]We assume Eqs. (2.1) are all of the second order or all of the first order. We use second order equations to illustrate the basic methods which we wish to develop.

[13]The interpretations of the x^i and their assigned index ranges for these specific types of dynamical equations will be made in the appropriate sections to follow.

[14]Velocity dependent constants of the motion are to be associated with Eqs. (2.1) which are of second order. See Footnote 12.

[15]Unless otherwise indicated the Einstein summation convention will be used. Also small italic indices will have the range 1 through n (unless otherwise indicated). A comma (,) indicates partial differentiation.

[16]Equations (3.14) will usually be in the form of polynomials in the x^i, in which case the symmetry conditions are easily obtained.

[17]The first term in the summation corresponding to $m=0$ will be denoted by P^i.

[18]For a general discussion of Lie derivatives, see K. Yanô, *The Theory of Lie Derivatives and Its Applications* (North-Holland, Amsterdam, 1957).

[19]Equations (5.15), (5.16) were derived in Ref. 7 by an alternative method. A detailed application of these equations to the Kepler problem and isotropic harmonic oscillator is also given in Ref. 7.

[20]Covariant differentiation with respect to Γ^i_{jk} is indicated by a semicolon (;).

[21]For applications of (5.18) to the problems mentioned in Ref. 19, see Ref. 7.

[22]Capital indices will have the range 1, 2, ..., $2n$.

[23]Here 0_n and I_n are nth order zero and identity matrices, respectively.

[24]An alternative derivation of (6.4) is given in Ref. 8, where it is shown that in general the dynamical mappings defined by (6.4) are not canonical. A similar equation (in which $\phi=0$) was obtained in Ref. 11 for the case of a time-dependent ξ^A.

[25]A similar constant of the motion with $\phi=0$ is given in Ref. 11.

[26]The notations ∂_A, $\bar{\partial}_A$ indicate partial differentiations with respect to x^A, $\bar{x}^A$, respectively.

[27]The deformation operator δ_α is defined in terms of the vector ξ^A_α.

[28]A general solution to (6.4) for this problem has been obtained and will be published elsewhere.

[29]In Sec. 8 the symmetry conditions (7.6) will be compared with the Noether symmetry condition (8.6).

[30]This choice will allow a comparison with the dynamical symmetry conditions based on the "Newtonian form" of the dynamical equations as discussed in Sec. 5 and will also be useful in our comparison with the Noether approach of Sec. 8.

[31]Note that Lagrange's equation (7.1) expands to the form (5.14) by use of (7.12).

[32]C.W. Kilmister, *Hamiltonian Dynamics* (American Elsevier, New York, 1965).

[33]In the remainder of this section we shall again assume all capital indices range $1, \ldots, 2n$ and employ the notation of Sec. 6.

[34]It can be shown that if the mapping (3.1), (3.3) is assumed to have the more general form $\bar{x}^i = x^i + \xi^i(x,t)\delta a$, $\bar{t} = t + \xi^0(x,t)\delta a$ and the Lagrangian L taken in the form $L = L(x, \dot{x}, t)$, then a similar derivation to that given above again leads to the Noether identity (8.5), but expressed in terms of the generalized mapping vectors.

[35]Note that the constant c is the same as appears in (3.3).

[36]Since the energy integral is independent of the mapping vector ξ^i this constant of the motion may be regarded as arising from a time-translation $\delta t = c\delta a$.

[37]A similar analysis of the Kepler problem shows that the Runge–Lenz vector to be given by (7.10) and the angular momenta given by the Noether formula (8.17). See Ref. 7.

[38]This Hamiltonian corresponds to Lagrangian (7.12).

[39]The question of mapping partial differential equation solutions into themselves has also been considered by R.L. Anderson, S. Kumei, and C.E. Wulfman, Phys. Rev. Lett. **28**, 988 (1972); J. Math. Phys. **14**, 1527 (1973).

Special relativity in general relativity?*

D. Hirondel†

Department of Physics, University of Delaware, Newark, Delaware
(Received 13 August 1974)

Starting from the metric in harmonic coordinates for a test particle m_1 around a heavy particle $m_2(m_2 \gg m_1)$ at rest, the EIH Lagrangian is recovered by making a Lorentz transformation, followed by a canonical transformation and an appropriate symmetrization in the two masses. This raises the question of a special relativity content in general relativity, a feature not directly implied by the general covariance.

1. INTRODUCTION

The laws of gravitation in Einstein's general relativity are generally covariant with respect to any change of coordinates; in any set of coordinates they will look like $R_{ij} - \frac{1}{2}g_{ij}R = 8\pi G c^{-4} T_{ij}$, where R_{ij} is a certain function of the g_{kl} and their derivatives, up to second order. $ds^2 = g_{ij}(q)\,dq^i dq^j$ is postulated to be an invariant.

General covariance was of much value to Einstein, and inspired the name of his theory. But it is now admitted that general covariance in itself is physically empty, to the extent that the equations of any theory could be cast in covariant form. This format invariance gives the equations of motion, for example, the same aspect in any set of coordinates, but the functional dependence of the accelerations on their arguments will be *a priori* different in each set of coordinates.

However, we observe that, for a certain family of coordinates, we do have functional invariance of the accelerations with respect to their arguments, under Lorentz transformations, for the case of two masses, up to order c^{-2}. Einstein, Infeld, and Hoffman (EIH)[1] and, later, Fock[2] showed there existed a common Lagrangian for the problem of two gravitating particles,

$$L = \frac{1}{2}m_1\mathbf{v}_1^2 + \frac{1}{8}m_1\frac{\mathbf{v}_1^4}{c^2} + \frac{1}{2}m_2\mathbf{v}_2^2 + \frac{1}{8}m_2\frac{\mathbf{v}_2^4}{c^2} + \frac{Gm_1m_2}{r}\left[1 + \frac{1}{2c^2}\left(3\mathbf{v}_1^2 + 3\mathbf{v}_2^2 - 7\mathbf{v}_1\cdot\mathbf{v}_2 - \frac{\mathbf{r}\cdot\mathbf{v}_1\,\mathbf{r}\cdot\mathbf{v}_2}{r^2}\right)\right] - \frac{G^2m_1m_2(m_1+m_2)}{2c^2r^2}.$$

On the other hand, Currie and Hill[3] have given the conditions on a dynamics:

$$\frac{d^2\mathbf{r}_1(t)}{dt^2} = \mathbf{a}_1\left[\mathbf{r}(t), \frac{d\mathbf{r}_1(t)}{dt}, \frac{d\mathbf{r}_2(t)}{dt}\right],$$

guaranteeing that, after a Lorentz transformation, it will look like

$$\frac{d^2\mathbf{r}_1'(t')}{dt'^2} = \mathbf{a}_1\left[\mathbf{r}'(t'), \frac{d\mathbf{r}_1'(t')}{dt'}, \frac{d\mathbf{r}_2'(t')}{dt'}\right],$$

where the **a**'s are the *same* functions of their arguments as before the transformation. World line invariance is postulated to obtain these conditions. They put all Lorentz frames on the same footing, and privilege none of them. This special relativistic covariance is a strong requirement, while general covariance is not.

It is a fact that the accelerations obtained from the EIH Lagrangian,

$$m_1\mathbf{a}_1 = -\frac{Gm_1m_2\mathbf{r}}{r^3} + \frac{Gm_1m_2}{c^2r^3}\times\left[\mathbf{r}\left(-\mathbf{v}_1^2 + 4\mathbf{v}_1\cdot\mathbf{v}_2 - 2\mathbf{v}_2^2 + \frac{3}{2}\frac{(\mathbf{r}\cdot\mathbf{v}_2)^2}{r^2}\right) + (\mathbf{v}_1 - \mathbf{v}_2)(4\mathbf{r}\cdot\mathbf{v}_1 - 3\mathbf{r}\cdot\mathbf{v}_2)\right] + \frac{G^2m_1m_2(5m_1+4m_2)}{c^2r^4}\mathbf{r},$$

satisfy Currie–Hill conditions up to order c^{-2}:

$$(\partial_{\mathbf{v}_1} + \partial_{\mathbf{v}_2})\mathbf{a}_1 = \frac{1}{c^2}\left[\mathbf{r}\mathbf{v}_2\cdot\partial_{\mathbf{r}}\mathbf{a}_1 + \mathbf{v}_1\mathbf{v}_1\cdot\partial_{\mathbf{v}_1}\mathbf{a}_1 + (\mathbf{v}_2\mathbf{v}_2 - \mathbf{r}\mathbf{a}_2)\cdot\partial_{\mathbf{v}_2}\mathbf{a}_1 - 2\mathbf{v}_1\mathbf{a}_1 - \mathbf{a}_1\mathbf{v}_1\right],$$

as one can check by direct calculation.

The original computations for obtaining the EIH Lagrangian being quite lengthy, the preceding result suggests how to recover it in a simple way by starting from the known metric for a test particle around a heavy particle at rest, in harmonic coordinates (the choice of this particular set will be justified), and making a Lorentz transformation for setting the heavy body into uniform motion. A canonical transformation will symmetrize the Lagrangian in the velocities of both particles. Symmetrization in the two masses, to yield a Lagrangian usable when the two masses are comparable, requires a certain care. The end result is the EIH Lagrangian.

2. THE EIH LAGRANGIAN FROM THE CASE OF ONE BODY AT REST

The metric for a test particle m_1 in the field of a spherically symmetric, heavy particle $m_2(m_2 \gg m_1)$ is, in harmonic coordinates ($r_0 \equiv Gm_2c^{-2}$),

$$ds^2 = c^2\,dt^2\,\frac{r - r_0}{r + r_0} - \frac{r + r_0}{r - r_0}\,dr^2 - (r + r_0)^2(d\theta^2 + \sin^2\theta\,d\varphi^2) = c^2\,dt^2\left\{\frac{r - r_0}{r + r_0} - \frac{1}{c^2}\left[\frac{r + r_0}{r - r_0}\left(\frac{\mathbf{r}\cdot\mathbf{v}_1}{r}\right)^2 + (r + r_0)^2\,\frac{\mathbf{v}_1^2 - (\hat{\mathbf{r}}\cdot\mathbf{v}_1)^2}{r^2}\right]\right\}.$$

The Lagrangian for particle 1 is, by expanding ds up to order $1/c^2$:

$$L_1 = m_1c\frac{ds}{dt} + m_1c^2$$

$$=m_1\left(\frac{\mathbf{v}_1^2}{2}+\frac{\mathbf{v}_1^4}{8c^2}\right)+\frac{Gm_1m_2}{r}\left(1+\frac{3}{2}\frac{\mathbf{v}_1^2}{c^2}\right)-\frac{G^2m_1m_2^2}{2c^2r^2}.$$

Since $1/c^2$ appears in the two dimensionless ratios v^2/c^2 and Gm_2/c^2r, keeping only the lowest powers of $1/c^2$ is an approximation valid for slow motion and sufficiently large separation. A cutoff at a given level in $1/c^2$ implies a corresponding cutoff of the series in G: The $1/c^2$ term in $\mathcal{L}_1$ is a polynomial in G containing no higher power than G^2. Later on, the expression "up to order G or G^2" will refer to the coefficient of $1/c^2$ in $\mathcal{L}_1$.

Now, we go into a frame in which the velocity of particle 2 is $\mathbf{v}_2$. The kinetic energy part of $\mathcal{L}_1$: $m_1(\mathbf{v}_1^2/2+\mathbf{v}_1^4/8c^2)$ is left unchanged by a Lorentz transformation as it corresponds to the invariance of the proper time $d\tau_1=(1-\mathbf{v}_1^2)^{1/2}\,dt$.

For the remainder, it is enough to use:

$$\mathbf{v}_1\to\mathbf{v}_1-\mathbf{v}_2,\quad \mathbf{r}\to\mathbf{r}+\frac{(\mathbf{r}\cdot\mathbf{v}_2)\mathbf{v}_2}{2c^2},$$

$$dt\to dt\left(1-\frac{\mathbf{v}_1\cdot\mathbf{v}_2}{c^2}+\frac{\mathbf{v}_2^2}{2c^2}\right),$$

where, after the Lorentz transformation, we made an instantaneity correction: $\mathbf{r}_2(t_2)=\mathbf{r}_2(t_1)-\mathbf{v}_2(t_1-t_2)$ to make the two particles simultaneous in the new frame. We obtain

$$\mathcal{L}_1=m_1\left(\frac{\mathbf{v}_1^2}{2}+\frac{\mathbf{v}_1^4}{8c^2}\right)+\frac{Gm_1m_2}{r}\left[1+\frac{1}{c^2}\left(-\frac{(\mathbf{r}\cdot\mathbf{v}_2)^2}{r^2}\right.\right.$$

$$\left.\left.+\frac{3}{2}(\mathbf{v}_1^2-2\mathbf{v}_1\cdot\mathbf{v}_2+\mathbf{v}_2^2)-\mathbf{v}_1\cdot\mathbf{v}_2+\frac{\mathbf{v}_2^2}{2}\right)\right]-\frac{G^2m_1m_2^2}{2c^2r^2}.$$

Adding

$$m_2\left(\frac{\mathbf{v}_2^2}{2}+\frac{\mathbf{v}_2^4}{8c^2}\right)+\frac{d}{dt}\left(\frac{Gm_1m_2}{2c^2}\,\frac{\mathbf{r}\cdot\mathbf{v}_2}{r}\right),$$

does not change the acceleration of particle 1, but yields a Lagrangian which, up to order G, is not symmetric in the indices of both particles, and can be used for both up to that order:

$$\mathcal{L}_1=m_1\left(\frac{\mathbf{v}_1^2}{2}+\frac{\mathbf{v}_1^4}{8c^2}\right)+m_2\left(\frac{\mathbf{v}_2^2}{2}+\frac{\mathbf{v}_2^4}{8c^2}\right)$$

$$+\frac{Gm_1m_2}{r}\left[1+\frac{1}{2c^2}\left(3\mathbf{v}_1^2+3\mathbf{v}_2^2-7\mathbf{v}_1\cdot\mathbf{v}_2-\frac{\mathbf{r}\cdot\mathbf{v}_1\mathbf{r}\cdot\mathbf{v}_2}{r^2}\right)\right]$$

$$-\frac{G^2m_1m_2(m_2)}{2c^2r^2}.$$

This last addition is the same as the one we would make to recover by a Lorentz transformation the Darwin–Breit Lagrangian from the Lagrangian of a test particle in the electromagnetic field of a heavy charged particle at rest:

$$\delta\int\left[-m_1c^2\left(1-\frac{\mathbf{v}_1^2}{c^2}\right)^{1/2}-\frac{e_1e_2}{r}\right]dt=0.$$

In this case, there is no term quadratic in the product of the two charges (contrasting with the gravitational case, where the G^2 term was the mark of the nonlinearity). The only effect of the Lorentz transformation is the appearance of a magnetic force; radiation is absent.

Now, by admitting that there exists a common Lagrangian up to order $1/c^2$, symmetric in the indices of the velocities and the mass of both particles,

$$\mathcal{L}=\text{kinetic energy}+\frac{Gm_1m_2}{r}\left[1+\frac{1}{c^2}[A(\mathbf{v}_1^2+\mathbf{v}_2^2)+B\mathbf{v}_1\cdot\mathbf{v}_2\right.$$

$$\left.+C\hat{r}\cdot\mathbf{v}_1\hat{r}\cdot\mathbf{v}_2]+D\left(\frac{Gm_1}{c^2r}+\frac{Gm_2}{c^2r}\right)\right]$$

(A,B,C,D being numerical coefficients), which we know from the work of Einstein, Infeld, and Hoffman, and Fock, then the only function symmetric in m_1 and m_2 which yields m_2 when $m_2\gg m_1$ is unambiguously m_1+m_2. This last step certainly modifies the acceleration of particle 1, but supplies us with a common Lagrangian which can be used for both particles up to order G^2, when both masses are comparable, and which coincides with the EIH Lagrangian.

Now, we have to give the reason why making a Lorentz transformation in harmonic coordinates was expected to lead to the correct result. A first practical reason is that Fock worked out the EIH Lagrangian in harmonic coordinates: Thus, if we wished to find the same result, we had to start from the same kind of coordinates (we will see that the harmonicity condition is left unchanged by a Lorentz transformation). Einstein, Infeld, and Hoffman defined their system of coordinates at each step of approximation, but their result is the same.

Another reason why, at least up to order $1/c^2$, Lorentz transformations were expected to play an important role is that the field equations in first approximation, and in harmonic coordinates, are

$$\Box^2(h_{ij}-\tfrac{1}{2}\delta_{ij}h)=-16\pi Gc^{-4}T_{ij}\quad(g_{ij}\approx\delta_{ij}+h_{ij}),$$

where $\Box^2$ is the d'Alembertian in flat space. Thus the field propagates along light cones, as in special relativity. At higher orders in c^{-2}, the characteristics for the propagation of the field will be curved, not straight, and their form will be codetermined with the configuration of the masses.

However, in spite of the loss of isotropy and uniformity, a theoretical reason to guess that Lorentz transformation will still play a major role to all orders in c^{-2} in Fock's proposition[2] that it is unlikely that there exists any system of coordinates, other than the harmonic set, determined uniquely apart from a Lorentz transformation, because they are characterized by the fact that they satisfy a linear, generally covariant equation:

$$\Gamma^i\equiv\frac{1}{|g|^{1/2}}\,\frac{\partial}{\partial x^j}\,(|g|^{1/2}g^{ji})\equiv\Box^2x^i=0.$$

Harmonic coordinates exclude all fictitious gravitational fields, and, in a way, they can be called the most inertiallike coordinates. Is that enough to entail world line invariance and functional invariance under Lorentz transformation to all orders?

CONCLUSION

We have recovered the EIH Lagrangian by a procedure whose simplicity contrasts with the lengthiness of the original computations. It raises the question of a Special Relativity content in General Relativity, apparently restricted to harmonic coordinates. As a test, it would be worth computing the accelerations up to order c^{-4} to see if they satisfy the Currie-Hill conditions.

ACKNOWLEDGMENTS

It is a pleasure to thank Dr. Kerner and Dr. Hill for discussions. Financial support from the Physics Department of the University of Delaware is acknowledged.

*Abstracted from Ph. D. Dissertation "Canonical Formulation of Instantaneous Electrodynamics. Gravitation", University of Delaware, submitted June 1973.

†Present address: 4, Rue Sédillot, Paris 7, France.

[1] L. Infeld and J. Peblanski, *Motion and Relativity* (Pergamon, New York, 1960). A. Einstein, L. Infeld, and B. Hoffman, Ann. Math. **39**, 65 (1938).

[2] V. Fock, *The Theory of Space, Time, and Gravitation* (Pergamon, New York, 1959), p. 288. Fock's remarks on the harmonicity condition, p. 350.

[3] R. N. Hill, J. Math. Phys. **8**, 201 (1967); D. G. Currie, Phys. Rev. **142**, 817 (1966).

Stability of stochastic functional differential equations

M. H. Chang, G. Ladde, and P. T. Liu

Department of Mathematics, University of Rhode Island, Kingston, Rhode Island
(Received 3 January 1973)

A system of functional differential equations with random retardation, $\dot{x}(t) = f(t, x_t)$, is studied, where $x_t(\theta) = x(t+\theta)$, $\eta(t,\omega) \le \theta \le 0$, $-r \le \eta(t,\omega) \le 0$, and $\eta(t,\omega)$ is a stochastic process defined on some probability space (Ω, μ, P). Some comparison theorems are stated and proved in details under suitable assumptions on $f(t, x_t)$. Sufficient conditions for stability in the mean for the trivial solution then follow. The usefulness of the sufficient conditions is illustrated by an example with two different Lyapunov functions.

1. INTRODUCTION

In the study of some sophisticated dynamical systems, it is always desirable to consider either stochastic differential equations or functional differential equations. Despite the amount of work that has been done on each of the two types of differential equations, very little effort seems to have been devoted to correlate them.

In Ref. 1, Lidskii investigated the problem of stability in the mean for the solutions of a system of linear differential equations with random delays. In that paper, he used the Lyapunov direct method, without proving its validity in the case of random delays, and obtained some sufficient conditions for stability of the trivial solution in the mean.

The concept of the Lyapunov function, together with the theory of differential inequalities provides a very general comparison theorem (see Ref. 2) by means of which a number of qualitative properties of solutions of differential equations may be studied in a unified way.

In the present paper, we consider a general class of stochastic functional differential equations. Our main purpose is to establish the stochastic version of the comparison theorem for the functional differential equation in Ref. 2. As in the deterministic case, sufficient conditions for stability (in the mean) follow easily from the comparison theorem.

2. NOTATIONS AND DEFINITIONS

Let $\{\eta(t, \cdot),\ t \in R^+\}$ be a stochastic process defined on a probability space (Ω, μ, P) and takes values in $[-r, 0]$, where $R^+ = [0, \infty)$. Let $\mathcal{C} = C([-r, 0], R^n)$ and, for $\phi \in \mathcal{C}$, we define $\|\phi\|_0 = \sup_{-r \le \theta \le 0} \|\phi(\theta)\|$, where $\|\cdot\|$ is an Euclidean norm.

Consider the system of stochastic functional differential equation

$$\dot{x}(t) = f(t, x_t) \tag{2.1}$$

with initial conditions

$$x_{t_0} = \phi_0 \in \mathcal{C}, \quad \eta(t_0) = \eta_0,$$

where

$$x_t = x(t+s), \quad \eta(t) \le s \le 0, \quad -r \le \eta(t) \le 0, \quad \text{and} \quad t_0 \in R^+.$$

From now on, we shall assume $f(t, 0) = 0$. To ensure that (2.1) has a solution $x(t_0, \phi_0, \eta_0)(t)$ on $[t_0, \infty)$ with initial data (t_0, ϕ_0, η_0), we shall assume that $f(t, x_t)$ is sufficiently smooth and $\eta(t, \cdot)$ takes on countably many values, say R, in $[-r, 0]$.

Also, consider the scalar differential equation

$$u(t) = g(t, u), \quad u(t_0) = u_0 \ge 0, \tag{2.2}$$

where $g \in C(R^+ \times R^+, R^+)$, $g(t, 0) = 0$ and $g(t, u)$ is concave in u for fixed t. A solution of (2.2) will be denoted by $u(t, t_0, u_0)$ defined on $[t_0, \infty)$.

Let us define some definitions of stability in mean of the trivial solution $x \equiv 0$ of (2.1) as following:

Definition 2.1: The trivial solution $x \equiv 0$ of (2.1) is said to be:

(i) equistable in mean, if for each $\epsilon > 0$, $t_0[0, \infty)$, there exists a positive function $\delta = \delta(t_0, \epsilon)$ that is continuous in t_0 for each $\epsilon > 0$ such that the inequality $\|\phi_0\|_0 \le \delta$ implies $M[\|x(t_0, \phi_0, \eta_0)(t)\| \mid \phi_0, \eta_0] < \epsilon$ for $t \ge t_0$.

(ii) uniformly stable in mean, if the δ in (i) is independent of t_0.

(iii) quasi-equi asymptotically stable in mean, if for each $\epsilon > 0$, $t_0 \in [0, \infty)$, there exist $\delta_0 = \delta_0(t_0) > 0$ and $T = T(t_0, \epsilon) > 0$ such that for $t \ge t_0 + T$ and $\|\phi_0\|_0 < \delta$ implies $M[\|x(t_0, \phi_0, \eta_0)(t)\| \mid \phi_0, \eta_0] < \epsilon$.

(iv) quasiuniformly asymptotically stable in mean, if δ_0 and T in (iii) are independent of t_0.

(v) equiasymptotically stable in mean if (i) and (iii), hold simultaneously; uniformly asymptotically stable in mean if (ii) and (iv) hold together.

Corresponding to Definition 2.1 (i), we can define the definition of stability of the trivial solution $u \equiv 0$ of (2.2) as the following:

Definition 2.2: The trivial solution $u = 0$ of (2.2) is said to be equi-stable, if for each $\epsilon > 0$, $t_0 \in [0, \infty)$, there exists a positive function $\delta = \delta(t_0, \epsilon)$ that is continuous in t_0 for each $\epsilon > 0$ such that

$$u(t, t_0, u_0) < \epsilon \quad t \ge t_0 \quad \text{whenever } u_0 \le \delta.$$

The following two definitions will be useful.

Definition 2.3: A function $a(r)$ is said to belong to the class $\mathcal{K}$ if $a \in C(R^+, R^+)$, $a(0) = 0$, and $a(r)$ is strictly monotone increasing in r.

Definition 2.4: A function $b(t, r)$ is said to belong to the class β, if $b \in C(R^+ \times R^+, R^+)$, $b(t, \cdot) \in \mathcal{K}$ for each $t \in R^+$.

3. COMPARISON THEOREMS

Let the Lyapunov function, $V(t, y, \eta(t))$, be defined as $V \in C([-r, \infty) \times R^n \times R, R^+)$. Define

$$D^+M[V(t, \phi(0), \eta(t); \phi)]$$

$$\equiv \overline{\lim_{h\to 0^+}}\, h^{-1}\{M[V(t+h,\phi(0)+hf(t,\phi),\eta(t+h))\mid \eta(t)] - V(t,\phi(0),\eta(t))\},\quad t>t_0, \tag{3.1}$$

where $\phi\in\mathcal{C}$, $M[A\mid B]$ is a conditional mathematical expectation.

$$\mathcal{A}_1\equiv\{\phi\in\mathcal{C}\mid \sup_{-r\le s\le 0} M[V(t+s,\phi(s),\eta(t+s))\mid\eta_0] = M[V(t,\phi(0),\eta(t))\mid\eta_0]\}, \tag{3.2}$$

and

$$\mathcal{A}_2\equiv\{\phi\in\mathcal{C}\mid \sup_{-r\le s\le 0} M[V(t+s,\phi(s),\eta(t+s))\mid\eta_0]A(t+s) = M[V(t,\phi(0),\eta(t))\mid\eta_0]A(t)\}, \tag{3.3}$$

where $A(t)>0$ is continuous on $[-r,\infty)$.

After introducing all those notation and definitions listed above, we shall state and prove in detail the stochastic version of the main comparison theorem (see Ref. 2, p. 83) as the following:

Theorem 3.1: Let $V\in C([-r,\infty)\times R^n\times R, R^+)$ and $V(t,\phi(0),\eta(t))$ be locally Lipschitzian in $\phi(0)$ uniformly in $\eta(t)$. Assume $D^+M[V(t,\phi(0),\eta(t);\phi)]$ defined as in (3.1), satisfies

$$D^+M[V(t,\phi(0),\eta(t);\phi)]\le g(t,V(t,\phi(0),\eta(t)))\quad \text{for } t>t_0, \tag{3.4}$$

where $\phi\in\mathcal{A}_1$; $g\in C(R^+\times R^+,R^+)$ and $g(t,u)$ is concave in u for fixed t.

Let $r(t)=r(t,t_0,u_0)$ be the maximal solution of the scalar differential equation

$$\dot u = g(t,u),\ u(t_0)=u_0\ge 0, \tag{3.5}$$

and $r(t)$ exists to the right of t_0.

If $\{x(t_0,\phi_0,\eta_0)(t,\omega)\}$ is the realization of solutions of (2.1) defined on $[t_0,\infty)$, satisfying

$$\sup_{-r\le s\le 0} V(t_0,\phi_0(s),\eta_0)\le u_0, \tag{3.6}$$

and $M[V(t,x(t_0,\phi_0,\eta_0)(t),\eta(t))\mid\phi_0,\eta_0]$ is differentiable in t, then

$$M[V(t,x(t_0,\phi_0,\eta_0)(t),\eta(t))\mid\phi_0,\eta_0]\le r(t,t_0,u_0),\ t\ge t_0. \tag{3.7}$$

Proof: Let $x(t_0,\phi_0,\eta_0)(t)$ be any solution of (2.1) with initial conditions (t_0,ϕ_0,η_0).

Define

$$m(t)\equiv M[V(t,x(t_0,\phi_0,\eta_0)(t),\eta(t))\mid\phi_0,\eta_0].$$

For $\epsilon>0$ sufficiently small consider the scalar differential equation

$$\dot u(t)=g(t,u)+\epsilon,\quad u(t_0)=u_0\ge 0 \tag{3.8}$$

whose solutions $u(t,\epsilon)=u(t,\epsilon,t_0,u_0)$ exists as far as $r(t,t_0,u_0)$ exists to the right of t_0. Since

$$\lim_{\epsilon\to 0} u(t,\epsilon)=r(t,t_0,u_0)\quad \text{for all } t\ge t_0.$$

Hence to show that the theorem is true, it is sufficient to show that

$$m(t)<u(t,\epsilon),\quad t\ge t_0. \tag{3.9}$$

Suppose (3.9) were not true (for contradiction). Let

$$t_1=\inf\{t\mid m(t)\ge u(t,\epsilon),\quad t>t_0\}.$$

Then the continuity of functions $m(t)$ and $u(t,\epsilon)$ implies the following:

$$m(t)\le u(t,\epsilon),\quad t_0\le t\le t_1, \tag{3.10a}$$

$$m(t_1)=u(t_1,\epsilon),\quad t=t_1. \tag{3.10b}$$

From (3.10 a, b) and the differentiability of $m(t)$, we have

$$\begin{aligned} Dm(t_1)=D^+m(t_1)=D_-m(t_1) &\equiv \overline{\lim_{h\to 0^-}}\, h^{-1}[m(t_1)-m(t_1+h)]\\ &= \overline{\lim_{h\to 0^-}}\, h^{-1}[u(t_1,\epsilon)-m(t_1+h)]\\ &\ge \lim_{h\to 0^-} h^{-1}[u(t_1,\epsilon)-u(t_1+h,\epsilon)]\\ &= \lim_{h\to 0} h^{-1}[u(t_1,\epsilon)-u(t_1+h,\epsilon)]\\ &= \dot u(t_1,\epsilon)=g(t_1,u(t_1,\epsilon))+\epsilon\\ &>0. \end{aligned} \tag{3.11}$$

Also, from (3.10a) and (3.11), we have

$$m(t_1+s)\le u(t_1+s,\epsilon)\quad \text{for } -r\le s\le 0;$$

hence

$$m(t_1)\le \sup_{-r\le s\le 0} m(t_1+s)\le \sup_{-r\le s\le 0} u(t_1+s,\epsilon),$$

$$=u(t_1,\epsilon)=m(t_1),$$

or consequently,

$$\sup_{-r\le s\le 0} M[V(t_1+s,x(t_0,\phi_0,\eta_0)(t_1+s),\eta(t_1+s))\mid\phi_0,\eta_0] = M[V(t_1,x(t_0,\phi_0,\eta_0)(t_1),\eta(t_1))\mid\phi_0,\eta_0],$$

which also implies

$x(t_0,\phi_0,\eta_0)(t_1+s)\in\mathcal{A}_1$, by the definition of $\mathcal{A}_1$;

$$\begin{aligned} D^+m(t_1) &= \overline{\lim_{h\to 0^+}}\, h^{-1}[m(t_1+h)-m(t_1)]\\ &= \overline{\lim_{h\to 0^+}}\, h^{-1}\{M[V(t_1+h,x(t_0,\phi_0,\eta_0)(t_1+h),\eta(t_1+h))\mid\phi_0,\eta_0]\\ &\qquad - M[V(t_1,\phi_0,\eta_0)(t_1),\eta(t_1))\mid\phi_0,\eta_0]\}\\ &= \overline{\lim_{h\to 0^+}}\, h^{-1}\{M[V(t_1+h,x(t_0,\phi_0,\eta_0)(t_1+h),\eta(t_1+h))\mid\phi_0,\eta_0]\\ &- M[V(t_1+h,x(t_0,\phi_0,\eta_0)(t_1)+hf(t_1,x_{t_1}),\eta(t_1+h))\mid\phi_0,\eta_0]\\ &+ M[V(t_1+h,x(t_0,\phi_0,\eta_0)(t_1)+hf(t_1,x_{t_1}),\eta(t_1+h))\mid\phi_0,\eta_0]\\ &- M[V(t_1,x(t_0,\phi_0,\eta_0)(t_1),\eta(t_1))\mid\phi_0,\eta_0]\}\\ &\le \overline{\lim_{h\to 0^+}}\, h^{-1}\{M[V(t_1+h,x(t_0,\phi_0,\eta_0)(t_1+h),\eta(t_1+h))\mid\phi_0,\eta_0]\\ &- M[V(t_1+h,x(t_0,\phi_0,\eta_0)(t_1)+hf(t_1,x_{t_1}),\eta(t_1+h))\mid\phi_0,\eta_0]\}\\ &+ \overline{\lim_{h\to 0^+}}\, h^{-1}\{M[V(t_1+h,x(t_0,\phi_0,\eta_0)(t_1)+hf(t_1,x_{t_1}),\\ &\qquad \eta(t_1+h))\mid\phi_0,\eta_0]\\ &- M[V(t_1,x(t_0,\phi_0,\eta_0)(t_1),\eta(t_1))\mid\phi_0,\eta_0]\}. \end{aligned} \tag{3.12}$$

But, since $V(t,\phi(0),\eta(t))$ is locally Lipschitzian uniformly in $\phi(0)$, the first term in the above expression becomes

$$\begin{aligned} &\overline{\lim_{h\to 0^+}}\, h^{-1}M[V(t_1+h,x(t_0,\phi_0,\eta_0)(t_1+h),\eta(t_1+h))\\ &- V(t_1+h,x(t_0,\phi_0,\eta_0)(t_1)+hf(t_1,x_{t_1}),\eta(t_1+h))]\mid\phi_0,\eta_0]\}\\ &= \overline{\lim_{h\to 0^+}}\, h^{-1}\{M[L(t_1+h)\|(x(t_1+h)-x(t_1))-f(t_1,x_{t_1})\|\mid\phi_0,\eta_0]\} \end{aligned}$$

$$= \overline{\lim_{h\to 0^+}} M[L(t_1+h)\|h^{-1}(x(t_1+h)-x(t_1))-f(t_1, x_{t_1})\| \mid \phi_0, \eta_0]$$
$$\leq 0.$$

Hence (3.12) becomes

$$D^+m(t_1) \leq \overline{\lim_{h\to 0^+}} h^{-1}\{M[M[V(t_1+h, x(t_0, \phi_0, \eta_0)(t_1)+hf(t_1, x_{t_1}),$$
$$\eta(t_1+h))-V(t_1, x(t_0, \phi_0, \eta_0)(t_1), \eta(t_1))\,|\,x(t_1), \eta(t_1)]\,|\,\phi_0, \eta_0]\}$$
$$\leq M[\overline{\lim_{h\to 0^+}} h^{-1}[M[V(t_1+h, x(t_0, \phi_0, \eta_0)(t_1)+hf(t_1, x_{t_1}), \eta(t_1+h))$$
$$-V(t_1, x(t_0, \phi_0, \eta_0)(t_1), \eta(t_1))\,|\,x(t_1), \eta(t_1)]\,|\,\phi_0, \eta_0]$$
$$= M[D^+M[V(t_1, x(t_0, \phi_0, \eta_0)(t_1), \eta(t_1); x_{t_1})]\,|\,\phi_0, \eta_0]$$
$$\leq M[g(t_1, V(t_1, x(t_0, \phi_0, \eta_0)(t_1), \eta(t_1)))\mid \phi_0, \eta_0]$$

or

$$D^+m(t_1) \leq M[g(t_1, V(t_0, \phi_0, \eta_0)(t_1), \eta(t_1))\,|\,\phi_0, \eta_0]. \tag{3.13}$$

Since $g(t, u)$ is concave in u, we apply Jensen's inequality (see Ref. 3, p. 33), to (3.13); we have

$$D^+m(t_1) \leq g(t_1, M[V(t_1, x(t_0, \phi_0, \eta_0)(t_1), \eta(t_1))\,|\,\phi_0, \eta_0])$$
$$= g(t_1, m(t_1)) = g(t_1, u(t_1, \epsilon)), \tag{3.14}$$

which contradicts (3.11), that

$$D^+m(t_1) \geq g(t_1, u(t_1, \epsilon)) + \epsilon \quad \text{for } \epsilon > 0.$$

Thus, $m(t) < u(t, \epsilon)$, $t \geq t_0$; or equivalently

$$M[V(t, x(t_0, \phi_0, \eta_0)(t), \eta(t))\,|\,\phi_0, \eta_0] \leq r(t, t_0, u_0),$$
$$t \geq 0. \qquad \text{QED}$$

The following by-product of Theorem 3.1 will be very useful.

Theorem 3.2: Assume that the hypotheses of Theorem 3.1 hold except that the inequality (3.4) is replaced by

$$A(t)D^+M[V(t, \phi(0), \eta(t); \phi)] + M[V(t, \phi(0), \eta(t))\,|\,\phi_0, \eta_0]D^+A(t)$$
$$\leq g(t, V(t, \phi(0), \eta(t))A(t)) \quad \text{for } t \geq t_0, \quad \phi \in \mathcal{A}_2, \tag{3.15}$$

where $A(t) > 0$ is continuous on $[-r, \infty)$.

Then, $\sup_{-r\leq s\leq 0} V(t_0, \phi_0(s), \eta(t_0))A(t_0) \leq u_0$ implies the estimate

$$M[A(t)V(t, x(t_0, \phi_0, \eta_0)(t), \eta(t))\,|\,\phi_0, \eta_0] \leq r(t, t_0, u_0), \quad t \geq t_0. \tag{3.16}$$

Proof: We set

$$K(t, \phi(0), \eta(t)) = A(t)V(t, \phi(0), \eta(t)).$$

Let $\phi \in \mathcal{A}_1$. For small $h > 0$, we have

$$K(t+h, \phi(0)+hf(t, \phi), \eta(t+h)) - K(t, \phi(0), \eta(t))$$
$$= V(t+h, \phi(0)+hf(t, \phi), \eta(t+h))\ [A(t+h)-A(t)]$$
$$+ A(t)[V(t+h, \phi(0)+hf(t, \phi), \eta(t+h)) - V(t, \phi(0), \eta(t))]. \tag{3.17}$$

From (3.17) and assumption (3.15), it follows that

$$D^+M[K(t, \phi(0), \eta(t); \phi)]$$
$$= \lim_{h\to 0^+} \sup h^{-1} M[K(t+h, \phi(0)+hf(t, \phi), \eta(t+h))\,|\,\eta(t), \phi(0)]$$
$$- K(t, \phi(0), \eta(t))\}$$
$$= A(t)\ D^+M[V(t, \phi(0), \eta(t); \phi)]$$
$$+ M[V(t, \phi(0), \eta(t))\,|\,\phi_0, \eta_0]D^+A(t)$$
$$\leq g(t, K(t, \phi(0), \eta(t))) \quad \text{for } t > t_0, \quad \phi \in \tilde{\mathcal{A}}, \tag{3.18}$$

where

$$\tilde{\mathcal{A}} = \{\phi \in C \mid \sup_{-r\leq s\leq 0} M[K(t+s, \phi(s), \eta(t+s))\,|\,\eta_0, \phi_0]$$
$$= M[K(t, \phi(0), \eta(t))\,|\,\eta_0, \phi_0]\}.$$

It is clear that $K(t, \phi(0), \eta(t))$ is uniformly locally Lipschitzian in $\phi(0)$, and, thus all the assumptions of Theorem 3.1 are satisfied, with $K(t, \phi(0), \eta(t))$ in place of $V(t, \phi(0), \eta(t))$. The conclusion is now immediate from Theorem 3.1 QED

4. SUFFICIENT CONDITIONS FOR STABILITY

The following theorem provides sufficient conditions of equistable in mean for the trivial solution $x \equiv 0$ of the stochastic functional differential equation (2.1).

Theorem 4.1: Let there exist functions $V(t, \phi(0), \eta(t))$ and $g(t, u)$ having the following properties:

(a) $g \in C(R^+ \times R^+, R^+)$, $g(t, u)$ is concave in u for fixed t and $g(t, 0) = 0$.

(b) $V \in C([-r, \infty) \times R^n \times R, R^+)$, $V(t, \phi(0), \eta(t))$ is locally Lipschitzian uniformly in $\phi(0)$ and $M[V(t, x(t_0, \phi_0, \eta_0)(t), \eta(t))\,|\,\phi_0, \eta_0]$ is differentiable in t, where $x(t_0, \phi_0, \eta_0)(t)$ is a solution of (2.1).

(c) $D^+M[V(t, \phi(0), \eta(t); \phi)] \leq g(t, V(t, \phi(0), \eta(t)))$, for $t > t_0$, $\phi \in \mathcal{A}_1$.

(d) There exist functions $b(t, r)$, and $a(r)$ where $b \in \beta$, $a \in \mathcal{K}$, and a is convex such that

$$a(\|\phi(0)\|) \leq V(t, \phi(0), \eta) \leq b(t, \|\phi\|_0)$$

for $\phi \in C$ and $(t, \phi(0), \eta) \in [-r, \infty) \times R^n \times R$.

Then, the trivial solution $x \equiv 0$ of (2.1) is equistable in mean if the trivial solution $u \equiv 0$ of (2.2) is equistable.

Proof: Let $x(t_0, \phi_0, \eta_0)(t)$ be any solution of (2.1). Choose

$$u_0 = b(t_0, \|\phi_0\|_0)$$

so that

$$\sup_{-r\leq s\leq 0} V(t_0, \phi_0(s), \eta_0) \leq u_0.$$

An application of Theorem 3.1 yields

$$M[V(t, x(t_0, \phi_0, \eta_0)(t), \eta(t))\,|\,\phi_0, \eta_0] \leq r(t, t_0, u_0), \quad t \geq t_0,$$

where $r(t, t_0, u_0)$ is the maximal solution of (2.2). Also, because of assumption (d), we have

$$a(\|x(t_0, \phi_0, \eta_0)(t)\|) \leq V(t, x(t_0, \phi_0, \eta_0)(t), \eta) \text{ for } a \in \mathcal{K}. \tag{4.1}$$

Since a is strictly monotone increasing and a is convex, it follows, from (4.1), immediately that

$$a(M[\|x(t_0, \phi_0, \eta_0)(t)\|\,|\,\phi_0, \eta_0])$$
$$\leq M[V(t, x(t_0, \phi_0, \eta_0)(t), \eta(t))\,|\,\phi_0, \eta_0]. \tag{4.2}$$

Now, let $\epsilon > 0$ and $t_0 \in R^+$ be given. Assume that the trivial solution $u \equiv 0$ of (2.2) is equi-stable. Then for these given $a(\epsilon) > 0$ and $t_0 \in R^+$, there exists a $\delta = \delta(t_0, \epsilon) > 0$ satisfying

$$u(t, t_0, u_0) < b(\epsilon) \quad \text{for } t \geq t_0 \text{ provided } u_0 \leq \delta. \tag{4.3}$$

Moreover, because of the continuity of $b(t, r)$ by (d), there exists a $\delta_1 = \delta_1(t_0, \epsilon) > 0$ such that

$$\|\phi_0\|_0 \leq \delta_1 \text{ implies } b(t_0, \|\phi_0\|_0) < \delta. \tag{4.4}$$

Combination of the facts (4.2), (4.3), and (4.4) implies

$$\begin{aligned} &a(M[\|x(t_0, \phi_0, \eta_0)(t)\| \mid \phi_0, \eta_0]) \\ &\leq M[V(t, x(t_0, \phi_0, \eta_0)(t), \eta(t)) \mid \phi_0, \eta_0] \\ &\leq r(t, t_0, u_0) < a(\epsilon) \quad \text{for } t \geq t_0, \end{aligned} \tag{4.5}$$

whenever

$$\|\phi_0\|_0 < \delta_1,$$

which, by the monotonicity of a, also implies

$$M[\|x(t_0, \phi_0, \eta_0)(t)\| \mid \phi_0, \eta_0] < \epsilon, \quad t \geq t_0,$$

whenever

$$\|\phi_0\|_0 < \delta_1. \qquad \text{QED}$$

As an extension of previous theorem, we shall have the following corollary which is pertaining to the sufficient conditions of uniformly stable in mean for the trivial solution $x \equiv 0$ of (2.1).

Corollary 4.1: Assume that all the hypothesis of Theorem 4.1 hold. Furthermore, assume that $b(t, r) \equiv b(r)$. Then the trivial solution $x \equiv 0$ of (2.1) is uniformly stable in mean, if the trivial solution $u \equiv 0$ of (2.2) is uniformly stable.

Proof: Following the proof of Theorem 4.1, we obtain (4.4). Note that δ_1 in (4.4) is independent of t_0. Thus $x \equiv 0$ is uniformly stable in mean. QED

An application of one of the comparison theorems (Theorem 3.2) yield the following useful result of the sufficient conditions of equiasymptotically stable in mean.

Theorem 4.2: Assume that there exist functions $V(t, \phi(0), \eta(t))$, $g(t, u)$ and $A(t)$ satisfying the following conditions:

(a) $A(t) > 0$ is continuous on $[-r, \infty)$, and $A(t) \to \infty$ as $t \to \infty$.

(b) $V \in C([-r, \infty) \times R^n \times R, R^+)$, $V(t, \phi(0), \eta(t))$ is uniformly locally Lipschitzian in $\phi(0)$, and $M[V(t, x(t_0, \phi_0, \eta_0)(t), \eta(t)) \mid \phi_0, \eta_0]$ is differentiable in t, where $x(t_0, \phi_0, \eta_0)(t)$ is a solution of (2.1).

(c) $g \in C(R^+ \times R^+, R^+)$, $g(t, 0) \equiv 0$, and $g(t, u)$ is concave in u, also

$$\begin{aligned} &A(t)D^+M[V(t, \phi(0), \eta(t); \phi)] + V(t, \phi(0), \eta(t))D^+A(t) \\ &\leq g(t, V(t, \phi(0), \eta(t))A(t)), \end{aligned}$$

for $t > t_0$ and $\phi \in \tilde{A}$, where

$$\begin{aligned} \tilde{A} = \{\phi \in C \mid &\sup_{-r \leq s \leq 0} M[V(t+s, \phi(s), \eta(t+s))A(t+s) \mid \eta_0, \phi_0] \\ &= M[V(t, \phi(0), \eta(t))A(t) \mid \eta_0, \phi_0]\}. \end{aligned}$$

(d) There exists functions $a(r) \in K$ and $b(t, r) \in \beta$ where a is convex, such that

$$a(\|\phi(0)\|) \leq V(t, \phi(0), \eta) \leq b(t, \|\phi\|_0)$$

for $\phi \in C$ and $(t, \phi(0), \eta) \in [-r, \infty) \times R^n \times R$.

Then, the trivial solution $x \equiv 0$ of (2.1) is equistable in mean and quasi-equiasymptotically stable in mean (hence equiasymptotically stable in mean) if the trivial solution $u \equiv 0$ of (2.2) is equistable.

Proof: If $x(t_0, \phi_0, \eta_0)(t)$ is any solution of (2.1) such that

$$A(t_0)a(t_0, \|\phi_0\|_0) = u_0,$$

we have, by Theorem 3.2,

$$M[A(t)V(t, x(t_0, \phi_0, \eta_0)(t), \eta(t)) \mid \phi_0, \eta_0] \leq r(t, t_0, u_0), \quad t \geq t_0.$$

Let $\epsilon > 0$ and $t_0 \in [0, \infty)$ be given. Let

$$\alpha = \min_{-r \leq t < \infty} A(t).$$

By assumption on $A(t)$, it is clear that $\alpha > 0$.

Set $C = \alpha a(\epsilon)$. Then proceeding as in the proof of Theorem 4.1 with this C instead of $a(\epsilon)$, it is easy to prove that the trivial solution of (2.1) is equistable in mean.

To prove quasi-equiasymptotically stable in mean, given $\rho > 0$, let $C^* = \alpha a(\rho)$. Let $\delta_1(t_0, \rho)$ be such that

$$\|\phi_0\|_0 < \delta_1$$

implies

$$M[\|x(t_0, \phi_0, \eta_0)(t)\| \mid \phi_0, \eta_0] < \rho,$$

for $t \geq t_0$.

This is possible by equistability in mean. Designate $\delta_0(t_0) = \delta_1(t_0, \rho)$, and suppose that $\|\phi_0\|_0 < \delta_0$.

Since $A(t) \to \infty$ as $t \to \infty$, there exists a positive number $T = T(t_0, \epsilon)$ such that

$$A(t)a(\epsilon) > C^*, \quad t \geq t_0 + T.$$

We then have, by Theorem 3.2, (4.2), the fact that $u(t, t_0, u_0) < C^*$ if $u_0 \leq \delta(t_0, \rho)$ and $A(t) > 0$,

$$\begin{aligned} &A(t)a(M[\|x(t_0, \phi_0, \eta_0)(t)\| \mid \phi_0, \eta_0]) \\ &\leq A(t)M[V(t, x(t_0, \phi_0, \eta_0)(t), \eta(t)) \mid \phi_0, \eta_0] \\ &\leq r(t, t_0, u_0) \\ &< C^* < A(t)a(\epsilon), \quad t \geq t_0 + T. \end{aligned} \tag{4.6}$$

Then it follows, from the foregoing enequality, that

$$M[\|x(t_0, \phi_0, \eta_0)(t)\| \mid \phi_0, \eta_0] < \epsilon, \quad t \geq t_0 + T$$

provided

$$\|\phi_0\|_0 < \delta_0. \qquad \text{QED}$$

5. EXAMPLES

In this very last section of this paper, we shall consider a scalar linear functional differential equation with retardation $\eta(t, \omega)$ and demonstrate the applicability of Theorem 4.1 by using two different Lyapunov functions.

(A) Consider the stochastic functional differential equation

$$\dot{x}(t) = \lambda(t) \int_{t-\eta(t)}^{t} x(s)ds, \tag{5.1}$$

where $\lambda(t) > 0$, $0 \leq \eta(t) \leq r$, and $\lambda(t)\eta(t) \leq k(t) \in L_1[0, \infty)$.

Let

$$A_1 = \{\phi \in C \mid \sup_{-r \leq s \leq 0} V(t+s, \phi(s), \eta(t+s))$$

$$= V(t, \phi(0), \eta(t))\} \tag{5.2}$$

Now choose a Lyapunov function

$$V(t, x, \eta) = |x|. \tag{5.3}$$

Therefore

$$\dot{V}(t, x, \eta) = \frac{d}{dt}|x(t)| \leq |\dot{x}(t)|,$$

$$\leq \lambda(t) \int_{t-\eta(t)}^{t} |x(s)|\, ds,$$

$$\leq \lambda(t)\eta(t) \sup_{-r \leq s \leq 0} |x(s)|. \tag{5.4}$$

Hence, from (5.2), we have

$$\dot{V}(t, x, \eta) \leq \lambda(t)\eta(t)V(t, x, \eta)$$

$$\leq k(t)V(t, x, \eta) \quad \text{for} \quad x_t \in \mathcal{A}_1. \tag{5.5}$$

Now choose

$$g(t, u) = k(t)u.$$

Therefore, the trivial solution $u \equiv 0$ of the scalar differential equation

$$\dot{u}(t) = g(t, u(t)) \equiv k(t)u, \quad u_0 \geq 0 \tag{5.6}$$

is uniformly stable, which in turn (by Theorem 4.1 and Corollary 4.1) implies the trivial solution $x \equiv 0$ of (5.1) is uniformly stable in mean.

(B) Again, consider the same stochastic functional differential equation as (5.1)

$$\dot{x}(t) = \lambda(t) \int_{t-\eta(t)}^{t} x(s)ds, \tag{5.7}$$

where $\lambda(t) > 0$, $0 \leq \eta(t) \leq r$ and we have further assumption on $\eta(t)$ and $\lambda(t)$ such as

$$\lambda(t)\eta(t) + [D\eta(t)/\eta(t) + 1] \leq h(t) \in L_1[0, \infty), \tag{5.8}$$

where $D\eta(t) \equiv \lim_{h \to 0} M[h^{-1}(\eta(t+h) - \eta(t)) \,|\, \eta(t) = \eta]$.

Notice that $D\eta(t)$ as defined above is a function of t and η. i.e., $D\eta(t) = f(t, \eta)$. Now, choose a Lyapunov function $V(t, x, \eta)$ as

$$V(t, x, \eta) = |x|[\eta(t) + 1]. \tag{5.9}$$

Then

$$D^+M[V(t, x(t), \eta(t))]$$

$$= \frac{d}{dt}|x(t)|[\eta(t) + 1] + |x| D\eta(t)$$

$$\leq |\dot{x}(t)|[\eta(t) + 1] + |x| D\eta(t)$$

$$\leq \lambda(t)\eta(t)V(t, x, \eta) + \{D\eta(t)/[\eta(t) + 1]\}\, V(t, x, \eta)$$

$$= (\lambda(t)\eta(t) + \{D\eta(t)/[\eta(t) + 1]\})\, V(t, x, \eta)$$

$$\leq h(t)V(t, x, \eta). \tag{5.10}$$

Choose $g(t, u) = h(t)u$.

The same conclusion we have, as in (A) that the trivial solution $u \equiv 0$ of

$$\dot{u}(t) = g(t, u) \equiv h(t)u, \; u_0 \geq 0 \tag{5.11}$$

is uniformly stable.

Hence, the trivial solution $x \equiv 0$ of (5.7) is uniformly stable in mean.

[1] E.A. Lidskii, Differ. Uravn. 1, 1 (1965).
[2] V. Lakshmikantham and S. Leela, *Differential and Integral Inequalities* (Academic, New York, 1969), Vol. II.
[3] J. L. Doob, *Stochastic Process* (Wiley, New York, 1953).
[4] J. Hale, *Functional Differential Equations* (Springer-Verlag, Berlin, 1971).

Application of cumulant techniques to multiplicative stochastic processes

Ronald Forrest Fox

School of Physics, Georgia Institute of Technology, Atlanta, Georgia 30332
(Received March 29 1974)

The use of cumulant techniques for analyzing time dependent, stochastic matrix expressions of the form $\langle \underset{\sim}{T} \exp[\int_0^t \tilde{\mathbf{B}}(S)dS] \rangle$ is explained. Because cumulants are complicated expressions when $\tilde{\mathbf{B}}(t)$ does not commute with itself at unequal times, we explicitly work out cumulant expressions up to fourth order. The fourth order terms can be used to demonstrate that noncommutivity prevents the generalization, to time-dependent, stochastic matrices which do not commute with themselves at unequal times, of the result which applies to commuting stochastic processes that states: If the stochastic process is Gaussian, then its cumulant expansion truncates after the second cumulant. Furthermore, it is argued that if the stochastic matrix process is both Gaussian *and* purely random then the cumulant expansion does truncate after the second cumulant, after all. The significance of this result with respect to the application of approximation involving cumulants is mentioned.

INTRODUCTION

It is the purpose of this paper to apply cumulant techniques to the analysis of multiplicative stochastic processes.[1] When the stochastic matrices, which appear in the differential equations defining a particular multiplicative stochastic process, have special properties, the use of cumulants can lead to significant simplifications in the analysis of solutions to the differential equations. In general, when no special conditions are imposed on the properties of the stochastic matrices, cumulant expressions contain all of the information of the original problem, but without any particular advantages for further analysis. Kubo[2] has introduced cumulant techniques for the study of multiplicative stochastic processes, and he has attempted to approach the problem in as general a setting as possible. Fox[1] has studied multiplicative stochastic processes without using cumulant techniques, and has restricted his approach to the problem to the use of stochastic matrices which are purely random and Gaussian. It will be demonstrated in this paper that the restriction to purely random, Gaussian, stochastic matrices corresponds with the special conditions alluded to earlier for which cumulant techniques are most useful.

In the course of this paper we will show how the utility of cumulant techniques is related to the Gaussian property of the stochastic matrices. When the stochastic matrices are time independent, the Gaussian property alone guarantees major simplifications in analysis if cumulants are used. When the stochastic matrices are time dependent and do not commute for unequal times, then the Gaussian property alone is insufficient, and must be augmented by the purely random property if efficacy is to be achieved using cumulant methods. It is this last point that requires emphasis when one is using approximation methods which rely upon cumulants expansions.

TIME ORDERING AND CUMULANTS

Consider a multiplicative stochastic process described by[1]

$$\frac{d}{dt}a_\alpha(t) = A_{\alpha\alpha'}a_{\alpha'}(t) + \tilde{A}_{\alpha\alpha'}(t)a_{\alpha'}(t) \tag{1}$$

in which $A_{\alpha\alpha'} = -A_{\alpha'\alpha}$, $\tilde{A}_{\alpha\alpha'}(t) = -\tilde{A}_{\alpha'\alpha}(t)$, $\tilde{A}_{\alpha\alpha'}(t)$ is a stochastic matrix with averaged value zero, $\alpha = 1, 2, \ldots, N$, and repeated indices are summed. If $\tilde{A}_{\alpha\alpha'}(t)$ had a nonzero averaged value, we could include its averaged value in the term $A_{\alpha\alpha'}$ and start over again with $\tilde{A}_{\alpha\alpha'}(t) - \langle \tilde{A}_{\alpha\alpha'}(t) \rangle$ as the stochastic matrix, where $\langle \cdots \rangle$ denotes stochastic averaging. We would then have a process such as described by (1) with the average value of $\tilde{A}_{\alpha\alpha'}(t)$ equal to zero. A useful manner for obtaining the solution to (1) is to use the transformation

$$a_\alpha(t) = [\exp(t\mathbf{A})]_{\alpha\alpha'}b_{\alpha'}(t). \tag{2}$$

Consequently, the b_α's satisfy

$$\begin{aligned}\frac{d}{dt}b_\alpha(t) &= [\exp(-t\mathbf{A})]_{\alpha\beta}\tilde{A}_{\beta\beta'}(t)[\exp(t\mathbf{A})]_{\beta'\alpha'}b_{\alpha'}(t) \\ &\equiv \tilde{B}_{\alpha\alpha'}(t)b_{\alpha'}(t)\end{aligned} \tag{3}$$

where the second equality defines $\tilde{\mathbf{B}}(t)$. Note that the averaged value of $\tilde{\mathbf{B}}(t)$ is zero since it is linear in the matrix elements of $\tilde{\mathbf{A}}(t)$. In general, $\mathbf{A}$ and $\tilde{\mathbf{A}}(t)$ will not commute, neither will $\tilde{\mathbf{A}}(t)$ and $\tilde{\mathbf{A}}(s)$ nor $\tilde{\mathbf{B}}(t)$ and $\tilde{\mathbf{B}}(s)$ for $t \neq s$.

The formal solution to (3) must be written with a time ordered exponential[3] defined by

$$\begin{aligned}b_\alpha(t) &= [\underset{\sim}{T} \exp(\int_0^t \tilde{\mathbf{B}}(s)\,ds)]_{\alpha\alpha'}b_{\alpha'}(0) \\ &\equiv \sum_{n=0}^{\infty} \int_0^t \int_0^{t_1} \int_0^{t_2} \cdots \int_0^{t_{n-2}} \int_0^{t_{n-1}} [\tilde{B}_{\alpha\mu_1}(t_1)\tilde{B}_{\mu_1\mu_2}(t_2)\cdots \\ &\quad \tilde{B}_{\mu_{n-2}\mu_{n-1}}(t_{n-1})\tilde{B}_{\mu_{n-1}\alpha'}(t_n)]\,dt_n\,dt_{n-1}\cdots dt_2\,dt_1 b_{\alpha'}(0)\end{aligned} \tag{4}$$

where the $n=0$ term is defined to be 1. If $\tilde{\mathbf{B}}(t)$ and $\tilde{\mathbf{B}}(s)$ commuted for $t \neq s$, then (4) would reduce to an ordinary exponential

$$\sum_{n=0}^{\infty} \frac{1}{n!} \int_0^t \cdots n \cdots \int_0^t [\tilde{B}_{\alpha\mu_1}(S_1)\cdots\tilde{B}_{\mu_{n-1}\alpha'}(S_n)]\,dS_n\cdots dS_1 b_{\alpha'}(0).$$

If $\tilde{\mathbf{B}}(t)$ is in fact time independent, then all time integrations are easily performed giving the ordinary exponential

$$\sum_{n=0}^{\infty} \frac{1}{n!} t^n [(\tilde{B})^n]_{\alpha\alpha'}b_{\alpha'}(0).$$

Cumulants arise when one wishes to perform the stochastic average of an expression like (4) or its more simple forms given above, at the end of the preceeding

paragraph. Historically speaking, cumulants arose in statistics and were known as Thiele semi-invariants.[4]

Suppose that we have a time independent stochastic variable $\tilde{X}$ and we are interested in calculating $\langle \exp(\tilde{X}) \rangle$, where $\tilde{X}$ may be either a scalar or a matrix stochastic quantity determined by the distribution function for its values. The cumulants for $\tilde{X}$ are defined by

$$\langle \exp(\tilde{X}) \rangle = \left\langle \sum_{n=0}^{\infty} \frac{1}{n!} (\tilde{X})^n \right\rangle = \sum_{n=0}^{\infty} \frac{1}{n!} \langle (\tilde{X})^n \rangle \equiv \exp\left(\sum_{n=1}^{\infty} \frac{1}{n!} \langle (\tilde{X})^n \rangle_c \right) \tag{5}$$

where $\langle \cdots \rangle_c$ denotes the cumulant average. Cumulant averaging is by definition a homogeneous process with respect to multiplication of $\tilde{X}$ by a constant, just as is the case with ordinary averaging. Therefore, $\langle (\lambda \tilde{X})^n \rangle_c = \lambda^n \langle (\tilde{X})^n \rangle_c$, and

$$\langle \exp(\lambda \tilde{X}) \rangle = \exp\left(\sum_{n=1}^{\infty} \frac{\lambda^n}{n!} \langle (\tilde{X})^n \rangle_c \right). \tag{6}$$

Differentiating (6) n times with respect to λ, followed by setting $\lambda = 0$, gives for $n = 1, 2, 3, 4, \cdots$

$$\begin{aligned}
&\langle \tilde{X} \rangle = \langle \tilde{X} \rangle_c, \\
&\langle (\tilde{X})^2 \rangle = \langle (\tilde{X})^2 \rangle_c + \langle \tilde{X} \rangle_c^2, \\
&\langle (\tilde{X})^3 \rangle = \langle (\tilde{X})^3 \rangle_c + 3\langle (\tilde{X})^2 \rangle_c \langle \tilde{X} \rangle_c + \langle \tilde{X} \rangle_c^3, \\
&\langle (\tilde{X})^4 \rangle = \langle (\tilde{X})^4 \rangle_c + 4\langle (\tilde{X})^3 \rangle_c \langle \tilde{X} \rangle_c + 3\langle (\tilde{X})^2 \rangle_c^2 + 6\langle (\tilde{X})^2 \rangle_c \langle \tilde{X} \rangle_c^2 + \langle \tilde{X} \rangle_c^4 \\
&\quad \cdot \qquad \cdot \\
&\quad \cdot \qquad \cdot \\
&\quad \cdot \qquad \cdot
\end{aligned} \tag{7}$$

Starting with the first line, these equations may be successively inverted to give expressions for cumulant averages in terms of ordinary stochastic averages:

$$\begin{aligned}
&\langle \tilde{X} \rangle_c = \langle \tilde{X} \rangle, \\
&\langle (\tilde{X})^2 \rangle_c = \langle (\tilde{X})^2 \rangle - \langle \tilde{X} \rangle^2, \\
&\langle (\tilde{X})^3 \rangle_c = \langle (\tilde{X})^3 \rangle - 3\langle (\tilde{X})^2 \rangle \langle \tilde{X} \rangle + 2\langle \tilde{X} \rangle^3, \\
&\langle (\tilde{X})^4 \rangle_c = \langle (\tilde{X})^4 \rangle - 4\langle (\tilde{X})^3 \rangle \langle \tilde{X} \rangle - 3\langle (\tilde{X})^2 \rangle^2 + 12\langle (\tilde{X})^2 \rangle \langle \tilde{X} \rangle^2 - 6\langle \tilde{X} \rangle^4 \\
&\quad \cdot \qquad \cdot \\
&\quad \cdot \qquad \cdot \\
&\quad \cdot \qquad \cdot
\end{aligned} \tag{8}$$

It is possible to define cumulants even when we are dealing with time-dependent stochastic matrices which do not commute at unequal times.[2] One gets time-ordered cumulants defined by

$$\left\langle \underset{\leftarrow}{T} \exp\left(\int_0^t \tilde{\mathbf{B}}(S)\, dS \right) \right\rangle \equiv \underset{\leftarrow}{T} \exp\left(\sum_{n=1}^{\infty} \int_0^t \int_0^{t_1} \int_0^{t_2} \cdots \int_0^{t_{n-2}} \int_0^{t_{n-1}} \times \langle \tilde{\mathbf{B}}(t_1) \tilde{\mathbf{B}}(t_2) \cdots \tilde{\mathbf{B}}(t_{n-1}) \tilde{\mathbf{B}}(t_n) \rangle_c \times dt_n\, dt_{n-1} \cdots dt_2\, dt_1 \right). \tag{9}$$

Because it is so easy to misinterpret the precise meaning of (9), we introduce here the following notational scheme which permits unambiguous rendering of (9). Define $g_{\alpha\beta}^{(n)}(t_1)$ by

$$g_{\alpha\beta}^{(n)}(t_1) \equiv \int_0^{t_1} \int_0^{t_2} \cdots \int_0^{t_{n-2}} \int_0^{t_{n-1}} \langle \tilde{B}_{\alpha\mu_1}(t_1) \tilde{B}_{\mu_1\mu_2}(t_2) \cdots \tilde{B}_{\mu_{n-2}\mu_{n-1}}(t_{n-1}) \tilde{B}_{\mu_{n-1}\beta}(t_n) \rangle_c\, dt_n\, dt_{n-1} \cdots dt_3\, dt_2. \tag{10}$$

Therefore,

$$\left\langle \underset{\leftarrow}{T} \exp\left(\int_0^t \tilde{\mathbf{B}}(S)\, dS \right) \right\rangle = \underset{\leftarrow}{T} \exp\left(\sum_{n=1}^{\infty} \int_0^t \mathbf{g}^{(n)}(t_1)\, dt_1 \right). \tag{11}$$

As before, we can introduce a parameter λ, and using (10) leads to

$$\left\langle \underset{\leftarrow}{T} \exp\left(\int_0^t \lambda \tilde{\mathbf{B}}(S)\, dS \right) \right\rangle = \underset{\leftarrow}{T} \exp\left(\sum_{n=1}^{\infty} \int_0^t \lambda^n \mathbf{g}^{(n)}(t_1)\, dt_1 \right). \tag{12}$$

Differentiating (12) with respect to λ, n times, followed by setting $\lambda = 0$, gives for $n = 1, 2, 3, 4, \cdots$

$$\int_0^t \langle \tilde{B}_{\alpha\beta}(t_1) \rangle\, dt_1 = \int_0^t g_{\alpha\beta}^{(1)}(t_1)\, dt_1 = \int_0^t \langle \tilde{B}_{\alpha\beta}(t_1) \rangle_c\, dt_1,$$

$$\begin{aligned}
\int_0^t \int_0^{t_1} \langle \tilde{B}_{\alpha\mu}(t_1) \tilde{B}_{\mu\beta}(t_2) \rangle\, dt_2\, dt_1 &= \int_0^t g_{\alpha\beta}^{(2)}(t_1)\, dt_1 + \int_0^t \int_0^{t_1} g_{\alpha\mu}^{(1)}(t_1) g_{\mu\beta}^{(1)}(t_2)\, dt_2\, dt_1 \\
&= \int_0^t \int_0^{t_1} \langle \tilde{B}_{\alpha\mu}(t_1) \tilde{B}_{\mu\beta}(t_2) \rangle_c\, dt_2\, dt_1 + \int_0^t \int_0^{t_1} \langle \tilde{B}_{\alpha\mu}(t_1) \rangle_c \langle \tilde{B}_{\mu\beta}(t_2) \rangle_c\, dt_2\, dt_1,
\end{aligned}$$

$$\begin{aligned}
\int_0^t \int_0^{t_1} \int_0^{t_2} &\langle \tilde{B}_{\alpha\mu_1}(t_1) \tilde{B}_{\mu_1\mu_2}(t_2) \tilde{B}_{\mu_2\beta}(t_3) \rangle\, dt_3\, dt_2\, dt_1 \\
&= \int_0^t g_{\alpha\mu}^{(3)}(t_1)\, dt_1 + \int_0^t \int_0^{t_1} g_{\alpha\mu}^{(2)}(t_1) g_{\mu\beta}^{(1)}(t_2) + \int_0^t \int_0^{t_1} g_{\alpha\mu}^{(1)}(t_1) g_{\mu\beta}^{(2)}(t_2)\, dt_2\, dt_1 \\
&\quad + \int_0^t \int_0^{t_1} \int_0^{t_2} g_{\alpha\mu_1}^{(1)}(t_1) g_{\mu_1\mu_2}^{(1)}(t_2) g_{\mu_2\beta}^{(1)}(t_3)\, dt_3\, dt_2\, dt_1 \\
&= \int_0^t \int_0^{t_1} \int_0^{t_2} \langle \tilde{B}_{\alpha\mu_1}(t_1) \tilde{B}_{\mu_1\mu_2}(t_2) \tilde{B}_{\mu_2\beta}(t_3) \rangle_c\, dt_3\, dt_2\, dt_1 + \int_0^t \int_0^{t_1} \int_0^{t_1} \langle \tilde{B}_{\alpha\mu_1}(t_1) \tilde{B}_{\mu_1\mu_2}(S_2) \rangle_c \langle \tilde{B}_{\mu_2\beta}(t_2) \rangle_c\, dt_2\, dS_2\, dt_1 \\
&\quad + \int_0^t \int_0^{t_1} \int_0^{t_2} \langle B_{\alpha\mu_1}(t_1) \rangle_c \langle B_{\mu_1\mu_2}(t_2) B_{\mu_2\beta}(t_3) \rangle_c\, dt_3\, dt_2\, dt_1 + \int^t \int^{t_1} \int^{t_2} \langle B_{\alpha\mu_1}(t_1) \rangle_c \langle B_{\mu_1\mu_2}(t_2) \rangle_c \langle B_{\mu_2\beta}(t_3) \rangle_c\, dt_3\, dt_2\, dt_1,
\end{aligned}$$

$$\begin{aligned}
\int_0^t \int_0^{t_1} \int_0^{t_2} \int_0^{t_3} &\langle \tilde{B}_{\alpha\mu_1}(t_1) \tilde{B}_{\mu_1\mu_2}(t_2) \tilde{B}_{\mu_2\mu_3}(t_3) \tilde{B}_{\mu_3\beta}(t_4) \rangle\, dt_4\, dt_3\, dt_2\, dt_1 \\
&= \int_0^t g_{\alpha\beta}^{(4)}(t_1)\, dt_1 + \int_0^t \int_0^{t_1} g_{\alpha\mu}^{(3)}(t_1) g_{\mu\beta}^{(1)}(t_2)\, dt_2\, dt_1 + \int_0^t \int_0^{t_1} g_{\alpha\mu}^{(1)}(t_1) g_{\mu\beta}^{(3)}(t_2)\, dt_2\, dt_1 \\
&\quad + \int_0^t \int_0^{t_1} g_{\alpha\mu}^{(2)}(t_1) g_{\mu\beta}^{(2)}(t_2)\, dt_2\, dt_1 + \int_0^t \int_0^{t_1} \int_0^{t_2} g_{\alpha\mu_1}^{(2)}(t_1) g_{\mu_1\mu_2}^{(1)}(t_2) g_{\mu_2\beta}^{(1)}(t_3)\, dt_3\, dt_2\, dt_1 \\
&\quad + \int_0^t \int_0^{t_1} \int_0^{t_2} g_{\alpha\mu_1}^{(1)}(t_1) g_{\mu_1\mu_2}^{(2)}(t_2) g_{\mu_2\beta}^{(1)}(t_3)\, dt_3\, dt_2\, dt_1 + \int_0^t \int_0^{t_1} \int_0^{t_2} g_{\alpha\mu_1}^{(1)}(t_1) g_{\mu_1\mu_2}^{(1)}(t_2) g_{\mu_2\beta}^{(2)}(t_3)\, dt_3\, dt_2\, dt_1 \\
&\quad + \int_0^t \int_0^{t_1} \int_0^{t_2} \int_0^{t_3} g_{\alpha\mu_1}^{(1)}(t_1) g_{\mu_1\mu_2}^{(1)}(t_2) g_{\mu_2\mu_3}^{(1)}(t_3) g_{\mu_3\beta}^{(1)}(t_4)\, dt_4\, dt_3\, dt_2\, dt_1 \\
&= \int_0^t \int_0^{t_1} \int_0^{t_2} \int_0^{t_3} \langle \tilde{B}_{\alpha\mu_1}(t_1) \tilde{B}_{\mu_1\mu_2}(t_2) \tilde{B}_{\mu_2\mu_3}(t_3) \tilde{B}_{\mu_3\beta}(t_4) \rangle_c\, dt_4\, dt_3\, dt_2\, dt_1
\end{aligned}$$

$$+\int_0^t\int_0^{t_1}\int_0^{s_2}\int_0^{t_1}\langle\tilde{B}_{\alpha\mu_1}(t_1)\tilde{B}_{\mu_1\mu_2}(s_2)\tilde{B}_{\mu_2\mu_3}(s_3)\rangle_c\langle\tilde{B}_{\mu_3\beta}(t_2)\rangle_c\,dt_2\,ds_3\,ds_2\,dt_1$$
$$+\int_0^t\int_0^{t_1}\int_0^{t_2}\int_0^{t_3}\langle\tilde{B}_{\alpha\mu_1}(t_1)\rangle_c\langle\tilde{B}_{\mu_1\mu_2}(t_2)\tilde{B}_{\mu_2\mu_3}(t_3)\tilde{B}_{\mu_3\beta}(t_4)\rangle_c\,dt_4\,dt_3\,dt_2\,dt_1$$
$$+\int_0^t\int_0^{t_1}\int_0^{t_1}\int_0^{t_2}\langle\tilde{B}_{\alpha\mu_1}(t_1)\tilde{B}_{\mu_1\mu_2}(s_2)\rangle_c\langle\tilde{B}_{\mu_2\mu_3}(t_2)\tilde{B}_{\mu_3\beta}(s_3)\rangle_c\,ds_3\,dt_2\,ds_2\,dt_1$$
$$+\int_0^t\int_0^{t_1}\int_0^{t_1}\int_0^{t_2}\langle\tilde{B}_{\alpha\mu_1}(t_1)\tilde{B}_{\mu_1\mu_2}(s_2)\rangle_c\langle\tilde{B}_{\mu_2\mu_3}(t_2)\rangle_c\langle\tilde{B}_{\mu_3\beta}(t_3)\rangle_c\,dt_3\,dt_2\,ds_2\,dt_1$$
$$+\int_0^t\int_0^{t_1}\int_0^{t_2}\int_0^{t_2}\langle\tilde{B}_{\alpha\mu_1}(t_1)\rangle_c\langle\tilde{B}_{\mu_1\mu_2}(t_2)\tilde{B}_{\mu_2\mu_3}(s_3)\rangle_c\langle\tilde{B}_{\mu_3\beta}(t_3)\rangle_c\,dt_3\,ds_3\,dt_2\,dt_1$$
$$+\int_0^t\int_0^{t_1}\int_0^{t_2}\int_0^{t_3}\langle\tilde{B}_{\alpha\mu_1}(t_1)\rangle_c\langle\tilde{B}_{\mu_1\mu_2}(t_2)\rangle_c\langle\tilde{B}_{\mu_2\mu_3}(t_3)\tilde{B}_{\mu_3\beta}(t_4)\rangle_c\,dt_4\,dt_3\,dt_2\,dt_1$$
$$+\int_0^t\int_0^{t_1}\int_0^{t_2}\int_0^{t_3}\langle\tilde{B}_{\alpha\mu_1}(t_1)\rangle_c\langle\tilde{B}_{\mu_1\mu_2}(t_2)\rangle_c\langle\tilde{B}_{\mu_2\mu_3}(t_3)\rangle_c\langle\tilde{B}_{\mu_3\beta}(t_4)\rangle_c\,dt_4\,dt_3\,dt_2\,dt_1. \tag{13}$$

Especially notice the limits of integration in several of the terms which contribute to the third and fourth order expressions. By successively inverting these equations we get expressions for the cumulants in terms of the ordinary stochastic averages:

$$\int_0^t\langle\tilde{B}_{\alpha\beta}(t_1)_c\,dt_1=\int_0^t\langle\tilde{B}_{\alpha\beta}(t_1)\rangle\,dt_1,$$
$$\int_0^t\int_0^{t_1}\langle\tilde{B}_{\alpha\mu}(t_1)\tilde{B}_{\mu\beta}(t_2)\rangle_c\,dt_2\,dt_1=\int_0^t\int_0^{t_1}\langle\tilde{B}_{\alpha\mu}(t_1)\tilde{B}_{\mu\beta}(t_2)\rangle\,dt_2\,dt_1-\int_0^t\int_0^{t_1}\langle\tilde{B}_{\alpha\mu}(t_1)\rangle\langle\tilde{B}_{\mu\beta}(t_2)\rangle\,dt_2\,dt_1,$$
$$\int_0^t\int_0^{t_1}\int_0^{t_2}\langle\tilde{B}_{\alpha\mu_1}(t_1)\tilde{B}_{\mu_1\mu_2}(t_2)\tilde{B}_{\mu_2\beta}(t_3)\rangle_c\,dt_3\,dt_2\,dt_1$$
$$=\int_0^t\int_0^{t_1}\int_0^{t_2}\langle\tilde{B}_{\alpha\mu_1}(t_1)\tilde{B}_{\mu_1\mu_2}(t_2)\tilde{B}_{\mu_2\beta}(t_3)\rangle\,dt_3\,dt_2\,dt_1-\int_0^t\int_0^{t_1}\int_0^{t_1}\langle\tilde{B}_{\alpha\mu_1}(t_1)\tilde{B}_{\mu_1\mu_2}(S_2)\rangle\langle\tilde{B}_{\mu_2\beta}(t_2)\rangle\,dt_2\,dS_2\,dt_1$$
$$-\int_0^t\int_0^{t_1}\int_0^{t_2}\langle\tilde{B}_{\alpha\mu_1}(t)\rangle\langle\tilde{B}_{\mu_1\mu_2}(t_2)\tilde{B}_{\mu_2\beta}(t_3)\rangle\,dt_3\,dt_2\,dt_1+\int_0^t\int_0^{t_1}\int_0^{t_1}\langle\tilde{B}_{\alpha\mu_1}(t_1)\rangle\langle\tilde{B}_{\mu_1\mu_2}(t_2)\rangle\langle\tilde{B}_{\mu_2\beta}(t_3)\rangle\,dt_3\,dt_2\,dt_1,$$
$$\int_0^t\int_0^{t_1}\int_0^{t_2}\int_0^{t_3}\langle\tilde{B}_{\alpha\mu_1}(t_1)\tilde{B}_{\mu_1\mu_2}(t_2)\tilde{B}_{\mu_2\mu_3}(t_3)\tilde{B}_{\mu_3\beta}(t_4)\rangle_c\,dt_4\,dt_3\,dt_2\,dt_1$$
$$=\int_0^t\int_0^{t_1}\int_0^{t_2}\int_0^{t_3}\langle\tilde{B}_{\alpha\mu_1}(t_1)\tilde{B}_{\mu_1\mu_2}(t_2)\tilde{B}_{\mu_2\mu_3}(t_3)\tilde{B}_{\mu_3\beta}(t_4)\rangle\,dt_4\,dt_3\,dt_2\,dt_1$$
$$-\int_0^t\int_0^{t_1}\int_0^{s_2}\int_0^{t_1}\langle\tilde{B}_{\alpha\mu_1}(t_1)\tilde{B}_{\mu_1\mu_2}(S_2)\tilde{B}_{\mu_2\mu_3}(S_3)\rangle\langle\tilde{B}_{\mu_3\beta}(t_2)\rangle\,dt_2\,dS_3\,dS_2\,dt_1$$
$$+\int_0^t\int_0^{t_1}\int_0^{t_1}\int_0^{t_1}\langle\tilde{B}_{\alpha\mu_1}(t_1)\tilde{B}_{\mu_1\mu_2}(S_2)\rangle\langle\tilde{B}_{\mu_2\mu_3}(S_3)\rangle\langle\tilde{B}_{\mu_3\beta}(t_2)\rangle\,dt_2\,dS_3\,dS_2\,dt_1$$
$$+\int_0^t\int_0^{t_1}\int_0^{s_2}\int_0^{t_1}\langle\tilde{B}_{\alpha\mu_1}(t_1)\rangle\langle\tilde{B}_{\mu_1\mu_2}(S_2)\tilde{B}_{\mu_2\mu_3}(S_3)\rangle\langle\tilde{B}_{\mu_3\beta}(t_2)\rangle\,dt_2\,dS_3\,dS_2\,dt_1$$
$$-\int_0^t\int_0^{t_1}\int_0^{t_1}\int_0^{t_1}\langle\tilde{B}_{\alpha\mu_1}(t_1)\rangle\langle\tilde{B}_{\mu_1\mu_2}(S_2)\rangle\langle\tilde{B}_{\mu_2\mu_3}(S_3)\rangle\langle\tilde{B}_{\mu_3\beta}(t_2)\rangle\,dt_2\,dS_3\,dS_2\,dt_1$$
$$-\int_0^t\int_0^{t_1}\int_0^{t_2}\int_0^{t_3}\langle\tilde{B}_{\alpha\mu_1}(t_1)\rangle\langle\tilde{B}_{\mu_1\mu_2}(t_2)\tilde{B}_{\mu_2\mu_3}(t_3)\tilde{B}_{\mu_3\beta}(t_4)\rangle\,dt_4\,dt_3\,dt_2\,dt_1$$
$$-\int_0^t\int_0^{t_1}\int_0^{t_1}\int_0^{t_2}\langle\tilde{B}_{\alpha\mu_1}(t_1)\tilde{B}_{\mu_1\mu_2}(S_2)\rangle\langle\tilde{B}_{\mu_2\mu_3}(t_2)\tilde{B}_{\mu_3\beta}(S_3)\rangle\,dS_3\,dt_2\,dS_2\,dt_1$$
$$+\int_0^t\int_0^{t_1}\int_0^{t_1}\int_0^{t_2}\langle\tilde{B}_{\alpha\mu_1}(t_1)\rangle\langle\tilde{B}_{\mu_1\mu_2}(S_2)\rangle\langle\tilde{B}_{\mu_2\mu_3}(t_2)\tilde{B}_{\mu_3\beta}(S_3)\rangle\,dS_3\,dt_2\,dS_2\,dt_1. \tag{14}$$

Again, especially notice the limits of integration in several of the terms which contribute to the third and fourth order expressions. By studying the special cases in which either $\tilde{\mathbf{B}}(t)$ commutes for unequal times, or $\tilde{\mathbf{B}}$ is time independent, it may be shown that (14) reduces to (8), and (13) reduces to (7). It should by now be clear how to obtain the higher order cumulants for $\tilde{\mathbf{B}}(t)$ when it does not commute with itself for unequal times, even though the expressions become quite complicated.

THE GAUSSIAN PROPERTY

Throughout this discussion of the Gaussian property, it is convenient to invoke the condition that the averaged value of $\tilde{\mathbf{B}}(t)$ is zero. This simplifies considerations without loss of generality, as was explained at the beginning of the last section. The analogous condition for $\tilde{X}$ is that $\langle\tilde{X}\rangle=0$.

Suppose that $\tilde{X}$ is Gaussian with averaged value zero. Then it is known that the moments of $\tilde{X}$ satisfy[5]

$$\langle(\tilde{X})^{2m+1}\rangle=0 \quad \text{for } m=0,1,2,\cdots, \tag{15}$$
$$\langle(\tilde{X})^{2m}\rangle=1\times3\times5\times\cdots\times(2m-1)\langle(\tilde{X})^2\rangle^m \quad \text{for } m=1,2,\cdots.$$

If we return to (8) we see that

$$\langle\tilde{X}\rangle_c=0, \quad \langle(\tilde{X})^2\rangle_c=\langle(\tilde{X})^2\rangle, \tag{16}$$
$$\langle(\tilde{X})^3\rangle_c=0, \quad \langle(\tilde{X})^4\rangle_c=0.$$

The last result follows from the cancellation of the two nonvanishing terms: $\langle(\tilde{X})^4\rangle$ and $-3\langle(\tilde{X})^2\rangle^2$, as may be seen using the second equation in (15). It may be proved generally that all cumulants of order higher than four also vanish. The proof may be performed by the method of induction. We assume it is true that except for the nonvanishing, second order cumulant, all cumulants up to and including order $2m$ for $m>2$ vanish. We shall now show that cumulants of order $2m+1$ and $2m+2$ also vanish.

The cumulant of order $2m+1$ will be expressible as an expansion in terms of the moments of $\tilde{X}$ up to order $2m+1$ as is indicated by (8). The leading term will be $\langle(\tilde{X})^{2m+1}\rangle$ which is zero according to (15). All other terms in the expansion will involve products of lower order moments in which at least one factor is an odd order, lower than $2m+1$ order moment. By (15), such moments vanish and we conclude that $\langle(\tilde{X})^{2m+1}\rangle_c=0$.

The consideration of the cumulant of order $2m+2$ is best pursued using the relationships given by (7). These relationships show that $\langle(\tilde{X})^{2m+2}\rangle$ may be expressed in terms of cumulants of order less than or equal to $2m+2$. The Gaussian properties given by (15) show that $\langle(\tilde{X})^{2m+2}\rangle$ is nonvanishing. Its expansion, according to

(7), in terms of cumulants begins with $\langle(\tilde{X})^{2m+2}\rangle_c$, which we wish to compute, and involves other terms, each of which is a product of lower order cumulants. By hypothesis, all lower order cumulant, except the second order and $2m+1$ order cumulants, vanish, and we have already seen that the cumulant of order $2m+1$ also vanishes. Therefore, all products vanish except the single term which is the product of $m+1$ second order cumulants. Therefore, it follows that $\langle(\tilde{X})^{2m+2}\rangle = \langle(\tilde{X})^{2m+2}\rangle_c + C\langle(\tilde{X})^2\rangle_c^{m+1}$ where C is a coefficient which has to be determined. Looking at (6), it may be seen that the $\langle(\tilde{X})^2\rangle_c^{m+1}$ term comes from the $m+1$ term in the expansion of the exponential on the right-hand side of (6). Such a term has a factor of $1/(m+1)!$ associated with it. Each factor of $\langle(\tilde{X})^2\rangle_c$ has associated with it a factor $1/2!$ as is directly evident in (6). Therefore, $\langle(\tilde{X})^2\rangle_c^{m+1}$ will have a factor, overall, of $[1/(m+1)!](1/2!)^{m+1}$. This must be multiplied by $(2m+2)!$ to get C as a result of the $2m+2$ fold differentiation of λ^{2m+2}. Consequently, $C=(2m+2)!/(m+1)!\,2^{m+1}$. Using (15), it is seen that $\langle(\tilde{X})^{2m+2}\rangle = 1\times3\times5\times\cdots\times(2m+1)\langle(\tilde{X})^2\rangle^{m+1}$. Therefore,

$$\begin{aligned}\langle(\tilde{X})^{2m+2}\rangle_c &= 1\times3\times5\times\cdots\times(2m+1)\langle(\tilde{X})^2\rangle^{m+1} \\ &\quad - \frac{(2m+2)!}{(m+1)!\,2^{m+1}}\langle(\tilde{X})^2\rangle_c^{m+1} \\ &= 0 \end{aligned} \tag{17}$$

since $\langle(\tilde{X})^2\rangle = \langle(\tilde{X})^2\rangle_c$. This completes the proof.

The conclusion is that if $\tilde{X}$ is Gaussian, with averaged value zero, then

$$\langle\exp[\tilde{X}]\rangle = \exp[\tfrac{1}{2}\langle(\tilde{X})^2\rangle] \tag{18}$$

which is especially simple. The cumulant expansion reduces to a single term!

Before attempting to generalize this result for consideration of stochastic matrices which do not commute for unequal times, it is worthwhile to observe that if $\tilde{X}$ is replaced by a scalar, time dependent stochastic quantity, $\tilde{F}(t)$, then the Gaussian property again leads to great simplifications. We make this digression in order to point out that time dependence, *per se*, does not lead to complications. The complications which will arise in our discussion of stochastic matrices, which depend upon time but which do not commute for unequal times, come from the noncommutivity.

Suppose $\tilde{F}(t)$ has averaged value zero, and is Gaussian. It can then be shown that[6]

$$\langle\exp[\textstyle\int_0^t \tilde{F}(S)\,dS]\rangle = \exp[\tfrac{1}{2}\int_0^t\int_0^t \langle\tilde{F}(S_1)\tilde{F}(S_2)\rangle\,dS_1\,dS_2]. \tag{19}$$

Equation (19) is the time dependent generalization of (18) for a time-dependent, scalar, stochastic process. Again, the cumulant expansion reduces to a single term.

The generalization of this result to time-dependent stochastic matrices which do not commute at unequal times cannot be made. The Gaussian property for $\tilde{\mathbf{B}}(t)$ is expressed by[1]

$$\langle\tilde{B}_{\mu_1\nu_1}(t_1)\cdots\tilde{B}_{\mu_{2m+1}\nu_{2m+1}}(t_{2m+1})\rangle = 0 \quad \text{for } m=0,1,2,\cdots, \tag{20}$$

$$\langle\tilde{B}_{\mu_1\nu_1}(t_1)\cdots\tilde{B}_{\mu_{2m}\nu_{2m}}(t_{2m})\rangle = \frac{1}{2^m m!} \times \sum_{p\in S_{2m}}\prod_{j=1}^{m}\langle\tilde{B}_{\mu_{p(2j-1)}\nu_{p(2j-1)}}(t_{p(2j-1)})\tilde{B}_{\mu_{p(2j)}\nu_{p(2j)}}(t_{p(2j)})\rangle$$
$$\text{for } m=1,2,\cdots,$$

where $\sum_{p\in S_{2m}}$ is the sum over all permutations p in the symmetric group of order $(2m)!$, S_{2m}. Even though we can think of a Gaussian $\tilde{\mathbf{B}}(t)$ in the sense of (20), the Gaussian property alone will prove insufficient for the reduction of (9) to a form analogous with (18) or (19). It is necessary that $\tilde{\mathbf{B}}(t)$ be Gaussian *and* purely random in order to write

$$\langle\underset{\leftarrow}{T}\exp[\textstyle\int_0^t \tilde{\mathbf{B}}(S)\,dS]\rangle = \underset{\leftarrow}{T}\exp[\int_0^t\int_0^{t_1}\langle\tilde{\mathbf{B}}(t_1)\tilde{\mathbf{B}}(t_2)\rangle\,dt_2\,dt_1]. \tag{21}$$

The purely random property is defined by[1]

$$\langle\tilde{B}_{\alpha\beta}(t)\tilde{B}_{\mu\nu}(S)\rangle = 2Q_{\alpha\beta\mu\nu}\delta(t-S) \tag{22}$$

in which the time delta function characterizes the purely random property. Defining $R_{\alpha\beta}$ by $R_{\alpha\beta}\equiv Q_{\alpha\theta\theta\beta}$, wherein summation over θ is implicit, Eq. (21) becomes

$$\langle\underset{\leftarrow}{T}\exp[\textstyle\int_0^t \tilde{\mathbf{B}}(S)\,dS]\rangle_{\alpha\beta} = [\exp(\mathbf{R}t)]_{\alpha\beta} \tag{23}$$

which is especially simple. We shall now proceed to demonstrate why Gaussianness alone is insufficient for the justification of (23) or (21).

That Gaussianness alone does not result in the vanishing of all but the second order cumulants of $\tilde{\mathbf{B}}(t)$ may be seen by studying (13) and (14). First of all, because $\langle\tilde{\mathbf{B}}(t)\rangle = 0$ is assumed, the expressions in (13) and (14) simplify greatly. We get immediately

$$\begin{aligned}&\int_0^t\langle\tilde{B}_{\alpha\beta}(t_1)\rangle_c\,dt_1 = 0, \\ &\int_0^t\int_0^{t_1}\langle\tilde{B}_{\alpha\mu}(t_1)\tilde{B}_{\mu\beta}(t_2)\rangle_c\,dt_2\,dt_1 \\ &= \int_0^t\int_0^{t_1}\langle\tilde{B}_{\alpha\mu}(t_1)\tilde{B}_{\mu\beta}(t_2)\rangle\,dt_2\,dt_1, \\ &\int_0^t\int_0^{t_1}\int_0^{t_2}\langle\tilde{B}_{\alpha\mu_1}(t_1)\tilde{B}_{\mu_1\mu_2}(t_2)\tilde{B}_{\mu_2\beta}(t_3)\rangle_c\,dt_3\,dt_2\,dt_1 = 0.\end{aligned} \tag{24}$$

Differentiation of these equations with respect to time leads to equations for the integrands. Difficulty arises with the fourth order term. From (14) we get

$$\begin{aligned}&\int_0^t\int_0^{t_1}\int_0^{t_2}\int_0^{t_3}\langle\tilde{B}_{\alpha\mu_1}(t_1)\tilde{B}_{\mu_1\mu_2}(t_2)\tilde{B}_{\mu_2\mu_3}(t_3)\tilde{B}_{\mu_3\beta}(t_4)\rangle_c \\ &\quad\times dt_4\,dt_3\,dt_2\,dt_1 \\ &= \int_0^t\int_0^{t_1}\int_0^{t_2}\int_0^{t_3}\langle\tilde{B}_{\alpha\mu_1}(t_1)\tilde{B}_{\mu_1\mu_2}(t_2)\tilde{B}_{\mu_2\mu_3}(t_3)\tilde{B}_{\mu_3\beta}(t_4)\rangle \\ &\quad\times dt_4\,dt_3\,dt_2\,dt_1 \\ &\quad - \int_0^t\int_0^{t_1}\int_0^{t_1}\int_0^{t_2}\langle\tilde{B}_{\alpha\mu_1}(t_1)\tilde{B}_{\mu_1\mu_2}(S_2)\rangle \\ &\quad\times\langle\tilde{B}_{\mu_2\mu_3}(t_2)\tilde{B}_{\mu_3\beta}(S_3)\rangle\,dS_3\,dt_2\,dS_2\,dt_1.\end{aligned} \tag{25}$$

It can be seen that this is nonvanishing if $\tilde{\mathbf{B}}(t)$ is not purely random by considering the special case in which

$$\langle\tilde{B}_{\alpha\beta}(t)\tilde{B}_{\mu\nu}(S)\rangle = 2Q_{\alpha\beta\mu\nu}\varphi(t-S), \tag{26}$$

wherein $\varphi(t-S)$ is not a delta function, but instead is nonvanishing for nonzero values of $|t-S|$. For example, $\varphi(t-S)$ may be proportional to $\exp(-\alpha|t-S|)$. Using the two symmetries $Q_{\alpha\beta\mu\nu}=Q_{\mu\nu\alpha\beta}$ and $\varphi(t-S)=\varphi(S-t)$, the Gaussian property, as expressed by (20), leads to the following expression for (25):

$$\begin{aligned}&\int_0^t\int_0^{t_1}\int_0^{t_2}\int_0^{t_3}\langle\tilde{B}_{\alpha\mu_1}(t_1)\tilde{B}_{\mu_1\mu_2}(t_2)\tilde{B}_{\mu_2\mu_3}(t_3)\tilde{B}_{\mu_3\beta}(t_4)\rangle_c \\ &\quad\times dt_4\,dt_3\,dt_2\,dt_1\end{aligned}$$

$$= 4\int_0^t \int_0^{t_1} \int_0^{t_2} \int_0^{t_3} \{Q_{\alpha\mu_1\mu_1\mu_2} Q_{\mu_2\mu_3\mu_3\beta} \varphi(t_1 - t_2)\varphi(t_3 - t_4)$$
$$+ Q_{\alpha\mu_1\mu_2\mu_3} Q_{\mu_1\mu_2\mu_3\beta} \varphi(t_1 - t_3)\varphi(t_2 - t_4)$$
$$+ Q_{\alpha\mu_1\mu_3\beta} Q_{\mu_1\mu_2\mu_2\mu_3} \varphi(t_1 - t_4)\varphi(t_2 - t_3)\} dt_4\, dt_3\, dt_2\, dt_1$$
$$- 4\int_0^t \int_0^{t_1} \int_0^{t_1} \int_0^{t_2} Q_{\alpha\mu_1\mu_1\mu_2} Q_{\mu_2\mu_3\mu_3\beta} \varphi(t_1 - S_2)\varphi(t_2 - S_3)$$
$$\times dS_3\, dt_2\, dS_2\, dt_1. \tag{27}$$

The factors of 4 come from 2^2 as indicated by (26). The negative term is the negative term in (25) replaced by (26). The three preceeding positive terms are the $m = 2$ case of the second equation in (20). There are 24 permutations to sum over, but only three distinct terms occur as a result of the two symmetries mentioned above. Only the first positive term in the right-hand side of (27) contains a matrix expression which matches the matrix expression of the negative term. The other two positive terms cannot be cancelled out at all! However, if $\varphi(t - S) \equiv \delta(t - S)$ then study of these extra two terms will show that their integrals vanish as a result of the time-ordered upper limits of integration,[7] and the remaining positive and negative two terms will exactly cancel! When $\varphi(t - S) \neq \delta(t - S)$, the above vanishing of integrals and cancellation of nonvanishing integrals fails to occur.

That the purely random property ultimately leads to (23) will not be proved here since it has been proved in another paper.[1] However, here it has been shown that without the purely random property, already the fourth order cumulant will not vanish for time dependent stochastic matrices which do not commute at unequal times. Consequently, the cumulant expansion will not truncate to a single term, but will involve all even order cumulants. Thus, no real simplification is achieved using cumulants.

APPROXIMATION PROCEDURES

Various physical problems involve computation of a quantity of the form

$$\langle \underset{\sim}{T} \exp[\int_0^t \mathbf{O}(S)\, dS] \rangle, \tag{28}$$

wherein $\mathbf{O}(S)$ is an operator parameterized by S, which may or may not be the time variable, and $\langle \cdots \rangle$ denotes some kind of averaging such as a canonical average over some or all of the variables $\mathbf{O}(S)$ depends upon, or a ground state expectation value, or some other "averaging." Many times one sees in the literature the introduction of cumulant techniques in order to handle the computation of (28). Often it is argued that to good approximation $\mathbf{O}(S)$ behaves as if it were Gaussian, and then the cumulant expansion is truncated after the second cumulant. However, the validity of such approximations also requires that it be demonstrated that to good approximation $\mathbf{O}(S)$ behaves as if it were also purely random. Then, and only then, the truncation of the cumulant expansion is a good approximation.

One application of this kind of approximation procedure in which both Gaussianness and pure randomness were considered has been presented by the author in another paper.[6]

ACKNOWLEDGMENT

The work was partially supported by NSF grant GP39309.

[1] R. F. Fox, J. Math Phys. 13, 1196 (1972).
[2] R. Kubo, "A Stochastic Theory of Line Shape and Relaxation," in *Fluctuation, Relaxation and Resonance in Magnetic Systems*, edited by ter Haar (Oliver and Boyd, Edinburgh, 1962), pp. 23–68.
[3] J. J. Sakurai, *Advanced Quantum Mechanics* (Addison-Wesley, Reading, Mass., 1967), Sec. 4-2.
[4] T. N. Thiele, *Theory of Observations* (London, 1903).
[5] M. C. Wang and G. E. Uhlenbeck, Rev. Mod. Phys. 17, 323 (1945), note III of Appendix.
[6] R. F. Fox, J. Math. Phys. (to be published).
[7] R. F. Fox, Ref. 1, Appendix

A classification of second-order raising operators for Hamiltonians in two variables

Charles P. Boyer* and Willard Miller Jr.†

Centre de recherches mathématiques, Université de Montréal, Montréal 101, P.Q., Canada
(Received 11 February 1974)

We develop a group theoretic method based on results of Winternitz *et al.* to compute and classify all first- and second-order raising and lowering operators admitted by Hamiltonians of the form $\underline{H} = -(1/2)\Delta_2 + V(x, y)$. The key to our results, which generalize to higher dimensions, is a proof that $\underline{H}$ admits a second-order raising operator only if the Schrödinger equation separates in Cartesian, polar, or elliptic coordinates.

INTRODUCTION

We call an operator $\underline{R}$ a raising operator for a Hamiltonian $\underline{H}$ if $[\underline{H}, \underline{R}] = \lambda \underline{R}$, where λ is a nonzero real constant. If ψ is an eigenvector of $\underline{H}$ with eigenvalue μ, $\underline{H}\psi = \mu\psi$, it follows easily that $\underline{H}(\underline{R}\psi) = (\mu + \lambda)\underline{R}\psi$. Thus, knowledge of $\underline{R}$ permits one to obtain new eigenvalues and eigenvectors of $\underline{H}$ from old ones.

In this paper we give a complete classification of all potentials occurring in the two-dimensional time independent Schrödinger equation $\underline{H}\psi = \mu\psi$ which admit first- and second-order raising operators. The classification of first-order operators is almost trivial, and it is only the second-order case which presents difficulties. Moreover, as one can see from the results of Secs. 2 and 3, there are very few potentials admitting second-order raising operators, and all such potentials are generalizations of the harmonic oscillator.

The principal interest in our results lies in the fact that they are exhaustive and in the method used to obtain them. Proceeding directly, one can show that a Hamiltonian admits a second-order raising operator if and only if the corresponding potential V satisfies the system (2.8)–(2.10) of second-order overdetermined partial differential equations. However, while one can easily find some solutions of these equations, it is extremely difficult to determine when one has found all solutions. We have not been able to solve these equations directly.

In order to solve (2.8)–(2.10) we have adopted an indirect method based on results of Winternitz *et al.*,[1] which relates this problem to the Euclidean group $E(2)$. In Ref. 1 the authors show that $\underline{H}$ admits a second-order symmetry operator if and only if the corresponding Schrödinger equation separates in Cartesian, polar, parabolic, or elliptic coordinates. In this paper we show in essence that if $\underline{H}$ admits a second-order raising operator, then it also admits a second-order symmetry operator, hence that the Schrödinger equation must separate in Cartesian, polar, or elliptic (but strangely, not in parabolic) coordinates. This means that we can restrict ourselves to a search for solutions of (2.8)–(2.10) which separate in one of these three coordinate systems. In this case (2.8)–(2.10) reduce to systems of ordinary differential equations which, though tedious to solve, are tractable. Thus we obtain a complete solution to our problem.

Our method can be generalized to the more interesting three-dimensional case[2] as well as to other types of differential equations, for example, wave equations or the time dependent Schrödinger equation.

The results of Refs. 1, 2, and this paper show the intimate connection between second-order raising and symmetry operators and the separation of the Schrödinger equation in some coordinate system. It appears that higher-order operators will not be of great interest unless and until one can find similar indirect means of characterizing them.

The paper is organized as follows: In Sec. 1 the problem of first-order raising operators is solved, while in Sec. 2 the problem for second-order operators is formulated as a system of overdetermined second-order partial differential equations. We then obtain some solutions, but not the most general class which must await the further development of the connection with separation of variables in Sec. 3, where we complete our classification of all solutions. Finally in Sec. 4, we give the action of the raising and lowering operators on a basis of eigenfunctions of the Schrödinger equation for each case.

1. FIRST-ORDER OPERATORS

Let $\underline{H}$ be the formal Hamiltonian

$$\underline{H} = -\tfrac{1}{2}(\partial_{xx} + \partial_{yy}) + V(x, y) \tag{1.1}$$

acting on the Hilbert space $L_2(R_2)$ of square integrable functions in the plane. Here $V(x, y)$ is a real-valued thrice-differentiable function of (x, y) to be determined. We search first for all Hamiltonians which admit a first-order raising operator $\underline{R}$, i.e., we look for all $\underline{H}$ which satisfy

$$[\underline{H}, \underline{R}] = \lambda \underline{R}, \tag{1.2}$$

where λ is a nonzero real constant and $\underline{R}$ is a first-order partial differential operator

$$\underline{R} = \alpha_1(xy)\partial_x + \alpha_2(x, y)\partial_y + \alpha_3(x, y), \quad |\alpha_1|^2 + |\alpha_2|^2 \not\equiv 0. \tag{1.3}$$

Without loss of generality, we can assume that $\underline{R}$ is real, i.e., that $\alpha_1, \alpha_2, \alpha_3$ are real-valued functions. Substituting (1.1) and (1.3) into (1.2) and equating coefficients of $\partial_{xx}, \partial_{xy}, \partial_{yy}, \partial_x, \partial_y, 1$ on both sides of the resulting expression, we obtain the conditions

$$\partial_x\alpha_1 = \partial_y\alpha_2 = 0, \quad \partial_x\alpha_2 + \partial_y\alpha_1 = 0, \tag{1.4}$$

$$(\partial_{xx} + \partial_{yy})\alpha_1 + 2\partial_x\alpha_3 + 2\lambda\alpha_1 = 0,$$

$$(\partial_{xx} + \partial_{yy})\alpha_2 + 2\partial_y\alpha_3 + 2\lambda\alpha_2 = 0, \tag{1.5}$$

$$(\partial_{xx}+\partial_{yy})\alpha_3+2\alpha_1\partial_x V+2\alpha_2\partial_y V+2\lambda\alpha_3=0. \tag{1.6}$$

It is easy to show that these equations have solutions if and only if

$$V(x,y) = f(bx-ay)+\tfrac{1}{2}\lambda^2(x^2+y^2)-\begin{cases}\lambda cx/a & \text{if } a\neq 0\\ \lambda cy/b & \text{if } a=0,\ b\neq 0.\end{cases} \tag{1.7}$$

Here, a, b, c are real constants with $a^2+b^2>0$ and f is an arbitrary real differentiable function. The raising operator is then

$$\underline{R}=a\partial_x+b\partial_y-\lambda(ax+by)+c. \tag{1.8}$$

By a simple translation and rotation of the (x,y) coordinates we can obtain new Cartesian coordinates X, Y in which

$$V(X,Y)=g(X)+\tfrac{1}{2}\lambda^2Y^2,\quad \underline{R}=\partial_Y-\lambda Y, \tag{1.9}$$

where $g(X)$ is arbitrary. In these coordinates the Schrödinger equation

$$\underline{H}\psi(X,Y)=\mu\psi(X,Y) \tag{1.10}$$

has solutions of the form

$$\psi_{\mu,n}=\exp(-|\lambda|Y^2/2)H_n(\sqrt{|\lambda|}Y)G(X),\quad n=0,1,2,\ldots, \tag{1.11}$$

where H_n is a Hermite polynomial[3] and $G(X)$ is a square integrable solution of the equation

$$G''-2g(X)G=[-2\mu+|\lambda|(2n+1)]G.$$

It follows easily that

$$\underline{R}\psi_{\mu,n}=\begin{cases}-\sqrt{|\lambda|}\,\psi_{\mu+\lambda,n+1} & \text{if } \lambda>0\\ 2n\sqrt{|\lambda|}\,\psi_{\mu+\lambda,n-1} & \text{if } \lambda<0.\end{cases} \tag{1.12}$$

2. SECOND-ORDER OPERATORS

Next we consider the more interesting problem of computing those Hamiltonians $\underline{H}$ which admit a second-order raising operator $\underline{R}$:

$$\underline{R}=\alpha_1\partial_{xx}+\alpha_2\partial_{xy}+\alpha_3\partial_{yy}+\alpha_4\partial_x+\alpha_5\partial_y+\alpha_6. \tag{2.1}$$

Here $\alpha_j(x,y)$ is a real function of (x,y) and $\alpha_1^2+\alpha_2^2+\alpha_3^2>0$. Substituting (1.1) and (2.1) into (1.2) and equating coefficients of the third-order and second-order derivatives, we obtain equations for $\alpha_1,\ldots,\alpha_5$ which can easily be solved to yield

$$\begin{aligned}&\alpha_1=-A_1y+A_4,\quad \alpha_2=A_1x+A_2y+A_3,\quad \alpha_3=-A_2x+A_5,\\ &\alpha_4=\lambda A_1xy-\lambda A_2y^2-\lambda A_3y-\lambda A_4x-A_6y+A_8,\\ &\alpha_5=-\lambda A_1x^2+\lambda A_2xy-\lambda A_5y+A_6x+A_7,\end{aligned} \tag{2.2}$$

where the A_i are real constants. The constraints on α_6 and V are obtained by equating coefficients of ∂_x, ∂_y, and 1 in (1.2):

$$\tfrac{1}{2}(\partial_{xx}+\partial_{yy})\alpha_4+\partial_x\alpha_6+2\alpha_1V_x+\alpha_2V_y=-\lambda\alpha_4, \tag{2.3}$$

$$\tfrac{1}{2}(\partial_{xx}+\partial_{yy})\alpha_5+\partial_y\alpha_6+\alpha_2V_x+2\alpha_3V_y=-\lambda\alpha_5, \tag{2.4}$$

$$\tfrac{1}{2}(\partial_{xx}+\partial_{yy})\alpha_6+\alpha_1V_{xx}+\alpha_2V_{xy}+\alpha_3V_{yy}+\alpha_4V_x+\alpha_5V_y=-\lambda\alpha_6. \tag{2.5}$$

Relations (2.3) and (2.4) yield

$$\begin{aligned}&\partial_x\alpha_6=-2\alpha_1V_x-\alpha_2V_y-\lambda\alpha_4+A_2\lambda,\\ &\partial_y\alpha_6=-\alpha_2V_x-2\alpha_3V_y-\lambda\alpha_5+A_1\lambda.\end{aligned} \tag{2.6}$$

Substituting (2.6) into (2.5), we obtain an expression for the multiplier α_6 in terms of V:

$$2\lambda\alpha_6=(A_1-2\alpha_5)V_y+(A_2-2\alpha_4)V_x+\lambda^2(A_1y+A_2x-A_4-A_5). \tag{2.7}$$

Equations (2.3) and (2.4) may not be consistent with (2.7). To guarantee consistency, we differentiate (2.7) to compute $\partial_x\alpha_6$, $\partial_y\alpha_6$ and substitute into (2.3), (2.4). This yields the consistency conditions for the potential:

$$\begin{aligned}&(A_2-2\alpha_4)V_{xx}+(A_1-2\alpha_5)V_{yx}+6\lambda\alpha_1V_x-2(\partial_x\alpha_5-\lambda\alpha_2)V_y\\ &\quad=-2\lambda^2\alpha_4+\lambda^2A_2,\end{aligned} \tag{2.8}$$

$$\begin{aligned}&(A_2-2\alpha_4)V_{yx}+(A_1-2\alpha_5)V_{yy}-2(\partial_y\alpha_4-\lambda\alpha_2)V_x+6\lambda\alpha_3V_y\\ &\quad=-2\lambda^2\alpha_5+\lambda^2A_1.\end{aligned} \tag{2.9}$$

Thus, corresponding to any choice of the constants $A_1,\ldots,A_8$, the Hamiltonian admits the raising operator $\underline{R}$, (2,1), (2,2), provided that V satisfies the partial differential equations (2.8) and (2.9). The multiplier α_6 for $\underline{R}$ is given by (2.7).

We can obtain another consistency relation for V by differentiating (2.8) with respect to x, differentiating (2.9) with respect to y, and subtracting the second equation from the first:

$$\begin{aligned}&(A_1x+A_2y+A_3)V_{xx}+2(A_1y-A_2x-A_4+A_5)V_{xy}\\ &\quad-(A_1x+A_2y+A_3)V_{yy}+3A_1V_x-3A_2V_y\\ &\quad=-\lambda(-3A_1\lambda x+3A_2\lambda y+\lambda A_3+2A_6).\end{aligned} \tag{2.10}$$

Although (2.10) is a consequence of (2.8) and (2.9), it is useful in its own right.

In conclusion, to find the potentials V admiting raising operators, we must solve the system (2.8)—(2.10) of overdetermined second order partial differential equations.

To simplify the solution of these equations, let us consider the action of the Euclidean group $E(2)$. Under the action of a Euclidean transformation the coordinates (x,y) go into new coordinates (x',y'), where

$$\begin{aligned}&x'=x\cos\phi+y\sin\phi+a,\quad \phi,a,b\in R,\\ &y'=-x\sin\phi+y\cos\phi+b.\end{aligned} \tag{2.11}$$

Since Euclidean transformations preserve the Laplace operator, we have

$$-\tfrac{1}{2}(\partial_{x'x'}+\partial_{y'y'})+V(x',y')=-\tfrac{1}{2}(\partial_{xx}+\partial_{yy})+V'(x,y),$$

where $V'(x,y)=V(x',y')$. Thus the Hamiltonian $\underline{H}$ is transformed into a new Hamiltonian $\underline{H}'=-\tfrac{1}{2}(\partial_{xx}+\partial_{yy})+V'(x,y)$. Similarly the raising operator $\underline{R}$ is transformed into a new raising operator $\underline{R}'$ satisfying $[\underline{H}',\underline{R}']=\lambda\underline{R}'$. Considering the set M of all pairs $\{V,\underline{R}\}$ which satisfy (1.2), we see that $E(2)$ acts on M as a transformation group. We will consider two solutions of (1.2) as *equivalent* if one solution can be obtained from the other by a transformation (2.11), i.e., if both solutions lie on the same $E(2)$ orbit. Clearly, it will be enough for us to find a solution, if one exists, corresponding to a single point on each orbit.

For the orbit analysis we make use of (2.1) and (2.2) to write a general raising operator as

$$\underline{R}=\sum_{j=1}^{8} A_j\underline{Q}_j+\alpha_6(x,y), \tag{2.12}$$

where

$$\underline{Q}_1=MP_1-\lambda xM,\quad \underline{Q}_2=-MP_2+\lambda yM,\quad \underline{Q}_3=P_1P_2-\lambda yP_1, \tag{2.13}$$

$$\underline{Q}_4=P_1^2-\lambda xP_1,\quad \underline{Q}_5=P_2^2-\lambda yP_2,\quad \underline{Q}_6=M,\quad \underline{Q}_7=P_2,\quad \underline{Q}_8=P_1.$$

Here,

$$P_1=\partial_x,\quad P_2=\partial_y,\quad M=x\partial_y-y\partial_x \tag{2.14}$$

are the basis operators for the Lie algebra action of $E(2)$. We see that the pure differential operator component of $\underline{R}$ is described by the vector $(A_1,\ldots,A_8)$ and that the action of $E(2)$ induces an orbit structure on the set of all such vectors. A direct computation shows that a rotation through the angle θ [Eqs. (2.11) with $a=b=0$] transforms (A_j) into (A_j') with

$$\begin{aligned}
&A_1'=\cos\theta A_1+\sin\theta A_2,\quad A_2'=-\sin\theta A_1+\cos\theta A_2,\\
&A_3'=\cos2\theta A_3+\sin2\theta\,(A_4-A_5),\\
&A_4'=-\sin\theta\cos\theta A_3+\cos^2\theta A_4+\sin^2\theta A_5,\\
&A_5'=\sin\theta\cos\theta A_3+\sin^2\theta A_4+\cos^2\theta A_5,\\
&A_6'=\lambda\sin^2\theta A_3-\lambda\sin\theta\cos\theta A_4+\lambda\sin\theta\cos\theta A_5+A_6,\\
&A_7'=\cos\theta A_7+\sin\theta A_8,\quad A_8'=-\sin\theta A_7+\cos\theta A_8.
\end{aligned} \tag{2.15}$$

Similarly, the translation $x\to x+a$ yields

$$\begin{aligned}
&A_1'=A_1,\quad A_2'=A_2,\quad A_3'=aA_1+A_3,\\
&A_4'=A_4,\quad A_5'=-aA_2+A_5,\\
&A_6'=-2a\lambda A_1+A_6,\quad A_7'=-a^2\lambda A_1+aA_6+A_7,\\
&A_8'=-a\lambda A_4+A_8,
\end{aligned} \tag{2.16}$$

and the translation $y\to y+b$ yields

$$\begin{aligned}
&A_1'=A_1,\quad A_2'=A_2,\quad A_3'=bA_2+A_3,\quad A_4'=-bA_1+A_4,\\
&A_5'=A_5,\quad A_6'=b\lambda A_2+A_6,\\
&A_7'=-b\lambda A_5+A_7,\quad A_8'=-b^2\lambda A_2-b\lambda A_3-bA_6+A_8.
\end{aligned} \tag{2.17}$$

Using these results, we will choose a point on each $E(2)$-orbit. We start with a general operator $\sum A_j\underline{Q}_j$. Noticing that $A_1^2+A_2^2$ is an $E(2)$-invariant, we see that there are three cases:

Case 1: $A_1^2+A_2^2>0$.

Case 2: $A_1=A_2=0,\quad A_3^2+A_4^2+A_5^2>0$.

Case 3: $A_1=A_2=\cdots=A_5=0,\quad A_6^2+A_7^2+A_8^2>0$.

In Case 3, $\underline{R}$ is first order and has already been treated in Sec. 1. In Case 1 we can perform a rotation so that $A_1'>0$, $A_2'=0$ and then translate so that $A_3''=A_4''=0$. Thus the vectors (A_j) of the form

$$(A_1,0,0,0,A_5,A_6,A_7,A_8),\quad A_1>0, \tag{2.18}$$

cut each Case 1 orbit exactly once. In Case 2 we can perform a rotation such that $A_3'=0$. Thus, vectors of the form

$$(0,0,0,A_4,A_5,A_6,A_7,A_8),\quad A_4^2+A_5^2>0, \tag{2.19}$$

cut each Case 2 orbit at least once.

These Case 2 solutions of (2.8)–(2.10) are easy to find. Indeed, assuming that $\underline{R}=\sum A_j\underline{Q}_j+\alpha_6$ and V are Case 2 solutions, we can use (2.19) to require $A_1=A_2=A_3=0$. Then (2.10) becomes

$$2(A_5-A_4)V_{xy}=-2\lambda A_6. \tag{2.20}$$

Suppose first that $A_5-A_4\neq0$. Then (2.20) has the general solution

$$V=-[\lambda A_6xy/(A_5-A_4)]+f(x)+g(y),$$

where f and g are arbitrary. Substituting this solution into (2.8) and (2.9), we find $A_6=0$ for consistency, and so $V=f(x)+g(y)$. If both A_4 and A_5 are nonzero, we can perform translations (2.16), (2.17) to achieve $A_7=A_8=0$. Thus Eqs. (2.8), (2.9) reduce to

$$xf''+3f'=\lambda^2x,\quad \lambda g''+3g'=\lambda^2y$$

with general solution

$$V(x,y)=\frac{\lambda^2}{8}(x^2+y^2)+\frac{a}{x^2}+\frac{b}{y^2}+c,\quad a,b,c\in R, \tag{2.21}$$

$$(A_j)=(0,0,0,A_4,A_5,0,0,0),\quad A_4,A_5\neq0.$$

If $A_4\neq0$, $A_5=0$, we can perform a translation (2.16) to achieve $A_8=0$. Then Eqs. (2.8), (2.9) reduce to

$$xf''+3f'=\lambda^2x,\quad (g''-\lambda^2)A_7=0$$

with solutions

$$V(x,y)=\frac{\lambda^2}{8}(x^2+4y^2)+\frac{a}{x^2}+by+c, \tag{2.22}$$

$$(A_j)=(0,0,0,A_4,0,0,A_7,0),\quad A_4,A_7\neq0$$

and

$$V(x,y)=\frac{\lambda^2x^2}{8}+\frac{a}{x^2}+g(y),\quad g(y)\text{ arbitrary}, \tag{2.23}$$

$$(A_j)=(0,0,0,A_4,0,0,0,0),\quad A_4\neq0.$$

The cases $A_5\neq0$, $A_4=0$ are identical to (2.22), (2.23) with x and y interchanged.

Finally, suppose $A_4=A_5\neq0$. Then (2.20) yields $A_6=0$, and by applying translations (2.16), (2.17) we can achieve $A_7=A_8=0$. Thus Eqs. (2.8), (2.9) reduce to

$$\frac{d}{dx}(xV_x+yV_y+2V)=\lambda^2x,\quad \frac{d}{dy}(xV_x+yV_y+2V)=\lambda^2y$$

with general solution

$$V(x,y)=\frac{\lambda^2}{8}(x^2+y^2)+\frac{g(x/y)}{y^2}+a=\frac{\lambda^2}{8}r^2+\frac{f(\theta)}{r^2}+a, \tag{2.24}$$

$$(A_j)=(0,0,0,A_4,A_4,0,0,0),\quad A_4\neq0.$$

Here g is arbitrary, $x=r\cos\theta$, $y=r\sin\theta$, $f(\theta)=g(\tan\theta)/\sin^2\theta$.

This completes the analysis of Case 2. However, Case 1, Eqs. (2.18), is much more difficult. We have not been able to discover a direct practical means of computing all solutions of (2.8)–(2.10) corresponding to this case. In the next section we develop an indirect group-theoretic procedure which not only enables us to solve these equations but also provides clear insight into the structure of second-order raising operators.

3. SEPARATION OF VARIABLES

Let us note that raising and lowering operators occur in pairs: If $\underline{R}$ is a raising operator for $\underline{H}$,

$$[\underline{H},\underline{R}]=\lambda\underline{R},\quad \lambda\neq 0,\ \lambda\in R \tag{3.1}$$

then, taking the formal adjoint, we have

$$[\underline{H},\underline{R}^+]=-\lambda\underline{R}^+ \tag{3.2}$$

so that $\underline{R}^+$ is a lowering operator (raising operator by $-\lambda$). In particular, if $\underline{R}$ takes the form (2.12), then

$$\underline{R}^+=\sum_{j=1}^{8}A_j\underline{Q}_j^+ +\alpha_6=\sum_{j=1}^{8}\tilde{A}_j\underline{Q}_j(-\lambda)+\tilde{\alpha}_6, \tag{3.3}$$

where

$$\begin{aligned}&\tilde{A}_j=A_j,\quad 1\leq j\leq 5,\\ &\tilde{A}_6=-A_6,\quad \tilde{A}_7=A_1-A_7,\quad \tilde{A}_8=A_2-A_8,\\ &\tilde{\alpha}_6=\alpha_6-\lambda(A_1y+A_2x-A_4-A_5).\end{aligned} \tag{3.4}$$

Here $\underline{Q}_j(-\lambda)$ is obtained from $\underline{Q}_j$, (2.13), by replacing λ with $-\lambda$. These results follow from (2.13) and the following facts:

$$\begin{aligned}&\underline{Q}_1^+=MP_1+\lambda xM+P_2-\lambda y,\quad \underline{Q}_2^+=-MP_2-\lambda yM+P_1-\lambda x,\\ &\underline{Q}_3^+=P_1P_2+\lambda yP_1,\quad \underline{Q}_4^+=P_1^2+\lambda xP_1+\lambda,\\ &\underline{Q}_5^+=P_2^2+\lambda yP_2+\lambda,\quad \underline{Q}_6^+=-M,\quad \underline{Q}_7^+=-P_2,\quad \underline{Q}_8^+=-P_1.\end{aligned} \tag{3.5}$$

Moreover, it follows from (3.1) and (3.2) that $[\underline{H},\underline{S}]=0$, where $\underline{S}=[\underline{R},\underline{R}^+]$, i.e., $\underline{S}$ is a symmetry of $\underline{H}$. We are concerned with the case where $\underline{R}$ and $\underline{R}^+$ are both second-order differential operators, so that we would expect that $\underline{S}$ was in general a third-order operator. However, we see from (2.12), (2.13), and (3.3) that the purely second-order terms in $\underline{R}$ and $\underline{R}^+$ are identical. This means that $\underline{S}$ is at most a second-order operator. Indeed for $\underline{S}=\mathcal{S}+\beta$, where $\mathcal{S}$ is a pure differential operator and β is a multiplier function, a straightforward computation yields

$$\begin{aligned}\mathcal{S}=&P_1^2(4\lambda A_4^2+2\lambda A_3^2+A_1^2-2A_1A_7+2A_3A_6)+P_2^2(4\lambda A_5^2+A_2^2\\ &-2A_2A_8-2A_3A_6)+P_1P_2(2A_1A_8-2A_1A_2+2A_2A_7+6\lambda A_3A_5\\ &+2\lambda A_3A_4-4A_4A_6+4A_5A_6)+M^2(4\lambda A_1^2+4\lambda A_2^2)+(MP_1\\ &+P_1M)(-\lambda A_1A_5+4\lambda A_1A_4-A_2A_6-3\lambda A_2A_3)+(MP_2\\ &+P_2M)(-A_1A_6+2\lambda A_1A_3-4\lambda A_2A_5+\lambda A_2A_4).\end{aligned} \tag{3.6}$$

At this point we can make use of the results of Ref. 1. There one studies differential operators

$$\begin{aligned}\underline{L}=&AP_1^2+BP_1P_2+CP_2^2+DM^2+E(P_1M+MP_1)\\ &+F(P_2M+MP_2)+\gamma(x,y)\end{aligned} \tag{3.7}$$

such that $[\underline{H},\underline{L}]=0$, where $\underline{H}$ is given by (1.1). A principal result of Ref 1 is essentially that if $\underline{H}$ commutes with a nontrivial $\underline{L}$, then the Schrödinger equation $\underline{H}\psi=\mu\psi$ separates in one of four orthogonal coordinate systems. More specifically the authors study the action of $E(2)$ on the set of all operators $\underline{L}$ via the coordinate transformations (2.11). They show that the $E(2)$-orbits are of five types.

$$\begin{aligned}&\text{I.}\quad P_1^2-P_2^2+a(P_1^2+P_2^2)+\beta,\\ &\text{II.}\quad P_1M+MP_1+a(P_1^2+P_2^2)+\beta,\\ &\text{III.}\quad M^2+a(P_1^2+P_2^2)+\beta,\\ &\text{IV.}\quad M^2+\tfrac{1}{2}l^2(P_1^2-P_2^2)+a(P_1^2+P_2^2)+\beta,\\ &\text{V.}\quad a(P_1^2+P_2^2)+\beta,\quad a,l\in R,\quad l>0.\end{aligned} \tag{3.8}$$

Every $\underline{L}$ lies on the same orbit as a constant multiple of exactly one of the elements I–V. (The term $P_1^2+P_2^2$ occurs with an arbitrary constant because the Hamiltonian always commutes with itself.) Thus by applying an appropriate $E(2)$ transformation we can always assume that $\underline{L}$ is equal to one of these five forms.

If $\underline{L}$ takes the form I, then, according to Ref. 1,

$$V(x,y)=f(x)+g(y), \tag{3.9}$$

and the Schrödinger equation separates in rectangular coordinates. If $\underline{L}$ takes the form II, then

$$V=\frac{f(\xi_1)+g(\xi_2)}{\xi_1^2+\xi_2^2},\quad x=\tfrac{1}{2}(\xi_1^2-\xi_2^2),\quad y=\xi_1\xi_2, \tag{3.10}$$

and the Schrödinger equation separates in parabolic coordinates, while if $\underline{L}$ takes form III,

$$V=f(r)+g(\theta)/r^2,\quad x=r\cos\theta,\quad y=r\sin\theta \tag{3.11}$$

and the equation separates in polar coordinates. If $\underline{L}$ takes form IV, then

$$V=\frac{f(\sigma)+g(\rho)}{\cos^2\sigma-\cosh^2\rho},\quad x=l\cosh\rho\cos\sigma,\quad y=l\sinh\rho\sin\sigma, \tag{3.12}$$

and the equation separates in elliptic coordinates. Finally, if $\underline{L}$ takes form V, then $\underline{L}$ is a multiple of $\underline{H}$ and there is no information about V.

The above results apply immediately to our study of the operator $\underline{S}$. First of all, by putting $\underline{R}$ in one of the forms (2.18), (2.19), we see from (3.6) that if $\underline{R}$ is truly second-order, then $\underline{S}$ is truly second-order (never first-order).

Note that the coefficient of M^2 in (3.6) is proportional to $A_1^2+A_2^2$. If this coefficient is nonzero, then $\underline{S}$ lies on a type III or IV orbit, i.e., the Schrödinger equation separates in either polar or elliptic coordinates. If $A_1=A_2=0$, then $\underline{S}$ lies on a type I, II, or V orbit.

We consider Case 2($A_1=A_2=0$) first. Then from (2.19) we can also require $A_3=0$, $A_4^2+A_5^2>0$. Substituting into (3.6), we find

$$\underline{S}=4\lambda A_4^2P_1^2+4\lambda A_5^2P_2^2+4A_6(A_5-A_4)P_1P_2+\beta. \tag{3.13}$$

It follows that type II orbits never appear, only type I and V orbits are possible. Moreover, our analysis of (2.20) has shown that we can find a potential V only if $A_6=0$. Thus $\underline{S}$ corresponds to a type I orbit if $A_4^2\neq A_5^2$ and to a type V orbit if $A_4^2=A_5^2$. The method of Ref. 1 yields no information for type V orbits but our direct approach in Sec. 2 has yielded the solutions (2.24), separation in polar coordinates, and the special case $A_4=-A_5$ of (2.21), separation in rectangular coordinates. For $A_4^2\neq A_5^2$ the results of Ref. 1 show that $\underline{H}$ lies on the same orbit as a Hamiltonian whose potential takes the form $V=f(x)+g(y)$. This agrees with the results (2.21)–(2.23).

So far we have merely verified previous results. However, the method of Ref. 1 now allows us to find all

solutions of (2.8)–(2.10) corresponding to Case 1. Indeed, if $A_1^2+A_2^2>0$, we know that $\underline{H}$ lies on the same orbit as a Hamiltonian with potential of the form (3.11) or of the form (3.12). Thus, we can find all Case 1 solutions of (2.8)–(2.10) by requiring that V take either the form (3.11) or (3.12). That is, every solution V lies on the same orbit as a V which separates in either polar or elliptic coordinates. This fact is of great importance for it allows us to separate variables in (2.8)–(2.10) and reduce these coupled partial differential equations to uncoupled ordinary differential equations for f and g.

At this point we have proved the following fact: If a Hamiltonian $\underline{H}$ admits a second-order raising operator then the Schrödinger equation $\underline{H}\psi=\mu\psi$ separates in either rectangular, polar or elliptic coordinates. Of course the converse is false.

To find all cases when $\underline{S}$ is type III we substitute the polar coordinate expression (3.11) into (2.8)–(2.10) and find all solutions which correspond to type III orbits. A tedious computation yields the single solution

$$V=\frac{\lambda^2r^2}{2}+\frac{a\sin\theta+b}{r^2\cos^2\theta}+c=\frac{\lambda^2}{2}(x^2+y^2)+\frac{ay}{x^2\sqrt{x^2+y^2}}+\frac{b}{x^2}+c, \tag{3.14}$$

$$(A_j)=(A_1,0,0,0,0,0,\tfrac{1}{2}A_1,0),\quad A_1\neq0.$$

Every type III solution lies on the same orbit as (3.14).

To find all cases when $\underline{S}$ is type IV we substitute the elliptic coordinate expression (3.12) into (2.8)–(2.10) and find all solutions which correspond to type IV orbits. We obtain

$$V=\frac{\tfrac{1}{2}\lambda^2(\cosh^2\rho+\cos^4\sigma-\cosh^4\rho-\cos^2\sigma)+b(1/\cosh^2\rho-1/\cos^2\sigma)}{\cos^2\sigma-\cosh^2\rho}$$

$$+c=\tfrac{1}{2}\lambda^2(x^2+y^2)+b/x^2+c, \tag{3.15}$$

$$(A_j)=(A_1,0,0,0,0,0,A_7,0),\quad A_1\neq0,\quad 2A_7\neq A_1,$$

$$x=\cosh\rho\cos\sigma,\quad y=\sinh\rho\sin\sigma,$$

$$V=\frac{\lambda^2}{2}\frac{(\cosh^2\rho+\cos^4\sigma-\cosh^4\rho-\cos^2\sigma)}{\cos^2\sigma-\cosh^2\rho}+c=\frac{\lambda^2}{2}(x^2+y^2)+c, \tag{3.16}$$

$$(A_j)=(A_1,0,0,0,0,0,A_7,A_8),\quad A_1,A_8\neq0.$$

The determination of all solutions of Eqs. (2.8)–(2.10) for elliptic coordinates is extremely tedious due to the complicated nature of the coefficients in the resultant coupled ordinary differential equations. Our method is to examine these equations in the vicinity of some convenient point which may or may not be a singularity of the potential. [This singularity cannot be essential since from (2.10) in elliptic coordinates one can see that if there is a singular point it is regular.] For example, examination of (2.8)–(2.10) about the points $\sin\sigma=0$ yields six differential equations involving only $g(\rho)$ which must be compatible. In this way one can proceed until all possibilities for the parameters A_j and potentials V are exhausted.

4. EXAMPLES

In this section we explicitly solve the Schrödinger equations corresponding to the above potentials and examine the action of our second-order raising operators. Without loss of generality we can assume $\lambda>0$ and set the additive constant c equal to zero for each potential. In each case we solve the equation $\underline{H}\psi=\mu\psi$ corresponding to appropriate choices of the potential parameters.

Consider first the potential (2.21),

$$V(x_1,x_2)=\frac{\lambda^2}{8}(x_1^2+x_2^2)+\frac{a_1}{x_1^2}+\frac{a_2}{x_2^2}.$$

Bound states exist for $a_i>-\frac{1}{8}$, and the normalized eigenfunctions are

$$\psi^{\pm\pm}_{k_1k_2}(x_1,x_2)=\frac{\lambda}{2}\prod_{i=1}^{2}\left(\frac{k_i!\,(\lambda^2/4)^{\pm\nu_i}}{\Gamma(k_i\pm\nu_i+1)}\right)^{1/2}\exp(-\lambda x_i^2/2)$$

$$\times x_i^{1/2\pm\nu_i}L_{k_i}^{\pm\nu_i}(\lambda x_i^2/2), \tag{4.1}$$

$$\nu_i=\tfrac{1}{2}(1+8a_i)^{1/2},\quad \mu^{\pm\pm}=\lambda(k_1+k_2+1)+\tfrac{1}{2}\lambda(\pm\nu_1\pm\nu_2),$$

$$k_i=0,1,2,\ldots.$$

For details on the degeneracies see Ref. 1.

Here $L_n^{\nu}(x)$ is a generalized Laguerre polynomial.[3] The raising operator in x_1 takes the form

$$\underline{R}=-\tfrac{1}{4}(2/\lambda\,\partial_{x_1x_1}-2x_1\partial_{x_1}+\tfrac{1}{2}\lambda x_1^2-4a_1/\lambda x_1^2-1) \tag{4.2}$$

with action

$$\underline{R}\psi_{k_1k_2}=\sqrt{(k_1+1)(k_1+\nu_1+1)}\,\psi_{k_1+1,k_2},$$

$$\underline{R}^+\psi_{k_1k_2}=\sqrt{k_1(k_1+\nu_1)}\,\psi_{k_1-1,k_2}. \tag{4.3}$$

There is a similar operator in x_2 which raises the k_2 index.

The potential (2.22),

$$V(x_1,x_2)=\tfrac{1}{8}\lambda^2(x_1^2+4x_2^2)+(a_1/x_1^2)+a_2x_2,$$

has bound states for $a_1>-\frac{1}{8}$, with eigenfunctions

$$\psi^{\pm}_{k_1k_2}(x_1,x_2)=\left(\frac{\lambda}{\pi}\right)^{1/4}\left(\frac{\lambda}{2}\right)^{(1\pm\nu)/2}\left(\frac{k_1!\,1}{k_2!\,2^{k_2}\Gamma(k_1\pm\nu+1)}\right)^{1/2}$$

$$\times\exp\left[-\frac{\lambda}{2}\left(x_2+\frac{a_2}{\lambda^2}\right)^2\right]H_{k_2}\left[\sqrt{\lambda}\left(x_2+\frac{a_2}{\lambda^2}\right)\right]$$

$$\times\exp\left(-\frac{\lambda}{4}x_1^2\right)x_1^{1/2\pm\nu}L_{k_1}^{\pm\nu}\left(\frac{\lambda}{2}x_1^2\right),$$

$$\nu=\tfrac{1}{2}(1+a_1)^{1/2},$$

$$\mu^{\pm}=\lambda\left[k_1+k_2+1\pm\frac{\nu}{2}-\frac{a_2^2}{2}\left(\frac{\lambda^2}{2}\right)^{-3/2}\right],\quad k_1,k_2=0,1,\ldots. \tag{4.4}$$

Details on the degeneracies can again be found in Ref. 1. This potential takes the form (1.7) with $x_1=x_1=x$, x_1, $x_2=y$, $a=0$ so that it admits a first-order raising operator (1.12) in x_2. It also admits the second-order operator (4.2) with action (4.3). Similarly the potential (2.23) admits a second-order raising operator in $x=x_1$ with the form (4.2).

The potential (2.24)

$$V=\tfrac{1}{8}\lambda^2r^2+f(\theta)/r^2$$

corresponds to eigenfunctions

$$\psi_{n,s}(r,\theta)=\left(\frac{\lambda}{2}\right)^{(s+1)/2}\left(\frac{2(n!)}{\Gamma(n+s+1)}\right)^{1/2}$$

$$\times\exp\left(-\frac{\lambda r^2}{4}\right)r^s L_n^s\left(\frac{\lambda r^2}{2}\right)\Theta_s(\theta),$$

$$n=0,1,\cdots. \tag{4.5}$$

where Θ_s is a solution of

$$\Theta_s''+[s^2-2f(\theta)]\Theta_s=0,$$

and $\mu=\lambda(n+\frac{1}{2}s+\frac{1}{2})$, $s>-1$. The raising operator takes the form

$$\underline{R}=-\frac{1}{4}\left[\partial_{rr}+\left(\frac{1}{r}-\lambda r\right)\partial_r-\frac{s^2}{r^2}+\frac{\lambda^2 r^2}{4}-\lambda\right]$$

with action

$$\underline{R}\psi_{n,s}=\tfrac{1}{2}\lambda[(n+s+1)(n+1)]^{1/2}\psi_{n+1,s},$$
$$R^*\psi_{n,s}=\tfrac{1}{2}\lambda[(n+s)n]^{1/2}\psi_{n-1,s}. \tag{4.7}$$

For the above potentials it was always possible to choose coordinates such that $\underline{R}$ could be expressed as a differential operator in a single variable. In the remaining three cases this is no longer possible and the action of the raising operator is more complicated.

The potential (3.14),

$$V=\frac{\lambda^2 r^2}{2}+\frac{a\sin\theta+b}{r^2\cos^2\theta},$$

has, for example, in the case $a=(\alpha^2-\beta^2)/4$, $b=-\frac{1}{8}+(\alpha^2+\beta^2)/4$, $\alpha>\frac{1}{2}$, $\beta>\frac{1}{2}$, normalized eigenfunctions

$$\psi_{n,k}(r,\theta)=\left(\frac{\lambda(n!)(k!)\Gamma(\alpha+\beta+k+1)(\alpha+\beta+2k+1)}{\Gamma(n+\frac{1}{2}(\alpha+\beta)+k+\frac{3}{2})\Gamma(\alpha+k+1)\Gamma(\beta+k+1)2^{\alpha+\beta}}\right)^{1/2}$$
$$\times\exp(-\lambda r^2/2)(\sqrt{\lambda}r)^{(\alpha+\beta)/2+k+1/2}L_n^{(\alpha+\beta+1+k)/2}(\lambda r^2)$$
$$\times(1+\sin\theta)^{\beta/2+1/4}(1-\sin\theta)^{\alpha/2+1/4}P_k^{\alpha,\beta}(\sin\theta),$$
$$n,k=0,1,2,\cdots, \tag{4.8}$$

and energy eigenvalues $\mu=\lambda[2n+k+\frac{1}{2}(\alpha+\beta)+1]$. Here, $P_k^{\alpha,\beta}(x)$ is Jacobi polynomial.[3] The raising operator is

$$\underline{R}=-\frac{\sin\theta}{r}\partial_{\theta\theta}+\cos\theta\,\partial_{r\theta}-\tfrac{1}{2}\sin\theta\partial_r-\left(\frac{\cos\theta}{2r}+\lambda r\cos\theta\right)\partial_\theta$$
$$+\frac{(\alpha^2-\beta^2)}{4r}\frac{(\sin^2\theta+1)}{\cos^2\theta}+\left(\frac{\alpha^2+\beta^2}{2}-\frac{1}{4}\right)\frac{\sin\theta}{r\cos^2\theta}+\frac{\lambda r}{2}\sin\theta \tag{4.9}$$

and its action takes the form

$$\underline{R}\psi_{n,k}=\gamma_{n,k}\psi_{n,k+1}+\xi_{n,k}\psi_{n+1,k-1},$$
$$\underline{R}^+\psi_{n,k}=\gamma_{n,k-1}\psi_{n,k-1}+\xi_{n-1,k+1}\psi_{n-1,k+1} \tag{4.10}$$

where γ, ξ are rather complicated real constants, nonzero in general. Thus, $\underline{R}$ no longer raises a single index n or k.

The potential (3.15),

$$V=\tfrac{1}{2}\lambda^2(x^2+y^2)+b/x^2,$$

is of the form (1.9) in y. Thus, it admits a first-order raising operator with action (1.11), (1.12). Furthermore, this potential corresponds to a special case of (2.21) so that its eigenfunctions are given by (4.1) with $a_1=b$, $a_2=0$, and it admits two second-order raising operators of the form (4.2). The potential is also a special case of (2.24) and (3.14) so that it admits the raising operators (4.6) and (4.9). However, the potential admits the further raising operator

$$\underline{R}=y(\partial_{xx}-\lambda^2x^2-2b/x^2-\lambda)-x(\partial_x-\lambda x)(\partial_x+\lambda y), \tag{4.11}$$

which is not admitted by the earlier mentioned potentials in their generality. In Cartesian coordinates the Hamiltonian has (unnormalized) eigenvectors

$$\psi_{k_1,k_2}^{\pm\nu}(x,y)=\exp(-\lambda r^2/2)(x)^{\pm\nu+1/2}L_{k_1}^{\pm\nu}(\lambda x^2)H_{k_2}(\sqrt{\lambda}y), \tag{4.12}$$

$$\nu=(2b+\tfrac{1}{4})^{1/4},\quad b>-\tfrac{1}{8},$$

and eigenvalues

$$\mu=\lambda(2k_1+k_2\pm\nu+1),\quad k_1,k_2=0,1,\cdots.$$

The action of $\underline{R}$ is given by

$$\underline{R}\psi_{k_1,k_2}^{\pm\nu}=-\lambda^{1/2}(2k_1\pm\nu+\tfrac{1}{2})\psi_{k_1,k_2+1}-4\lambda^{1/2}(k_1+1)k_2\psi_{k_1+1,k_2-1}, \tag{4.13}$$

$$\underline{R}^+\psi_{k_1,k_2}^{\pm\nu}=-2\lambda^{1/2}(k_1\pm\nu)\psi_{k_1-1,k_2+1}-2\lambda^{1/2}k_2(2k_1\pm\nu+\tfrac{3}{2})\psi_{k_1,k_2-1}.$$

The potential (3.16), isotropic harmonic oscillator, is a special case of all previous potentials except (2.22) and it admits all of the raising operators allowed by these potentials.

The raising operator $Q_1+\alpha_6$ admitted by potentials (3.15) and (3.16) implies via our procedure that the corresponding Schrödinger equations separate in elliptic coordinates. Thus one might expect that the action of these raising operators would be simplest in elliptic coordinates. This is not the case. The elliptic coordinate solutions of the harmonic oscillator Hamiltonian are Ince polynomials,[4] but the corresponding polynomial solutions for (3.15) in elliptic coordinates appear not to have been studied in any detail. In any event, the action of the raising operator on an elliptic basis is not transparent.

In conclusion, we remark that Refs. 5 and 6 contain results related to our work.

ACKNOWLEDGMENTS

We wish to thank P. Winternitz and E. Kalnins for helpful discussions, and the members of the Centre de Recherches Mathématiques for extending their hospitality to us. One of us (C. P. B.) would like to acknowledge the Eisenstadt Foundation for financial support.

[1]P. Winternitz, Y. Smorodinskiĭ, M. Uhlir, and I. Fris, Sov. J. Nucl. Phys. **4**, 444 (1967).
[2]A. Makarov, I. Smorodinskii, H. Valiev, and P. Winternitz, Nuovo Cimento A **52**, 1061 (1967).
[3]A. Erdelyi, *et al.*, *Higher Transcendental Functions*, *Vol.* 2 (McGraw-Hill, New York, 1953).
[4]F. Arscott, *Periodic Differential Equations* (Pergamon, Oxford, England, 1964).
[5]S. Komy and L. O'Raifeartaigh, J. Math. Phys. **9**, 738 (1968).
[6]Y. Dothan, Phys. Rev. D **2**, 2944 (1970).

Structure of the 12j and 15j coefficients in the Bargmann approach

Cheng-Sheng Huang* and Alfred C. T. Wu

Department of Physics, University of Michigan, Ann Arbor, Michigan 48104
(Received 19 November 1973; final revised manuscript received 4 March 1974)

Generating functions of the 12j and 15j angular momentum recoupling coefficients are computed explicitly in the Bargmann formalism. Symmetry properties are deduced therefrom. A geometrical Möbius strip representation (originally due to Ord-Smith for the 12j case), which can be generalized to all n, suggests a $4n$-fold symmetry for the $3nj$ coefficients ($n \geq 4$).

I. INTRODUCTION

The structure of the angular momentum 9j coefficient[1-3] has been studied in the Bargmann approach.[4,5] It is the purpose of this note to extend some of the considerations to higher $3nj$ coefficients.

(A) The generating functions for the 12j and 15j coefficients are explicitly computed in the Bargmann scheme. It is a tribute to the powerful Bargmann lemmas on the Laplacian integrals[4] that those seemingly complicated $6n$-fold integrals can in fact be systematically carried out. Thus in principle the generating functions for the $3nj$ coefficients are computable for arbitrary n in the Bargmann approach. Alternatively, the generating functions can also be found in the algebraic recursive scheme of Schwinger.[3] For $n=4$ and 5, they have been verified; the answers are essentially the same apart from a difference in an over-all phase factor.[6]

(B) Symmetry relations of the $3nj$ coefficients ($n=4, 5$) are here deduced on the basis of the explicit knowledge of their generating functions. They turn out to confirm the $4n$-fold symmetry ($n=4, 5$). For $n=4$, this was first discussed by Ord-Smith[7] using (i) a geometrical Möbius strip picture which incorporates the basic 3j triangular relations and (ii) an reduction formula (attributed to J. P. Elliott) of the 12j coefficient as a sum over products of four Racah coefficients.

The Möbius strip picture can be properly generalized to all n. (There is a slight technical difference between even or odd n cases.) Thus a basic $4n$-fold symmetry is expected to hold for arbitrary n. The situation may be summarized as in Table I. Lower order coefficients (for various reasons such as looser structure) are seen to possess larger symmetry. We find it gratifying that for $n \geq 4$, the symmetry for the $3nj$ coefficients becomes more systematic. [Note, however, the remark (b) below].

(C) Explicit expressions for the 12j and 15j coefficients can be extracted from their generating functions. However, in view of the excessively large numbers of summations involved [namely, $(2^{n+1}-1-3n)$—fold], we shall not write them down here. The reduction formulas[6,10] of $3nj$ coefficients in terms of $3(n-1)j$ coefficients on one hand, and in terms of the Racah coefficients on the other, are probably more useful in practice.

The following remarks are made in view of the extensive work on the theory of angular momentum by A. P. Jucys *et al.*,[10] although the present undertaking is entirely independent of their approach.

(a) Jucys *et al.* have adopted a graphical method of their own; they were able to do calculations with the aid of their graphical method. Our emphasis, however, is on the explicit calculation of the generating functions.

(b) There is a proliferation in the definition of the $3nj$ coefficients. As the number of j's goes up, there are obviously various different recoupling schemes. Thus Jucys *et al.* have defined several kinds of $3nj$ coefficients. The ones we discussed here in this paper, the canonical ones, correspond to what they call *the first kind.* We shall not be concerned with those other than the first kind here.

(c) We have independently rediscovered a set of recursion formulas for the $3nj$ coefficients (i) in terms of $3(n-1)j$ coefficients and (ii) in terms of 6j coefficients.[6] These are known to Jucys *et al.* The basic $4n$-fold symmetry is also implicit in their work. However, we wish to emphasize that the methodology used are quite different, especially in regard to the symmetry. Our emphasis in this paper is to carry out the explicit calculation of the generating functions. From what we learn from the previously known cases, we adopt the viewpoint that all the symmetry of the $3nj$ coefficients is contained in the generating functions. The symmetry should be transparent and unambiguous in the Bargmann approach. What we have found is that (i) from our study of the generating functions comes the basic $4n$-fold symmetry ($n=4, 5$); (ii) the symmetry operations can be transcribed to those on a suitably defined Möbius strip; and (iii) this geometrical picture and the $4n$ fold symmetry is obviously valid for arbitrary $n \geq 4$.

II. THE 12j COEFFICIENT

A. Definition

In analogy with the previously discussed $n \leq 3$ cases,[4,5] we express the 12j coefficient (which is the recoupling coefficient involved in adding five angular momenta to a total j, or adding six angular momenta to zero) in terms of sums of products of eight 3j coefficients. We adopt the following labeling in Eq. (1) for the twelve j's, which is a slight modification of that of Ord-Smith[7]:

$$\{12j\} \equiv \begin{Bmatrix} j_{30} & j_{01} & j_{12} & j'_{23} \\ & j_{00} & j_{11} & j_{22} & j_{33} \\ j'_{30} & j'_{01} & j'_{12} & j'_{23} \end{Bmatrix} \tag{1}$$

TABLE I.

$3nj$ recoupling coefficients		Symmetry relations
$n=2$	6j Racah	144 [Refs. 8, 4]
$n=3$	9j	72 [Refs. 2, 9, 5]
$n=4$	12j	16 [Ref. 7]
$n \geq 4$	$3nj$ general	$4n$ [this work and Ref. 10]

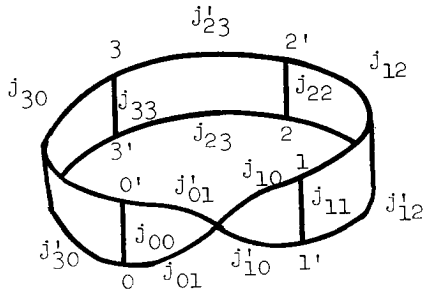


FIG. 1. Möbius strip representation for the 12j coefficient.

$$= \sum_{m,m'} \begin{pmatrix} j_{30} & j_{00} & m'_{01} \\ m_{30} & m_{00} & j'_{01} \end{pmatrix} \begin{pmatrix} j'_{01} & j_{11} & j'_{12} & j_{12} & j_{22} & j'_{23} \\ m'_{01} & m_{11} & m'_{12} & m_{12} & m_{22} & m'_{23} \end{pmatrix}$$

$$\times \begin{pmatrix} j_{23} & j_{33} & j_{03} \\ m_{23} & m_{33} & m_{03} \end{pmatrix} \begin{pmatrix} m'_{30} & m_{00} & j_{01} \\ j'_{30} & j_{00} & m_{01} \end{pmatrix} \begin{pmatrix} m_{01} & m_{11} & m_{12} \\ j_{01} & j_{11} & j_{12} \end{pmatrix}$$

$$\times \begin{pmatrix} m'_{12} & m_{22} & m_{23} \\ j'_{12} & j_{22} & j_{23} \end{pmatrix} \begin{pmatrix} m'_{23} & m_{33} & m_{30} \\ j'_{23} & j_{33} & j_{30} \end{pmatrix}, \tag{2}$$

where

$$j'_{pq} \equiv j_{qp}, \quad m'_{pq} \equiv m_{qp} \tag{3a}$$

except

$$\begin{aligned} &j'_{01} \equiv j'_{10}, \qquad j_{01} \equiv j_{10} \qquad \text{(even-}n\text{ rule)}, \\ &m'_{01} \equiv m'_{10}, \qquad m_{01} \equiv m_{10}. \end{aligned} \tag{3b}$$

It is clear that there are triangle relations governing in each of the eight 3j coefficients. In the present notation, each 3j factor calls for a set of consecutive triplet indices $(p-1\,q,\ pq,\ p+1\,q)$ or $(p\,q-1,\ pq,\ p\,q+1)$, $(p,q=0,1,\ldots,n-1,\ \mathrm{mod}\ n)$. It is convenient to label a set of vertices p,p' accordingly. This results in the Möbius strip representation[7] (see Fig. 1). Note that the index convention is as follows: (i) j_{pq} connects from vertices p to q; (ii) j gets primed if the first index is primed; the prime on the second index is suppressed [except for those for t_{pq}, see Eq. (10)]; (iii) rules (3a) and (3b) are to be obeyed for even n cases [cf. Eq. (22) for odd n]

B. Generating function of the 12j coefficients

The generating function is defined as follows:

$$G^{(12)}(t,t') \equiv \sum_{k,k'} N_4^{-1} \{12j\} \prod_{p,q} t_{pq}^{k_{pq}} t'^{k'_{pq}}, \tag{4}$$

where the normalization factor is given by

$$N_4 \equiv \left[\prod_{p=0}^{3} (J_p+1)!(J'_p+1)! \Big/ \left(\prod_{p,q} k_{pq}!\, k'_{pq}! \right) \right]^{1/2}. \tag{5}$$

For a triplet of indices $(p-1,p,\ p+1)$, we define

$$J_p \equiv j_{p\,p-1} + j_{pp} + j_{p\,p+1} = \sum_q j_{pq}, \tag{6a}$$

$$J'_p \equiv j_{p-1\,p} + j_{pp} + j_{p+1\,p} = \sum_q j_{qp} = \sum_q j'_{pq}, \tag{6b}$$

and

$$k_{pq} \equiv J_p - 2j_{pq}, \tag{6c}$$

$$k'_{pq} \equiv J'_p - 2j'_{pq}. \tag{6d}$$

In a manner which is perfectly parallel to the known cases $n \leq 3$,[4,5] the generating function can be converted into the following integral:

$$G^{(12)}(t,t') = \int d\mu_{24}(\xi) \exp\left(\sum_{p=0}^{3} (D_p + \bar{D}'_p) \right), \tag{7}$$

where

$$d\mu_N(\zeta) \equiv \pi^{-N} \exp(-\bar{\zeta}\cdot\zeta)\, d^N\zeta, \quad \zeta \equiv \xi + i\eta,$$

$$D_p \equiv t_p \times \xi_p \cdot \eta_p, \quad p=0,1,\ldots,3, \tag{8a}$$

$$\bar{D}'_p \equiv t'_p \times \bar{\xi}'_p \cdot \eta'_p \tag{8b}$$

denote 3×3 determinants formed by components of the indicated 3-vectors. The components of t_p are labeled by the triplets $(t_{p\,p-1},\ t_{pp},\ t_{p\,p+1})$; likewise for t'_p. For ξ_p and η_p, a distinction has to be made involving the index 0, namely for $l \neq 0$, ξ, η, have components labeled by $(l-1\,l,\ ll,\ l\,l+1)$; likewise for $\bar{\xi}'_l$ and $\bar{\eta}'_l$. On the other hand, for $p=0$, the components are

$$\begin{aligned} &\xi_0 \equiv (\xi_{03}, \xi_{00}, \bar{\eta}'_{01}); \quad \eta_0 \equiv (\eta_{03}, \eta_{00}, -\bar{\xi}'_{01}), \\ &\bar{\xi}'_0 \equiv (\bar{\xi}'_{03}, \bar{\xi}'_{00}, -\eta_{01}), \ \bar{\eta}'_0 \equiv (\bar{\eta}'_{03}, \bar{\eta}'_{00}, \xi_{01}). \end{aligned} \tag{9}$$

This complication comes about because two of the 3j coefficients in Eq. (1) (namely those involving the 0 and 0′ vertices) appear in a mixed conjugate fashion. In (8b) and (9), we have

$$\xi'_{pq} \equiv \xi_{qp}, \ \eta'_{pq} \equiv \eta_{qp} \tag{10}$$

while $t'_{pq} \equiv t_{pq'}$ are distinct from t_{qp}.

The 24-fold integration in (7) can be carried out in four steps. The calculation is straightforward with the aid of the Bargmann lemmas on the Laplacian integrals.[4] A slight extension leads to the following formula which turns out to be quite useful[6]:

$$\begin{aligned} &\int d\mu_3(\xi)\, d\mu_3(\eta) \exp(t\times\xi\cdot\eta + t'\times\bar{\xi}\cdot\bar{\eta} + c\cdot\bar{\xi} + d\cdot\bar{\eta}) \\ &\quad = (1 - t\cdot t')^{-1} \exp[(t\times c\cdot d)\,(1 - t\cdot t')^{-1}]. \end{aligned} \tag{11}$$

The final answer for the generating function (7) is

$$G^{(12)}(t,t') = (1 - a_1 - a_2 - a_3 - a_4 - a_5)^{-2}, \tag{12}$$

where

$$\begin{aligned} a_1 &= \hat{t}_{01}\hat{t}_{32} + \hat{t}_{10}\hat{t}_{23} + \hat{t}_{30}\hat{t}_{21} + \hat{t}_{03}\hat{t}_{12}, \\ a_2 &= t_{00}t_{11}(t'_{22}t'_{32}t_{30} + t'_{33}t'_{23}t_{21}) + t_{11}t'_{22}t_{33}t'_{01}t_{03} \\ &\quad + t_{22}t'_{33}t_{00}t_{12}t'_{10} - t_{pq} \longleftrightarrow t'_{pq}, \\ a_3 &= -\hat{t}_{00}\hat{t}_{12}\hat{t}_{32} - \hat{t}_{11}\hat{t}_{23}\hat{t}_{03} + \hat{t}_{22}\hat{t}_{30}\hat{t}_{10} + \hat{t}_{33}\hat{t}_{01}\hat{t}_{21}, \\ a_4 &= t'_{00}t_{10}t'_{12}t_{21}t'_{23}t_{32}t'_{30} + t'_{11}t'_{21}t_{23}t'_{32}t_{30}t'_{03}t_{01} \\ &\quad + t'_{22}t'_{32}t_{30}t'_{03}t_{01}t_{10}t'_{12} + t'_{33}t'_{03}t_{01}t_{10}t'_{12}t_{21}t'_{23} \\ &\quad - t_{pq} \longleftrightarrow t'_{pq}, \\ a_5 &= \hat{t}_{00}\hat{t}_{11}\hat{t}_{22}\hat{t}_{33} - \hat{t}_{00}\hat{t}_{11}\hat{t}_{23}\hat{t}_{32} - \hat{t}_{11}\hat{t}_{22}\hat{t}_{03}\hat{t}_{30} \\ &\quad + \hat{t}_{22}\hat{t}_{33}\hat{t}_{01}\hat{t}_{10} - \hat{t}_{33}\hat{t}_{00}\hat{t}_{12}\hat{t}_{21} + \hat{t}_{30}\hat{t}_{03}\hat{t}_{12}\hat{t}_{21} \\ &\quad - \hat{t}_{01}\hat{t}_{10}\hat{t}_{23}\hat{t}_{32}. \end{aligned} \tag{13}$$

with

$$\hat{t}_{pq} \equiv t_{pq}t'_{pq}. \tag{14}$$

C. Consistency check

Setting one of the appropriate angular momentum to

be zero should reduce the 12j coefficient to a 9j coefficient, and this implies that $G^{(12)}$ should reduce to $G^{(9)}$, which is known.[5] Our expression (12) satisfies this test. [To be precise, there are some sign difference among some of the corresponding terms and this is attributed to a difference in the choice of phase in going from $3nj$ to $3(n-1)j$ coefficients.]

D. Symmetry (even n case)

(a) Define the operation $P_{\mathrm{I}}^{(t)}$ which carries $t_{pq} \longleftrightarrow t'_{pq}$ and the operation $P_{\mathrm{I}}^{(k)}$ which carries $k_{pq} \longleftrightarrow k'_{pq}$ (15)

It is easily verified that the generating function $G^{(12)}(t, t')$ is invariant under $P_{\mathrm{I}}^{(t)}$. From Eq. (4), it follows that the 12j coefficient is invariant under $P_{\mathrm{I}}^{(k)}$ which carries $j_{pq} \longleftrightarrow j'_{pq}$.

(b) Define the operation $P_{--}^{(t)}$ = permutation $\binom{0\,1\,2\,3}{1\,0\,3\,2}$ among the t_{pq} indices. Likewise $P_{--}^{(k)}$ among the the k_{pq} indices. (16)

$G^{(12)}(t, t')$ is readily seen to be invariant under $P_{--}^{(t)}$. This implies that the 12j coefficient is invariant under $P_{--}^{(k)}$.

(c) Define the operation $P_{\times}^{(t)}$ that carries

$$t_{pq} \to t_{p-1\,q-1},$$
$$t'_{pq} \to t'_{p-1\,q-1}$$

except

$$t_{1q} \to -t_{0\,q-1}, \quad q \neq 1,$$
$$t_{11} \to t_{00}, \tag{17a}$$
$$t'_{1q} \to t'_{0\,q-1},$$

and $P_{\times}^{(k)}$ that carries

$$k_{pq} \to k'_{p+1\,q+1}, \qquad k'_{pq} \to k_{p+1\,q+1} \tag{17a}$$

except

$$k_{0q} \to k_{1\,q+1}, \qquad k'_{0q} \to k'_{1\,q+1}. \tag{17b}$$

Then $G^{(12)}(t, t')$ is invariant under $P_{\times}^{(t)}$. The 12j coefficient is left unchanged apart from a phase:

$$P_{\times}^{(k)}\{12j\} = (-1)^2 j_{11}\{12j\}. \tag{18}$$

In terms of the Möbius strip picture, the above three operations correspond to the following:

$P_{\mathrm{I}}^{(t)}$: up–down symmetry of the Möbius strip: two fold symmetry,

$P_{--}^{(t)}$: left–right symmetry of the Möbius strip: two fold symmetry,

$P_{\times}^{(t)}$: moving the "twist" between p and $p+1$ vertices: n-fold symmetry

Thus the combined symmetry is $4n$-fold ($n \geq 4$). (19)

The fact that the 6j and 9j coefficients in fact posses larger symmetry than the basic $4n$-fold symmetry discussed here might be attributed to the looser structure of their corresponding Möbius networks. (We emphasize the lines rather than the surface.) For $n \leq 3$ (i.e., with at most three vertical lines), it is possible to interchange the roles of horizontal and vertical lines, thereby resulting in enlarged symmetry. We claim that this is no longer possible for a Möbius network with four (or more) vertical lines.

III. THE 15j COEFFICIENT

A. Definition

Parallel to the discussion of the 12j case, we take

$$\{15j\} \equiv \begin{Bmatrix} j'_{40} & j'_{01} & j_{12} & j'_{23} & j_{34} \\ & j_{00} \quad j_{11} \quad j_{22} \quad j_{33} \quad j_{44} & \\ j_{40} & j_{01} & j'_{12} & j_{23} & j'_{34} \end{Bmatrix} \tag{20}$$

$$= \sum_{m,m'} \begin{pmatrix} m'_{40} & m'_{00} & m'_{10} \\ j'_{40} & j'_{00} & j'_{10} \end{pmatrix} \begin{pmatrix} m'_{01} & m'_{11} & m'_{21} \\ j'_{01} & j'_{11} & j'_{21} \end{pmatrix}$$
$$\times \begin{pmatrix} m'_{12} & m'_{22} & m'_{32} \\ j'_{12} & j'_{22} & j'_{32} \end{pmatrix} \begin{pmatrix} m'_{23} & m'_{33} & m'_{43} \\ j'_{23} & j'_{33} & j'_{43} \end{pmatrix} \begin{pmatrix} m'_{34} & m'_{44} & m'_{04} \\ j'_{34} & j'_{44} & j'_{04} \end{pmatrix}$$
$$\times \begin{pmatrix} j_{40} & j_{00} & j_{10} \\ m_{40} & m_{00} & m_{10} \end{pmatrix} \begin{pmatrix} j_{01} & j_{11} & j_{21} \\ m_{01} & m_{11} & m_{21} \end{pmatrix} \begin{pmatrix} j_{12} & j_{22} & j_{32} \\ m_{12} & m_{22} & m_{32} \end{pmatrix}$$
$$\times \begin{pmatrix} j_{23} & j_{33} & j_{43} \\ m_{23} & m_{33} & m_{43} \end{pmatrix} \begin{pmatrix} j_{34} & j_{44} & j_{04} \\ m_{34} & m_{44} & m_{04} \end{pmatrix}, \tag{21}$$

where

$$j'_{pq} \equiv j_{qp}, \quad m'_{pq} \equiv m_{qp}$$
(odd-n rule), $p = 0, 1, \ldots, 4.$ (22)

The remarks following Eq. (3) for the 12j coefficient apply here also with Eq. (22) replacing Eq. (3). The Möbius strip picture for the 15j is shown in Fig. 2.

B. Generating function of the 15j coefficient

As an obvious generalization from Eq. (4), we have

$$G^{(15)}(t, t') \equiv \sum_{p,q} N_5^{-1}\{15j\} \prod_{p,q} t_{pq}^{k_{pq}} t'^{k'_{pq}} \tag{23}$$

where

$$N_5 \equiv \left[\prod_{p=0}^{4} (J_p+1)!(J'_p+1)! \Big/ \left(\prod_{p,q} k_{pq}!\, k'_{pq}!\right)\right]^{1/2} \tag{24}$$

with the k, k', J, J' defined as in Eq. (6) now for $p, q = 0, 1, \ldots, 4$. As before, Eq. (23) is converted into the following integral:

$$G^{(15)}(t, t') = \int d\mu_{15}(\xi)\, d\mu_{15}(\eta) \exp\left(\sum_{p=0}^{4} (D_p + \bar{D}'_p)\right), \tag{25}$$

where D_p and $\bar{D}'_p$ are defined as in (8) now for p

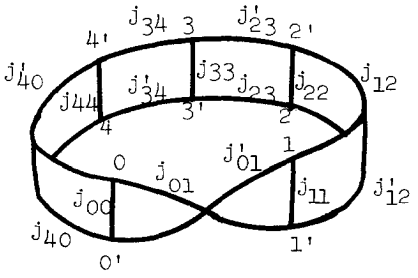


FIG. 2. Möbius strip representation for the 15j coefficient.

$=0,1,\ldots,4$, t_p having components $(t_{p-1\,p}, t_{pp}, t_{p+1\,p})$ (mod 5), labeled by a set of triplet indices. Likewise for ξ_p and η_p. All this is quite parallel to the 12j case except that the even-n complication (9) is absent here. Furthermore, ξ'_{pq}, η'_{pq} satisfy Eq. (10).

After performing the 30-fold integration in (25), the final answer reads:

$$G^{(15)}(t,t')=(1-b_1-b_2-b_3-b_4)^{-2}, \tag{26}$$

where b_i consists of polynomials of degree $2(i+1)$ in t and t', namely

$$b_1=\sum_{p=0}^{4}\hat{t}_{p\,p+1}\hat{t}_{p-1\,p-2}$$

$[\hat{t}_{pq}$ defined in (14)],

$$b_2=\sum_{p=0}^{4}[\hat{t}_{pp}\hat{t}_{p+1\,p+2}\hat{t}_{p-1\,p-2}+(t'_{pp}t_{p+1\,p+1}\times t'_{p+2\,p+2}t_{p+3\,p+3}t'_{p-1\,p}t_{p-1\,p-2}+t\longleftrightarrow t')],$$

$$\begin{aligned}b_3 &= \sum_{p=0}^{4}\{(\hat{t}_{pp}\hat{t}_{p+1\,p+1}-\hat{t}_{p\,p+1}\hat{t}_{p+1\,p})\hat{t}_{p+2\,p+3}\hat{t}_{p-1\,p-2}\\ &\quad+[t'_{pp}(t_{p+1\,p+1}t_{p+2\,p+1}t'_{p+2\,p+3}\\ &\quad+t_{p+2\,p+2}t_{p+1\,p+2}t'_{p+1\,p})\,t_{p+3\,p+2}t'_{p-2\,p-1}\\ &\quad\times t_{p-1\,p-2}t'_{p-1\,p}+t\longleftrightarrow t']\},\end{aligned}$$

$$b_4=\sum_{p=0}^{4}\hat{t}_{pp}(\hat{t}_{p+1\,p+1}\hat{t}_{p+2\,p+2}-\hat{t}_{p+1\,p+2}\hat{t}_{p+2\,p+1})\times\hat{t}_{p-2\,p-1}\hat{t}_{p-1\,p-2}-\prod_{p=0}^{4}[\hat{t}_{pp}+(t'_{p\,p+1}t_{p+1\,p}+t\longleftrightarrow t')]. \tag{27}$$

C. Consistency check

The statement made under Sec. IIC for the 12j case is valid also for the 15j case.

D. Symmetry (odd-n case)

(a) Define the operation $P_1^{(t)}$ which carries $t_{pq}\longleftrightarrow t'_{pq}$; correspondingly for $P_1^{(k)}$: $k_{pq}\longleftrightarrow k'_{pq}$. It is obvious that $G^{(15)}(t,t')$ is invariant under (28). This implies that the 15j coefficient is invariant under P_1^k.

(28)

(b) Define the operation $P_-^{(t)}=$ permutation $\binom{0\,1\,2\,3\,4}{1\,0\,4\,3\,2}$ on t_{pq} (recall $t'_{pq}\equiv t_{pq'}$). Correspondingly for $P_-^{(k)}$ on k_{pq}. We have $G^{(15)}(t,t')$ invariant under $P_-^{(t)}$, thus the 15j coefficient is invariant under $P_-^{(k)}$.

(29)

(c) Define $P_{\mathfrak{t}}^{(t)}$:

$$t_{pq}\to t'_{p-1\,q-1},\quad t'_{pq}\to t_{p-1\,q-1}\ \text{(mod 5)} \tag{30}$$

and $P_{\mathfrak{k}}^{(k)}$:

$$k_{pq}\to k'_{p+1\,q+1},\quad k'_{pq}\to k_{p+1\,q+1}. \tag{31}$$

Since $G^{(15)}(t,t')$ is invariant under (30), we have the invariance of $\{15j\}$ under (31). The remark following Eq. (18) holds here for $n=5$.

IV. CONCLUDING REMARKS

What we have done is to demonstrate by explicit calculations that the study of the properties of higher-order 3nj angular momentum recoupling coefficients can be carried out in principle for all n. The algebraic complexities, though increasing rapidly with n, turn out still to be controlable. Extraction of the explicit expansion forms for the 3nj coefficients are in principle possible from the generating functions.

The 3nj coefficients ($n\geq 4$) are seen to possess a $4n$-fold symmetry. Visualization of some of the structural properties of 3nj coefficients are greatly enhanced with the aid of a geometric Möbius network representation.

Note added in proof: For graphical method for angular momentum, see also E. El Baz and B. Castel, *Graphical Methods of Spin Algebras* (Dekker, New York, 1972).

*Based in part on a dissertation submitted by C. S. Huang in partial fulfillment of the requirements for the Ph. D. degree at the University of Michigan, 1973 (unpublished).
[1]For an excellent introduction to and collection of the classics, see L. C. Biedenharn and H. Van Dam, Eds., *Quantum Theory of Angular Momentum* (Academic, New York, 1968).
[2]E. P. Wigner, famous unpublished 1940 manuscript, reprinted in Ref. 1.
[3]J. Schwinger, famous unpublished 1952 AEC Report, reprinted in Ref. 1.
[4]V. Bargmann, Rev. Mod. Phys. 34, 829 (1962).
[5]A. C. T. Wu, J. Math. Phys. 13, 84 (1972).
[6]The algebraic details are contained in C. S. Huang, Ph. D. thesis University of Michigan, 1973 (unpublished).
[7]R. J. Ord-Smith, Phys. Rev. 94, 1227 (1954).
[8]T. Regge, Nuovo Cimento 11, 116 (1959).
[9]H. A. Jahn and J. Hope, Phys. Rev. 93, 318 (1954).
[10]A. P. Jucys, I. B. Levinson, and V. V. Vanagas, *Mathematical Apparatus of the Theory of Angular Momentum*, translated from Russian by Israel Program for Scientific Translation, 1962. The authors wish to thank Professor B. G. Wybourne for reminding them of this reference.

Variational formulation of the relativistic theory of microelectromagnetism

Gérard A. Maugin

Université de Paris (VI), Département de Mécanique Théorique, Tour 66 - 4 Place Jussieu, 75005 Paris, France

A. Cemal Eringen

Princeton University, Department of Aerospace and Mechanical Sciences, Princeton, New Jersey, 08540
(Received 6 December 1972)

By adjoining a set of adequate potentials to the classical electromagnetic potential, it is possible to formulate a variational principle that yields the equations of the micromorphic EM theory proposed by Eringen and Kafadar [J. Math. Phys. 11, 1984 (1970)]. The energy-momentum law for micromorphic EM fields is obtained and constitutive equations are derived for relativistic EM-elastic fields.

I. INTRODUCTION

Recently Eringen and Kafadar[1] proposed a set of basic laws for relativistic microelectromagnetism and, as a particular case, introduced the field equations of the microelectromagnetic theory of grade I. This theory was intended for the prediction of physical phenomena involving ferromagnetism, micromagnetism (in the sense of Brown[2]), electrets,[3] microwave propagations, and other related microelectromechanical effects for which the classical continuum hypothesis is violated. This situation arises when the length scale associated with the exciting agents become comparable to the average dimension of "grains" (microelements) in bodies and/or the average distance between grains. Then it is believed that the classical field theories do not contain the necessary mechanism to take into account the local degrees of freedom. The theory of microelectromagnetism is so constructed as to provide the necessary description, still within the frame of continuum formalism (i.e., with "nice" differential equations), the new degrees of freedom being accounted for by new field equations referred to as "moment" equations. In fact, Eringen[4] has shown that: (i) The basic balance laws of phenomenological micromagnetism such as given by, among others, Brown,[2] Tiersten,[5] and Maugin and Eringen,[6] could be derived from microelectromagnetism theory by assuming certain approximations (quasimagnetostatics in a *rigid* body); (ii) in different approximations, the theory contains London's phenomenological equations of superconductivity (there, the new degree of freedom is represented via a local "superconductivity" field $\mathcal{K}$).

By analogy with the mechanical theory developed earlier by Eringen and his co-workers,[7,8] this theory may preferably be referred to as micromorphic EM theory. Indeed Eringen and Kafadar constructed their theory in the frame of the four-dimensional formalism of Minkowskian space–time, and introduced averages of Maxwell's equations in matter over well-defined volumes or hypersurfaces in a manner very similar to that used for constructing the mechanical theory of micromorphic media.[7,8] Though in the latter the concepts of new kinds of forces and couples and of energy arose quite naturally since it was a mechanical theory, the equivalent result for the energy equation in the micromorphic EM theory was not obtained. As a result of this and the lack of thermodynamical considerations, the theory lacked constitutive equations. This fact is not due to indifference or oversimplification; rather, these authors postponed the solution of a problem which involves difficulties inherent to all theories dealing with electromagnetism in matter such as the difficult problem of defining the ponderomotive force and couple, and the energy density and/or the electromagnetic stress–energy–momentum tensor. This, of course, requires in general the consideration of interactions of the electromagnetic fields with a deformable medium. The construction of a complete micromorphic EM theory with mechanical interactions based on the statement of global or local balance laws (direct approach) necessitates the introduction of a *model* for the above-cited interactions. This difficulty may be bypassed if, instead of this general approach, one is satisfied with the study of nondissipative phenomena and thus, considers a variational formulation. Such a formulation is in general possible and yields results in complete agreement with those obtained from a direct approach. For instance, the direct approach of Eringen and Suhubi[7] and the variational one of Maugin[9] lead to comparable results for the theory of nondissipative micromorphic media. The same holds true for the treatments of the classical continuum theory for the interactions of EM fields with deformable bodies provided by Grot and Eringen[10] on the one hand and by Grot[11] on the other. We must, however, note that the identification of the results was possible in the two different approaches because some expressions resulting from the direct approach were known. For the same reason, it has been possible to grant a physical significance to quantities resulting from the variational formulation. The situation is somewhat more annoying for the micromorphic EM theory, for no direct approach is known so far that includes mechanical interactions. It is therefore along this blind alley that we try to construct a simple theory for the interactions of micromorphic electromagnetic fields with deformable bodies, by using a variational principle as starting point and invariance principles as the main tools.

In all rationality, the deformable body considered should be micromorphic of grade I in the mechanical sense[8] too. This would, however, result in cumbersome algebra. Thus, for the sake of simplicity, we shall assume that the material in question has a very simple mechanical behavior, namely that it is hyperelastic, i.e., nonlinear elastic with constitutive equations derivable from a potential, the relativistic strain–energy function.

In this paper we deduce from the variational principle (i) the field equations of micromorphic EM theory of grade I, (ii) the equations of conservation of energy-momentum and of moment of energy—momentum—this follows from the application of Noether's theorem—(iii) constitutive equations for all constitutive variables. At some point, we shall emphasize the analogies with the mechanical micromorphic theory.

II. BACKGROUND

The background of the subsequent developments is the space—time continuum of Minkowski, M^4, equipped with the Lorentz metric of signature $(+,+,+,-)$. The Greek subscripts and superscripts assume the values 1, 2, 3, 4. Small and capital Latin indices assume the values 1, 2, and 3. x^α and X^K, $K=1, 2, 3$, denote, respectively, a curvilinear system of coordinates in M^4 and the Lagrangian coordinates of the reference state in Euclidean space E^3. In an inertial frame, the square of the element of arc in M^4 is given by

$$ds^2=dz_\alpha dz^\alpha,\quad (z^1,z^2,z^3,z^4)=(x,y,z,ict),\quad i\equiv(-1)^{1/2}, \tag{2.1}$$

where z^α are rectangular coordinates, t is the time, and c is the velocity of light in vacuum. Referred to curvilinear coordinates x^α, we have

$$ds^2=g_{\alpha\beta}dx^\alpha dx^\beta, \tag{2.2}$$

where $g_{\alpha\beta}$ is the metric tensor which is normal hyperbolic.

Partial and covariant differentiation with respect to x^α are denoted, respectively, by commas and semicolons or symbols ∇_α. Indices are raised and lowered by the metric tensor $g_{\alpha\beta}$ and its reciprocal $g^{\alpha\beta}$. The summation convention is used throughout the paper. Parentheses around a set of indices denote symmetrization and brackets denote alternation. $\epsilon_{\alpha\beta\gamma\delta}$ is the permutation symbol.

We refer the reader to Grot and Eringen,[10] Kafadar and Eringen,[12] and Maugin and Eringen[13] for a complete description of kinematics of relativistic continua. The elements of kinematics sufficient for the present exposé are the following ones: With τ the proper time of the particle (X^K) initially at the coordinates X^K in E^3, the motion of (X^K) along its worldline $(\mathscr{C}_X\kappa)$ in M^4 is entirely described by the set of relations of class C^2

$$x^\alpha=x^\alpha(X^K,\tau). \tag{2.3}$$

Conversely, we have

$$X^K=X^K(x^\alpha),\quad \tau=\tau(x^\alpha). \tag{2.4}$$

From (2.3) we compute

$$x^\alpha{}_{,K}=\frac{\partial x^\alpha}{\partial X^K},\quad u^\alpha=\frac{\partial x^\alpha}{\partial\tau} \tag{2.5}$$

with

$$g_{\alpha\beta}u^\alpha u^\beta+c^2=0,\quad \frac{\partial}{\partial\tau}\equiv u^\alpha\nabla_\alpha, \tag{2.6}$$

and define the operator of projection $P^\alpha{}_\beta$, onto the hyperplane M_1^3 orthogonal to $(\mathscr{C}_X\kappa)$ at x^α, by the relation

$$P^\alpha{}_\beta=\delta^\alpha_\beta+(1/c^2)u^\alpha u_\beta, \tag{2.7}$$

with

$$P^\alpha{}_\beta P^\beta{}_\gamma=P^\alpha{}_\gamma,\quad P^\alpha{}_\beta u^\beta=0.$$

Finally the direct and inverse deformation gradients of the motion are given by the definitions

$$x^\alpha{}_K=P^\alpha{}_\beta x^\beta{}_{,K},\quad X^K{}_{,\alpha}=\frac{\partial X^K}{\partial x^\alpha}. \tag{2.8}$$

A material body $(B)\subset E^3$ [(B_R) in the reference configuration] of boundary (∂B) sweeps out the tube $(\mathscr{B})\subset M^4$ as time goes on. $(\partial\mathscr{B})$, $(\mathring{\mathscr{B}})=(\mathscr{B})-(\partial\mathscr{B})$, $(\bar{\mathscr{B}})$, $(M^4-\bar{\mathscr{B}})$ denote, respectively, the boundary of $(\mathscr{B})$, the open set that corresponds to $(\mathscr{B})$, the closure of $(\mathscr{B})$, and the complement of $(\bar{\mathscr{B}})$ in M^4. A discontinuity three-dimensional hypersurface (Γ) of unit oriented normal n_α may split $(\mathscr{B})$ into two parts. The familiar symbolism $[\cdots]$ denotes the jump across (Γ) or, sometimes, across $(\partial\mathscr{B})$. The outward unit normal of $(\partial\mathscr{B})$ is denoted by N_α.

Here we briefly recall the salient results of Kafadar and Eringen.[1] The field equations for the micromorphic EM theory of grade I read

$$G^{\alpha\beta}{}_{;\beta}=(1/c)J^\alpha \text{ in } (\mathring{\mathscr{B}}-\Gamma),\quad [G^{\alpha\beta}]n_\beta=(1/c)K^\alpha \text{ on } (\Gamma), \tag{2.9}$$

$$G^{\alpha\beta\lambda}{}_{;\beta}+G^{\alpha\lambda}-\mathscr{G}^{\alpha\lambda}=(1/c)J^{\alpha\lambda} \text{ in } (\mathring{\mathscr{B}}-\Gamma);$$

$$[G^{\alpha\beta\lambda}]n_\beta=(1/c)K^{\alpha\lambda} \text{ on } (\Gamma), \tag{2.10}$$

with the definitions

$G^{\alpha\beta}\equiv$ electric displacement tensor,
$G^{\alpha\beta\lambda}\equiv$ first moment of the electric displacement tensor,
$J^\alpha\equiv$ volume current 4-vector,
$J^{\alpha\lambda}\equiv$ first moment of the volume current,
$\mathscr{G}^{\alpha\lambda}\equiv$ average local electric displacement,
$K^\alpha\equiv$ surface current 4-vector,
$K^{\alpha\lambda}\equiv$ first moment of the surface current.

The first set of Maxwell's equations (2.9) is thus supplemented by the new "moment" equations (2.10). If the symbol $\langle\cdots\rangle_{(n)}$ indicates the average over a manifold of dimension n, then the *moments* $G^{\alpha\beta\lambda}$ and $J^{\alpha\lambda}$ and the field $\mathscr{G}^{\alpha\lambda}$ are defined as[1]

$$G^{\alpha\beta\lambda}\equiv\langle G^{\alpha\beta}\xi^\lambda\rangle_{(2)},\quad J^{\alpha\lambda}\equiv\langle J^\alpha\xi^\lambda\rangle_{(3)},\quad \mathscr{G}^{\alpha\lambda}\equiv\langle G^{\alpha\lambda}\rangle_{(3)}$$

where ξ^λ is the 4-vector that joins the "center" of the manifold element to any point of this manifold element.

The second set of micromorphic field equations is

$$\hat{F}^{\alpha\beta}{}_{;\beta}=0,\quad \hat{\mathfrak{F}}^{\alpha\beta}{}_{;\beta}=0,$$
$$\hat{F}^{\alpha\beta\lambda}{}_{;\beta}+\hat{F}^{\alpha\lambda}-\hat{\mathfrak{F}}^{\alpha\lambda}=0, \tag{2.11}$$

where we have used the definition

$$\hat{\mathbf{F}}=\text{dual}(\mathbf{F}),$$

Thus, in the index notation (2.11) is equivalent to

$$\epsilon^{\alpha\beta\gamma\delta}F_{\gamma\delta;\beta}=0,\quad \epsilon^{\alpha\beta\gamma\delta}\mathfrak{F}_{\gamma\delta;\beta}=0,$$

$$\epsilon^{\alpha\beta\gamma\delta}(F_{\gamma\delta}-\mathfrak{F}_{\gamma\delta})+\epsilon^{\alpha\lambda\gamma\delta}F_{\gamma\delta}{}^\beta{}_{;\lambda}=0,$$

in which

$F_{\alpha\beta}$ ≡ magnetic flux density tensor,
$\mathfrak{F}_{\alpha\beta}$ ≡ average local magnetic flux density tensor,
$F_{\mu\lambda}{}^{\beta}$ ≡ first moment of the magnetic flux density tensor,

with

$$\mathfrak{F}_{\alpha\beta} \equiv \langle F_{\alpha\beta} \rangle_{(3)}, \quad F_{\mu\lambda}{}^{\beta} = \langle F_{\mu\lambda} \xi^{\beta} \rangle_{(2)}.$$

Here again $(2.11)_{2,3}$ are the micromorphic supplements of the Maxwell's equations $(2.11)_1$.

Equations $(2.9)_1$, $(2.10)_1$, and (2.11) are supplemented by the conservation of charge laws

$$J^{\alpha}{}_{;\alpha} = 0, \quad J^{\mu\lambda}{}_{;\mu} + J^{\lambda} - \mathscr{I}^{\lambda} = 0. \tag{2.12}$$

One can show that the following equations hold:

$$\begin{aligned} &\mathscr{E}^{[\mu\lambda]}{}_{;\lambda} = (1/c)(J^{\mu} - J^{[\mu\lambda]}{}_{;\lambda}), \\ &\mathscr{E}^{(\mu\lambda)} = -(1/c) J^{(\mu\lambda)}, \quad \mathscr{E}^{\mu\lambda}{}_{;\mu} = (1/c)\mathscr{I}^{\lambda}, \end{aligned} \tag{2.13}$$

of which the latter provides a definition for $\mathscr{I}^{\lambda}$.

The structure of Eqs. (2.11) shows that we can introduce three sets of potentials to satisfy these equations identically. This is achieved by setting[14]

$$\begin{aligned} &F_{\mu\lambda} = 2\nabla_{[\mu} A_{\lambda]}, \quad \mathfrak{F}_{\mu\lambda} = 2\nabla_{[\mu} \mathfrak{A}_{\lambda]}, \\ &F_{\mu\lambda}{}^{\beta} = 2\nabla_{[\mu} A_{\lambda]}{}^{\beta} + 2\delta^{\beta}_{[\mu}(A_{\lambda]} - \mathfrak{A}_{\lambda]}), \end{aligned} \tag{2.14}$$

where A_{λ} and $\mathfrak{A}_{\lambda}$ are two 4-potentials of which the former is the potential of classical EM theory, and $A_{\lambda}{}^{\beta}$ is a second-order tensor potential. In a Euclidean frame of reference (inertial frame at which the 3-velocity $\mathbf{v} = 0$), these potentials assume the decompositions

$$\begin{aligned} &A_{\mu} = (A_k, i\phi), \quad \mathfrak{A}_{\lambda} = (\mathfrak{A}_k, i\psi), \\ &A_{\mu\lambda} = \begin{pmatrix} A_{kl} & iA_{k(4)} \\ iA_{(4)k} & -A \end{pmatrix}. \end{aligned} \tag{2.15}$$

We note from the relations (2.14) that the potential introduced are not entirely independent. As a consequence, we certainly need not introduce all these potentials in a variational formulation; in fact, hereafter we shall use a set of potentials different from those introduced above and all independent.

III. THE VARIATIONAL PRINCIPLE

A. The Lagrangian density

With the open $(\mathring{\mathscr{B}})$ of M^4 and $(M^4 - \bar{\mathscr{B}})$, we associate the following actions

$$\begin{aligned} &\mathcal{A} = -\int_{(\mathscr{B})} \rho\Psi d^4v + \int_{(\mathscr{B})} \tfrac{1}{4} F_{\alpha\beta} F^{\beta\alpha} d^4v, \\ &\bar{\mathcal{A}} = \int_{(M^4 - \bar{\mathscr{B}})} \tfrac{1}{4} F_{\alpha\beta} F^{\beta\alpha} d^4v \end{aligned} \tag{3.1}$$

Thus, in general, the Lagrangian density is

$$\mathfrak{L} = -\rho\Psi + \tfrac{1}{4} F_{\alpha\beta} F^{\beta\alpha}, \tag{3.2}$$

where ρ is the invariant relativistic density of matter, Ψ is called the relativistic specific internal energy, and $\frac{1}{4} F_{\alpha\beta} F^{\beta\alpha}$ represents the density of the free magnetic field. In the absence of matter, Ψ vanishes. Therefore Ψ represents the interactions of matter with matter and matter with electromagnetic fields. The crucial point here is the selection of the arguments on which Ψ depends. By analogy with the mechanical micromorphic theory[9] and the theory of couple-stresses,[15] and following the guide provided by Mie's theory of electrodynamics[16,17] (that we generalize in a certain sense), we postulate that Ψ depends on the set of basic arguments constituted by the motion x^{α}, the electromagnetic potential A_{α}, and a supplementary set of four 4-potentials $A^{(\xi)}_{\alpha}$, $(\xi) = 1, 2, 3, 4$. Here, the number (ξ) has no tensorial character.[18] For a theory of hyperelastic media and for micromorphism of first order, Ψ will also depend on an adequate set of arguments derived from the basic arguments, namely their first gradients. For instance, we may choose the following functional dependence:

$$\Psi = \Psi(x^{\alpha}, x^{\alpha}{}_{,K}, A_{\alpha}, A_{\alpha;\beta}, A^{(\xi)}_{\alpha}, A^{(\xi)}_{\alpha;\beta}). \tag{3.3}$$

However, to restrict this form, we use three principles commonly accepted nowadays: (i) Ψ must be Lorentz invariant; (ii) the quantities which describe the deformation field must reduce to their classical analogues in a rest frame; (iii) we require Ψ to be gauge invariant, for we are dealing with electromagnetism. The last requirement overcomes the difficulty appearing in Mie's theory.[17] The requirement (ii) indicates that we must consider $x^{\alpha}{}_{K}$ instead of $x^{\alpha}{}_{,K}$. The requirement (iii) which is satisfied if Ψ is invariant under the following group of gauge transformations (Φ: arbitrary scalar function),

$$a_{\alpha} \longrightarrow a_{\alpha} + \nabla_{\alpha}\Phi, \quad \text{for every 4-potential } a_{\alpha},$$

rules out the explicit dependence on A_{α} and $A^{(\xi)}_{\alpha}$. Furthermore Ψ can depend on $A_{\alpha;\beta}$ and $A^{(\xi)}_{\alpha;\beta}$ only through the combinations $A_{[\alpha;\beta]}$ and $A^{(\xi)}_{[\alpha;\beta]}$. We shall set

$$F_{\alpha\beta} \equiv 2\nabla_{[\alpha} A_{\beta]}, \quad \mathfrak{F}^{(\xi)}_{\alpha\beta} \equiv 2\nabla_{[\alpha} A^{(\xi)}_{\beta]}. \tag{3.4}$$

It follows that the two following equations are identically satisfied:

$$\epsilon^{\alpha\beta\gamma\delta} F_{\gamma\delta;\beta} = 0, \quad \epsilon^{\alpha\beta\gamma\delta} \mathfrak{F}^{(\xi)}_{\gamma\delta;\beta} = 0. \tag{3.5}$$

With the requirement (i), the invariance of Ψ under space–time translations rules out the explicit dependence on x^{α}. Finally, as it has been shown that using $x^{\alpha}{}_{K}$ or $X^{K}{}_{,\alpha}$ was equivalent,[10] we consider a relativistic internal energy of the form

$$\Psi = \Psi(X^{K}{}_{,\alpha}, F_{\alpha\beta}, \mathfrak{F}^{(\xi)}_{\alpha\beta}). \tag{3.6}$$

In absence of electromagnetic micromorphism, i.e., for $A^{(\xi)}_{\alpha} = 0$, for every (ξ), we recognize the function used by Grot[11] and Maugin[19] in special and general relativity.

It remains to study the invariance of (3.6) under the rotation members of the Lorentz group. An infinitesimal Lorentz transformation in M^4 is described by the mapping

$$x^{*\alpha} = (\delta^{\alpha}_{\beta} + \epsilon Q^{\alpha}{}_{\beta}) x^{\beta} + d^{\alpha}, \tag{3.7}$$

where ϵ, d^{α}, and $Q^{\alpha\beta}$ are, respectively, an infinitesimally small constant, an infinitesimal constant 4-vector, and a second-order skew-symmetric constant tensor.

For $d^{\alpha} = 0$, we get from Eq. (3.7)

$$\delta^{\alpha} = \epsilon Q^{\alpha}{}_{\beta} x^{\beta}, \quad \delta M^{\alpha\beta} = 2\epsilon Q^{[\alpha}{}_{\gamma} M^{|\gamma|\beta]}, \tag{3.8}$$

where δ indicates the variation resulting from such a transformation. Here $(3.8)_2$ gives the infinitesimal variation of a second order skew-symmetric tensor $M^{\alpha\beta}$ under infinitesimal 4-rotations in M^4. The invariance of Ψ is thus written

$$\Psi(X^K{}_{,\alpha}, F_{\alpha\beta}, \mathcal{F}^{(\xi)}_{\alpha\beta}) = \hat{\Psi}(X^K{}_{,\alpha} + \delta X^K{}_{,\alpha}, F_{\alpha\beta} + \delta F_{\alpha\beta}, \mathcal{F}^{(\xi)}_{\alpha\beta} + \delta\mathcal{F}^{(\xi)}_{\alpha\beta}). \tag{3.9}$$

Using the relations (3.7) in the right-hand-side of Eq. (3.9) and noting that the left-hand side does not depend on $Q^{\alpha\beta}$, we have

$$\frac{\partial\hat{\Psi}}{\partial Q^{\alpha\beta}} = 0,$$

i.e.,

$$\frac{\partial\Psi}{\partial X^{K,[\gamma}} X^K{}_{,\delta]} + 2\frac{\partial\Psi}{\partial F^{[\gamma|\beta|}} F^{\beta}_{\delta]} + 2\sum_{(\xi)} \frac{\partial\Psi}{\partial\mathcal{F}^{(\xi)[\gamma|\beta|}} \mathcal{F}^{(\xi)\beta}_{\delta]} = 0. \tag{3.10}$$

B. The variational principle

According to the general scheme for a variational principle in continuum physics[20] and following the tradition established by Lagrange and Piola, we introduce indeterminate multipliers for each term that can arise in varying the basic arguments in Ψ (i.e., x^α, A_α, $A^{(\xi)}_\alpha$) in $(\mathring{\mathscr{B}})$, on $(\partial\mathscr{B})$ and on (Γ). We therefore express the proposed variational principle as

$$\delta\mathcal{A} + \delta\bar{\mathcal{A}} + \delta W = 0, \tag{3.11}$$

where

$$\begin{aligned}\delta W = &\int_{(\mathring{\mathscr{B}}-\Gamma)} \rho f^\alpha \delta x_\alpha \, d^4v + \int_{(\partial\mathscr{B}-\Gamma)} T^\alpha \delta x_\alpha \, d^3s + \int_{(\mathring{\mathscr{B}}-\Gamma)} \\ &\times (1/c) J^\alpha \hat{\delta} A_\alpha d^4v \\ &- \int_{(\Gamma)} (1/c) K^\alpha \hat{\delta} A_\alpha \, d^3s_\Gamma - \int_{(\Gamma)} (1/c) \\ &\times \Big(\sum_{(\xi)} K^{(\xi)\alpha} \hat{\delta} A_{(\xi)\alpha}\Big) d^3s_\Gamma \\ &+ \int_{(\mathring{\mathscr{B}}-\Gamma)} \Big(\sum_{(\xi)} \mathcal{J}^{(\xi)\alpha} \hat{\delta} A_{(\xi)\alpha}\Big) d^4v.\end{aligned} \tag{3.12}$$

Here we have considered a discontinuity hypersurface (Γ) in $(\mathscr{B})$. The physical significance granted to the multipliers is the following:

$f^\alpha \equiv$ applied specific body-force 4-vector,
$T^\alpha \equiv$ applied surface-traction 4-vector,
$J^\alpha \equiv$ volume 4-current,
$K^\alpha \equiv$ surface 4-current prescribed on (Γ),
$\mathcal{J}^{(\xi)\alpha} \equiv$ volume 4-microcurrent relative to the field $\mathcal{F}^{(\xi)}_{\alpha\beta}$,
$K^{(\xi)\alpha} \equiv$ surface 4-microcurrent relative to the field $\mathcal{F}^{(\xi)}_{\alpha\beta}$.

f^α and T^α are not due to electromagnetic causes, e.g., f^α and T^α may stand for gravity and a mechanical stress vector due to pressure respectively. For the process to be nondissipative, J^α must be due to convection currents only. That is, in a rest frame, the Joule term $\mathbf{J}\cdot\mathbf{E}$ is zero (for example, this is the case when the conduction current is a homogeneous function of degree one in $\mathbf{E}$). A similar condition must hold for $\mathcal{J}^{(\xi)\alpha}$; however, we must admit that the physical significance of such a requirement is not clear. The notation $\hat{\delta}$ used in Eq. (3.12) is explained below.

C. The variation

We assume that the variation (3.11) is to be carried out under the constraints (3.5) which can be written in integral form using the language of exterior calculus[21] as

$$\begin{aligned}&\int_{(\mathcal{S}^2)} F_{\alpha\beta}\, \mathbf{d}x^\alpha \wedge \mathbf{d}x^\beta = \int_{(\partial\mathcal{S}^2)} A_\alpha \mathbf{d}x^\alpha, \\ &\int_{(\mathcal{S}^2)} \mathcal{F}^{(\xi)}_{\alpha\beta}\, \mathbf{d}x^\alpha \wedge \mathbf{d}x^\beta = \int_{(\partial\mathcal{S}^2)} A^{(\xi)}_\alpha \mathbf{d}x^\alpha,\end{aligned} \tag{3.13}$$

for all $\mathbf{A}^{(\xi)}$, (ξ) fixed, where $(\mathcal{S}^2)$ is an arbitrary two-dimensional hypersurface whose boundary is $(\partial\mathcal{S}^2)$. $\mathbf{d}$ indicates the one-form basis. Equations (3.13) express that $F_{\alpha\beta}$ and $\mathcal{F}^{(\xi)}_{\alpha\beta}$ are closed forms. Following Weiss[22] and Grot,[11] it is not difficult to show that Eqs. (3.13) yield the following variations of the fields:

$$\begin{aligned}&\delta F_{\alpha\beta} = 2(\hat{\delta}A_{[\beta})_{;\alpha]} - 2F_{\gamma[\beta}(\delta x^\gamma)_{;\alpha]}, \\ &\delta\mathcal{F}^{(\xi)}_{\alpha\beta} = 2(\hat{\delta}A^{(\xi)}_{[\beta})_{;\alpha]} - 2\mathcal{F}^{(\xi)}_{\gamma[\beta}(\delta x^\gamma)_{;\alpha]},\end{aligned} \tag{3.14}$$

where the Weiss-gauge-invariant variation[22] $\hat{\delta}A_\alpha$ (similarly for $\hat{\delta}A^{(\xi)}_\alpha$) has been defined by

$$\hat{\delta}A_\alpha = \delta A_\alpha - A_{\gamma;\alpha}\delta x^\gamma.$$

The variation (3.11) is thus carried out by varying x^α and $F_{\alpha\beta}$ and $\mathcal{F}^{(\xi)}_{\alpha\beta}$, the latters being according to Eqs. (3.14). Finally, in order to preserve the identity of a material particle (X^K), the following obvious constraint is imposed:

$$\delta X^K = 0 \quad \text{in } (\mathscr{B}).$$

The fields A_α and $A^{(\xi)}_\alpha$ are assumed to be of class C^1 throughout $(\mathscr{B})$. The following necessary intermediate variations have been computed elsewhere[11,13]:

$$\begin{aligned}&\delta X^K{}_{,\alpha} = -X^K{}_{,\beta}(\delta x^\beta)_{;\alpha}, \\ &\delta\rho = -\rho P^\alpha{}_\beta (\delta x^\beta)_{;\alpha}, \\ &\delta(d^4v) = \delta^\alpha_\beta (\delta x^\beta)_{;\alpha}.\end{aligned} \tag{3.15}$$

Furthermore, we define the following quantities:

$$t^{\beta\alpha} \equiv -\rho\frac{\partial\Psi}{\partial X^K{}_{,\beta}} X^{K,\alpha}, \tag{3.16}$$

$$T^{\alpha\beta}_{(\text{em.v})} \equiv -F^\alpha{}_\gamma F^{\gamma\beta} + \tfrac{1}{4}F_{\mu\nu}F^{\nu\mu}g^{\alpha\beta} \tag{3.17}$$

$$[= T^{\alpha\beta} \text{ in } (M^4 - \bar{\mathscr{B}})],$$

$$G^{\alpha\beta} \equiv F^{\alpha\beta} + \rho\frac{\partial\Psi}{\partial F_{\alpha\beta}}, \tag{3.18}$$

$$\mathfrak{C}^{\alpha\beta}_{(\xi)} \equiv \rho\frac{\partial\Psi}{\partial\mathcal{F}^{(\xi)}_{\alpha\beta}}, \tag{3.19}$$

$$T^{\alpha\beta} = (1/c^2)\rho\Psi u^\alpha u^\beta - t^{\beta\alpha} - F^{\alpha\gamma}G_\gamma{}^\beta + \tfrac{1}{4}F_{\mu\nu}F^{\nu\mu}g^{\alpha\beta} - \sum_{(\xi)} \mathcal{F}^{(\xi)\alpha\gamma}\mathfrak{C}_{(\xi)\gamma}{}^\beta. \tag{3.20}$$

They represent, respectively, the relativistic stress tensor, the electromagnetic stress–energy–momentum tensor in vacuum, the electric displacement tensor, the *microelectric* displacement tensor, and the *total* stress–energy–momentum tensor of the material

medium in $(\mathscr{B})$.

By using (3.14)—(3.16), the variation $\delta\mathcal{A}+\delta\bar{\mathcal{A}}$ can be expressed as

$$\delta\mathcal{A}+\delta\bar{\mathcal{A}}=\int_{(\mathscr{B})}\{T^{\alpha\beta}(\delta x_\alpha)_{;\beta}+G^{\alpha\beta}(\hat{\delta}A_\alpha)_{;\beta}+\sum_{(\xi)}\mathbb{G}_{(\xi)}^{\ \alpha\beta}(\hat{\delta}A_\alpha^{(\xi)})_{;\beta}\}d^4v+\int_{(M^4-\bar{\mathscr{B}})}\{T^{\alpha\beta}_{(\mathrm{em.v})}(\delta x_\alpha)_{;\beta}+F^{\alpha\beta}(\hat{\delta}A_\alpha)_{;\beta}\}d^4v. \tag{3.21}$$

Integrating (3.21) by parts and using the generalized Stokes' theorem[23] [in order to take account of the presence of (Γ)], we obtain the following expression for the variational principle (3.11):

$$\begin{aligned}\int_{(\mathring{\mathscr{B}}-\Gamma)}&\Big\{\Big(-T^{\alpha\beta}{}_{;\beta}+\rho f^\alpha\Big)\delta x_\alpha-\Big(G^{\alpha\beta}{}_{;\beta}-\frac{1}{c}J^\alpha\Big)\hat{\delta}A_\alpha\\&-\sum_{(\xi)}\Big(\mathbb{G}_{(\xi)}{}^{\alpha\beta}{}_{;\beta}-\frac{1}{c}\mathscr{I}_{(\xi)}^{\alpha}\Big)\hat{\delta}A_\alpha^{(\xi)}\,d^4v\\&+\int_{(M^4-\bar{\mathscr{B}})}\Big(-T^{\alpha\beta}_{(\mathrm{em.v});\beta}\,\delta x_\alpha-F^{\alpha\beta}{}_{;\beta}\hat{\delta}A_\alpha\Big)d^4v\\&-\int_{(\partial\mathscr{B}-\Gamma)}\Big\{\Big([T^{\alpha\beta}]N_\beta-T^\alpha\Big)\delta x_\alpha+[G^{\alpha\beta}]N_\beta\hat{\delta}A_\alpha\\&+\sum_{(\xi)}\mathbb{G}_{(\xi)}^{\alpha\beta}N_\beta\hat{\delta}A_\alpha^{(\xi)}\Big\}d^3s\\&+\int_{(\Gamma)}\Big\{[T^{\alpha\beta}]n_\beta\,\delta x_\alpha+\Big([G^{\alpha\beta}]n_\beta-\frac{1}{c}K^\alpha\Big)\hat{\delta}A_\alpha\\&+\sum_{(\xi)}\Big([\mathbb{G}_{(\xi)}^{\alpha\beta}]n_\beta-\frac{1}{c}K_{(\xi)}^{\alpha}\Big)\hat{\delta}A_\alpha^{(\xi)}\Big\}d^3s_\Gamma=0.\end{aligned} \tag{3.22}$$

D. Field equations

1. Electromagnetic field equations

We posit Eq. (3.22) to be valid for any variations $\hat{\delta}A_\alpha$ and $\hat{\delta}A_\alpha^{(\xi)}$ and for any region in $(\mathring{\mathscr{B}})$ and any hypersurface $(\partial\mathscr{B})$ and (Γ). Thus we have

$$\begin{aligned}&G^{\alpha\beta}{}_{;\beta}=(1/c)J^\alpha\ \text{ in }(\mathring{\mathscr{B}}-\Gamma),\\&[G^{\alpha\beta}]N_\beta=0\ \text{ on }(\partial\mathscr{B}-\Gamma),\\&F^{\alpha\beta}{}_{;\beta}=0\ \text{ in }(M^4-\bar{\mathscr{B}}),\\&[G^{\alpha\beta}]n_\beta=(1/c)K^\alpha\ \text{ on }(\Gamma),\end{aligned} \tag{3.23}$$

and

$$\begin{aligned}&\mathbb{G}_{(\xi)}^{\alpha\beta}{}_{;\beta}=(1/c)\mathscr{I}_{(\xi)}^{\alpha}\ \text{ in }(\mathring{\mathscr{B}}-\Gamma),\\&[\mathbb{G}_{(\xi)}^{\alpha\beta}]N_\beta=0\ \text{ on }(\partial\mathscr{B}-\Gamma),\\&[\mathbb{G}_{(\xi)}^{\alpha\beta}]n\beta=(1/c)K_{(\xi)}^{\alpha}\ \text{ on }(\Gamma),\end{aligned} \tag{3.24}$$

for every $(\xi)=1,2,3,4$. Equations (3.23) are the usual Maxwell's equations in matter and vacuum. It remains to show that we can deduce from Eqs. (3.24) the "moment" equations which supplement Eqs. (3.23) in micromorphic EM theory. This is dealt with in Sec. 4.

2. Dynamical field equations

We now apply Noether's theorem for the group of infinitesimal Lorentz transformations (3.7) by selecting the special variations

$$\delta x^\alpha=d^\alpha\ \text{ and }\ \delta x^\alpha=\epsilon Q^{\alpha\beta}x_\beta. \tag{3.25}$$

For any region in $(\mathscr{B})$ and any hypersurface $(\partial\mathscr{B})$ and (Γ), $(3.25)_1$ yields the field equations which express the conservation of momentum and energy

$$\begin{aligned}&T^{\alpha\beta}{}_{;\beta}=\rho f^\alpha\ \text{ in }(\mathring{\mathscr{B}}-\Gamma),\\&T^{\alpha\beta}_{(\mathrm{em.v})\,;\beta}=0\ \text{ in }(M^4-\bar{\mathscr{B}}),\\&[T^{\alpha\beta}]N_\beta=T^\alpha\ \text{ on }(\partial\mathscr{B}-\Gamma),\\&[T^{\alpha\beta}]n_\beta=0\ \text{ on }(\Gamma).\end{aligned} \tag{3.26}$$

Considering $(3.25)_2$ for any skew-symmetric $Q^{\alpha\beta}$, constant throughout $(\mathscr{B})$, and taking account of $(3.26)_1$, we obtain[24]

$$T^{[\alpha\beta]}=0\ \text{ in }(\mathring{\mathscr{B}}). \tag{3.27}$$

By construction, this is also true in $(M^4-\bar{\mathscr{B}})$, cf. Eq. (3.17). Upon using Eqs. (3.16) through (3.20), it is shown that Eq. (3.27) is nothing but Eq.[25] (3.10) $(\rho\neq0)$.

We must adjoin Eq. (3.5) and appropriate equations for the conservation of charge. Equations, (3.26)—(3.27) are supplemented with the well-known continuity equations

$$\begin{aligned}&(\rho u^\alpha)_{;\alpha}=0\ \text{ in }(\mathring{\mathscr{B}}-\Gamma),\\&[\rho u^\alpha]n_\alpha=0\ \text{ on }(\Gamma).\end{aligned} \tag{3.28}$$

We have thus obtained the full set of field equations that govern the behavior of *nondissipative elastic materials* which are *micromorphic of grade I from the electromagnetic point of view.*

IV. THEORY OF ERINGEN AND KAFADAR

In this section we transform the sets of equations (3.24) and (3.5) in order to arrive at the formulation given by Eringen and Kafadar (Ref. 1, Sec. 5). For instance, consider Eq. (3.24), multiply each member by $A^{(\xi)\gamma}$ and sum over (ξ). After integration by parts and addition of the vanishing quantity $G^{\alpha\gamma}-G^{\alpha\gamma}$ to the left-hand side, we obtain

$$G^{\alpha\beta\gamma}{}_{;\beta}+G^{\alpha\gamma}-\mathscr{G}^{\alpha\gamma}=(1/c)J^{\alpha\gamma}, \tag{4.1}$$

in which we have set

$$G^{\alpha\beta\gamma}\equiv\sum_{(\xi)}\mathbb{G}_{(\xi)}^{\alpha\beta}A^{(\xi)\gamma},\quad G^{\alpha\beta\gamma}=-G^{\beta\alpha\gamma}, \tag{4.2}$$

$$J^{\alpha\gamma}\equiv\sum_{(\xi)}\mathscr{I}_{(\xi)}^{\alpha}A^{(\xi)\gamma}, \tag{4.3}$$

$$\mathscr{G}^{\alpha\gamma}\equiv G^{\alpha\gamma}+\sum_{(\xi)}\mathbb{G}_{(\xi)}^{\alpha\beta}A^{(\xi)\gamma}_{;\beta}. \tag{4.4}$$

The skew-symmetric and symmetric parts of Eq. (4.1) are then written

$$\begin{aligned}&G^{[\alpha\ |\beta|\gamma]}{}_{;\beta}+G^{\alpha\gamma}-\mathscr{G}^{[\alpha\gamma]}=(1/c)J^{[\alpha\gamma]},\\&\mathscr{G}^{(\alpha\gamma)}=-(1/c)J^{(\alpha\gamma)}.\end{aligned} \tag{4.5}$$

Taking the divergence of Eq. $(4.5)_1$ with respect to x^γ yields on account of Eq. $(3.23)_1$

$$\mathscr{G}^{[\alpha\gamma]}{}_{;\gamma}=1/c\,(J^\alpha-J^{[\alpha\gamma]}{}_{;\gamma}), \tag{4.6}$$

which is Eq. $(2.13)_1$.

If we assume that each microcurrent 4-vector satisfies the usual law of conservation of charge,

$$\mathscr{J}^{(\xi)\alpha}{}_{;\alpha}=0, \quad (\xi)=1,2,3,4, \tag{4.7}$$

and multiply this equation by $A_{(\xi)}^{\gamma}$ and sum over (ξ), we obtain

$$J^{\alpha\gamma}{}_{;\alpha}+J^{\gamma}-\mathscr{J}^{\gamma}=0, \tag{4.8}$$

where

$$\mathscr{J}^{\gamma}\equiv J^{\gamma}+\sum_{(\xi)}\mathscr{T}^{(\xi)\alpha}A_{(\xi)}{}^{\gamma}{}_{;\alpha}. \tag{4.9}$$

With Eqs. (4.1) and (4.8), this gives

$$\mathscr{G}^{\gamma\alpha}{}_{;\gamma}=-(1/c)\mathscr{J}^{\alpha}. \tag{4.10}$$

A transformation similar to that performed on Eq. $(3.24)_1$ can be achieved for Eq. $(3.5)_2$, leading to

$$\hat{F}^{\alpha\beta\lambda}{}_{;\beta}+\hat{F}^{\alpha\lambda}-\hat{\mathscr{F}}^{\alpha\lambda}=0, \tag{4.11}$$

where we have defined

$$\hat{F}^{\alpha\beta\lambda}\equiv\sum_{(\xi)}\hat{\mathscr{F}}_{(\xi)}^{\alpha\beta}A^{(\xi)\lambda}, \quad F_{\alpha\beta}{}^{\lambda}=-F_{\beta\alpha}{}^{\lambda},$$

$$\hat{\mathscr{F}}^{\alpha\lambda}=\hat{F}^{\alpha\lambda}+\sum_{(\xi)}\hat{\mathscr{F}}^{(\xi)\alpha\beta}A_{(\xi)}{}^{\lambda}{}_{;\beta}. \tag{4.12}$$

From (4.11) it follows that

$$\hat{\mathscr{F}}^{\lambda\alpha}{}_{;\lambda}=0. \tag{4.13}$$

Finally, since the $A_{(\xi)}^{\gamma}$ are assumed to be of class C^1 throughout $(\mathscr{B})$, Eqs. $(3.24)_{2-3}$ can be written in the forms

$$[G^{\alpha\beta\lambda}]N_{\beta}=0 \text{ on } (\partial\mathscr{B}-\Gamma),$$

$$[G^{\alpha\beta\lambda}]n_{\beta}=(1/c)K^{\alpha\lambda} \text{ on } (\Gamma), \tag{4.14}$$

with the definition

$$K^{\alpha\lambda}\equiv\sum_{(\xi)}K_{(\xi)}^{\alpha}A^{(\xi)\lambda}. \tag{4.15}$$

All equations obtained in the present section are in agreement with those of Eringen and Kafadar. Moreover, constitutive equations have been given for all fields deduced from the variational formulation. We remark that, in absence of electromagnetic micromorphism, i.e., for $A^{(\xi)}_{\alpha}\equiv 0$, $(\xi)=1,2,3,4$, everywhere in $(\mathscr{B})$, the tensor fields $\mathscr{G}^{\alpha\beta}$, $\mathscr{J}^{\alpha}$, and $\mathscr{F}^{\alpha\beta}$ reduce to the classical Maxwellian fields $G^{\alpha\beta}$, J^{α}, and $F^{\alpha\beta}$ while the "moment" fields $G^{\alpha\beta\lambda}$, $J^{\alpha\gamma}$, and $F^{\alpha\beta\lambda}$ vanish identically, thus yielding the theory given by Grot.[11]

V. SPECIAL CASE

As pointed out in Footnote 18, we may consider that only one new 4-potential is needed to describe the micromorphism of the electromagnetic field. Let this be a_{α}. And further we take it to be of constant unit amplitude, i.e., $g^{\alpha\beta}\,a_{\alpha}\,a_{\beta}=1$ throughout $(\mathscr{B})$. This means that we are interested only in the variations of direction of the 4-vector a_{α}. By analogy with the mechanical theories of micropolar media where the tensor fields describing the micromotion are of constant magnitude,[26] we can say that we are now dealing with the micropolar theory of electromagnetism. All sums disappear in the definitions of the "moment" fields and the total stress–energy–momentum tensor defined by (3.20) may be written as

$$T^{\alpha\beta}=(1/c^2)\,\rho\,\Psi u^{\alpha}u^{\beta}-t^{\beta\alpha}-F^{\alpha\gamma}G_{\gamma}{}^{\beta}-F^{\alpha\gamma}{}_{\lambda}G_{\gamma}{}^{\beta\lambda}+\tfrac{1}{4}F_{\mu\nu}F^{\nu\mu}g^{\alpha\beta}, \tag{5.1}$$

thus yielding an explicit form of the momentum and energy densities of the micropolar electromagnetic fields.

[1] A. C. Eringen and C. B. Kafadar, J. Math. Phys. **11**, 1984 (1970).
[2] W. F. Brown, *Magnetoelastic Interactions* (Springer-Verlag, New York, 1966).
[3] F. Gutman, Rev. Mod. Phys. **20**, 457 (1948).
[4] A. C. Eringen, J. Math. Phys. **12**, 1353 (1971).
[5] H. F. Tiersten, J. Math. Phys. **5**, 1298 (1964); J. Math. Phys. **6**, 779 (1965).
[6] G. A. Maugin and A. C. Eringen, J. Math. Phys. **13**, 143 (1972); J. Math. Phys. **13**, 1334 (1972).
[7] A. C. Eringen and E. S. Suhubi, Int. J. Eng. Sci. 2, 189 (1964).
[8] A. C. Eringen, Int. J. Eng. Sci. **8**, 819 (1970).
[9] G. A. Maugin, C.R. Acad. Sci. A **271**, 807 (1970).
[10] R. A. Grot and A. C. Eringen, Int. J. Eng. Sci. **4**, 611 (1966); Int. J. Eng. Sci. **4**, 638 (1966).
[11] R. A. Grot, J. Math. Phys. **11**, 109 (1970).
[12] C. B. Kafadar and A. C. Eringen, Int. J. Eng. Sci. **9**, 307 (1971).
[13] G. A. Maugin and A. C. Eringen, J. Math. Phys. **13**, 1777 (1972).
[14] C. B. Kafadar, private communication.
[15] R. A. Toupin, Arch. Ration. Mech. Anal. **17**, 85 (1964).
[16] G. Mie, Ann. Phys. (Leipz.) **37**, 511 (1912); Ann. Phys. (Leipz.) **39**, 1 (1912);
[17] K. Sen, *Fields and/or Particles* (Academic, New York, 1968), pp. 41–43.
[18] We could introduce the 4-potentials as follows. Let a_{α} be the 4-potential in a submanifold $m^4 \subset (\mathscr{B})$. At point $x_{\alpha}+y_{\alpha}$ (where x_{α} is the "center" of m^4), we have $a_{\beta}=a_{\beta}(x^{\alpha})+a_{\beta}^{(\xi)}\,y_{\xi}+a^{(\xi)(\zeta)}\,y_{\xi}y_{\zeta}+\cdots$. Set $a_{\beta}(x^{\alpha})\equiv A_{\beta}$, $a_{\beta}^{(\xi)}=A_{\beta}^{(\xi)}$, and disregard higher terms in the expansion. If we only take $(\xi)=1$, this means that the expansion is not performed in a submanifold m^4 but along a line in M^4; this is quite natural given the integral representation (3.13).
[19] G. A. Maugin, Ann. Inst. Henri Poincare **15**, 275 (1971).
[20] G. A. Maugin, Ph.D. thesis (Princeton, 1971).
[21] H. Flanders, *Differential Forms with Applications to Physical Sciences* (Academic, New York, 1963).
[22] P. Weiss, Proc. R. Soc. A **169**, 119 (1939).
[23] A. C. Eringen, Ed., *Continuum Physics* (Academic, New York, 1971), Vol. 1, p. 125.
[24] If we use a statement of local balance laws to start with, we shall require the *angular momentum density* $m_{\alpha\beta\gamma}$ defined by $m_{\alpha\beta\gamma}=2T_{\alpha[\beta}x_{\gamma]}$, to satisfy the following conservation law (in absence of spin) $m^{\alpha}{}_{\beta\gamma;\alpha}=L_{\beta\gamma}$, $L_{\beta\gamma}=2\rho f_{[\beta}x_{\gamma]}$. Upon using $(3.26)_1$, this yields (3.27).
[25] An apparent discrepancy by a factor 2 will appear in the two last terms of (3.10). This is because we considered Ψ to be a skew-symmetric function of $F_{\alpha\beta}$ and $\mathscr{F}_{\alpha\beta}^{(\xi)}$ in Sec. 3.3.
[26] A. C. Eringen, "Micropolar Elasticity," in *Fracture, Vol. II*, edited by H. Liebowitz (Academic, New York 1968).

Lattice Green's function of the body-centered cubic lattice at arbitrary points

Shigetoshi Katsura and Hiroshi Watanabe*

Department of Applied Physics, Tohoku University, Sendai, Japan
(Received 15 January 1973)

Lattice Green's function for the body-centered cubic lattice at arbitrary points outside and inside the band is evaluated by the method of analytic continuation using Mellin–Barnes type integral.

1. INTRODUCTION

Recently a method of the analytic continuation using Mellin-Barnes type integral was developed in the calculation of the lattice Green's functions. Body-centered cubic lattice[1] and simple cubic lattice[2] at the origin, square, and rectangular lattices,[3] simple cubic lattice, and tetragonal lattice at arbitrary points,[4] were studied and calculated by this method. Similar problems were studied also by the method of complete elliptic integrals.[5-7]

In this paper, the lattice Green's function of the body-centered cubic lattice at arbitrary points

$$I(a;l,m,n)=\frac{1}{\pi^3}\iiint_0^\pi \frac{\cos lx\cos my\cos nz\,dxdydz}{a-i\epsilon-\cos x\cos y\cos z} \tag{1}$$

is considered. Here $\mathbf{r}\equiv(l,m,n)$ and $\mathbf{k}\equiv(x,y,z)$ denote the lattice point, and momentum. $I(a;l,m,n)$ do not vanish when l,m,n are all even or all odd. The inside and the outside the band are specified by $|a|<1$ and $|a|>1$, respectively. Earlier works on (1) were reviewed in Ref. 8. Recently Joyce discussed $I(a;l,m,n)$ and expressed[7] $I(a;0,2m,2n)$ as a product of ${}_2F_1$ functions and expressed[8] $I(1;l,m,n)$ for $0\leq l,m,n\leq 8$ in terms of the complete elliptic integrals.

In Secs. 2, 3, and 4 of the present paper, $I(a;l,m,n)$ is calculated by the method of the Mellin–Barnes integral. The value inside the band is obtained by the analytic continuation from that outside the band. Results are expressed in terms of generalized hypergeometric functions.

2. OUTSIDE THE BAND

Now we express the integral (1) in terms of the generalized hypergeometric function ${}_pF_q$. First consider the case $a>1$, and l,m,n are all even. We assume $l,m,n\geq 0$ without loss of generality and we put $l\to 2l$, $m\to 2m$, $n\to 2n$. Expanding the integrand in powers of $1/a$, the term-by-term integration gives

$$I(a;2l,2m,2n)=\frac{1}{\pi^3 a}\sum_{p=0}^{\infty}\int_0^\pi \cos 2lx\cos^p x\,dx\int_0^\pi \cos 2my\cos^p y\,dy\int_0^\pi \cos 2nz\cos^p z\,dz\left(\frac{1}{a}\right)^p$$

$$=\frac{\Gamma^3(2M+1)}{(2^{2M})^3\Gamma(M+l+1)\Gamma(M-l+1)\Gamma(M+m+1)\Gamma(M-m+1)\Gamma(M+n+1)\Gamma(M-n+1)}\frac{1}{a^{2M+1}} \tag{2}$$

$$\times {}_7F_6\left[\begin{matrix}M+1,\ M+1,\ M+1,\ M+\frac{1}{2},\ M+\frac{1}{2},\ M+\frac{1}{2},\ 1;\\ M+l+1,\ M-l+1,\ M+m+1,\ M-m+1,\ M+n+1,\ M-n+1\end{matrix}\ a^{-2}\right], \tag{3}$$

where $M=\max(l,m,n)$. Equation (3) which is reduced to ${}_6F_5$ is a closed form outside the band.

For $l=m=n=0$, Eq. (3) is reduced to a known form

$$I(a;0,0,0)=a^{-1}\,{}_3F_2(\tfrac{1}{2},\tfrac{1}{2},\tfrac{1}{2};1,1;a^{-2}).$$

For the case $l=0$, ${}_7F_6$ in (3) reduced to ${}_4F_3$. With use of a formula given in Eq. (3.1) in Ref. 3, ${}_4F_3$ thus reduced is transformed into F_4 which can be factored as a product of ${}_2F_1$. We assume $m\geq n$.

$$I(a;0,2m,2n)=Ca^{-2m-1}\,{}_4F_3\left[\begin{matrix}m+\frac{1}{2},\ m+\frac{1}{2},\ m+\frac{1}{2},\ m+1;\\ 2m+1,\ m+n+1,\ m-n+1\end{matrix}\ a^{-2}\right],$$

$$=Ca^{-2m-1}F_4(m+\tfrac{1}{2},\ m+\tfrac{1}{2};\ m+n+1,\ m-n+1;\ 1/4a^2,\ 1/4a^2),$$

$$=Ca^{-2m-1}\,{}_2F_1\left[\begin{matrix}m+\frac{1}{2},m+\frac{1}{2};\\ m+n+1\end{matrix}\ k^2\right]{}_2F_1\left[\begin{matrix}m+\frac{1}{2},m+\frac{1}{2};\\ m-n+1\end{matrix}\ k^2\right], \tag{3'}$$

where

$$C=\frac{[(2m)!]^2}{2^{6m}(m!)^2(m+n)!(m-n)!}$$

$$k^2=\tfrac{1}{2}-\tfrac{1}{2}(1-a^{-2})^{1/2}.$$

Equation (3′) agrees with the result by Joyce.[7]

For the case where l,m,n are all odd, we have in a similar way,

$$I(a;\,2l+1,\,2m+1,\,2n+1) = \frac{\Gamma^3(2M+2)}{(2^{2M+1})^3\Gamma(M+l+2)\Gamma(M-l+1)\Gamma(M+m+2)\Gamma(M-m+1)\Gamma(M+n+2)\Gamma(M-n+1)}\frac{1}{a^{2M+2}}$$

$$\times {}_7F_6\left[\begin{matrix} M+1,\ M+1,\ M+1,\ M+\frac{3}{2},\ M+\frac{3}{2},\ M+\frac{3}{2},\ 1;\ a^{-2} \\ M+l+2,\ M-l+1,\ M+m+2,\ M-m+1,\ M+n+2,\ M-n+1 \end{matrix}\right], \tag{4}$$

where $M=\max(l, m, n)$.

3. INSIDE THE BAND

Mellin–Barnes integral representation of (3) is given by

$$I=\frac{\pi^{-3/2}}{a^{2M+1}}\cdot\frac{1}{2\pi i}$$

$$\times\int_{-\delta-i\infty}^{-\delta+i\infty}\frac{\Gamma^3(t+M+\frac{1}{2})\Gamma^3(t+M+1)}{\Gamma(t+M+l+1)\Gamma(t+M-l+1)\Gamma(t+M+m+1)}$$

$$\times\frac{\Gamma(t+1)\Gamma(-t)(-1/a^2)^t}{\Gamma(t+M-m+1)\Gamma(t+M+n+1)\Gamma(t+M-n+1)}\,dt, \tag{5}$$

where δ is a positive infinitesimal and $|\arg(-1/a^2)|<\pi$. In considering the case inside the band, a is understood to be $a=|a|e^{-i\epsilon}$, hence $-1/a^2=(1/|a^2|)\exp(-i\pi)$.

The path in Eq. (4) is shifted to the left, i.e., put $s=t+M+1$, then we have

$$I=-\pi^{-3/2}a\frac{1}{2\pi i}\int_{1-\delta-i\infty}^{1-\delta+i\infty}$$

$$\frac{\Gamma^3(s-\frac{1}{2})\Gamma^3(s)\Gamma(1-s)\Gamma(s)(-1/a^2)^s}{\Gamma(s+l)\Gamma(s-l)\Gamma(s+m)\Gamma(s-m)\Gamma(s+n)\Gamma(s-n)}\,ds. \tag{6}$$

The poles of $\Gamma^3(s-\frac{1}{2})$ are triple poles at $s=\frac{1}{2}-q$, $q=0, 1, 2, \cdots$. The quadruple poles of $\Gamma^4(s)$ at $s=-q$, $q=0, 1, 2, \ldots$, are partly canceled by the poles of the denominator, and the resulting poles in the integrand are

(1) $s=-q$, $q=0, 1, \cdots, \min(l, m, n)-1$, simple pole,

(2) $s=\frac{1}{2}-q$, $q=0, 1, 2, \cdots$, triple pole.

These poles are shown in Fig. 1. The contribution of these poles is denoted by $I_i(a;\,2l, 2m, 2n)$ $(i=1, 2,)$.

The calculation of the residues of the simple poles is straightforward and the result is

$$I_1(a;\,2l, 2m, 2n)=(-)^{l+m+n}8lmn\,a$$

$$\times{}_6F_5\left[\begin{matrix} 1+l,\ 1-l,\ 1+m,\ 1-m,\ 1+n,\ 1-n;\ a^2 \\ 1,\ 1,\ \frac{3}{2},\ \frac{3}{2},\ \frac{3}{2} \end{matrix}\right]. \tag{7}$$

4. INSIDE THE BAND—CONTINUED

In this section we consider the contribution of the triple poles, i.e., $I_2(a;\,2l, 2m, 2n)$. The residue of the integrand (6) at $s=\frac{1}{2}-q$ is given by

$$-\pi^{-3/2}a\left|\frac{1}{2!}\frac{d^2}{ds^2}\left[(s-\tfrac{1}{2}+q)^3\right.\right.$$

$$\left.\left.\times\frac{\Gamma^3(s-\frac{1}{2})\Gamma^3(s)(\pi/\sin s\pi)(-1/a^2)^s}{\Gamma(s+l)\Gamma(s-l)\Gamma(s+m)\Gamma(s-m)\Gamma(s+n)\Gamma(s-n)}\right]\right|_{s=1/2-q}, \tag{8}$$

$$=-\pi^{-3/2}a\frac{(-1)^q}{2\pi^2}\frac{d^2}{ds^2}[h((s-\tfrac{1}{2}+q)\pi)g(s)]_{s=1/2-q}, \tag{9}$$

where

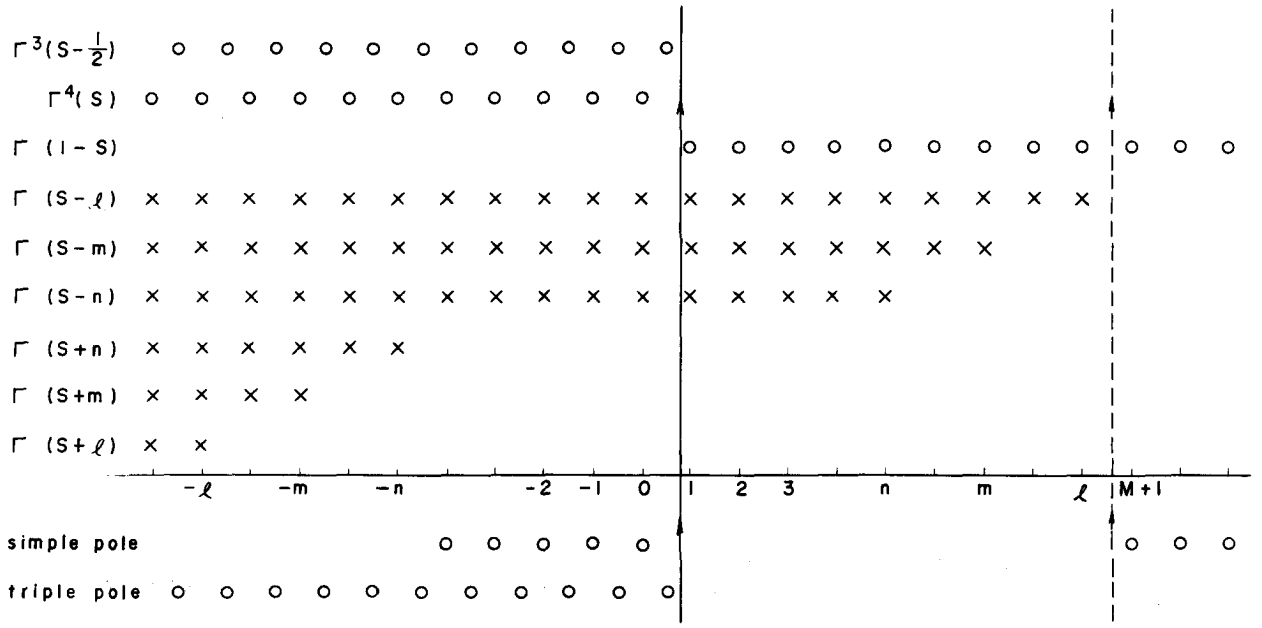


FIG. 1. The poles of the integrand of Eq. (6). Abscissa denotes Re s. ○ pole of each factor in the denominator. × pole of each factor in the nominator. The last two lines below the horizontal line indicate resulting poles in the integrand.

$$g(s)=\frac{\Gamma(1-l-s)\Gamma(1+l-s)\Gamma(1-m-s)\Gamma(1+m-s)\Gamma(1-n-s)\Gamma(1+n-s)\sin^2 s\pi(-1/a^2)^s}{\Gamma^3(\frac{3}{2}-s)\Gamma^3(1-s)} \tag{10}$$

$$h(s)=(\pi s/\sin\pi s)^3.$$

Taking

$$\lim_{s\to 0}h(s)=1,$$

$$\lim_{s\to 0}\frac{dh(s)}{ds}=0,$$

$$\lim_{s\to 0}\frac{d^2h(s)}{ds}=\pi^2$$

into account, we have

$$\text{rhs of (9)}=-\pi^{-3/2}a[(-1)^q/2\pi^2][\pi^2 g(s)+g''(s)]_{s=1/2-q} \tag{11}$$

where

$$\begin{aligned}g'(s)=g(s)[&-\psi(1-l-s)-\psi(1+l-s)-\psi(1-m-s)\\&-\psi(1+m-s)-\psi(1-n-s)\\&-\psi(1+n-s)+2\pi\cot s\pi+\log(-1/a^2)+3\psi(\tfrac{3}{2}-s)\\&+3\psi(1-s)],\end{aligned} \tag{12}$$

$$\begin{aligned}g''(s)=g(s)\{&[-\psi(1-l-s)-\psi(1+l-s)-\psi(1-m-s)\\&-\psi(1+m-s)-\psi(1-n-s)-\psi(1+n-s)\\&+2\pi\cot s\pi+\log(-1/a^2)+3\psi(\tfrac{3}{2}-s)+3\psi(1-s)]^2\\&+[\psi'(1-l-s)+\psi'(1+l-s)+\psi'(1-m-s)\\&+\psi'(1+m-s)+\psi'(1-n-s)+\psi'(1+n-s)\\&-2\pi^2\csc^2 s\pi-3\psi'(\tfrac{3}{2}-s)-3\psi'(1-s)]\}.\end{aligned} \tag{13}$$

Here we define three functions ${}_p\tilde{F}_q$, ${}_pG_q$, and ${}_pH_q$ as

$$\begin{aligned}&{}_6\tilde{F}_5\left[\begin{matrix}a_1,a_2,a_3,a_4,a_5,a_6;\\ b_1,b_2,b_3,b_4,b_5\end{matrix}\,z\right]\\&=\frac{1}{\pi}\sum_{k=0}^{\infty}\frac{(a_1)_k(a_2)_k(a_3)_k(a_4)_k(a_5)_k(a_6)_k}{k!(b_1)_k(b_2)_k(b_3)_k(b_4)_k(b_5)_k}\\&\times z^k[\psi(1+k)+\psi(b_1+k)+\cdots\\&+\psi(b_5+k)-\psi(a_1+k)\\&-\psi(a_2+k)-\cdots-\psi(a_6+k)\quad-\log z],\end{aligned} \tag{14}$$

$$\begin{aligned}&{}_6G_5\left[\begin{matrix}a_1,a_2,a_3,a_4,a_5,a_6;\\ b_1,b_2,b_3,b_4,b_5\end{matrix}\,z\right]\\&=\frac{1}{2\pi^2}\sum_{k=0}^{\infty}\frac{(a_1)_k(a_2)_k(a_3)_k(a_4)_k(a_5)_k(a_6)_k}{k!(b_1)_k(b_2)_k(b_3)_k(b_4)_k(b_5)_k}z^k[\psi(1+k)\\&+\psi(b_1+k)+\cdots+\psi(b_5+k)-\psi(a_1+k)-\psi(a_2+k)\\&-\cdots-\psi(a_6+k)-\log z]^2,\end{aligned} \tag{15}$$

$$\begin{aligned}&{}_6H_5\left[\begin{matrix}a_1,a_2,a_3,a_4,a_5,a_6;\\ b_1,b_2,b_3,b_4,b_5\end{matrix}\,z\right]\\&=\frac{1}{2\pi^2}\sum_{k=0}^{\infty}\frac{(a_1)_k(a_2)_k(a_3)_k(a_4)_k(a_5)_k(a_6)_k}{k!(b_1)_k(b_2)_k(b_3)_k(b_4)_k(b_5)_k}z^k[\psi'(1+k)\\&+\psi'(b_1+k)+\cdots+\psi'(b_5+k)-\psi'(a_1+k)-\psi'(a_2+k)\\&-\cdots-\psi'(a_6+k)].\end{aligned} \tag{16}$$

Then the contribution of residues of triple poles is given by

$$\begin{aligned}I_2(a;\,2l,\,2m,\,2n)=&(-)^{l+m+n}[{}_6\tilde{F}_5+i({}_6G_5-{}_6H_5-{}_6F_5)]\\&\times\left[\begin{matrix}\frac{1}{2}-l,\frac{1}{2}+l,\frac{1}{2}-m,\frac{1}{2}+m,\frac{1}{2}-n,\frac{1}{2}+n;\\ 1,\ 1,\ \frac{1}{2},\ \frac{1}{2},\ \frac{1}{2}\end{matrix}\,a^2\right]\end{aligned} \tag{17}$$

where the arguments

$$\left[\begin{matrix},&,&,&,&,&;\\ ,&,&,&,&\end{matrix}\,a^2\right]$$

of ${}_6\tilde{F}_5$, ${}_6G_5$, ${}_6H_5$, and ${}_6F_5$ are all same.

The final result is given by

$$\begin{aligned}I(a;\,2l,\,2m,\,2n)=&(-1)^{l+m+n}8lmn\,a\,{}_6F_5\\&\times\left[\begin{matrix}1+l,1-l,1+m,1-m,1+n,1-n;\\ 1,\ 1,\ \frac{3}{2},\ \frac{3}{2},\ \frac{3}{2}\end{matrix}\,a^2\right]\\&+(-1)^{l+m+n}[{}_6\tilde{F}_5+i({}_6G_5-{}_6H_5-{}_6F_5)]\\&\times\left[\begin{matrix}\frac{1}{2}-l,\frac{1}{2}+l,\frac{1}{2}-m,\frac{1}{2}+m,\frac{1}{2}-n,\frac{1}{2}+n;\\ 1,\ 1,\ \frac{1}{2},\ \frac{1}{2},\ \frac{1}{2}\end{matrix}\,a^2\right].\end{aligned} \tag{18}$$

In a similar way we have for the case where l, m, n are odd

$$\begin{aligned}&I(a;\,2l+1,\,2m+1,\,2n+1)\\&=-(-1)^{l+m+n}{}_6F_5\left[\begin{matrix}1+l,-l,1+m,-m,1+n,-n;\\ 1,\ 1,\ \frac{1}{2},\ \frac{1}{2},\ \frac{1}{2}\end{matrix}\,a^2\right]\\&+(-)^{l+m+n}(2l+1)(2m+1)(2n+1)a[{}_6\tilde{F}_5+i({}_6G_5-{}_6H_5-{}_6F_5)]\\&\times\left[\begin{matrix}\frac{1}{2}-l,\frac{3}{2}+l,\frac{1}{2}-m,\frac{3}{2}+m,\frac{1}{2}-n,\frac{3}{2}+n;\\ 1,\ 1,\ \frac{3}{2},\ \frac{3}{2},\ \frac{3}{2}\end{matrix}\,a^2\right].\end{aligned} \tag{19}$$

We obtain the leading term for $a\to 0$ from the term of $k=0$. For the even case we have

$$\begin{aligned}&I(a;\,2l,\,2m,\,2n)\\&=(-1)^{l+m+n}[8lmn\,a+(1/\pi)(-6\gamma-6\log 2-\psi(\tfrac{1}{2}-l)\\&-\psi(\tfrac{1}{2}+l)-\psi(\tfrac{1}{2}-m)-\psi(\tfrac{1}{2}+m)-\psi(\tfrac{1}{2}-n)-\psi(\tfrac{1}{2}+n)\\&-\log a^2)\\&+i\{-\tfrac{1}{2}+(1/2\pi^2)(-6\gamma-6\log 2-\psi(\tfrac{1}{2}-l)-\psi(\tfrac{1}{2}+l)\\&-\psi(\tfrac{1}{2}-m)-\psi(\tfrac{1}{2}+m)-\psi(\tfrac{1}{2}-n)-\psi(\tfrac{1}{2}+n)\\&-\log a^2)^2\}]+O(a^2),\end{aligned} \tag{20}$$

where the term $O(a^2)$ contains those of $a^2(\log a)^2$ and $a^2(\log a)$. Thus the nature at $a\sim 0$ is of the logarithmic divergence for the even case. For the odd case,

$$\begin{aligned}&I(a;\,2l+1,\,2m+1,\,2n+1)\\&=(-1)^{l+m+n}[-1+(2l+1)(2m+1)(2n+1)(a/\pi)(3\psi(1)\end{aligned}$$

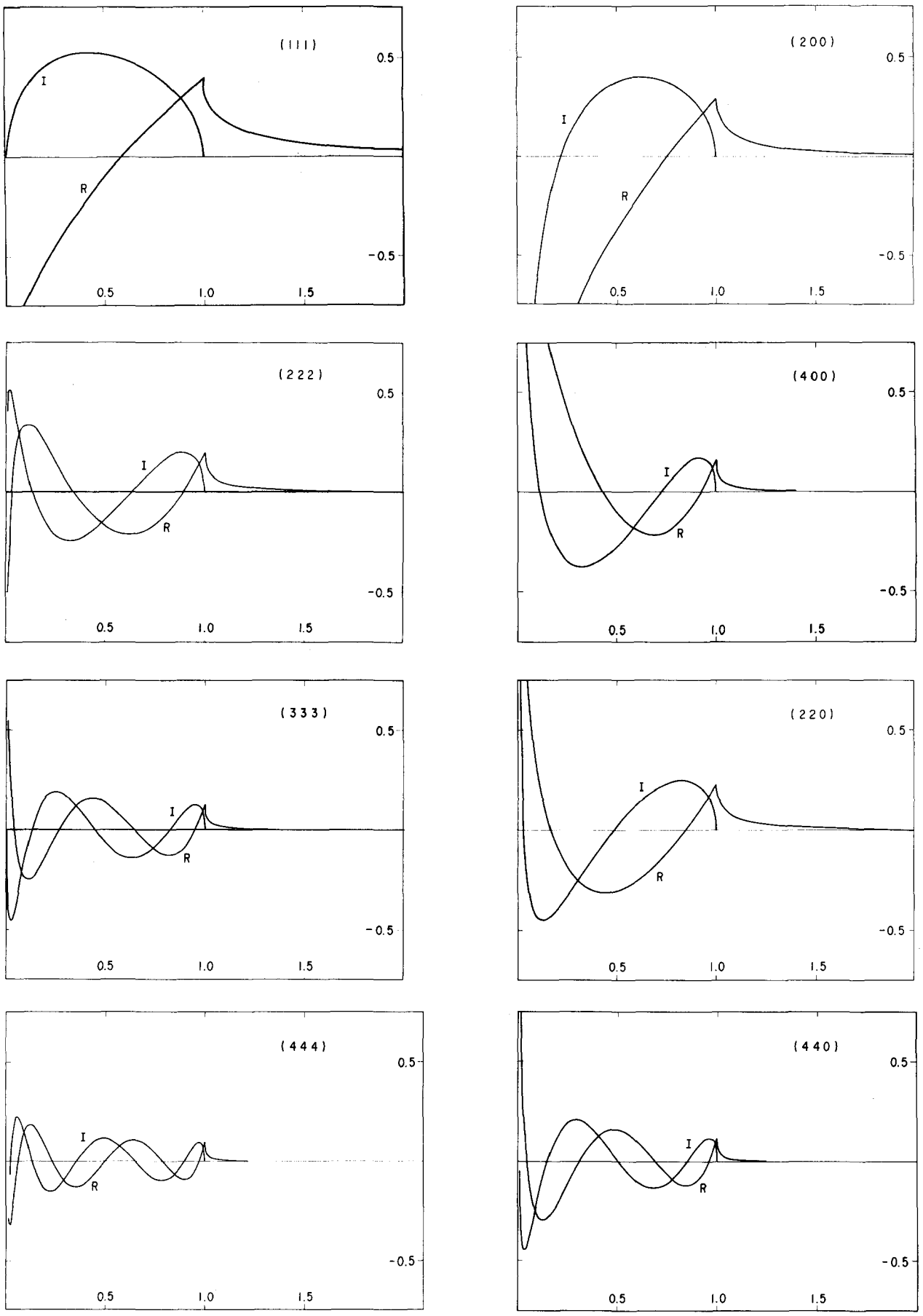


FIG. 2. Values of $I(a;\ l,\ m,\ n)$. R: Real part. I: Imaginary part. The abscissa denotes a.

$$+ 3\psi(\tfrac{3}{2}) - \psi(\tfrac{1}{2}-l) - \psi(\tfrac{3}{2}+l) - \psi(\tfrac{1}{2}-m) - \psi(\tfrac{3}{2}+m)$$
$$- \psi(\tfrac{1}{2}-n) - \psi(\tfrac{3}{2}+n) - \log a^2) + (ia/2\pi^2)(3\psi(1)$$
$$+ 3\psi(\tfrac{3}{2}) - \psi(\tfrac{1}{2}-l) - \psi(\tfrac{3}{2}+l) - \psi(\tfrac{1}{2}-m) - \psi(\tfrac{3}{2}+m)$$
$$- \psi(\tfrac{1}{2}-n) - \psi(\tfrac{3}{2}+n) - \log a^2)^2 - (3\psi'(1) + 3\psi'(\tfrac{3}{2})$$
$$- \psi'(\tfrac{1}{2}-l) - \psi'(\tfrac{3}{2}+l) - \psi'(\tfrac{1}{2}-m) - \psi'(\tfrac{3}{2}+m)$$
$$- \psi'(\tfrac{1}{2}-n) - \psi'(\tfrac{3}{2}+n)) - 2\pi^2\}] + O(a^2), \quad (21)$$

then, we see

$$\mathrm{Re}I(0;\ 2l+1,\ 2m+1,\ 2n+1) = (-1)^{l+m+n+1},$$
$$\mathrm{Im}I(0;\ 2l+1,\ 2m+1,\ 2n+1) = 0.$$

An expansion of $I(a;\ l,\ m,\ n)$ near $a=1$ can be obtained in a similar way as in the case of square lattice.[3] The values $I(1;l,m,n)$ for $0 \leq l, m, n \leq 8$ have been expressed by combinations of elliptic integrals by Joyce.[6] The leading singularity at $a \sim 1$ have been given by Morita and Horiguchi.[9]

Equations (3), (4), (18) (19) supply simple, general, and rapid subroutines for the calculation of the lattice Green's functions of the body-centered lattice at arbitrary points inside and outside the band.

Figure 2 shows some of $\mathrm{Re}I(a;\ l,\ m,\ n)$ and $\mathrm{Im}I(a;\ l,\ m,\ n)$ calculated. Parts near $a \sim 0$, which show very sharp changes, are omitted. In some subfigures, they have another maximum or minimum and tend to $\pm$ infinity with opposite directions. Tables and figures of $I(a;\ l,\ m,\ n)$ for $0 \leq l,\ m,\ n \leq 5$ giving $10D$ values will be supplied on request.

5. CONCLUSION

In this paper lattice Green's functions for the body-centered cubic lattice at arbitrary points were calculated. The value for the outside of the band is continued analytically and gives that for the inside of the band. The method of Mellin–Barnes integrals is adopted. It gives general expressions and subroutines for arbitrary l, m, and n, and does not require repetitious use of recurrence relations of $I(a;l,\ m,\ n)$.

ACKNOWLEDGMENTS

The authors would like to thank Dr. Inawashiro and Y. Abe for their helpful discussions.

*On leave of absence of Miyagi Technical College, Natori, Miyagi, Japan.

[1]S. Katsura and T. Horiguchi, J. Math. Phys. 12, 230 (1971).
[2]S. Katsura, S. Inawashiro, and Y. Abe, J. Math. Phys. 12, 895 (1971); S. Inawashiro, S. Katsura, and Y. Abe, J. Math. Phys. 14, 560 (1973).
[3]S. Katsura and S. Inawashiro, J. Math. Phys. 12, 1622 (1971).
[4]Y. Abe and S. Katsura, Ann. Phys. 75, 348 (1973).
[5]T. Morita and T. Horiguchi, J. Math. Phys. 12, 981 (1971); T. Morita, J. Math. Phys. 12, 1744 (1971); T. Horiguchi, J. Phys. Soc. Japan 30, 1261 (1971).
[6]G.S. Joyce, J. Math. Phys. 12, 1390 (1971).
[7]G.S. Joyce, J. Phys. C 1, 1510 (1971).
[8]S. Katsura, T. Morita, S. Inawashiro, T. Horiguchi, and Y. Abe, J. Math. Phys. 12, 892 (1971).
[9]T. Morita and T. Horiguchi, J. Phys. A 5, 67 (1972).

Multigroup replication property for external, spherically symmetric problems of transport theory

R. Żelazny and T. Duracz

Institute of Nuclear Research, Świerk by Otwock, Poland
(Received 5 February 1973)

The replication property for multigroup spherically symmetric external problems in the transport theory is derived and applied to reduce the system of multigroup integral transport equations to a system of planelike singular integral equations, which can be solved by means of well-known methods.

1. INTRODUCTION

In this paper we propose a method of derivation of planelike singular integral equations for spherically symmetric, external problems in transport theory. By external problems we understand problems in which the concentric internal sphere is filled by a black (completely absorbing) material. The outer shell extends from the radius of the black sphere R to the outer radius R_1. For the simplicity we consider that the scattering of particles is isotropic, the medium of the outer shell is homogeneous and there are no sources of particles in it. Presented results can be applied as well in the neutron transport theory as in the astrophysical problems, wherever linear transport is applicable. The proposed method is an extension of the known and already applied method of replication property for the integral transport equation to new and up to now not explored situations. The replication property has been introduced for internal spherical and cylindrical problems.[1,2] This method, as concerns final results, is equivalent with the method of integral transforms, used originally only for internal problems.[3] Recently after deriving our results we have found that the integral transform method has also been used for particular external problems[4] ($R_1=\infty$, monoenergetic case).

Some analogies to our ideas presented here one may find also in the paper of Sahni.[5] However the explicite formulation of the idea of "replication property" for spherically symmetric external problems seems to be presented for the first time in this paper and, moreover, from the very beginning for the multigroup case.

2. THE REPLICATION PROPERTY

The N-group integral transport equation in the case of isotropic scattering and a homogeneous medium, extended from the radius of the internal black sphere R to the outer radius R_1, has the form

$$n_i(r)=\sum_{j=1}^{N}C_{ij}\int_R^{R_1}dr'n_j(r')\int_0^1\frac{d\nu}{\nu}\left(\exp\left(-\frac{\sigma_i}{\nu}|r-r'|\right)-\exp\left\{-\frac{\sigma_i}{\nu}[(r^2-R^2)^{1/2}+(r'^2-R^2)^{1/2}]\right\}\right),\quad i=1,\ldots,N, \tag{1}$$

where $n_i(r)/r$ is the density and σ_i is the total cross section for the ith group, and C_{ij} describes the transfer from the jth group to the ith one. The groups are numbered as follows (assuming $1/\sigma_N$ as a unit length): $\sigma_1>\sigma_2>\ldots>\sigma_N=1$. Inserting into the rhs of Eq. (1)

$$n_j(r)=\eta_j e^{r/\mu} \tag{2}$$

we obtain

$$\text{rhs}=[(a)+(b)-(c)]\sum_{j=1}^{N}C_{ij}\eta_j,$$

where

$$(a)=\int_R^r dr'e^{r'/\mu}\int_0^1\frac{d\nu}{\nu}e^{(\sigma_i/\nu)(r'-r)},$$

$$(b)=\int_r^{R_1}dr'\,e^{r'/\mu}\int_0^1\frac{d\nu}{\nu}e^{(\sigma_i/\nu)(r-r')},$$

$$(c)=\int_R^{R_1}dr'\,e^{r'/\mu}\int_0^1(d\nu/\nu)\exp\{-(\sigma_i/\nu)[(r^2-R^2)^{1/2}+(r'^2-R^2)^{1/2}]\}.$$

Integrating over r' in (a) and (b), we find

$$(a)=e^{r/\mu}\int_0^1\frac{d\nu}{\nu}\left(\frac{\sigma_i}{\nu}+\frac{1}{\mu}\right)^{-1}-\int_0^1 d\nu\exp\left[-(\sigma_i/\nu)r\right]\frac{1}{\nu}\left(\frac{\sigma_i}{\nu}+\frac{1}{\mu}\right)^{-1}\exp\left[R\left(\frac{\sigma_i}{\nu}+\frac{1}{\mu}\right)\right],$$

$$(b)=\int_0^1 d\nu\,e^{(\sigma_i/\nu)r}\frac{1}{\nu}\left(\frac{1}{\mu}-\frac{\sigma_i}{\nu}\right)^{-1}\exp\left[R_1\left(\frac{1}{\mu}-\frac{\sigma_i}{\nu}\right)\right]-e^{r/\mu}\int_0^1\frac{d\nu}{\nu}\left(\frac{1}{\mu}-\frac{\sigma_i}{\nu}\right)^{-1}.$$

For $-1/\sigma_i\le\mu\le 0$ in (a) and $0\le\mu\le 1/\sigma_i$ in (b) the integrals are singular; they will be understood in the principal value sense.

Since [Ref. 6, Sec. 4, 17, Eq. (8)]

$$\exp[-(\sigma_i/\nu)(r^2-R^2)^{1/2}]=\exp[-(\sigma_i/\nu)r]+\sigma_i R\int_0^{\nu}\frac{d\epsilon}{\epsilon}\times\frac{I_1[(R\sigma_i/\nu\epsilon)(\nu^2-\epsilon^2)^{1/2}]}{(\nu^2-\epsilon^2)^{1/2}}e^{-(\sigma_i/\epsilon)r}$$

(I_1 is the modified Bessel function of the first kind), we have

$$(c)=(d)+(e),$$

where

$$(d)=\int_0^1 d\nu\exp[-(\sigma_i/\nu)r]\frac{1}{\nu}\int_R^{R_1}dr'\exp[(r'/\mu)-(\sigma_i/\nu)(r'^2-R^2)^{1/2}],$$

$$(e)=\sigma_i R\int_0^1\frac{d\nu}{\nu}\int_R^{R_1}dr'\exp[(r'/\mu)-(\sigma_i/\nu)(r'^2-R^2)^{1/2}]\times\int_0^{\nu}\frac{d\epsilon}{\epsilon}\frac{I_1[(R\sigma_i/\nu\epsilon)(\nu^2-\epsilon^2)^{1/2}]}{(\nu^2-\epsilon^2)^{1/2}}\times\exp[-(\sigma_i/\epsilon)r].$$

Interchanging the order of integration in (e) we obtain

$$(e)=\int_0^1 d\epsilon \exp[-(\sigma_i/\epsilon)r]\frac{\sigma_i R}{\epsilon}\int_\epsilon^1 \frac{d\nu}{\nu}\frac{I_1[(R\sigma_i/\nu\epsilon)(\nu^2-\epsilon^2)^{1/2}]}{(\nu^2-\epsilon^2)^{1/2}}$$
$$\times\int_R^{R_1} dr' \exp[r'/\mu-\sigma_i/\nu(r'^2-R^2)^{1/2}].$$

Finally we find

$$\sum_{j=1}^N C_{ij}\int_R^{R_1} dr'\,\eta_j \exp[(r'/\mu)]\int_0^1 (d\nu/\nu)[\exp\{-(\sigma_i/\nu)|r-r'|\}$$
$$-\exp\{-(\sigma_i/\nu)[(r^2-R^2)^{1/2}+(r'^2-R^2)^{1/2}]\}]$$
$$=\{\alpha_i(\mu)\exp[(r/\mu)]+\int_0^1 d\nu[\beta_i(\nu,\mu)\exp[(\sigma_i/\nu)r]$$
$$+\gamma_i(\nu,\mu)\exp\{-(\sigma_i/\nu)r\}]\}$$
$$\times\sum_{j=1}^N C_{ij}\eta_j, \tag{3}$$

where

$$\alpha_i(\mu)=\int_0^1 \frac{d\nu}{\nu}\left[\left(\frac{\sigma_i}{\nu}+\frac{1}{\mu}\right)^{-1}+\left(\frac{\sigma_i}{\nu}-\frac{1}{\mu}\right)^{-1}\right],$$

$$\beta_i(\nu,\mu)=\frac{1}{\nu}\left(\frac{1}{\mu}-\frac{\sigma_i}{\nu}\right)^{-1}\exp[R_1(1/\mu-\sigma_i/\nu)], \tag{4}$$

$$\gamma_i(\nu,\mu)=-\frac{1}{\nu}\left(\frac{1}{\mu}+\frac{\sigma_i}{\nu}\right)^{-1}\exp[R(1/\mu+\sigma_i/\nu)]$$
$$-\omega_i(\nu,\mu), \tag{5}$$

where

$$\omega_i(\nu,\mu)=\omega_i'(\nu,\mu)+\omega_i''(\nu,\mu),$$

$$\omega_i'(\nu,\mu)=\frac{1}{\nu}\int_R^{R_1} dx\exp[x/\mu-\sigma_i/\nu(x^2-R^2)^{1/2}],$$

$$\omega_i''(\nu,\mu)=\frac{\sigma_i R}{\nu}\int_\nu^1\frac{d\epsilon}{\epsilon}\frac{I_1[(R\sigma_i/\nu\epsilon)(\epsilon^2-\nu^2)^{1/2}]}{(\epsilon^2-\nu^2)^{1/2}}$$
$$\times\int_R^{R_1} dx\exp[x/\mu-\sigma_i/\epsilon(x^2-R^2)^{1/2}].$$

In (4) for $0\le\mu\le 1/\sigma_i$ and in (5) for $-1/\sigma_i\le\mu\le 0$, respectively, $\beta_i(\nu,\mu)$ and $\gamma_i(\nu,\mu)$ are treated as principal values, in accordance with the previous remark about singular integrals. The formula (3) describes "the replication property": expression (2) inserted into the rhs of Eq. (1) "is replicated" as a sum of expressions of the same type (with appropriate coefficients).

3. APPLICATION FOR DERIVATION OF PLANELIKE SINGULAR INTEGRAL EQUATIONS

The replication property suggest that we look for a solution of Eq. (1) of the form:

$$n_i(r)=\sum_{\mu\in D} a_i(\mu)e^{r/\mu}+\int_0^1 d\mu\, b_i(\mu)e^{r/\mu}+\int_0^1 d\mu\, d_i(\mu)e^{-r/\mu}, \tag{6}$$

where D denotes a set of complex numbers. Inserting the expansion (6) in Eq. (1) and applying the replication property we obtain:

$$\sum_{\mu\in D} a_i(\mu)e^{r/\mu}+\int_0^1 d\mu\, b_i(\mu)e^{r/\mu}+\int_0^1 d\mu\, d_i(\mu)e^{-r/\mu}$$
$$=\sum_{\mu\in D}(\alpha_i(\mu)e^{r/\mu}+\int_0^1 d\nu\{\beta_i(\nu,\mu)\exp[(\sigma_i/\nu)r]$$
$$+\gamma_i(\nu,\mu)\exp[-(\sigma_i/\nu)r]\}\sum_{j=1}^N C_{ij}a_j(\mu)$$
$$+\int_0^1 d\mu[\alpha_i(\mu)\exp[-r/\mu]+\int_0^1 d\nu[\beta_i(\nu,\mu)\exp(\sigma_i r/\nu)$$
$$+\gamma_i(\nu,\mu)\exp(-\sigma_i r/\nu)]]\sum_{j=1}^N C_{ij}b_j(\mu)$$
$$+\int_0^1 d\mu(\alpha_i(-\mu)e^{-(r/\mu)}+\int_0^1 d\nu\{\beta_i(\nu,-\mu)\exp[(\sigma_i/\nu)r]$$
$$+\gamma_i(\nu,-\mu)\exp[-(\sigma_i/\nu)r]\})\sum_{j=1}^N C_{ij}d_j(\mu).$$

Equating now coefficients of $e^{r/\mu}$ we find

$$a_i(\mu)=\alpha_i(\mu)\sum_{j=1}^N C_{ij}a_j(\mu),\quad \mu\in D, \tag{7}$$

$$b_i(\mu)=\sum_{\nu\in D}\sigma_i\beta_i(\sigma_i\mu,\nu)\sum_{j=1}^N C_{ij}a_j(\nu)$$
$$+\alpha_i(\mu)\sum_{j=1}^N C_{ij}b_j(\mu)+\int_0^1 d\nu\,\sigma_i\beta_i(\sigma_i\mu,\nu)\sum_{j=1}^N C_{ij}b_j(\nu)$$
$$+\int_0^1 d\nu\,\sigma_i\beta_i(\sigma_i\mu,-\nu)\sum_{j=1}^N C_{ij}d_j(\nu),\quad \mu\in[0,1], \tag{8}$$

$$d_i(\mu)=\sum_{\nu\in D}\sigma_i(\sigma_i\mu,\nu)\sum_{j=1}^N C_{ij}a_j(\nu)+\int_0^1 d\nu\,\sigma_i\gamma_i(\sigma_i\mu,\nu)$$
$$\times\sum_{j=1}^N C_{ij}b_j(\nu)+\alpha_i(-\mu)\sum_{j=1}^N C_{ij}d_j(\mu)$$
$$+\int_0^1 d\nu\,\sigma_i\gamma_i(\sigma_i\mu,-\nu)\sum_{j=1}^N C_{ij}d_j(\nu),\quad \mu\in[0,1], \tag{9}$$

where

$$\beta_i(\nu,\mu)=\gamma_i(\nu,\mu)=0 \text{ for } |\nu|>1,\ |\mu|<1$$

[definitions (4) and (5) are for $|\nu|<1,\ |\mu|<1$].

Let us introduce the notation

$$\theta_i(x)=\begin{cases}1, & |x|\le 1/\sigma_i,\\ 0, & |x|>1/\sigma_i.\end{cases}$$

$\mathbf{a}$ is a vector with components a_i. $\hat{A}$ is a matrix with elements A_{ij}.

Defining

$$\Omega_{ij}(\mu)=\delta_{ij}-\alpha_i(\mu)C_{ij},$$
$$P_{ij}(\mu,\nu)=\sigma_i\beta_i(\sigma_i\mu,\nu)C_{ij}=[\nu/(\mu-\nu)]\times\exp[R_1(1/\nu$$
$$-1/\mu)C_{ij}\theta_i(\mu),$$
$$Q_{ij}(\mu,\nu)=\sigma_i\gamma_i(\sigma_i\mu,\nu)C_{ij}=-\{[\nu/(\mu+\nu)]\exp[R(1/\nu$$
$$+1/\mu)+\sigma_i\omega_i(\sigma_i\mu,\nu)\}C_{ij}\theta_i(\mu),$$

we can write Eqs. (7)—(9) as

$$\hat{\Omega}(\mu)\bar{a}(\mu)=0,\quad \mu\in D, \tag{10}$$

$$\hat{\Omega}(\mu)\bar{b}(\mu)=\int_0^1 d\nu[\hat{P}(\mu,\nu)\bar{b}(\nu)+\hat{P}(\mu,-\nu)\bar{d}(\nu)]$$
$$+\sum_{\nu\in D}P(\mu,\nu)\bar{a}(\nu),\quad \mu\in[0,1], \tag{11}$$

$$\hat{\Omega}(\mu)\bar{d}(\mu)=\int_0^1 d\nu[\hat{Q}(\mu,-\nu)\bar{d}(\nu)+\hat{Q}(\mu,\nu)\bar{b}(\nu)]$$
$$+\sum_{\nu\in D}\hat{Q}(\mu,\nu)\bar{a}(\nu),\quad \mu\in[0,1]. \tag{12}$$

Equations (10)—(12) for coefficients of the expansion (6) form a system of singular integral equations (11) and (12) and a system of linear equations (10). The condition of existence of nontrivial solutions of the equation (10) determines the set D:

$$\det\hat{\Omega}(\mu)=0,\quad \mu\in D. \tag{13}$$

Note, that

$$\int_0^1 d\nu\,[b_i(\nu)\,e^{r/\nu} + d_i(\nu)\,e^{-r/\nu}] < \infty \quad \text{for } R \le r \le R_1$$

and let us introduce new functions

$$b'_i(\mu) = b_i(\mu)\,\mu \exp[(R_1/\mu), \quad d'_i(\mu) = d_i(\mu)\,\mu \exp[-(R/\mu),$$

which are integrable now.

Let us finally rewrite the system of singular integral equations (11) and (12) in the more convenient form

$$\hat{\Omega}(\mu)\overline{b}'(\mu) + \hat{S}(\mu)\int_0^1 d\nu\,\frac{\overline{b}'(\nu)}{\nu-\mu} + \overline{\mathrm{Reg}}^b(\mu) = 0, \tag{14}$$

$$\hat{\Omega}(\mu)\overline{d}'(\mu) + \hat{S}(\mu)\int_0^1 d\nu\,\frac{\overline{d}'(\nu)}{\nu-\mu} + \overline{\mathrm{Reg}}^d(\mu) = 0, \tag{15}$$

where

$$S_{ij}(\mu) = C_{ij}\theta_i(\mu)\mu$$

and "regular" terms are the following:

$$\mathrm{Reg}_i^b(\mu) = \theta_i(\mu)\mu\sum_{j=1}^N C_{ij}\int_0^1 \frac{d\nu}{\mu+\nu}\exp[-(1/\nu)(R_1-R)]d'_j(\nu),$$

$$+\ \theta_i(\mu)\mu\sum_{j=1}^N C_{ij}\sum_{\nu\in\Omega}\frac{\nu}{\nu-\mu}\exp[(R_1/\nu)]a_j(\nu),$$

$$\mathrm{Reg}_i^d(\mu) = \theta_i(\mu)\mu\sum_{j=1}^N C_{ij}\int_0^1 d\nu\left[\left(\frac{1}{\nu+\mu}\exp[-(1/\nu)(R_1-R)]\right.\right.$$

$$\left.+\frac{\sigma_i}{\nu}\omega_i(\sigma_i\mu,\nu)\exp[-(R_1/\nu)-(R/\mu)]\right)b'_j(\nu)$$

$$\left.+\frac{\sigma_i}{\nu}\omega_i(\sigma_i\mu,-\nu)\exp[R(1/\nu-1/\mu)\,d'_j(\nu)\right]$$

$$+\ \theta_i(\mu)\mu\sum_{j=1}^N C_{ij}\sum_{\nu\in\Omega}\left(Te^{R/\nu}\frac{\nu}{\nu+\mu}+\sigma_i\omega_i(\sigma_i\mu,\nu)\right.$$

$$\left.\times\exp[-(R/\mu)]\right)a_j(\nu).$$

4. CONCLUSIONS

Matrices $\Omega(\mu)$ and $\hat{S}(\mu)$ of dominant parts of singular integral equations (14) and (15) as well as Eq. (10) are identical with those appearing for plane problems.[7,8] One may then say, that in this sense the spherical problems have been reduced to corresponding, equivalent plane problems. Thus, many of the considerations concerning plane problems, e.g.,[9,10] are relevant also to Eqs. (14) and (15). In the particular cases $N=1$ and 2 there exist vast amount of references, which are well known and will be not quoted in a form of a complete list. Interested readers are refered to the already quoted monograph[2] (mainly cases $N=1$), references connected with two-group problems may be found in Refs. 9 and 10. Methods developed for these particular cases may be adopted easily to solve Eqs. (14) and (15).

As the particular cases we get for $R=0$ an internal problem, for $R_1<\infty$ a critical problem, and for $R_1=\infty$ the Milne problem.

[1] K.M. Case in *Developments in Transport Theory*, edited by E. Inönü and P.F. Zweifel (Academic, New York, 1967).
[2] K.M. Case and P.F. Zweifel, *Linear Transport Theory* (Addison-Wesley, Reading, Mass., 1967).
[3] F.J. Mitsis, "Transport Solutions to the Monoenergetic Critical Problems," Report No. ANL-6787, Argonne National Laboratory, Argonne, Ill. (1963).
[4] O.J. Sheaks, J. Math. Phys. 13, 203 (1972).
[5] D.C. Sahni, J. Nucl. Energy 20, 915 (1966).
[6] *Tables of Integral Transforms* (McGraw-Hill, New York, 1954), Vol. I.
[7] R. Zelazny and A. Kuszell, "Physics of Fast and Intermediate Reactors," *Proceedings of a Seminar* (International Atomic Energy Agency, Vienna, 1962), Vol. II, p. 55.
[8] T. Yoshimura and S. Katsuragi, Nucl. Sci. Eng. 33, 297 (1968).
[9] E.E. Burniston, C.E. Siewert, P. Silvennoinen, and P.F. Zweifel, Nucl. Sci. Eng. 45, 331 (1971).
[10] P. Silvennoinen and P.F. Zweifel, J. Math. Phys. 13, 1114 (1972).

Existence and uniqueness of solutions to Low's problem

I. P. Nedelkov

Institute of Physics at the Bulgarian Academy of Sciences, Sofia, Bulgaria
(Received 24 August 1971; final revised manuscript received 21 August 1972)

In the framework of S-matrix theory, the partial scattering amplitudes are sought as a solution of a certain problem involving analyticity, unitarity, and crossing symmetry. This problem, with a condition of analyticity which is weaker than the usual one, is called Low's problem in this paper. By means of the fixed-point theorems of Schauder and Banach–Cacciopoli, conditions for the existence and uniqueness of solutions to Low's problem are given.

I. INTRODUCTION

In this paper we shall discuss the question of the existence and uniqueness of the solutions of a certain problem from S-matrix theory. The problem is to find N functions $h^\alpha(z)$, $\alpha=1,2,\ldots,N$ (the partial scattering amplitudes) of the complex variable $z=x+iy$ which satisfy the following conditions: (a) analyticity in some subregion of the plane z; (b) unitarity; (c) crossing symmetry; (d) reality; (e) a condition on behavior at infinity. This problem (a)–(e), which will be formulated more precisely below, is called Low's problem in this paper. It is a generalization of the problem solved by means of the integral equations of Low,[1] Chew and Mandelstam,[2] Shirkov[3,4] and the like.

We shall make use of the fact that the problem (a)–(e) can be re-formulated as the algebraic system (5). Although this system is nonlinear and infinite, it is in some respects sufficiently simple and can be investigated by means of the fixed-point theorems.[5,6] Following this method we shall prove, with the help of Schauder's theorem, the existence of solutions of (5). After imposing additional limitations it will be shown on the basis of the Banach—Cacciopoli theorem that these solutions are unique.

Recently several authors have shown interest in similar questions. For instance, Warnock[7] and MacDaniel and Warnock[8,9] have studied the conditions under which there exist solutions of Low's integral equation, while in Refs. 10 and 11 Atkinson has made a detailed mathematical analysis of the integral equation of Chew and Mandelstam, and of Shirkov *et al.*, respectively.

These authors examine the question of the existence and uniqueness of the solutions $h^\alpha(z)$ of the integral equations with the assumption that $h^\alpha(z)$ have at most one pole in the cut plane z.

Some of the results they obtained are less general than those obtained in the present work, because here it is supposed that $h^\alpha(z)$ may have not only poles but also more complicated singularities, e.g., cuts.

The approach in this work differs from the usual approach by the way in which the analytical functions are represented. For instance, in the integral equation of Low, the functions $h^\alpha(z)$ are represented through the Cauchy integral, while here Laurent's series are being used.

The algebraic approach has some peculiarities which manifest themselves both in the theoretical studies and in the numerical calculations (Refs. 12–14).

Because of the specific features of the algebraic system (5) it is appropriate to use the conventional methods of nonlinear functional analysis, such as Newton's method and the principle of contracting mapping[5,6] as, for example, applied in the Low amplitude method.[13,14] On the other hand, the integral equations of dispersive type are solved numerically exclusively by means of the N/D method or the inverse Low amplitude method,[8] which techniques are specific for that class of problems. The theorems proved in this work are a more precise version of the theorems of Ref. 15. They justify the applicability of the numerical methods of Refs. 12–14.

In Sec. II, the precise formulation of Low's problem is given. Besides that, it is shown that under certain conditions it is equivalent to the algebraic system (5). In Sec. III, by means of Schauder's theorem, the existence of solutions of the system (5) is proved. In Sec. IV, using the theorem of Banach—Cacciopoli, other conditions have been found guaranteeing both the existence and the uniqueness of the solutions of (5).

II. FORMULATION OF LOW'S PROBLEM

Here we shall summarize the basic results from the papers,[13,14] which we shall need in our further work.

By Low's problem we mean the problem in which N functions $h^\alpha(z)$, $\alpha=1,2,\ldots,N$ of the complex variable $z=x+iy$ are sought which obey the following conditions:

(a) *Analyticity*: $h^\alpha(z)$ are analytic in $p-s_I^\alpha$, where the region p is the plane z from which the points belonging to the cuts $-\infty\leqslant x\leqslant -1$ and $1\leqslant x\leqslant\infty$ have been taken away, and the closed regions s_I^α is a subregion of the region p.

(b) *Unitarity*: $\mathrm{Im}h^\alpha(x)=f(x)|h^\alpha(x)|^2$, $1\leqslant x\leqslant\infty$, where $f(x)$ is a real function the properties of which are specified below.

(c) *Crossing symmetry*: $h^\alpha(-z)=\sum_{\beta=1}^{N} C^{\alpha\beta}h^\beta(z)$, where the crossing matrix $C^{\alpha\beta}$ is equal to the square root of the unit N-dimensional matrix, but otherwise is arbitrary.

(d) *Reality*: $h^{*\alpha}(z)=h^\alpha(z^*)$.

(e) *Behavior at infinity*: The integrals in (1) converge. The contribution of the contour integrals $\int h^\alpha(z)dz/z$ taken on a semicircle with an infinite radius in the upper half-plane is zero.

The problem (a)–(e) is a generalization of the problem which is solved by means of Low's integral equation.[1,8,9]

$$h^{\alpha}(z)=\frac{\lambda^{\alpha}}{z}+\frac{1}{\pi}\int_{1}^{\infty}dz' f(z')\left(\frac{|h^{\beta}(z')|^2}{z'-z}+\frac{\Sigma_{\beta=1}^{N}C^{\alpha\beta}|h^{\beta}(z)|^2}{z'+z}\right), \tag{1}$$

where $C^{\alpha\beta}$, $\alpha,\beta=1,2,\ldots,N$ is the crossing matrix and $\lambda_{\alpha}=-\Sigma_{\beta=1}^{N}C^{\alpha\beta}\lambda_{\beta}$ are numbers proportional to the coupling constant f^2.

With an appropriate choice of λ_{α} and $C^{\alpha\beta}$ one could describe by means of (1) the partial scattering amplitudes of various processes, e.g., of the $\pi-N$ scattering.[1,16,17]

By the conformal mapping

$$z=2Z/1+Z^2, \tag{2}$$

where $Z=Y+iY=Re^{i\varphi}$, the cut plane p goes over into the interior P of the unit circle C_0 of the Z plane, the functions $h^{\alpha}(z)$ are transformed into the functions $H^{\alpha}(Z)$, the regions s_i^{α} into the regions S_i^{α}.

The regions S_i^{α} contain all singularities of $H^{\alpha}(Z)$ which lie inside C_0. By analogy we shall denote by S_e^{α} closed regions which contain all singularities of $H^{\alpha}(Z)$ lying outside C_0. [All singularities in S_e^{α} were situated on the second sheet of $h^{\alpha}(z)$ before the conformal mapping. Some of them correspond to the resonances, if any.]

Let the curves dS_i^{α} and dS_e^{α} denote the boundaries of S_i^{α} and S_e^{α}, respectively. The functions $H^{\alpha}(Z)$ are analytic in the annular regions D^{α} which are bounded from the inside by the curves dS_i^{α} and from the outside by the curves dS_e^{α}.

For several purposes instead of regions D^{α} their subregions D_c^{α} are preferred. The D_c^{α} are defined as the circular rings $R_i^{\alpha}<|Z|<R_e^{\alpha}$, $R_i^{\alpha}\leq 1$, $R_e^{\alpha}\geq 1$, the R_i^{α} being the radii of the circles $|Z|=R_i^{\alpha}$ which are tangent to the curves dS_i^{α} and R_e^{α}, the radii of the circles $|Z|=R_e^{\alpha}$ which are tangent to the curves dS_e^{α}.

After the conformal transformation the problem (a)–(e) turns into a problem for the transformed functions $H^{\alpha}(Z)$. This problem, after some generalization, will be formulated in the following way.

Find the functions $H^{\alpha}(Z)$, $\alpha=1,2,\ldots,N$ which satisfy the conditions:

(A) *Analyticity*: $H^{\alpha}(Z)$, $Z\in D^{\alpha}$ are analytic.

(B) *Unitarity*: $\mathrm{Im}H^{\alpha}(Z)=F(\varphi)|H^{\alpha}(\varphi)|^2$, where $F(\varphi)=f(1/\cos\varphi)$, $-\frac{1}{2}\pi\leq\varphi\leq\frac{1}{2}\pi$ and $H^{\alpha}(\varphi)=H^{\alpha}(e^{i\varphi})$.

(C) *Crossing symmetry*: $H^{\alpha}(\varphi+\pi)=\Sigma_{\beta=1}^{N}C^{\alpha\beta}H^{\beta}(\varphi)$, $-\pi\leq\varphi\leq\pi$.

(D) *Reality*: $H^{\alpha *}(Z)=H^{\alpha}(Z^*)$.

Further on we shall suppose that $H^{\alpha}(\varphi)$ are Hölder continuous. Under this notion we shall mean functions $H^{\alpha}(\varphi)$, which satisfy the conditions:

$$H^{\alpha}(\varphi+2\pi)=H^{\alpha}(\varphi),\quad \alpha=1,2,\ldots,N,$$
$$|H^{\alpha}(\varphi_2)-H^{\alpha}(\varphi_1)|\leq K|\varphi_2-\varphi_1|^{\epsilon},\quad \text{where } K>0 \text{ is a suitable constant and } 0<\epsilon\leq 1 \text{ and } -\pi-\eta\leq\varphi_1,$$
$$\varphi_2\leq\pi+\eta,\ \eta>0. \tag{3}$$

Under this hypothesis the functions $\mathrm{Re}H^{\alpha}(\varphi)$ and $\mathrm{Im}H^{\alpha}(\varphi)$ are Hölder continuous and coincide with their Fourier series. This is sufficient to assert that $H^{\alpha}(Z)$, $|Z|=1$ can be expanded in Laurent series

$$H^{\alpha}(Z)=\sum_{n=-\infty}^{\infty}H_n^{\alpha}Z^n,\quad |Z|=1. \tag{4}$$

Taking into account the conditions (B), (C), and (D), one can derive the following algebraic system which is to be satisfied by the unknown coefficients H_n^{α}

$$H_{\nu}^{\alpha}=H_{-\nu}^{\alpha}+\sum_{m,k=-\infty}^{\infty}F(\nu,k)E_{\nu}^{\alpha}(H_m;H_{m+k}),\quad \alpha=1,2,\ldots,N,\quad \nu=1,2,\ldots,\infty, \tag{5}$$

where

$$F(\nu,k)=\pi^{-1}\int_{-\pi/2}^{\pi/2}d\nu\sin\nu\varphi\cos k\varphi\,F(\varphi) \tag{6}$$

and

$$E_{\nu}^{\alpha}(H_m;H_{m+k})=H_m^{\alpha}H_{m+k}^{\alpha}+(-1)^{\nu}\sum_{\beta=1}^{N}C^{\alpha\beta}H_m^{\beta}H_{m+k}^{\beta}.$$

The system (5) has been derived in Ref. 14.

The following theorem is based on the corresponding theorem in Ref. 14. It clarifies the equivalence between the analytical formulation of Low's problem through the conditions (A), (B), (C), (D) and its algebraic formulation given by the system (5).

Theorem 1: Let the functions $H^{\alpha}(\varphi)$, $\alpha=1,2,\ldots,N$ satisfy the conditions (3), (B), (C), and (D) and let $F(\varphi)$ satisfy the condition

$$|F(\varphi_2)-F(\varphi_1)|\leq K_1|\varphi_2-\varphi_1|^{\epsilon_1},\quad -\tfrac{1}{2}\pi\leq\varphi_1,\varphi_2\leq\tfrac{1}{2}\pi,$$
where $K_1>0$ is a suitable constant and $0<\epsilon_1\leq 1$
$$F(\pm\tfrac{1}{2}\pi)=0. \tag{7}$$

Then the coefficients of the series (4) H_n^{α}, $\alpha=1,2,\ldots,N$, $n=0,\pm1,\pm2,\ldots,\pm\infty$ will satisfy the algebraic system (5).

With certain modifications of the theorem the opposite assertion is also true:

Let the system (5) have real roots H_n^{α}, $\alpha=1,2,\ldots,N$, $n=0,\pm1,\pm2,\ldots,\pm\infty$ satisfying the following conditions:

The series $\sum_{n=1}^{\infty}H_n^{\alpha}\sin n\varphi$ and $\sum_{n=1}^{\infty}H_{-n}^{\alpha}\sin n\varphi$, $\alpha=1,2,\ldots,N$, converge on the whole interval $-\pi\leq\varphi\leq\pi$ to certain functions $V_{+}^{\alpha}(\varphi)$ and $V_{-}^{\alpha}(\varphi)$, respectively, which are known to satisfy the Hölder condition with the exponent ϵ, $0<\epsilon<1$ on the interval $[-\pi-\eta,\pi+\eta]$, where η is some positive number (8)

$$H_n^{\alpha}=(-1)^n\sum_{\beta=1}^{N}C^{\alpha\beta}H_n^{\beta},\quad \alpha=1,2,\ldots,N,\quad n=0,-1,-2,\ldots,-\infty, \tag{9}$$

and let $F(\varphi)$ satisfy the condition (7).

Then the series (4) converge to the functions $H^{\alpha}(Z)$, $\alpha=1,2,\ldots,N$ which satisfy the conditions (3), (B), (C), (D).

If besides that the roots of (5) satisfy the condition

$$|H^\alpha_{-n}| \leq H(R^\alpha_i)^n, \quad \alpha=1,2,\ldots,N, \quad n=0,1,2,\ldots,\infty, \tag{10}$$
$$|H^\alpha_n| \leq H(R^\alpha_e)^{-n}, \quad \alpha=1,2,\ldots,N, \quad n=1,2,\ldots,\infty,$$

where H is a positive constant and R^α_i and R^α_e are the inner and the outer radii of the annual region D^α_c. Then the functions $H^\alpha(Z)$, $Z \in D^\alpha_c$ are analytic and satisfy the conditions (B), (C), (D).

Remark 1: Using a result of Ref. 18, Chap. II, Sec. 3, we conclude that if $|H^\alpha_n| \leq$ const. $[1/|n|^{(1+\epsilon)}]$ then condition (8) is automatically fulfilled.

Remark 2: Condition (10) induces the analyticity of $H^\alpha(Z)$ in the region D^α_c which is a subregion of D^α. This condition is introduced because it is convenient for the proof of the existence theorems in Sec. III and Sec. IV.

In what follows it is advisable instead of system (5) to investigate its equivalent system[14]

$$t = A(t), \tag{11}$$

where $t \to t^\alpha_\nu$, $\alpha=1,2,\ldots,N$, $\nu=1,2,\ldots,\infty$ is an element of the metric space, and the operator A is defined by the right-hand side of the system:

$$t^\alpha_\nu = \sum_{\lambda,\mu} F(\nu;\lambda-\mu) E^\alpha_\nu(t_\lambda;t_\mu) + 2\sum_{\lambda,\mu} F(\nu;\lambda-\mu) E^\alpha_\nu(\tau_\mu;t_\lambda)$$
$$+ 2\sum_{\xi,\lambda} F(\nu;\xi+\lambda) E^\alpha_\nu(R_{-\xi};t_\lambda) + 2\sum_{\xi,\lambda} F(\nu;\xi+\lambda) E^\alpha_\nu(R_{-\xi};\tau_\lambda)$$
$$+ \sum_{\lambda,\mu} F(\nu;\lambda-\mu) E^\alpha_\nu(\tau_\lambda;\tau_\mu) + \sum_{\xi,\eta} F(\nu;\xi-\eta) E^\alpha_\nu(R_{-\xi};R_{-\eta})$$
$$+ R^\alpha_{-\nu} - \tau^\alpha_\nu. \tag{12}$$

In (12), as well as in the following, $\xi, \eta, \lambda, \mu, \nu$, and α are indices. Furthermore, ξ and η take the values $0,1,2,\ldots,\infty$; λ, μ, ν take the values $1,2,\ldots,\infty$ and $\alpha = 1,2,\ldots,N$, unless stated otherwise. In (12) the values $R_{-\xi}$ and τ_λ are known, and the values t^α_ν are sought. Moreover, $R^\alpha_{-\xi}$ denotes $H^\alpha_{-\xi}$, and $t^\alpha_\nu + \tau^\alpha_\nu$ is equal to H^α_ν. The values $H^\alpha_{-\xi} = R^\alpha_{-\xi}$ are considered to be known. For example, $R^\alpha_{-1} = \frac{1}{2}\lambda_\alpha$, where λ_α is the baryon pole residue, which is written explicitly in (1).

It is supposed that approximate values are known for H^α_ν which are denoted by τ^α_ν. Therefore, in (12) the small corrections t^α_ν to the approximate values τ^α_ν are sought.

III. APPLICATION OF SCHAUDER'S THEOREM FOR PROVING THE EXISTENCE OF SOLUTIONS OF LOW'S PROBLEM

System (12) is very convenient for numerical determination of the solutions of Low's problem.[13,14] In the present paragraph we shall use it in order to prove the existence of such solutions. For this purpose, we shall make use of one of the fixed-point theorems—Schauder's theorem.

Schauder's theorem is formulated in the following way[5]:

Let the operator A from (11) have the properties:

(1) A maps the bounded, closed convex set U belonging to the Banach space B into itself, i.e., if $t \in U$, then $A(t) \in U$.

(2) A is a completely continuous operator.

Then at least one element of the set U exists, which is a solution of (11).

The application of Schauder's theorem to Low's problem is facilitated by making use of the function

$$\chi_k(n) = \chi(j_k;n) = |n|^{-j_k}, \quad n = \pm 1, \pm 2, \ldots, \pm\infty$$
$$\chi_k(0) = \chi(j_k;0) = 1.$$

In our case j_k, $k=1,2,3,4$ are numbers larger than 1.

The sets U_t, U_τ, and U_R which we use below are defined, respectively, by the inequalities

$$|t^\alpha_\lambda| \leq t^* \chi(j_1,\lambda), \quad \alpha=1,2,\ldots,N, \quad \lambda=1,2,\ldots,\infty, \tag{13}$$
$$|\tau^\alpha_\mu| \leq \tau^* \chi(j_2,\mu), \quad \alpha=1,2,\ldots,N, \quad \mu=1,2,\ldots,\infty, \tag{14}$$
$$|R^\alpha_{-\xi}| \leq R^* \chi(j_3,\xi), \quad \alpha=1,2,\ldots,N, \quad \xi=0,1,2,\ldots,\infty. \tag{15}$$

In (13), (14), and (15), t^*, τ^*, and R^* are positive numbers.

The function χ is convenient for the estimation of $F(\nu,\xi^*)$ also. This expression is defined by the integral (6), which in this case is conveniently put down in the form

$$F(\nu;\xi^*) = \frac{1}{2\pi}\int_{-\pi/2}^{\pi/2} d\varphi [\sin(\nu+\xi^*)\varphi + \sin(\nu-\xi^*)\varphi] F(\varphi).$$

Further on we shall suppose that

$$|F'(\varphi_2) - F'(\varphi_1)| \leq \text{const}\,|\varphi_2-\varphi_1|^\epsilon, \quad -\tfrac{1}{2}\pi \leq \varphi_1, \ \varphi_2 \leq \tfrac{1}{2}\pi, \quad 0<\epsilon\leq 1$$
$$F(\pm\tfrac{1}{2}\pi) = F'(\pm\tfrac{1}{2}\pi) = 0, \quad F' = \frac{dF}{d\varphi}. \tag{16}$$

Let us consider the auxiliary function

$$\tilde{F}(\varphi) = F(\varphi), \quad -\tfrac{1}{2}\pi \leq \varphi \leq \tfrac{1}{2}\pi,$$
$$\tilde{F}(\varphi) = 0, \qquad \tfrac{1}{2}\pi \leq \varphi \leq 3(\pi/2).$$

It is obvious that $\tilde{F}'(\varphi)$, $-\pi \leq \varphi \leq \pi$ is Hölder continuous with an exponent ϵ, $0<\epsilon\leq 1$. This means that the Fourier coefficients of the function $\tilde{F}(\varphi)$ obey condition $\tilde{F}_n = O(1/|n|^{(1+\epsilon)})$ (for proof, see Sec. III in Ref. 19).

Hence $F(\nu,\xi^*)$ can be majorized by the inequality

$$F(\nu;\xi^*) \leq \text{const}\,[(\nu+\xi^*)^{-j_4} + (\nu-\xi^*)^{-j_4}], \quad j_4>1 \text{ if } \nu\pm\xi^* \neq 0,$$
$$F(\nu;\xi^*) = \text{const}, \quad \text{if } \nu\pm\xi^* = 0.$$

Using the function $\chi(j_4,n)$ introduced above, at $n = \nu+\xi^*$ and $n=\nu-\xi^*$, and choosing an appropriate positive constant F we obtain the inequality

$$F(\nu;\xi^*) \leq F[\chi_4(\nu+\xi^*) + \chi_4(\nu-\xi^*)]. \tag{17}$$

By means of (13), (14), (15), and (17), Eq. (12) is majorized by inequalities containing χ. In order to simplify these inequalities it is convenient to use the formula

$$\sum_{n=\infty}^{\infty} \chi_4(n)\chi_2(n+m) < K_{12}\chi_1(m) + K_{21}\chi_2(m), \tag{18}$$

where

$$K_{12} = (2^{j_1}+1)\zeta(j_2)+1; \quad K_{21} = (2^{j_2}+1)\zeta(j_1)+1;$$
$$j_1 > i; \ j_2 > 1;$$

and $\zeta(j_1)$ and $\zeta(j_2)$ are the Riemann ζ-functions from the theory of numbers.

When proving (18) it is convenient to proceed from the expression

$$\sum_{n=-\infty}^{\infty} \chi_1(n)\chi_2(n-m),$$

which is numerically equal to the expression

$$\sum_{n=-\infty}^{\infty} \chi_1(n)\chi_2(n+m).$$

The inequality (10) is proved by majorizing for $m \geq 2$ the right-hand side of the equality

$$\sum_{n=-\infty}^{\infty} \chi_1(n)\chi_2(n-m) = S_1 + \chi_1(0)\chi_2(-m) + S_2 + S_3 + \chi_1(m)\chi_2(0) + S_4,$$

where

$$S_1 = \sum_{n=-\infty}^{-1} |n|^{-j_2} < m^{-j_2} \sum_{k=-\infty}^{-1} |k|^{-j_1} = \chi_2(m)\zeta(j_2),$$

$$S_2 = \sum_{n=1}^{n'} n^{-j_1} |n-m|^{-j_2} < m^{-j_2} \sum_{k=-\infty}^{-1} |k|^{-j_1} = \chi_2(m)\zeta(j_2),$$

$n' = 0.5\,m$ is even and $n' = 0.5\,(m+1)$ if m is uneven

$$S_3 = \sum_{n=n'}^{m-1} n^{-j_1}(n-m)^{-j_2} < 2^{j_1} m^{-j_1} \sum_{n=n'}^{m-1} |n-m|^{-j_2},$$

$$< 2^{j_1} m^{-j_1} \sum_{k=1}^{\infty} |k|^{-j_2} = 2^{j_1}\chi_1(m)\zeta(j_2),$$

$$S_4 = \sum_{n=m+1}^{\infty} n^{-j_2} < m^{-j_2} \sum_{k=1}^{\infty} |k|^{-j_2} = \chi_1(m)\zeta(j_2).$$

For $m = 0$ and 1 the inequality (18) is immediately confirmed. For $m < 0$ the proofs are analogous.

In order to satisfy the first condition for the operator A we substitute in (12) t_λ^α, τ_μ^α, R_μ^α, and $F(\nu, \xi^*)$ with the expressions from (13), (14), (15), and (17).

Having the inequality (18) we can easily apply the Schauder's theorem to Low's problem.

For this purpose we choose the Banach space to be a subspace of the space of the bounded sequences of numbers.[5] More precisely, we use a space Y, the elements y of which are the sequences of numbers y_ξ, $|y_\xi| \leq A\,\chi(j, \xi)$, $\xi = -\infty \ldots -2, -1, 0, 1, \ldots, \infty$, $j > 1$, the norm being defined by equality

$$\|y\| = \sup_\xi |y_\xi|.$$

As (18) holds only for $j > 1$ in the following we shall suppose in (13), (14), (15), and (17), $j_1 > 1$, $j_2 > 1$, $j_3 > 1$, and $j_4 > 1$, respectively.

To satisfy the condition $A(t) \in t$, it is enough to put down

$$(1+NC)Ft^* \sum_{\lambda,\mu} [\chi_4(\nu+\lambda-\mu) + \chi_4(\nu-\lambda+\mu)]\chi_1(\lambda)\chi_1(\mu)$$
$$+ 2(1+NC)Ft^*\tau^* \sum_{\lambda,\mu} [\chi_4(\nu+\lambda-\mu) + \chi_4(\nu-\lambda+\mu)]\chi_2(\mu)\chi_1(\lambda)$$
$$+ 2(1+NC)Ft^*R^* \sum_{\lambda,\xi} [\chi_4(\nu+\lambda+\xi) + \chi_4(\nu-\lambda-\xi)]\chi_3(\xi)\chi_1(\lambda)$$
$$+ 2(1+NC)F\tau^*R^* \sum_{\lambda,\xi} [\chi_4(\nu+\xi+\lambda) + \chi_4(\nu-\xi-\lambda)]\chi_3(\xi)\chi_2(\lambda)$$
$$+ (1+NC)F\tau^{*2} \sum_{\lambda,\mu} [\chi_4(\nu+\lambda-\mu) + \chi_4(\nu-\lambda+\mu)]\chi_2(\mu)\chi_2(\lambda)$$
$$+ (1+NC)FR^{*2} \sum_{\xi,\eta}^{\lambda,\mu} [\chi_4(\nu+\xi-\eta) + \chi_4(\nu-\xi+\eta)]\chi_3(\xi)\chi_3(\eta)$$
$$+ R^*\chi_3(\nu) + \tau^*\chi_2(\nu) < t^*\chi_1(\nu),$$

where $C = \max C^{\alpha,\beta}$, $\alpha, \beta = 1, 2, \ldots, N$. When deducing the latter inequality it is advisable to suppose at first that $N = 1$, and $C^{\alpha\beta} = 0$. In this case in the inequality we would have 1 instead of the factors $1 + NC$. In the last expression NC accounts for the contribution of the term

$$(-1)^\nu \sum_{\beta=1}^{N} C^{\alpha\beta} H_m^\beta H_{m+k}^\beta$$

in the formula which defines E_ν^α.

Summing over all indices from $-\infty$ to $+\infty$ and using (18), we obtain

$$2(1+NC)Ft^{*2}[K_{41}^2\chi_4(\nu) + (K_{41}K_{14} + 2K_{14}K_{11})\chi_1(\nu)]$$
$$+ 2(1+NC)F\tau^{*2}[K_{42}^2\chi_4(\nu) + (K_{42}K_{24} + 2K_{24}K_{22})\chi_2(\nu)]$$
$$+ 2(1+NC)FR^{*2}[K_{43}^2\chi_4(\nu) + (K_{43}K_{34} + 2K_{34}K_{33})\chi_3(\nu)]$$
$$+ 4(1+NC)Ft^*\tau^*[K_{42}K_{41}\chi_4(\nu) + (K_{43}K_{24} + K_{34}K_{23})\chi_2(\nu)$$
$$+ K_{34}K_{32}\chi_3(\nu)]$$
$$+ 4(1+NC)F\tau^*R^*[K_{43}K_{42}\chi_4(\nu) + (K_{43}K_{24} + K_{34}K_{23})\chi_2(\nu)$$
$$+ K_{34}K_{32}\chi_3(\nu)]$$
$$+ 4(1+NC)FR^*t^*[K_{43}K_{41}\chi_4(\nu) + (K_{43}K_{14} + K_{34}K_{13})\chi_3(\nu)$$
$$+ K_{34}K_{31}\chi_1(\nu)]$$
$$+ \tau^*\chi_2(\nu) + R^*\chi_3(\nu) < t^*\chi_1(\nu), \quad \nu = 1, 2, \ldots, \infty.$$

We suppose that $j_2 \geq j_1$; $j_3 \geq j_1$; and $j_4 \geq j_1$. Under this assumption $\chi_2(\nu) \leq \chi_1(\nu)$; $\chi_3(\nu) \leq \chi_1(\nu)$; and $\chi_4(\nu) \leq \chi_1(\nu)$. If we put $\tau^* = pt^*$ and $R^* = qt^*$ and suppose that $p + q < 1$, the above inequality is transformed into the inequality

$$t^* = t_1^* < \frac{1-p-q}{2(1+NC)F(U_1 + pU_2 + qU_3 + p^2U_4 + q^2U_5 + pqU_6)}, \tag{19}$$

where

$$U_1 = K_{41}^2 + K_{41}K_{14} + 2K_{14}K_{11},$$
$$U_2 = 2(K_{42}K_{41} + K_{42}K_{14} + K_{24}K_{21} + K_{24}K_{12}),$$
$$U_3 = 2(K_{43}K_{41} + K_{43}K_{14} + K_{34}K_{31} + K_{34}K_{13}),$$
$$U_4 = K_{42}^2 + K_{42}K_{24} + 2K_{24}K_{22},$$
$$U_5 = K_{43}^2 + K_{43}K_{34} + 2K_{34}K_{33},$$
$$U_6 = 2(K_{43}K_{42} + K_{43}K_{24} + K_{34}K_{24} + K_{34}K_{32}).$$

Let us suppose that t^* is so chosen that inequality (19) is satisfied. In respect to (13) that means that the absolute values of the left-hand side of the system (12) are less than the absolute values of those on the right-hand side. In other words, if inequality (19) is satisfied, the set $U_1 = U_t$ is such that $A(U_1) \in U_1$. And because by (13) U_1 is a bounded and convex set, it follows that condition 1 of Schauder's theorem has been satisfied. The second condition of Schauder's theorem demands that A should be a completely continuous operator. Let us recall the definition of a completely continuous operator[5]: the operator A is completely continuous on the set U_1 if it is continuous on U_1 and compact on U_1, i.e., when A maps every bounded subset of U_1 into a compact one. The operator A is continuous on U_1. This is easily proved with regard to formula (23) from the next section. The compactness of U_1 is proved when taking into consideration that according to its definition U_1 is compact.[20]

Therefore, if condition (19) is satisfied, which with an

appropriate choice of the parameters t^*, p and q can always be achieved, then all the requirements for the applicability of Schauder's theorem are also satisfied. This result is expressed in the following theorem:

Theorem 2:

Let (16) be satisfied. Let sequences of numbers

$R^\alpha_{-\xi}$, $\alpha = 1, 2, \ldots, N$; $\xi = 0, 1, 2, \ldots, \infty$

are known such that $|R^\alpha_{-\xi}| \leqslant \text{const}\, \chi(j_3; \xi)$. (20)

Then the algebraic system (12) has at least one solution h^α_ν, $\alpha = 1, 2, \ldots, N$, $\nu = 1, 2, \ldots, \infty$ such that $h^\alpha_\nu = O[1/|\nu|^{j_1}]$, $j_1 > 1$.

Let in addition the condition (9) be satisfied.

Then the series (4) converge to the functions $H^\alpha(Z)$, $\alpha = 1, 2, \ldots, N$, which satisfy the conditions (3), (B), (C), (D). In the particular case when the sequences H^α_n, $n = 0, -1, -2$, are finite, the functions $H^\alpha(Z)$, $Z \in D^\alpha_c$ ($1 = R^\alpha_e > |Z| \geqslant R^\alpha_i = 0$) are analytic.

Remark 1: Condition (16) can be replaced by the stronger condition:

The function $F'(\varphi)$, $-\frac{1}{2}\pi \leqslant \varphi \leqslant \frac{1}{2}\pi$ is bounded and

$$\lim_{\varphi \to \pi/2} [F(\varphi)(\tfrac{1}{2}\pi - \varphi)^{(1+\epsilon)}] \neq \infty. \tag{16'}$$

Proof: In addition to the above motivation in proving the theorem we remark that $j_3 > 1$ and $j_4 > 1$, the first because of (20), and the second because $F^\alpha_n = O[1/|n|^{(1+\epsilon)}]$, $0 < \epsilon \leqslant 1$. Then choosing $j_2 \geqslant j_1$ and $j_1 > 1$, we can write $j_4 \geqslant j_1$, $j_3 \geqslant j_1$ and $j_2 \geqslant j_1$, which was supposed in deriving (19). So we complete the proof of the first part of Theorem 2.

To prove the second part of the theorem, it is sufficient to demonstrate that the conditions of the inverse part of Theorem 1 are satisfied:

Condition (8) is indeed fulfilled. This is a consequence of the fact that according to the first part of Theorem 2 $|H^\alpha_n| \leqslant \text{const}\, \chi(j_1, n)$ $n = 0, 1, 2, \ldots, \infty$. Condition (20) means that $|H^\alpha_n| \leqslant \text{const}\, \chi(j_3, n)$. From here, in connection with a theorem of Ref. 18, Chap. II, Sec. 3, it follows that $V^\alpha_+(\varphi)$ and $V^\alpha_-(\varphi)$ exist and satisfy the Hölder condition with exponent ϵ, $0 < \epsilon \leqslant 1$, on the interval $[-\pi - \eta, \pi + \eta]$, where η is some positive number.

Condition (19) of Theorem 1 is also fulfilled because it figures in Theorem 2 as well.

Condition (7) of Theorem 1 is satisfied because it is a consequence of condition (16) of Theorem 2.

With this the proof of Theorem 2 is completed.

Let us consider the application of this theorem in two special cases.

(a) Suppose that $R^\alpha_{-\xi} = 0$, $\xi = 2, 3, \ldots, \infty$ and $R^\alpha_{-1} = \lambda_\alpha/2$

With these assumptions and the appropriate choice of $C^{\alpha\beta}$ Low's problem corresponds to the problem resolved by means of the integral equation of Chew and Low. The existence of solution to this problem depends mainly on the properties of the cut-off function.

So in the case of the G. Salzman and F. Salzman's choice of cutoff function $f(x) = [(x-1)^{3/2}/12\pi] \exp[-(x^2 - 1)/4m_\pi]$, where m_π is the meson mass, passing from $f(x)$ to $F(\varphi)$, we conclude that the condition of Remark 1 to Theorem 2 is satisfied. Hence the problem has at least one solution $H(Z)$ which is analytic at least in region $0 < |Z| < 1$. The existence of at least one solution to (1) was proved by Warnock[7] through its direct investigation.

(b) Suppose that $R^\alpha_{-\xi} = 0$, $\xi = 1, 2, 3, \ldots$, $\alpha = 1, 2, \ldots, N$, i.e., the partial scattering amplitudes have no pole at the origin. For $N = 3$ and with the appropriate choice of $C^{\alpha\beta}$ this problem is equivalent to the integral equation of Shirkov *et al.*[3,4] for $\pi - \pi$ scattering in the low-energy region. If instead of the function $f(x)$ from Ref. 10 we use the function $f(x)e^{-kx}$, $k \to 0$ the results of Theorem 2 could be transferred directly to that case. If we conjecture that, we can put in the solution $k = 0$, we may conclude that the Shirkov equation has at least one solution.

In particular cases as, for instance, in the case of the applications (a) and (b) of Theorem 2, the condition of analyticity (A) is satisfied. But in general Theorem 2 does not guarantee the fulfilment of this condition. More general conditions assuring the fulfilment of the four conditions are defined in the next theorem.

Theorem 3:

Let (16) be satisfied. Let sequences of numbers

$R^\alpha_{-\xi}$, $\alpha = 1, 2, \ldots, N$, $\xi = 0, 1, 2, \ldots, \infty$ are known such that $|R^\alpha_{-\xi}| \leqslant \text{const}\, (R^\alpha_i)^\xi$, where $1 > R^\alpha_i > 0$ are constants. (21)

Then the algebraic system (12) has at least one solution h^α_ν, $\alpha = 1, 2, \ldots, N$; $\nu = 1, 2, \ldots, \infty$ such that $h^\alpha_\nu = O[1/\nu^{j_1}]$, $j_1 > 0$.

Let in addition condition (9) be satisfied.

Then the series (4) converge for $1 \geqslant |Z| > R^\alpha_i$ to the functions $H^\alpha(Z)$ $\alpha = 1, 2, \ldots, N$, which are analytic for $Z \in D^\alpha_c (1 = R^\alpha_e > |Z| > R^\alpha_i)$ and satisfy the conditions (B), (C), (D).

Proof: Having in view condition (21) we introduce instead of $\chi(j_3, n)$ the function $(R^\alpha_i)^{-|n|}$.

We remark that the relations (13) and (14) hold also for Theorem 3 if relation (15) is substituted by (15')

$$|R^\alpha_{-\xi}| \leqslant R'^*(R^\alpha_i)^\xi, \quad \alpha = 1, 2, \ldots, N; \ \xi = 0, 1, 2, \ldots, \infty. \tag{15'}$$

The proof of Theorem 3 can be carried out merely as a literal repetition of the proof of Theorem 2. For this purpose the relation (12) must be substituted with an analogous relation for the expression

$$\sum_{n=-\infty}^{\infty} \chi_1(n)(R^\alpha_i)^{-|n+m|}.$$

This is easily achieved, observing that for K large enough $(R^\alpha_i)^{-|n|} \leqslant K\chi(j_3, n)$ so that we get the relation

$$\sum_{n=-\infty}^{\infty} \chi_1(n)(R^\alpha_i)^{-|n+m|} \leqslant K'_{13}\chi_1(m) + K'_{31}\chi_3(m), \tag{12'}$$

where

$$K'_{13} = KK_{13}, \quad K'_{31} = KK_{31}.$$

With (15′) and (18′) instead of (15) and (18) we repeat the reasoning leading to the proof of Theorem 2 and get the proof of Theorem 3. In Theorem 3 an extra moment is the proof of the analyticity of $H^{\alpha}(Z)$, which is trivial.

IV. CONDITIONS UNDER WHICH THE SOLUTIONS OF LOW'S PROBLEM EXIST AND ARE UNIQUE

With a certain modification of the conditions of Theorem 2, one could guarantee not only the existence but also the uniqueness of the solutions of Low's problem. This can be achieved by means of the Banach–Cacciopoli contraction mapping principle. The latter is another variant of the fixed-point theorems which give not only the existence of the solution but also uniqueness and a method of calculating that solution.[5,6]

The Banach–Cacciopoli theorem reads[5,6]:

Let A be an operator defined on the complete metric space X satisfying the following conditions:

(a) A maps the space X into itself, i.e., $A(X)\in X$.

(b) In X, A is a contracting operator, i.e., if t' and t'' are two elements of X and if the distance in X is denoted by ρ then

$$\rho[A(t''),A(t')]\leq\lambda\rho[t'',t'],$$

where $\lambda<1$ is a constant.

Then Eq. (11) has one and only one solution, which can be obtained numerically by the method of successive approximations.

The first condition of the theorem is satisfied if inequality (19) is satisfied. This assertion needs some justifications. In Sec. III, where inequality (19) was derived, t_{ν}^{α}, $\alpha=1,2,\ldots,N$, $\nu=1,2,\ldots,\infty$ was supposed to be an element of the Banach space. This conclusion, however, does not depend on the choice of the functional space. On that account (19) is also used in the present section where the weaker assumption is made, viz. that t_{ν}^{α} is an element of some complete metric space X.

The choice of X is not unique. The proofs are, however, simplified if X represents the closed set U_t, defined by (13) where the distance between the elements t'' and t' is given by the expression $\rho(t'',t')=\sup_{\alpha,\nu}|t_{\nu}''^{\alpha}-t_{\nu}'^{\alpha}|$, $\alpha=1,2,\ldots,N$; $\nu=1,2,\ldots,\infty$. It is obvious that X is a subspace of the space C_0 of the converging numerical sequences.[6]

The proof of the second condition of the Banach–Cacciopoli theorem is reduced to evaluation of the distance $\rho[A(t''),A(t')]$.

Here again, as in the previous section, the conclusions are to be made with the assumption that $\alpha=N=1$ and $C^{\alpha\beta}=0$. The transfer to the general case $N\neq0$ and $C^{\alpha\beta}\neq0$ in the final result follows immediately.

We have

$$\rho[A(t''),A(t')]=\max_{\nu}\sum_{\lambda,\mu}F(\nu;\lambda-\mu)(t_{\lambda}''t_{\mu}''-t_{\lambda}'t_{\mu}')$$
$$+2\sum_{\lambda,\mu}F(\nu;\lambda-\mu)\tau_{\mu}(t_{\lambda}''-t_{\lambda}')+2\sum_{\xi,\lambda}F(\nu;\xi+\lambda)R_{-\xi}(t_{\nu}''-t_{\nu}').$$

Setting $t_{\lambda}''=t_{\lambda}'+\varphi_{\lambda}$, we obtain for $\rho[A(t''),A(t')]$

$$\rho[A(t''),A(t')]$$
$$=\max_{\nu}\Big(\sum_{\lambda,\mu}F(\nu;\lambda-\mu)[(t_{\mu}'+\tfrac{1}{2}\varphi_{\mu})\varphi_{\lambda}+(t_{\lambda}'+\tfrac{1}{2}\varphi_{\lambda})\varphi_{\mu}]$$
$$+2\sum_{\lambda,\mu}F(\nu;\lambda-\mu)\tau_{\mu}\varphi_{\lambda}+2\sum_{\xi,\lambda}F(\nu;\xi+\lambda)R_{-\xi}\varphi_{\lambda}\Big).$$

Or, exchanging the indices in the last term of the bracketed expression on the right-hand side,

$$\rho[A(t''),A(t')]=\max_{\nu}\Big(2\sum_{\lambda,\mu}F(\nu;\lambda-\mu)(t_{\mu}'+\tfrac{1}{2}\varphi_{\mu})\varphi_{\lambda}$$
$$+\sum_{\lambda,\mu}F(\nu;\lambda-\mu)\tau_{\mu}\varphi_{\lambda}+2\sum_{\xi,\lambda}F(\nu;\xi+\lambda)R_{-\xi}\varphi_{\lambda}\Big).$$

From the definition of $\chi(j_1;n)$ it follows that

$$\varphi_{\mu}\leq2t^{*}\chi_1(\mu)\quad\text{and}\quad\varphi_{\lambda}\leq2t^{*}\chi_1(\lambda).$$

We denote by ϕ the largest of the numbers φ_{λ}, $\lambda=1,2,\ldots,\infty$. Obviously, $\phi=\rho(t'',t')$. Then for $\rho[A(t''),A(t')]$ we obtain

$$\rho[A(t''),A(t')]\leq\max_{\nu}\Big(4t^{*}\sum_{\lambda,\mu}|F(\nu;\lambda-\mu)|\chi_1(\mu)$$
$$+2\tau^{*}\sum_{\lambda,\mu}|F(\nu;\lambda-\mu)|\chi_2(\mu)+2R^{*}\sum_{\xi,\lambda}|F(\nu;\xi+\lambda)|\chi_3(\xi)\Big)\phi.\tag{22}$$

Introducing (17) into (22) we have

$$\rho[A(t''),A(t')]$$
$$<\max_{\nu}\Big(4t^{*}F\sum_{\lambda,\mu}[\chi_4(\nu-\lambda+\mu)+\chi_4(\nu+\lambda-\mu)]\chi_1(\mu)$$
$$+2\tau^{*}F\sum_{\lambda,\mu}[\chi_4(\nu-\lambda+\mu)+\chi_4(\nu-\lambda+\mu)]\chi_2(\mu)$$
$$+2R^{*}F\sum_{\xi,\lambda}[\chi_4(\nu+\xi+\lambda)+\chi_4(\nu-\xi-\lambda)]\chi_3(\xi)\Big)\phi.$$

Strengthening the inequality we sum on μ and ξ from $-\infty$ to $+\infty$. Making repeated use of (18), we obtain

$$\rho[A(t''),A(t')]\leq\max_{\nu}\sum_{\lambda}\Big(4t^{*}F[K_{41}\chi_4(\nu-\lambda)+K_{14}\chi_1(\nu-\lambda)$$
$$+K_{41}\chi_4(\nu+\lambda)+K_{14}\chi_1(\nu+\lambda)]+2\tau^{*}F[K_{42}\chi_4(\nu-\lambda)$$
$$+K_{24}\chi_2(\nu-\lambda)+K_{42}\chi_4(\nu+\lambda)+K_{24}\chi_2(\nu+\lambda)]$$
$$+2R^{*}F[K_{43}\chi_4(\nu+\lambda)+K_{34}\chi_3(\nu+\lambda)$$
$$+K_{43}\chi_4(\nu-\lambda)+K_{34}\chi_3(\nu-\lambda)]\Big)\phi.$$

There remains the sum over λ. Strengthening the inequality we sum from $-\infty$ to ∞. For instance, let us consider the sum

$$\sum_{\lambda=-\infty}^{\infty}\chi_4(\nu-\lambda).$$

With the substitution $\nu-\lambda=\bar{\lambda}$ we have

$$\sum_{\lambda=-\infty}^{\infty}\chi_4(\nu-\lambda)=\sum_{\lambda=-\infty}^{-1}\frac{1}{|\bar{\lambda}|^{j_4}}+1+\sum_{\bar{\lambda}=1}^{\infty}\frac{1}{(\bar{\lambda})^{j_4}}=2\zeta_4+1,$$

where ζ_4 means $\zeta(j_4)$. Making repeated use of this formula in the latter inequality, for $\rho[A(t''),A(t')]$ we obtain

$$\rho[A(t''),A(t')]<F(t^{*}V_1+\tau^{*}V_2+R^{*}V_3)\phi,$$

where

$$V_1=8K_{41}(2\zeta_4+1)+8K_{14}(2\zeta_1+1),$$
$$V_2=4K_{42}(2\zeta_4+1)+4K_{24}(2\zeta_2+1),$$
$$V_3=4K_{43}(2\zeta_4+1)+4K_{34}(2\zeta_3+1).$$

Passing to the general case $N>1$ and $C^{\alpha\beta}\neq 0$, as in the former section, we obtain the formula

$$\rho[A(t''),A(t')]<(1+NC)F(t^*V_1+\tau^*V_2+R^*V_3)\phi. \quad (23)$$

It follows from (23) that for all the elements of the set U_t the operator $A(t)$ is continuous since at $\phi\to 0$, $\rho[A(t''),A(t')]\to 0$, a result which was used in the former section. As $\phi=\rho(t'',t')$, it follows from (23) that the second condition of the Banach–Cacciopoli theorem requires that the inequality

$$(1+NC)F(t^*V_1+\tau^*V_2+R^*V_3)=\gamma<1$$

is satisfied. Or, with the notations $\tau^*=pt^*$ and $R^*=qt^*$

$$t^*<t_2^*=[(1+NC)F(V_1+pV_2+qV_3)]^{-1}. \quad (24)$$

The first condition of the Banach–Cacciopoli theorem is satisfied if in (19) we choose $t^*<t_1^*$. For the second condition it is necessary for t^* to be less than the number t_2^*, defined in (24). For that reason the Banach–Cacciopoli theorem can be applied to Low's problem if the smaller of the numbers t_1^* and t_2^* is chosen for t^*. That gives us grounds for formulating the following theorem:

Theorem 4:

Let (16) and (20) be satsified. Let t^* be the smaller of the numbers t_1^* and t_2^* defined in (19) and (24), respectively. (25)

Then the algebraic system (12) has one and only one solution t_ν^α, $\alpha=1,2,\ldots,N$, $\nu=1,2,\ldots,\infty$ which satisfies the inequalities $|t_\nu^\alpha|\leq t^*\chi(j_1,\nu)$.

Let in addition the condition (9) be satisfied.

Then the series (4) converge for $|Z|=1$ to one and only one set of functions $H^\alpha(Z)$, $\alpha=1,2,\ldots,N$, which satisfy conditions (3), (B), (C), (D).

The proof of the theorem was actually made above. We have only to add that in the Theorem 2 the first two conditions ensure the inequalities $j_4,j_3,j_2\geq j_1>1$, which are necessary for the deduction of (19). Similarly as in Theorem 2 the last condition is necessary to guarantee the correspondence between the algebraic and abstract version of the problem. The proof of the last assertion of Theorem 4 follows immediately from Remark 1 to Theorem 1.

As can be proved, Theorem 3 guarantees not only the existence but also the uniqueness of solutions to a problem (A), (B), (C), (D) corresponding to the Chew and Low equation and as we conjecture to that of Shirkov. As in Sec. III, in the latter case we must replace $f(x)$ by $f(x)e^{-kx}$, $k\to 0$, and then put $k=0$ in the solution.

If we are interested in the case when $H(Z)$ is analytic in a circular ring, then Theorem 4 must be replaced by the theorem

Theorem 5: Let (16), (21), and (25) be satisfied. Then the algebraic system (12) has at least one solution h_ν^α, $\alpha=1,2,\ldots,N$; $\nu=1,2,\ldots,\infty$ such that $h_\nu^\alpha=O[1/|\nu|^{j_1}]$, $j_1>1$. If in addition the condition (9) is satisfied, then the series (4) converge for $1\leq|Z|>R_i^\alpha$ to one and only one set of functions $H^\alpha(Z)$, $\alpha=1,2,\ldots,N$ which satisfy the conditions (B), (C), (D) and are analytic in the region $D_c^\alpha(R_e^\alpha=1>|Z|>R_i^\alpha>0)$.

The proof of Theorem 5 is analogous to that of Theorem 4, except at a few points where, as in Theorem 3, the function $(R_i^\alpha)^{-|n|}$ is introduced instead of the function $\chi(j_3,n)$.

V. CONCLUSION

In this paper a generalization of Low's integral equation (1) was studied. The generalized problem under the name of Low's problem in Sec. II, is formulated as the problem (A), (B), (C), (D) and the algebraic problem (5). The conditions under which the two formulations are equivalent are specified by Theorem 1. The conditions for existence and uniqueness of solutions to Low's problem in its abstract and algebraic formulations are given in Theorems 2, 3, 4, and 5.

The results obtained in this paper partially coincide with the results in Refs. 7–11. In some cases they are more general and in others less so.

So, for example, the methods developed in Refs. 7 and 8 can be used for the examination of a kind of Low's problem which is more general than the integral equation (1) but less so than Low's problem considered here. The generalization consists in the replacement in (1) of the inhomogeneous term λ_α/z, by the series

$$\frac{\lambda_\alpha}{z}+\sum_{n=2}^{\infty}\frac{R_{-n}^\alpha}{z^n}.$$

In this case $h^\alpha(z)$ could have not only a single pole at $z=0$ but also singularities in a larger domain. More precisely the region which is the point $Z=0$ in the case of equation (1) after the generalization of (1) would become a circle C with the center $z=0$ and radius $r<1$. Circles C with $r>1$ are excluded on the following ground: When C has a radius $r>1$, the coefficients R_{-n}^α would contain information of the unknowns $h^\alpha(z)$. This would complicate the problem in the case $r>1$ to such an extent that the treatment of the integral equation would not be possible by simple generalization of the methods in Refs. 7 and 8. For the algebraic method which was developed here s_I^α is not restricted in such a way—it can, in principle, coincide even with the cut plane z.

There are, however, other problems where the approach of the integral equations is more efficient. These are, for example, the proof of the existence of resonant solutions and the proof of their multiplicity. Although in both cases the investigation on the basis of the algebraic system is possible, it would give less interesting results than the direct analysis of the integral equations.

Interesting results may be expected in studying the behavior of the solution of Low's problem for $\varphi\to\pm\frac{1}{2}\pi$. Investigations in this direction are now in progress.

ACKNOWLEDGMENTS

A large part of this work was done at the author's home institution, The Institute of Physics at the Bulgarian Academy of Sciences, Sofia, Bulgaria. The opportunity to complete the work was found at the International Centre for Theoretical Physics, Trieste, Italy. The author is very grateful to Professor Abdus

Salam and Professor P. Budini, as well as to the International Atomic Energy Agency and UNESCO for hospitality at the ICTP, Trieste. He also wishes to thank Professor L. Fonda and Professor D. Shirkov for interesting discussions, and Dr. S. Coleman for reading the manuscript. Finally, he wishes to express his deep gratitude to Professor R. L. Warnock for his stimulating critical remarks.

[1]S. S. Schweber, *An Introduction to Relativistic Quantum Field Theory* (Row, Peterson, New York, 1961).
[2]G. Chew and S. Mandelstam, Phys. Rev. **119**, 467 (1960).
[3]V. V. Serebryakov and D. V. Shirkov, Fortschr. Phys. **13**, 227 (1965).
[4]D. V. Shirkov, V. V. Serebryakov, and V. A. Meshcheryakov, *Dispersion Theories of Strong Interaction at Low Energies* (North-Holland, Amsterdam, 1969).
[5]L. A. Liusternik and V. J. Sobolev, *Elements of Functional Analysis* (in Russian) (Nauka, Moscow, 1965), 2nd ed.
[6]L. A. Liusternik and V. J. Sobolev, *Elements of Functional Analysis* (English translation from the first Russian edition) (Ungar, New York, 1961).
[7]R. L. Warnock, Phys. Rev. **170**, 1323 (1968).
[8]H. McDaniel and R. L. Warnock, Nuovo Cimento **64**, 905 (1969).
[9]H. McDaniel and R. L. Warnock, Phys. Rev. **180**, 1433 (1969).
[10]D. Atkinson, J. Math. Phys. **8**, 2281 (1967).
[11]D. Atkinson, in *Developments in High Energy Physics*, Suppl. VII of Acta Phys. Austr. (Springer Berlin, 1970).
[12]I. P. Nedialkov, JINR, Dubna, Preprint E-1294, 1963 (unpublished).
[13]I. P. Nedialkov, ICTP, Trieste, Preprint IC/71/86, 1971 (published as Ref. 14).
[14]I. P. Nedelkov, Phys. Rev. D **6**, 2842 (1972).
[15]I. P. Nedelkov, ICTP, Trieste, Preprint IC/71/91, 1971 (unpublished).
[16]R. L. Warnock, Lect. Theor. Phys. **11D**, (1969).
[17]G. Salzman and F. Salzman, Phys. Rev. **108**, 1619 (1957).
[18]N. K. Bari, *Trigonometrical series* (in Russian) (State Publishing House for Physical and Mathematical Literature, Moscow, 1961).
[19]G. H. Hardy and W. W. Rogosinski, *Fourier Series* (Cambridge U. P., London, 1965).
[20]The compactness of U_1 is proved in the same way as the compactness of the Hilbert cube u (Refs. 5 and 6).

Properties of two-phase "cell materials"

William Fuller Brown Jr.

Department of Electrical Engineering, University of Minnesota, Minneapolis, Minnesota 55455
(Received 13 December 1972)

"Symmetric cell materials" and "asymmetric cell materials" were defined by Miller in connection with the physical properties (such as the effective dielectric constant) of two-phase solid mixtures. It is shown here that while the "symmetric cell material" is self-consistent, the "asymmetric cell material" is not: The postulated three-point probabilities do not add up to the correct one-point probabilities. A self-consistent generalization of the "symmetric cell material", based on the requirement that a certain integral must reduce to an integral over a finite region, is developed, and one construction procedure for producing such a material is described.

1. INTRODUCTION

In a study of the physical properties of two-phase solid mixtures, Miller[1,2] introduced the concept of a "symmetric cell material". For such a material, the upper and lower bounds[3] to such quantities as the effective dielectric constant are somewhat closer than for a general two-phase mixture. The symmetric cell material can be defined most simply by describing a mathematical procedure for constructing one: first divide the space by some random procedure into statistically equivalent[4] cells, then assign each cell randomly and independently to material A or to material B with probabilities p and $q=1-p$, respectively. It follows that, given that two points are in different cells, each of the two points has, independently of the other, probability p of being in material A and probability q of being in material B.

Miller[1] also introduced the concept of an "asymmetric cell material." For it, the cells of one component material are not statistically equivalent (as regards shape or size) to the cells of the other. Consequently some other definition than that given above is necessary. Miller specifies that "the material property ϵ of a cell is statistically independent of the material property of any other cell." It is not obvious that the postulated statistical independence is compatible with distinguishability of the two types of cell, or that Miller's recipes for constructing such a material will indeed lead to one that satisfies his definition. To get a clear idea of what is meant, one must examine the specific three-point probabilities postulated.

In Sec. 2 it will be shown that Miller's "asymmetric cell material" is in fact not self-consistent: The three-point probabilities do not add up to the correct one-point probabilities. In Sec. 3 a self-consistent generalization of the "symmetric cell material" will be developed. The property required of it is that, like the "symmetric cell material," it must reduce a certain integral to an integral over a finite region; this is the property that leads to the improvement in upper and lower bounds mentioned earlier.

In Sec. 4 a specific model, based on a definite construction procedure, will be described and will be shown to possess the general properties derived in Sec. 3. In Sec. 5 some conclusions will be drawn.

2. MILLER'S "ASYMMETRIC CELL MATERIAL"

In this section Miller's notation will be used, with some small modifications. A point chosen at random has probability ϕ of being in material A (Miller's material 1) and probability $1-\phi$ of being in material B (his material 2). For three points 1, 2, 3, with position vectors $\mathbf{r}_1$, $\mathbf{r}_2$, $\mathbf{r}_3$, the various joint and conditional probabilities, because of the assumed statistical homogeneity, depend only on the two relative position vectors $\mathbf{r}=\mathbf{r}_2-\mathbf{r}_1$ and $\mathbf{s}=\mathbf{r}_3-\mathbf{r}_1$ (Miller therefore takes $\mathbf{r}_1=0$); and because of the assumed statistical isotropy, these probabilities are invariant to a rigid rotation of the triangle formed by the three points. They may therefore be taken to be functions of $r=|\mathbf{r}|$, $s=|\mathbf{s}|$, and a third variable. We shall choose as this third variable the third side of the triangle, $t=|\mathbf{r}_{23}|=|\mathbf{r}_3-\mathbf{r}_2|=|\mathbf{s}-\mathbf{r}|$ (see Fig. 1).

The variables r, s, and t are in a certain sense not completely independent, since they are subject to triangular inequalities. For our purposes, however, the important property of r, s, and t is that whatever values they have, the value of one of them can always be changed (though perhaps in only one direction) without changing the others; for example, t can be changed at constant r and s by rotating $\mathbf{r}$ or $\mathbf{s}$ about point 1. From the independence of r, s, and t in this sense, it follows that if $f(r)=g(s)$, then $f(r)=\text{const}$; for $f(r)$ is not affected by a change of s, $g(s)$ is not affected by a change of r, and therefore the common value of these two quantities is not affected by a change of r or of s.

We shall encounter three types of function, for which certain abbreviated notations are convenient. (In the following illustrations, the letters f, v, and u may be replaced by other letters.)

(1) A function $f(r)$ of the distance r between points 1 and 2. Our abbreviated notation will be

$$f\equiv f(r),\quad f'\equiv f(s),\quad f''\equiv f(t). \tag{2.1}$$

(2) A function $v(r;s,t)$ of the three distances r, s, and t, invariant to an interchange of s and t but not to an interchange of r and s. We shall use the abbreviated notation

$$v(r;)\equiv v(r;s,t)=v(r;t,s) \tag{2.2}$$

and the still more abbreviated notation

$$v\equiv v(r;),\quad v'\equiv v(s;),\quad v''\equiv v(t;). \tag{2.3}$$

(3) A function $u(r,s,t)$ invariant to all permutations of r, s, and t:

$$u\equiv u(r,)\equiv u(r,s,t)=u(r,t,s)=u(s,r,t). \tag{2.4}$$

The primed notation is that used by Miller; the more explicit notation $f(r)$, $v(r;)$, $u(r,)$ is helpful in applica-

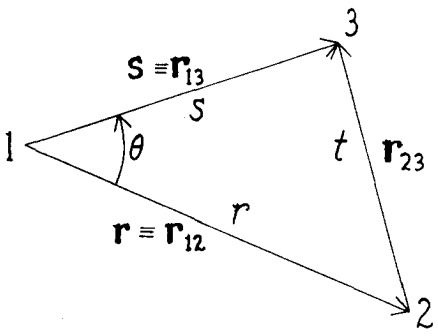


FIG. 1. Notation for describing the relative positions of three points 1, 2, 3. In most of this article, the independent variables chosen are r, s, and t.

tions of the theorem of the preceding paragraph.

Miller's three-point probabilities are listed and defined in Table I. The possible states are listed in column 1, where, for example, an entry A in the subcolumn headed 1 means that point 1 is in material A, and brackets enclose points that are in the same cell. (The brackets in "$A]B[A$" are to be interpreted as enclosing the two A's. Two points in the same cell are automatically in the same material.) The corresponding probabilities are listed in column 2. They are expressed by Miller in terms of the conditional probabilities defined in columns 3 and 4. Here column 3 shows the condition under which the conditional probability in column 4 is defined; the state of which it is the conditional probability is the state or group of states in column 1. Thus g_1 is the conditional probability that all three points are in a single A cell, given that point 1 is in A; the condition can equally be that point 2 or point 3 is in A, but for simplicity these alternates have not been indicated in the table. Again, h_1 is the conditional probability that points 1 and 2 are in a single A cell, given that point 1 (or equivalently point 2) is in A. The other g and the other h's are defined similarly. Finally, Z is the probability that the three points are in three separate cells; its value, in order that the total probability may be unity, is

$$Z = 1 - \phi(g_1 + h_1 + h_1' + h_1'') - (1-\phi)(g_2 + h_2 + h_2' + h_2''). \tag{2.5}$$

From the forms of the probabilities in column 2, it is evident that the "statistical independence" of the definition has been interpreted as follows: The conditional probability that point 1 is in A or B, given that points 2 and 3 are not in the same cell with it, is independent of whether the other points are in a single cell or in two different cells and of which materials they are in, and is ϕ for A, $1-\phi$ for B. Thus the probability of $B\,[A\,A]$ has been found by multiplying the probability that points 2 and 3 are in a single A cell, namely $\phi h_1''$, by $1-\phi$.

Is the independence thus postulated consistent with the assumption of different functions g_1 and g_2, etc., for the two materials?

The test of the consistency of a set of three-point probabilities is that they must give the correct one-point probabilities; it is not sufficient that they give the correct total probability 1. In this case the probability that any one point shall be in A must be ϕ; and in B, $1-\phi$.

On adding all the probabilities for states in which point 1 is in A and equating the sum to ϕ, we get

$$\phi[g_1 + h_1 + h_1' + \phi h_1'' + (1-\phi)h_2'' + Z] = \phi, \tag{2.6}$$

whence (except in the trivial case $\phi = 0$)

$$g_1 + h_1 + h_1' + \phi h_1'' + (1-\phi)h_2'' + Z = 1. \tag{2.7}$$

Similarly we get for point 2

$$g_1 + h_1 + h_1'' + \phi h_1' + (1-\phi)h_2' + Z = 1. \tag{2.8}$$

Subtraction gives

$$(1-\phi)(h_1' - h_1'' + h_2'' - h_2') = 0, \tag{2.9}$$

whence

$$h_1' - h_2' = h_1'' - h_2''. \tag{2.10}$$

Now the conditional probability that 1 and 2 are both in the same A cell, given that 1 is in A, is

$$f_1 = f_1(r) = g_1 + h_1 \tag{2.11}$$

and is a function of r only. Hence we may (as Miller did) express the asymmetric function $h_1(r;\,)$ in terms of the symmetric function $g_1(r,\,)$ and the function $f_1(r)$ of a single variable:

$$h_1 = f_1 - g_1. \tag{2.12}$$

Similarly $h_2 = f_2 - g_2$; also $h_1' = f_1' - g_1$, $h_1'' = f_1'' - g_1$, $h_2' = f_2' - g_2$, $h_2'' = f_2'' - g_2$. With these substitutions, (2.10) becomes

$$f_1' - f_2' = f_1'' - f_2'' \tag{2.13}$$

or

$$f_1(s) - f_2(s) = f_1(t) - f_2(t). \tag{2.14}$$

By the theorem of the second paragraph of this section, it follows from Eq. (2.14) that $f_1(s) - f_2(s) = \text{const}$. We suppose that the cells are of finite size, or at least have a distribution of sizes such that the probability of maximum linear cell dimension L approaches zero as $L \to \infty$; then $f_1(s)$ and $f_2(s)$ approach zero as $s \to \infty$, and therefore the constant is zero, and

$$f_2 = f_1,\ f_2' = f_1',\ f_2'' = f_1''. \tag{2.15}$$

Returning now to (2.7) and inserting (2.12) and (2.15), we get

$$(1-\phi)(g_1 - g_2) = 0 \tag{2.16}$$

and hence

$$g_2(r,\,) = g_1(r,\,). \tag{2.17}$$

From (2.15) and (2.17) it follows that $h_2 = h_1$, etc. Thus all the functions are the same for materials A and B. That is, the three-point probabilities postulated by Miller are consistent only if the material is symmetric; if $g_1 \neq g_2$ and so on, they do not add up to the correct one-point probabilities.

The same result can be obtained by setting the total proability that a point is in B equal to $1-\phi$. This condition and the one used are not independent, because

TABLE I. Three-point probabilities and conditional probabilities according to Miller.

State 1	2	3	Probability	Condition 1	2	3	Conditional Probability
[A	A	A]	ϕg_1	A			$g_1 \equiv g_1(r,)$
[B	B	B]	$(1-\phi)g_2$	B			$g_2 \equiv g_2(r,)$
[A	A]	A	$\phi^2 h_1$	A			$h_1 \equiv h_1(r;)$
[A	A]	B	$\phi(1-\phi)h_1$				
[B	B]	A	$\phi(1-\phi)h_2$	B			$h_2 \equiv h_2(r;)$
[B	B]	B	$(1-\phi)^2 h_2$				
A]	A	[A	$\phi^2 h_1'$	A			$h_1' \equiv h_1(s;)$
A]	B	[A	$\phi(1-\phi)h_1'$				
B]	A	[B	$\phi(1-\phi)h_2'$	B			$h_2' \equiv h_2(s;)$
B]	B	[B	$(1-\phi)^2 h_2'$				
A	[A	A]	$\phi^2 h_1''$	A			$h_1'' \equiv h_1(t;)$
B	[A	A]	$\phi(1-\phi)h_1''$				
A	[B	B]	$\phi(1-\phi)h_2''$	B			$h_2'' \equiv h_2(t;)$
B	[B	B]	$(1-\phi)^2 h_2''$				
A	A	A	$\phi^3 Z$				
A	A	B	$\phi^2(1-\phi)Z$				
A	B	A	$\phi^2(1-\phi)Z$				
A	B	B	$\phi(1-\phi)^2 Z$	—			$Z \equiv Z(r,)$ $= 1-\phi(g_1+h_1+h_1'+h_1'')$ $-(1-\phi)(g_2+h_2+h_2'+h_2'')$
B	A	A	$\phi^2(1-\phi)Z$				
B	A	B	$\phi(1-\phi)^2 Z$				
B	B	A	$\phi(1-\phi)^2 Z$				
B	B	B	$(1-\phi)^3 Z$				

Z has been given a value that insures that the sum of the two probabilities in question is unity.

It might seem that we could rescue the model by interpreting the ϕ of the three-point probabilities as a quantity (perhaps a function of r, s, and t) different from the one-point probability (or volume fraction) p. But then the probability that points 1 and 2 are both in the same A cell becomes $\phi(g_1+h_1)=\phi f_1=\phi(r,s,t)f_1(r)$. Since this must be a function only of r, ϕ must be independent of s and t. By a similar argument for points 1 and 3, ϕ must be independent of r and t. Hence ϕ must be a constant, equal to its value at $r=s=t=\infty$. But there the only states with nonvanishing probabilities are the last eight in Table I, whose probabilities are clearly p^3, $p^2(1-p)$, etc.; the g's and h's are all zero, $Z=1$, and we have $\phi^3=p^3$, $\phi^2(1-\phi)=p^2(1-p)$, etc., i.e. $\phi=p$. Thus we return to the original model.

3. A SELF-CONSISTENT ASYMMETRIC GENERALIZATION OF THE "SYMMETRIC CELL MATERIAL"

Since the "asymmetric cell material" as defined by Miller is not self-consistent, the asymmetric generalization of the "symmetric cell material" must take some other form. Such a generalization is the topic of this section.

We define "cells" as nonoverlapping regions, each of which contains only one of the individually homogeneous component materials. Since a cell as thus defined can always be divided into smaller cells, the division into cells is not unique. It could be made so by requiring that no two cells of the same material have a surface of contact, but for our purposes such a requirement would not be helpful. We shall suppose merely that the division has been made in some definite way, and that the cells are of finite size (or at least have a size distribution such that the probability of maximum dimension L approaches zero as L becomes infinite).

According to this definition, all two-phase mixtures may be considered to be composed of cells. In Miller's "symmetric cell material," the distribution of the two materials among the cells is statistically independent of the distribution of any other property that distinguishes different cells; in consequence, certain integrals over the specimen or over space reduce to integrals over a cell. The generalization undertaken here is a generalization *of this property* to a less restricted class of two-phase materials. Accordingly, we shall first consider an arbitrary two-phase mixture, and shall find the general form that the one-, two-, and three-point probabilities must take; we shall then determine what conditions may be imposed in order that the material may possess the desired property.

In this section, we shall use joint probabilities instead of conditional probabilities and shall use the notation p and q instead of ϕ and $1-\phi$.

We consider first the one-point probabilities. These are very simple: probability p that a point $\mathbf{r}$ picked at random is in material A, probability q that it is in B; in order that the total probability may be unity, p and q must satisfy the constraint

$$p+q=1. \tag{3.1}$$

Since we assume statistical homogeneity, p and q are constants, independent of $\mathbf{r}$.

We consider next the two-point probabilities. These are shown in Table II for points 1 and 2. We have used a single symbol R for the probability that 1 is in A and 2 in B and for the probability that 1 is in B and 2 in A. Their equality follows from the postulated statistical homogeneity and isotropy; these enable us, without alteration of the probability, to translate points 1 and 2 rigidly until 1 is where 2 was, and then to rotate $\mathbf{r}_{12}$ rigidly about the new position of 1 until 2 is where 1 was. The homogeneity and isotropy also insure that all the probabilities listed are functions of r only, and not of $\mathbf{r}_1$ and $\mathbf{r}_2$ separately or of the direction of $\mathbf{r}_{12}$. The corresponding probabilities for points 1 and 3 are $P'=P(s)$, etc.; for points 2 and 3, $P''=P(t)$, etc.

These probabilities must satisfy the constraints: total probability that 1 is in $A=p$, total probability that 1 is in $B=q$; and similarly for point 2. Explicitly,

$$P+P^*+R=p, \tag{3.2}$$

$$Q+Q^*+R=q. \tag{3.3}$$

TABLE II. Most general two-point probabilities.

State 1	2	Probability	Value at $r=0$	Value at $r=\infty$
[A	A]	$P \equiv P(r)$	P	0
[B	B]	$Q \equiv Q(r)$	q	0
A	A	$P^* \equiv P^*(r)$	0	p^2
A	B	$R \equiv R(r)$	0	pq
B	A	$R \equiv R(r)$	0	pq
B	B	$Q^* \equiv Q^*(r)$	0	q^2

TABLE III. Most general three-point probabilities.

State 1	State 2	State 3	Probability
[A	A	A]	$S_A \equiv S_A(r,)$
[B	B	B]	$S_B \equiv S_B(r,)$
[A	A]	A	$T_{AA} \equiv T_{AA}(r;)$
[A	A]	B	$T_{AB} \equiv T_{AB}(r;)$
[B	B]	A	$T_{BA} \equiv T_{BA}(r;)$
[B	B]	B	$T_{BB} \equiv T_{BB}(r;)$
A]	A	[A	$T'_{AA} \equiv T_{AA}(s;)$
A]	B	[A	$T'_{AB} \equiv T_{AB}(s;)$
B]	A	[B	$T'_{BA} \equiv T_{BA}(s;)$
B]	B	[B	$T'_{BB} \equiv T_{BB}(s;)$
A	[A	A]	$T''_{AA} \equiv T_{AA}(t;)$
B	[A	A]	$T''_{AB} \equiv T_{AB}(t;)$
A	[B	B]	$T''_{BA} \equiv T_{BA}(t;)$
B	[B	B]	$T''_{BB} \equiv T_{BB}(t;)$
A	A	A	$U_{AAA} \equiv U_{AAA}(r,)$
A	A	B	$U_{AAB} \equiv U_{AAB}(r;)$
A	B	A	$U_{ABA} \equiv U'_{AAB} \equiv U_{AAB}(s;)$
A	B	B	$U_{ABB} \equiv U''_{BBA} \equiv U_{BBA}(t;)$
B	A	A	$U_{BAA} \equiv U''_{AAB} \equiv U_{AAB}(t;)$
B	A	B	$U_{BAB} \equiv U'_{BBA} \equiv U_{BBA}(s;)$
B	B	A	$U_{BBA} \equiv U_{BBA}(r;)$
B	B	B	$U_{BBB} \equiv U_{BBB}(r,)$

These enable us to express two of the two-point probabilities in terms of the other three.

The third and fourth columns of Table II show the limiting values of P, Q, etc. at $r=0$ and at $r=\infty$. At $r=0$, the probability that one point is in A or B is p or q, respectively, and the other is then certain to be in the same cell and in the same material. At $r=\infty$, the two points have zero probability of being in the same cell, and their material probabilities are independent. [If the A cells have a maximum linear dimension L_A and the B cells a maximum linear dimension L_B, then P vanishes for $r>L_A$ and Q for $r>L_B$. When $r>L_A$ and L_B, both P and Q vanish, and the constraints (3.2) and (3.3) enable us to express P^* and Q^* in terms of R, though not to replace P^* by p^2 and so on.]

Finally, we consider the three-point probabilities. These are shown in Table III. The first column is the same as in Table I. The second column lists the symbols that will be used for the probabilities, in the abbreviated notation S_A, T_{AA}, T'_{AA}, etc. and in the less abbreviated notation $S_A(r,)$, $T_{AA}(r;)$, $T_{AA}(s;)$, etc. Although there are 22 different states, the probabilities can be expressed by means of only 10 different functions S_A, S_B, T_{AA}, T_{AB}, T_{BA}, T_{BB}, U_{AAA}, U_{BBB}, U_{AAB}, U_{BBA} by permuting the arguments. Among the probabilities of states with the points in three separate cells, U_{AAA} and U_{BBB} are symmetric in all three variables, whereas U_{AAB} and U_{BBA} are symmetric only in s and t.

The values of the ten different functions in limiting cases are shown in Table IV. These are based on the principles (1) that when two points coincide, they are certain to be in the same cell and the same material; and (2) that when one point is infinitely distant from the other two points, its probabilities and theirs are independent. Under the conditions listed, the three-point probabilities can be expressed simply in terms of the two-point. Such cases as $r=s=0$ and $r=s=\infty$, $t=0$ or ∞ need not be listed specially; they follow from the entries in Table IV by use of the further limiting values in Table II. The limiting values of such functions as T'_{AB} can be found by appropriate permutation of the variables r, s, t.

The probabilities in Table III are subject to the constraints that they must add up to the correct two-point probabilities of Table II. Thus the sum of the probabilities of all states with points 1 and 2 in a single A cell, namely $S_A + T_{AA} + T_{AB}$, must equal the probability P that points 1 and 2 are in a single A cell: $S_A(r,) + T_{AA}(r;) + T_{AB}(r;) = P(r)$. By carrying out this addition for each state of points 1 and 2, we get the following six constraints:

$$\begin{aligned} &S_A + T_{AA} + T_{AB} = P, \\ &S_B + T_{BA} + T_{BB} = Q, \end{aligned} \tag{3.4}$$

$$\begin{aligned} &T'_{AA} + T''_{AA} + U_{AAA} + U_{AAB} = P^*, \\ &T'_{BB} + T''_{BB} + U_{BBB} + U_{BBA} = Q^*, \end{aligned} \tag{3.5}$$

$$\begin{aligned} &T'_{AB} + T''_{BA} + U'_{AAB} + U''_{BBA} = R, \\ &T'_{BA} + T''_{AB} + U''_{AAB} + U'_{BBA} = R. \end{aligned} \tag{3.6}$$

The last two of these are not independent; one can be obtained from the other by interchanging the arguments s and t. The constraints obtained by considering the probabilities for points 1 and 3, and for points 2 and 3, need not be written separately; they can be obtained from the above by permutation of arguments.

We can solve Eqs. (3.4) for T_{AB} and T_{BA}, and Eqs. (3.5) for U_{AAB} and U_{BBA}. Substitution of appropriately permuted forms of the results in either of Eqs. (3.6) then gives a constraint which, by use of the two-point constraints (3.2) and (3.3), can be simplified to

$$\begin{aligned} S_A + S_B + (T_{AA} + T'_{AA} + T''_{AA}) + (T_{BB} + T'_{BB} + T''_{BB}) \\ + U_{AAA} + U_{BBB} = 1 - (R + R' + R''). \end{aligned} \tag{3.7}$$

In this relation between three-point and two-point functions, those functions that are not themselves symmetric in r, s, and t occur in symmetric combinations such as $R + R' + R''$.

If the functions that appear in this constraint are assigned arbitrary values consistent with it and with the limiting values at 0 and ∞, the other three- and two-point functions can be found from the other con-

TABLE IV. Values of the three-point probabilities in limiting cases.

Function	Value at $r=0$, $s=t$	Value at $t=0$, $s=r$	Value at $r=s=\infty$, t finite	Value at $s=t=\infty$, r finite
$S_A(r,)$	$P(s)$	$P(r)$	0	0
$S_B(r,)$	$Q(s)$	$Q(r)$	0	0
$T_{AA}(r;)$	$P^*(s)$	0	0	$P(r)p$
$T_{AB}(r;)$	$R(s)$	0	0	$P(r)q$
$T_{BA}(r;)$	$R(s)$	0	0	$Q(r)p$
$T_{BB}(r;)$	$Q^*(s)$	0	0	$Q(r)q$
$U_{AAA}(r,)$	0	0	$pP^*(t)$	$P^*(r)p$
$U_{AAB}(r;)$	0	0	$pR(t)$	$P^*(r)q$
$U_{BBA}(r;)$	0	0	$qR(t)$	$Q^*(r)p$
$U_{BBB}(r,)$	0	0	$qQ^*(t)$	$Q^*(r)q$

straints. The assigned values must also, of course, be consistent with the general requirement that a probability may not be less than 0 or greater than 1.

The constraint (3.7) can be put into a physically more illuminating form. The probability that all three points are in material A, without regard to their distribution among cells, is

$$p_{123}^{(3)} = S_A + T_{AA} + T'_{AA} + T''_{AA} + U_{AAA}. \tag{3.8}$$

Similarly, the probability that all three points are in material B is

$$q_{123}^{(3)} = S_B + T_{BB} + T'_{BB} + T''_{BB} + U_{BBB}. \tag{3.9}$$

The constraint (3.7) may therefore be written

$$p_{123}^{(3)} + q_{123}^{(3)} = 1 - (R + R' + R''). \tag{3.10}$$

Since Eq. (3.7) contains six three-point functions S_A, S_B, T_{AA}, T_{BB}, U_{AAA}, and U_{BBB} and one two-point function R, and since the only other constraints on these functions are the conditions at 0 and ∞ and the limitation to the interval $(0,1)$, it is clear that there is wide latitude in choosing them. Our aim, however, is to find a simple generalization of the "symmetric cell material".

The desirable property of the "symmetric cell material" is that it reduces the integral[3]

$$I = \frac{1}{16\pi^2}\frac{1}{\bar{\epsilon}^2}\int\int \frac{\partial^2\langle \epsilon_1'\epsilon_2'\epsilon_3'\rangle}{\partial z_2 \partial z_3}\frac{\mathbf{r}_{12}\cdot\mathbf{r}_{13}}{r_{12}^3 r_{13}^3}\,dv_3 dv_2 \tag{3.11}$$

to an integral over a single cell, in some cases simply related to the cell geometry. In Eq. (3.11), $\epsilon_i' = \epsilon_i - \bar{\epsilon}$, where ϵ_i is the dielectric constant (or other material constant) at point i and $\bar{\epsilon}$ is the volume or ensemble average of ϵ_i; $\langle\ \rangle$ denotes an ensemble average. For a two-phase mixture, $\epsilon_i = \epsilon_A$ when point i is in A and $= \epsilon_B$ when it is in B; $\bar{\epsilon} = p\epsilon_A + q\epsilon_B$; and if

$$\delta' \equiv \epsilon_A - \epsilon_B, \tag{3.12}$$

then

$$\langle \epsilon_1'\epsilon_2'\epsilon_3'\rangle = \delta'^3\{p_{123}^{(3)} - p[p_{12}^{(2)} + p_{13}^{(2)} + p_{23}^{(2)}] + 2p^3\}, \tag{3.13}$$

where $p_{123}^{(3)}$ is as before and $p_{12}^{(2)}$ is the probability that points 1 and 2 are both in material A. From Table II and Eq. (3.2),

$$p_{12}^{(2)} = P + P^* = p - R, \tag{3.14}$$

and the integral I reduces to

$$I = \frac{\delta'^3}{\bar{\epsilon}^2}\left(\frac{1}{16\pi^2}\int\int \frac{\partial^2 p_{123}^{(3)}}{\partial z_2 \partial z_3}\frac{\mathbf{r}_{12}\cdot\mathbf{r}_{13}}{r_{12}^3 r_{13}^3}\,dv_3 dv_2 - \tfrac{1}{3}p^2 q\right). \tag{3.15}$$

The term $-\frac{1}{3}p^2q$ is contributed by the term containing $p_{23}^{(2)}$ in (3.13) (see Appendix).

In general, the integral in (3.15) extends over all space; for the "symmetric cell material", however (Table I with $g_2 = g_1 = g$, etc.), in Miller's notation,

$$\begin{aligned} p_{123}^{(3)} &= \phi g + \phi^2(h + h' + h'') + \phi^3 Z \\ &= \phi g + \phi^2(h + h' + h'') + \phi^3(1 - g - h - h' - h''). \end{aligned} \tag{3.16}$$

Since h can be replaced by $f - g$, $p_{123}^{(3)}$ can be expressed in terms of two-point functions (whose contributions to I can be intergrated in closed form) and the single three-point function g, which vanishes when any one of r, s, and t exceeds the maximum linear dimension L of a cell. The same result can be obtained by using $q_{123}^{(3)}$ instead of $p_{123}^{(3)}$.

Our aim is to accomplish, for an asymmetric material, a similar reduction of the region of integration to a finite region.

In the formulas (3.8) and (3.9) for $p_{123}^{(3)}$ and $q_{123}^{(3)}$, the only terms that have the desired property of vanishing when any one of r, s, and t exceeds L_A and L_B are the terms S_A and S_B. Since the constraint (3.7) and the conditions at 0 and ∞ leave us considerable freedom in the choice of the functions, let us try to impose on the functions a further constraint that will reduce $p_{123}^{(3)}$ to an expression containing only S_A and S_B and the two-point functions R, R', and R''. Such a constraint is

$$\begin{aligned} T_{AA} + T'_{AA} + T''_{AA} + U_{AAA} = k_1 S_A + k_2 S_B \\ + k_3(R + R' + R'') + k_0, \end{aligned} \tag{3.17}$$

where the k's are constants to be determined. If (3.17) is satisfied, it follows from (3.7) that an analogous equation of the form

$$\begin{aligned} T_{BB} + T'_{BB} + T''_{BB} + U_{BBB} = l_1 S_A + l_2 S_B \\ + l_3(R + R' + R'') + l_0 \end{aligned} \tag{3.18}$$

must also be satisfied. To keep the calculation symmetric, we shall impose both (3.17) and (3.18) and shall later determine the relations between the k's and the l's by imposing (3.7). We shall call the constraints (3.17) and (3.18) *imposed* constraints and the earlier ones *natural* constraints.

The left member of (3.17) or (3.18) is the probability that the three points are all in A or in B, respectively, but not all in the same cell.

Equations (3.17) and (3.18) take the following forms in limiting cases:

At $r = 0$, $s = t$:

$$\begin{aligned} P^*(s) &= k_1 P(s) + k_2 Q(s) + 2k_3 R(s) + k_0, \\ Q^*(s) &= l_1 P(s) + l_2 Q(s) + 2l_3 R(s) + l_0; \end{aligned} \tag{3.19}$$

at $r = s = \infty$, t finite:

$$\begin{aligned} pP(t) + pP^*(t) &= k_3[R(t) + 2pq] + k_0, \\ qQ(t) + qQ^*(t) &= l_3[R(t) + 2pq] + l_0. \end{aligned} \tag{3.20}$$

(The forms at $t = 0$, etc., can be found by permuting the variables.) These in turn take the following forms in further limiting cases:

At argument 0:

$$\begin{aligned} 0 &= k_1 p + k_2 q + k_0, \\ 0 &= l_1 p + l_2 q + l_0, \\ p^2 &= k_3 \cdot 2pq + k_0, \\ q^2 &= l_3 \cdot 2pq = l_0; \end{aligned} \tag{3.21}$$

at argument ∞:

$$\begin{aligned} p^2 &= 2k_3 \cdot pq + k_0, \\ q^2 &= 2l_3 \cdot pq + l_0, \end{aligned}$$

$$p^3 = k_3 \cdot 3pq + k_0, \qquad q^3 = l_3 \cdot 3pq + l_0. \tag{3.22}$$

These eight equations consist of four involving the k's and four involving the l's; in each group, two equations are identical. The three independent k equations give

$$k_1 p + k_2 q = -p^2(1+2q), \quad k_3 = -p, \quad k_0 = p^2(1+2q), \tag{3.23}$$

and the three independent l-equations give

$$l_1 p + l_2 q = -q^2(1+2p), \quad l_3 = -q, \quad l_0 = q^2(1+2p). \tag{3.24}$$

Once k_1 and l_2 are assigned values, the constants are all determined. Equations (3.19) and (3.20) and the two-point constraints (3.2) and (3.3) then enable us to express P, P^*, Q, and Q^* in terms of R if we so desire.

On expressing k_2 in terms of k_1 and l_1 in terms of l_2, substituting the constants in (3.17) and (3.18), and inserting the results in (3.8) and (3.9), we get

$$p_{123}^{(3)} = (1+k_1)S_A + q^{-1}[-k_1 p - p^2(1+2q)]S_B - p(R+R'+R'') + p^2(1+2q), \tag{3.25}$$

$$q_{123}^{(3)} = p^{-1}[-l_2 q - q^2(1+2p)]S_A + (1+l_2)S_B - q(R+R'+R'') + q^2(1+2p). \tag{3.26}$$

Insertion of these in the natural constraint (3.10) gives, after considerable algebra (in which the relation $p+q=1$ must be frequently used),

$$[(p-q)(1+pq) + (k_1 p - l_2 q)][qS_A - pS_B] = 0. \tag{3.27}$$

If Eq. (3.27) is satisfied, our aim is accomplished. It can be satisfied by two methods.

The first method is to set

$$S_A/p = S_B/q \equiv g(r,s,t); \tag{3.28}$$

then $S_A = pg$, $S_B = qg$, and

$$k_1 S_A + k_2 S_B = (k_1 p + k_2 q)g = -p^2(1+2q)g, \quad l_1 S_A + l_2 S_B = (l_1 p + l_2 q)g = -q^2(1+2p)g, \tag{3.29}$$

by (3.23) and (3.24). Only these combinations of the constants are physically significant, not k_1 and k_2 separately or l_1 and l_2 separately. By virtue of the other Eqs. (3.23) and (3.24), the imposed constraints (3.17) and (3.18) become

$$T_{AA} + T'_{AA} + T''_{AA} + U_{AAA} = p^2(1+2q)(1-g) - p(R+R'+R''), \tag{3.30}$$

$$T_{BB} + T'_{BB} + T''_{BB} + U_{BBB} = q^2(1+2p)(1-g) - q(R+R'+R''). \tag{3.31}$$

That the three-point probabilities should be those of Miller's "symmetric cell material" is a sufficient but not a necessary condition for satisfaction of (3.30) and (3.31); thus we can, if we wish, generalize the "symmetric" material somewhat by this method. Of more interest, however, is the case of a material for which $S_A/p \neq S_B/q$.

The second method of satisfying (3.27), available when $g_1 \equiv S_A/p$ and $g_2 \equiv S_B/q$ are not equal, is to set

$$k_1 p - l_2 q = -(p-q)(1+pq). \tag{3.32}$$

If k_1 and l_2 are chosen so that this equation is satisfied, we find from Eqs. (3.25) and (3.15), with use of (3.14), that

$$I = \frac{1}{16\pi^2}\frac{\delta'^3}{\bar{\epsilon}^2}\{C_A G_A + C_B G_B\}, \tag{3.33}$$

where

$$C_A = 1 + k_1, \quad C_B = q^{-1}[-k_1 p - p^2(1+2q)], \tag{3.34}$$

and

$$G_A = \int\int \frac{\partial^2 S_A}{\partial z_2 \partial z_3} \frac{\mathbf{r}_{12}\cdot\mathbf{r}_{13}}{r_{12}^3 r_{13}^3}\, dv_3 dv_2, \tag{3.35}$$

$$G_B = \int\int \frac{\partial^2 S_B}{\partial z_2 \partial z_3} \frac{\mathbf{r}_{12}\cdot\mathbf{r}_{13}}{r_{12}^3 r_{13}^3}\, dv_3 dv_2. \tag{3.36}$$

Alternatively, we can use instead of (3.13) the equivalent formula

$$\langle \epsilon'_1 \epsilon'_2 \epsilon'_3 \rangle = -\delta'^3\{q_{123}^{(3)} - q[q_{12}^{(2)} + q_{13}^{(2)} + q_{23}^{(2)}] + 2q^3\}, \tag{3.37}$$

where the q's have the same meaning for material B that the p's have for material A. This gives formual (3.33) but with

$$C_A = p^{-1}[l_2 q + q^2(1+2p)], \quad C_B = -(1+l_2). \tag{3.38}$$

The pairs of formulas (3.34) and (3.38) are equivalent by virtue of (3.32).

From Eq. (3.32), we may set

$$p(k_1 + 1 + pq) = q(l_2 + 1 + pq) \equiv Kpq. \tag{3.39}$$

Then

$$k_1 = Kq - 1 - pq, \quad l_2 = Kp - 1 - pq, \tag{3.40}$$

and K remains an arbitrary constant, subject only to the condition that the resulting probabilities must all lie in the interval (0,1).

The relation of G_A and G_B to Miller's G_1 and G_2 is

$$G_A = 16\pi^2 p G_1, \quad G_B = 16\pi^2 q G_2. \tag{3.41}$$

4. ILLUSTRATIVE MODEL

In the previous section, the properties of our composite material were defined in a formal manner. In this section, it will be shown that such a material can in fact be constructed by a straightforward procedure.

The procedure consists of the following steps: (1) Divide the space by some random procedure into statistically equivalent cells. (2) Assign each cell randomly and independently to material A or to material B with probabilities ϕ and $1-\phi$, respectively (we use the symbol ϕ at this stage, reserving the symbols p and q for the ultimate one-point probabilities). (3) Divide each A cell independently, by some random process, into two subcells. (4) In each of the original A cells independently, select one of the subcells by a random and unbiased method and change its material to B. The random processes used must guarantee statistical isotropy and homogeneity.

TABLE V. Two-point probabilities for illustrative model.

State 1	State 2	Probability	Number of states after step (3)	Number of states after step (4)	
[(• •)]		$\zeta \equiv \zeta(r)$	2	2+1	=3
[(•)	(•)]	$\eta \equiv \eta(r)$	2	2+1	=3
[(•)]	[(•)]	$\theta \equiv \theta(r)$	4	4+2+2+1	=9
			8		15

The calculation of probabilities is facilitated by modifying the construction procedure to the following equivalent one: (1) Divide the space by some random procedure into statistically equivalent cells. (2) Divide every cell independently, by some random process, into two subcells. (3) Assign each of the original cells randomly and independently to material A or to material B with probabilities ϕ and $1-\phi$ respectively. (4) In each of the A cells independently, select one of the subcells by a random and unbiased method and change its material to B. (5) In each of the B cells, remove the boundaries between the subcells.

After step (2) of the modified procedure, the two-point probabilities are those shown in Table V, and the three-point probabilities are those shown in Table VI.[5] Brackets indicate cells and parentheses subcells: thus $[(\bullet\,\bullet)][(\circ)]$ means that points 1 and 2 are in the same subcell of one cell and point 3 is in a different cell, whereas $[(\cdot)(\circ)][(\cdot)]$ means that points 1 and 2 are in different subcells of one cell and point 3 is in a different cell. The two-point functions are shown only for points 1 and 2; for points 1 and 3 we write $\zeta' \equiv \zeta(s)$, etc., and for points 2 and 3 $\zeta'' \equiv \zeta(t)$, etc. The two-point probabilities are subject to the constant

$$\zeta + \eta + \theta = 1 \tag{4.1}$$

and the three-point probabilities to the constraints

$$\begin{aligned} &u + v + w = \zeta, \\ &v' + v'' + x = \eta, \\ &w' + w'' + x' + x'' + y = \theta \end{aligned} \tag{4.2}$$

and their permutations; the constraints (4.2) are obtained by equating the entries in Table V to appropriate sums of entries in Table VI. If we choose as independent functions ζ, η, u, and v, we can solve Eqs. (4.1) and (4.2) and their permutations for the other functions; this gives

$$\begin{aligned} &\theta = 1 - \zeta - \eta, \\ &w = \zeta - u - v, \\ &x = \eta - v' - v'', \\ &y = 1 - (\zeta + \sim) - (\eta + \sim) + 2u + 2(v + \sim) \end{aligned} \tag{4.3}$$

and permutations of these equations. Here we have abbreviated the symmetric sums $\zeta + \zeta' + \zeta''$, etc., to $\zeta + \sim$, etc.

The third column in Tables V and VI shows the number of states after step (3). In Table V, for example, the state $[(\bullet\,\bullet)]$ becomes, in this step, either $[(AA)]$ or $[(BB)]$, but the state $[(\cdot)][(\cdot)]$ has the four possibilities $[(A)][(A)]$, $[(A)][(B)]$, $[(A)][(B)]$, $[(B)][(A)]$, and $[(B)][(B)]$. The conditional probabilities of these states, given the initial states, are ϕ and $1-\phi$ in the first example; ϕ^2, $\phi(1-\phi)$, $\phi(1-\phi)$, and $(1-\phi)^2$ in the second.

In step (4) there is a further splitting: thus $[(AA)]$ remains unchanged or changes to $[(BB)]$ with probabilities $\frac{1}{2}$; $[(A)][(A)]$ has the four possibilities $[(A)][(A)]$, $[(A)][(B)]$, $[(B)][(A)]$, and $[(B)][(B)]$, each of probability $\frac{1}{4}$; $[(B)][(B)]$ remains unchanged with probability 1. The fourth column in Tables V and VI shows the number of states into which each of the states enumerated in column 3 splits, and the resulting total number of states. Thus in Table V, the initial state $[(\bullet\,\bullet)]$ has become, after step (3), either $[(AA)]$ or $[(BB)]$; after step (4), $[(AA)]$ has remained unchanged or changed to $[(BB)]$, with probabilities $\frac{1}{2}$, and $[(BB)]$ has remained unchanged with probability 1; the number of final states is indicated as $2+1=3$. Initial state $[(\cdot)][(\cdot)]$ has become, after step (3), $[(A)][(A)]$, $[(A)[[(B)]$, $[(B)][(A)]$, or $[(B)][(B)]$. After step (4), $[(A)][(A)]$ has become one of the four states $[(A)][(A)]$, $[(A)][(B)]$, $[(B)][(A)]$, $[(B)][(B)]$ with probabilities $\frac{1}{4}$; $[(A)][(B)]$ has become one of the two states $[(A)][(B)]$, $[(B)][(B)]$ with probabilities $\frac{1}{2}$; $[(B)][(A)]$ has become $[(B)][(A)]$ or $[(B)][(B)]$ with probabilities $\frac{1}{2}$; and $[(B)][(B)]$ has remained unchanged. Hence the entry in column 4 is $4+2+2+1=9$. The last entry in column 4 of Table VI, $8+3\cdot4+3\cdot2+1$, is short for $8+4+4+4+2+2+2+1$.

Step (5) consists merely of removal of the parentheses in B cells.

The complete two-point table, after step (4) or (5), contains 15 rows; the complete three-point table contains 93. Since the construction of the tables is straightforward, they will not be given here. The probability for each row is found by multiplying the probabilities in Tables V and VI by the two subsequent conditional probabilities [steps (3) and (4)]; then the various probabilities for each final state must be added together. The results, initially expressed in terms of ϕ, can be reexpressed in terms of the final one-point probabilities p and q by noting that

$$\tfrac{1}{2}\phi = p, \quad \tfrac{1}{2}\phi + (1-\phi) = q, \quad 1-\phi = -(p-q); \tag{4.4}$$

it is these combinations and powers of them that occur in the formulas.

TABLE VI. Three-point probabilities for illustrative model.

State 1	State 2	State 3	Probability	Number of states after step (3)	Number of states after step (4)	
[(•	∘	∘)]	$u \equiv u(r,)$	2	2+1	=3
[(∘	∘)	(∘)]	$v \equiv v(r;)$	2	2+1	=3
[∘)	(∘)	(∘]	$v' \equiv v(s;)$	2	2+1	=3
[(∘)	(∘	∘)]	$v'' \equiv v(t;)$	2	2+1	=3
[(∘	∘)]	[(∘)]	$w \equiv w(r;)$	4	4+2+2+1	=9
∘)]	[(∘)]	[(∘	$w' \equiv w(s;)$	4	4+2+2+1	=9
[(∘)]	[(∘	∘)]	$w'' \equiv w(t;)$	4	4+2+2+1	=9
[(∘)	(∘)]	[(∘)]	$x \equiv x(r;)$	4	4+2+2+1	=9
(∘)]	[(∘)]	[(∘)	$x' \equiv x(s;)$	4	4+2+2+1	=9
[(∘)]	[(∘)	(∘)]	$x'' \equiv x(t;)$	4	4+2+2+1	=9
[(∘)]	[(∘)]	[(∘)]	$y \equiv y(r,)$	8	8+3∘4+3∘2+1	=27
				40		93

The results of this calculation are as follows:

$$
\begin{aligned}
&P=p\zeta,\\
&Q=p\zeta-(p-q)(\zeta+\eta),\\
&P^*=p^2\theta,\\
&Q^*=q^2\theta,\\
&R=p\eta+pq\theta;
\end{aligned}
\tag{4.5}
$$

$$
\begin{aligned}
&S_A=pu,\\
&S_B=qu-(p-q)(v+\sim),\\
&T_{AA}=p^2w,\\
&T_{AB}=pv+pqw,\\
&T_{BA}=pv+pqw-p(p-q)x,\\
&T_{BB}=q^2w-q(p-q)x,\\
&U_{AAA}=p^3y,\\
&U_{AAB}=p^2(x'+x'')+p^2qy,\\
&U_{BBA}=pq(x'+x'')+pq^2y,\\
&U_{BBB}=q^3y.
\end{aligned}
\tag{4.6}
$$

It may be verified that these functions satisfy the natural constraints (3.2), (3.3), (3.4), (3.5), (3.6), and (3.7). They also satisfy the imposed constraints (3.17) and (3.18), with

$$
\begin{aligned}
&k_1=q^2-1, && l_1=-q^2,\\
&k_2=-p^2, && l_2=p^2-1,\\
&k_3=-p, && l_3=-q,\\
&k_0=p^2(1+2q), && l_0=q^2(1+2p).
\end{aligned}
\tag{4.7}
$$

The last statement may be proved by expressing each member of (3.17) and (3.18) in terms of the independent functions ζ, η, u, and v by means of (4.3); the two members are equal if and only if the coefficients of corresponding functions are equal, and these conditions are satisfied if the constants have the values (4.7).

The values of the k's and l's are those of Sec. 3, with $K=1$ in Eq. (3.40).

With this model, since $p=\frac{1}{2}\phi$, p is limited to the range $0\leq p\leq\frac{1}{2}$. To get values of p in the range $\frac{1}{2}$ to 1, we may interchange the roles of the two materials, splitting the B cells instead of the A cells. The formulas can be obtained from the preceding ones by interchanging A and B, p and q, k_1 and l_2, k_2 and l_1, k_3 and l_3, and k_0 and l_0. Equations (4.6) change, but Eqs. (4.7) are still valid, so that again $K=1$. For $p=\frac{1}{2}$, the two materials are treated alike: to construct the model in this special case, one divides the space into cells, divides every cell into two subcells, and then in each cell, randomly and independently, assigns one subcell to A and the other to B.

Comparison of Eqs. (4.6) with Table I shows distinct differences from the probabilities assumed by Miller. According to Miller, U_{AAA}: U_{AAB}: U_{BBA}: $U_{BBB}=p^3$: p^2q: pq^2: q^3; our U_{AAB} and U_{BBA} each contain an extra term not present in U_{AAA} and U_{BBB}. According to Miller, T_{AB}: $T_{AA}=q$: p, and T_{BA}: $T_{BB}=p$: q; our T_{AB} and T_{BA} each contain an extra term pv. It is interesting that these differences persist even when $p=\frac{1}{2}$, even though in that case $S_B=S_A$, $T_{BB}=T_{AA}$, $T_{BA}=T_{AB}$, $U_{BBB}=U_{AAA}$, and $U_{BBA}=U_{AAB}$; that is, even though the material is then completely symmetric.

This example demonstrates that the type of material defined in Sec. 3 is not a mere formal abstraction but can in fact be constructed by definite procedures.

5. DISCUSSION

The foregoing sections demonstrate that although Miller's generalization of his "asymmetric cell material" is not self-consistent, a self-consistent generalization, based on the requirement that a certain integral reduce to one over a finite range, is possible. The illustrative model of Sec. 4 is a specific realization of that generalization. It is subject to this criticism: It may owe its success to its having started with material-independent cell probabiliites. There has been no demonstration that the desired reduction of the integral is possible when the construction procedure begins with geometrically different processes for forming cells of different materials: for example, with the process described by Miller[1], Sec. 2D, in which spherical A cells and aspherical B cells grow from randomly distributed seeds.

To the extent that this generalization succeeds, it suggests that two-phase mixtures in which the integration range reduces to a finite one owe that property not to the fact that they are "cell materials," but to the fact that their two- and three-point probability functions satisfy certain rather special relations.

APPENDIX: INTEGRATION OF TERMS IN / DEPENDENT ON TWO-POINT FUNCTIONS

Because of the presence of the operator $\partial^2/\partial z_2\partial z_3$ in (3.11), the only terms in (3.13) that contribute to I are the three-point term $\delta'^3p^{(3)}_{123}$ and the two-point term $-\delta'^3pp^{(2)}_{23}$. We therefore consider the integral

$$
I_f=\int\int\frac{\partial^2 f(r_{23})}{\partial z_2\partial z_2}\frac{\mathbf{r}_{12}\cdot\mathbf{r}_{13}}{r_{12}^3r_{13}^3}\,dv_3dv_2. \tag{A1}
$$

The following integration procedure is not the shortest (cf. Miller[1], Appendix B), but it is straightforward and requires no special tricks beyond the usual ones of integral calculus.

In fixed xyz axes, with center at point 1, let the spherical coordinates of point 2 be (r, Θ, Φ); then

$$
dv_2=r^2\sin\Theta\,dr d\Theta\,d\Phi. \tag{A2}
$$

Let $x'y'z'$ axes, with center at point 1, be oriented with Euler angles[6] Φ, Θ, 0, so that $\mathbf{r}_{12}$ is along z'. In these axes, let the spherical coordinates of point 3 be (s, θ, ϕ); then for given $\mathbf{r}_{12}$

$$
dv_3=s^2\sin\theta\,ds d\theta d\phi, \tag{A3}
$$

where θ is the angle between $\mathbf{r}_{12}$ and $\mathbf{r}_{13}$ in Fig. 1. The factor $\mathbf{r}_{12}\cdot\mathbf{r}_{13}/r_{12}^3r_{13}^3$ in (A1) is $(\cos\theta)/r^2s^2$; the factor $\partial^2 f(r_{23})/\partial z_2\partial z_3$ is

$$
\frac{\partial^2 f(r_{23})}{\partial z_2\partial z_3}=-\left[f_{tt}(t)\frac{z_{23}^2}{t^2}+\frac{f_t(t)}{t}\left(1-\frac{z_{23}^2}{t^2}\right)\right], \tag{A4}
$$

where

$$t^2 = r_{23}^2 = r^2 + s^2 - 2rs\cos\theta \tag{A5}$$

and

$$z_{23} = z_3 - z_2 = (-x_3'\sin\Theta + z_3'\cos\Theta) - r\cos\Theta$$
$$= s(-\sin\Theta\sin\theta\cos\phi + \cos\Theta\cos\theta) - r\cos\Theta; \tag{A6}$$

a subscript t indicates differentiation with respect to t. Since the limits (0 to ∞ for r and s, 0 to π for Θ and θ, 0 to 2π for Φ and ϕ) are all independent of the variables, the integrations over the six variables chosen can be performed in any order.

The integrations can be carried out as follows: (1) Integrate over Φ, Θ, and ϕ; the result is insertion of a factor $8\pi^2$ and replacement of z_{23}^2 by $\frac{1}{3}t^2$. (2) For given r and s, change from θ to t as variable of integration; t goes from $|r-s|$ to $r+s$. (3) Invert the order of integration over t and s, and carry out the integration over s; it goes from $|r-t|$ to $r+t$, and the result is zero when $t>r$, so that t may be integrated from 0 to r rather than from 0 to ∞. (4) Invert the order of integration over t and r, and carry out the integration over r; the result of this step is

$$I_f = -\tfrac{16}{3}\pi^2\int_0^\infty [tf_{tt}(t) + 2f_t(t)]\,dt. \tag{A7}$$

(5) Integrate the first term by parts. If $f(t)$ and $f_t(t)$ are continuous and if $tf_t(t)\to 0$ as $t\to 0$ or ∞, this gives

$$I_f = -(16/3)\pi^2[f(\infty) - f(0)]. \tag{A8}$$

In the application to (3.11), $f(r_{23}) = -\delta'^3 p p_{23}^{(2)}$, so that $f(\infty) = -\delta'^3 p^3$, $f(0) = -\delta'^3 p^2$, $f(\infty) - f(0) = \delta'^3 p^2 q$; this leads to the last term in (3.15).

If $f(t)$ and $f_t(t)$ are zero for $t>R$ and continuous for $t<R$, and if $tf_t(t)\to 0$ as $t\to 0$, the term $f(\infty)$ in (A8) is replaced by $Rf_t(R-0) + f(R-0)$. For validity of Eq. (A8) in this case, $Rf_t(R-0) + f(R-0)$ must vanish; this will be the case if, for example, both $f(t)$ and $f_t(t)$ approach zero as t approaches R from below. Throughout the text, it has been assumed that one or another of the sets of conditions necessary for validity of Eq. (A8) is satisfied; if this is not so, the results may require revision.

[1]M.N. Miller, J. Math. Phys. 10, 1988—2004 (1969).
[2]M.N. Miller, J. Math. Phys. 10, 2005—2013 (1969).
[3]M. Beran, Nuovo Cimento 38, 771—782 (1965).
[4]Obviously the statistical properties of boundary cells must differ from those of internal cells, since the orientation of the boundary imposes geometric constraints. We assume throughout that the specimen is so large that the special properties of boundary regions are unimportant.
[5]The notation in Tables V and VI differs from that in Tables I—III in one respect: Brackets have been placed around a point that is in a different cell from the other point or points. The purpose is to make explicit the relation of cells to subcells.
[6]E.T. Whittaker, *Analytical Dynamics* (Cambridge U.P., Cambridge, 1927), 3rd Ed., pp. 9—10.

Some unitarity bounds for finite matrices

G. Mennessier and J. Nuyts*

Laboratoire de Physique Theorique et Hautes Energies, Orsay, France†
(Received 24 October 1972)

We study the problem of obtaining the bounds on the modulus of one element of a finite unitary matrix once the values of the moduli of a set of other elements are given. The problem is solved in simple cases, and indications of more general cases are given. This question is interesting from a physical point of view since it leads to direct inequalities between sets of partial waves. It would also be useful to extend this kind of results to the continuous matrices where inequalities on cross sections could be obtained.

I. INTRODUCTION

Unitarity imposes strong restrictions on the S matrix elements, but due to its nonlinear character these restrictions are usually difficult to obtain. In this article we have tried to obtain bounds on the modulus of one element of S once some set of moduli of other elements are known. Only few results have yet been obtained in this field where the S matrix is infinite dimensional.

As a first attempt to explore the problem we have focussed our attention on the simpler case of a finite unitary matrix. This, in itself, is already a rather difficult task. We have succeeded in solving it in a few cases only. Some of them, we hope, may become physically interesting.

We have restricted ourselves to matrices which are symmetrical (i.e., equal to their transpose). Indeed time-reversal invariance is probably a good symmetry of strong interactions and implies that

$$\langle f_{\text{out}}|i_{\text{in}}\rangle=\langle f_{\text{in}}|S|i_{\text{in}}\rangle=\langle i_{\text{in}}{}^{T}|S|f_{\text{in}}{}^{T}\rangle=\langle i_{\text{out}}{}^{T}|f_{\text{in}}{}^{T}\rangle, \quad \text{(I. 1)}$$

where the state $|a^T\rangle$ is the time-reversed state (spin and momenta reversed) of $|a\rangle$. By choosing the basic states $|a+a^T\rangle$ and $i|a-a^T\rangle$ one sees easily that the S matrix is symmetrical due to (I. 1). Henceforth we will always work in the latter basis. Thus

$$S=S^t \quad (t=\text{transposed}). \quad \text{(I. 2)}$$

II. LAGRANGE FORMULATION OF THE PROBLEM. NOTATION

Let S be a N-dimensional unitary symmetric matrix with elements s_{ij}

$$SS^+=S^+S=1 \quad (+=\text{Hermitian conjugate}) \quad \text{(II. 1)}$$

$$S=S^t \quad \text{(II. 2)}$$

The problem we want to solve is to obtain the bounds on the modulus of one given element s_{pq} of S once the moduli of a set $\mathcal{J}'$ of J-1 elements of S are given,

$$a_{ij}=|s_{ij}|, \quad (i,j)\in\mathcal{J}'. \quad \text{(II. 3)}$$

This problem can be formulated by using the method of Lagrange multipliers. Let Λ be the matrix of Lagrange multipliers related to the unitarity condition (II. 1). It can be chosen to be Hermitian

$$\Lambda^+=\Lambda. \quad \text{(II. 4)}$$

Let Ξ be the complex matrix of Lagrange multipliers related to the symmetry condition (II. 2). It can be chosen to be skewsymmetric.

$$\Xi^t=-\Xi. \quad \text{(II. 5)}$$

Let finally Y be the real symmetric matrix of Lagrange multipliers related to the moduli conditions (II. 3). Explicitly

$$\begin{aligned} &y_{ij}=y_{ji} \text{ real arbitrary for } (i,j)\in\mathcal{J}', \\ &y_{pq}=y_{qp}=1 \text{ for the special element } (p,q), \\ &y_{ij}=0 \text{ for all other elements.} \end{aligned} \quad \text{(II. 6)}$$

For convenience the set of elements for which y_{ij} can be nonzero will be denoted by $\mathcal{J}$

$$\mathcal{J}=\mathcal{J}+(p,q), \quad \text{(II. 7)}$$

then J is the number of elements of $\mathcal{J}$.

To find the extremal values of $|s_{pq}|$, one has to extremize the following action:

$$\begin{aligned} \mathcal{A}=&\operatorname{Tr}(\Lambda(S^+S-1))+\operatorname{Tr}(\Xi^*(S-S^t))+\operatorname{Tr}(\Xi(S^*-S^+)) \\ &+\sum_{(i,j)} y_{ij}(|s_{ij}|^2-a_{ij}^2) \end{aligned} \quad \text{(II. 8)}$$

(* = complex conjugate).

Introduce the symmetric matrix X with elements x_{ij}

$$x_{ij}=y_{ij}s_{ij} \quad (y_{ij} \text{ real}), \quad \text{(II. 9a)}$$

$$X=X^t. \quad \text{(II. 9b)}$$

Note that (II. 9a) is not in the form of a matrix product. The derivative of $\mathcal{A}$ with respect to s_{ij} or $s_{ij}{}^*$ gives the matrix equation

$$S\Lambda-2\Xi+X=0. \quad \text{(II. 10)}$$

The symmetrical part of this equation is

$$S\Lambda+\Lambda^tS+2X=0, \quad \text{(II. 11)}$$

which in turn implies

$$SX^+=XS^+ \quad \text{(II. 12a)}$$

or equivalently

$$S^+X=X^+S. \quad \text{(II. 12b)}$$

The Ξ matrix can then be chosen to be zero. Then

$$\Lambda=-S^+X, \quad \Xi=0. \quad \text{(II. 13)}$$

The crucial step to be performed is to obtain a solution of the equations (II. 12) and (II. 1-2-3) with the restrictions imposed by (II. 9), namely that x_{ij} and s_{ij} have the "same phase", i.e., they differ by multiplicative real numbers.

The conditions imposed by the Lagrange method can be divided into two classes.

The first class contains the conditions of symmetry and unitarity of S, of symmetry of X as well as equation (II. 12). These four sets of equations are invariant under transformations of the form

$$X=U\tilde{X}U^t, \quad S=U\tilde{S}U^t, \tag{II. 14}$$

where U is an arbitrary unitary $N\times N$ matrix.

The second class of conditions contains the restrictions on the moduli of S (II. 3) and the phase condition (II. 9). These equations are in general clearly not invariant under the transformations of equation (II. 14). However if one restricts oneself to the special unitary matrices which are diagonal

$$U_{jk}=\delta_{jk}\exp(i\psi_j). \tag{II. 15}$$

The phase and the moduli conditions are also invariant. This freedom will be used subsequently since it allows to choose some phases arbitrarily.

It will often be convenient to state our results very symmetrically by considering the J-dimensional space $\mathcal{E}$ spanned by the $a_{ij}=|s_{ij}|$ $[(i,j)\in\mathcal{G}]$. Let $\mathcal{V}$ be the J-dimensional volume of $\mathcal{E}$ defined as follows: a point a_{ij} $((i,j)\in\mathcal{G})$ belongs to $\mathcal{V}$ if, and only if, there exist both a system of phases φ_{ij}

$$s_{ij}=a_{ij}\exp(i\varphi_{ij}), \quad (i,j)\in\mathcal{G} \tag{II. 16}$$

and a set of complex numbers

$$s_{kl}, \quad (k,l)\notin\mathcal{G} \tag{II. 17}$$

such that the full S matrix is symmetrical and unitary. In this language our problem is equivalent to finding the $(J-1)$-dimensional surface $\mathcal{B}$ which is the boundary of volume $\mathcal{V}$. As will become clear later, this surface $\mathcal{B}$ will often be composed of several different smooth pieces which intersect each other on lower dimensional surfaces.

It will be useful to subdivide the matrices $X, S, \ldots$ into submatrices characterized by partitions $\{n_i\}$ of N. Using upper indices to specify those submatrices, we write for example

$$X\begin{pmatrix} n_1, & n_2, & \cdots \\ m_1, & m_2, & \cdots \end{pmatrix}=\begin{pmatrix} X^{11} & X^{12} & X^{13} & \cdots \\ X^{21} & X^{22} & X^{23} & \cdots \\ X^{31} & X^{32} & X^{33} & \cdots \\ \cdot & & & \\ \cdot & & & \\ \cdot & & & \end{pmatrix} \tag{II. 18}$$

where X^{11} has dimension $(n_1\times m_1)$, X^{12} dimension $(n_1\times m_2)$, and X^{ij} dimension $(n_i\times m_j)$.

$$N=\Sigma\, n_i=\Sigma\, m_i. \tag{II. 19}$$

When $n_i=m_i$, only one set of indices will be used, $X(n_1, n_2, \ldots)$. In the next section the case in which all $X(m,n,p)$ except X^{12} is zero will be treated explicitly.

III. FIRST CASE

A. Presentation of the problem

Let all matrices involved in the problem have a decomposition (m,n,p). In particular

$$S(m,n,p)=\begin{pmatrix} S^{11} & S^{12} & S^{13} \\ S^{12t} & S^{22} & S^{23} \\ S^{13t} & S^{23t} & S^{33} \end{pmatrix}. \tag{III. 1}$$

The problem we shall try to solve explicitely in this section is to obtain the bounds on the modulus of one element of S^{12} once all the remaining moduli of S^{12} are given and all the other elements of S are completely arbitrary or unknown.

According to (II. 6) and (II. 9) the general form of X is

$$X(m,n,p)=\begin{pmatrix} 0 & X^{12} & 0 \\ X^{12t} & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}, \tag{III. 2}$$

which means that the set $\mathcal{G}$ of elements (i,j) for which y may be nonzero is given by

$$\mathcal{G}=\{(i,j)\}: 1\leq i\leq m, \quad m+1\leq j\leq m+n. \tag{III. 3}$$

B. Solution of the first-class conditions

We now present the solution of the first class of conditions. Since we have in mind the second part of the problem it is useful to restrict ourselves to the special transformations (II. 14),

$$U(m,n,p)=\begin{pmatrix} U^{11} & 0 & 0 \\ 0 & U^{22} & 0 \\ 0 & 0 & U^{33} \end{pmatrix}, \tag{III. 4}$$

which have the nice property to leave the set $\mathcal{G}$ $(X^{12}\neq 0)$ globally invariant. As the detailed proof of the resulting form of $\tilde{X}^{12}$ and $\tilde{S}^{12}$ is rather involved, we have deferred it to Appendix B. The relevant results only will be given here.

To be definite let us take $n\geq m$. The matrices $\tilde{X}^{12}$ and $\tilde{S}^{12}$ with m lines and n columns can be "pseudodiagonalized", i. e., their first m columns form a square diagonal matrix while their $(n-m)$ last columns are identically zero. The diagonal elements of $\tilde{X}^{12}(x_i)$ are real. The diagonal elements of $\tilde{S}^{12}(s_i)$ are the real positive square roots of the real positive eigenvalues of the matrix $(S^{12}S^{12+})$.

It is easy to see that the volume $\mathcal{V}$ in the $J=nm$-dimensional space $\mathcal{E}$ defined in Sec. II can then be characterized by the following requirement: a point a_{ij} belongs to $\mathcal{V}$ if, and only if, there exists a system of phases φ_{ij}

$$s_{ij}^{12}=a_{ij}\exp(i\varphi_{ij}) \tag{III. 5}$$

such that all the eigenvalues of $S^{12}S^{12+}$ are smaller than or equal to one:

$$s_j^2\leq 1. \tag{III. 6}$$

Thus all diagonal elements of $\tilde{S}^{12}$ must be larger than or equal to zero, and smaller than or equal to one. If S^{12}

defined by (III. 5) satisfies (III. 6), it is clear that the matrix

$$S'^{12} = \lambda S^{12} \quad (0 \leq \lambda \leq 1) \tag{III. 7}$$

leads to eigenvalues

$$s'^2_j = \lambda^2 s^2_j. \tag{III. 8}$$

Thus

Proposition 0: If a_{ij} is a point of $\mathcal{V}$ then λa_{ij} $(0 \leq \lambda \leq 1)$ belongs also to $\mathcal{V}$.

It is also shown in Appendix B that the x_i and the s_i are correlated in the following manner: for the diagonal elements of $\tilde{S}^{12}$ which are equal to one the corresponding values of $\tilde{X}^{12}$ are arbitrary. For all the diagonal elements of $\tilde{S}^{12}$ which have the common eigenvalue s $(0 < s < 1)$ the values of $\tilde{X}^{12}$ are either zero or appear by pairs of opposite signs. Finally, for the elements of $\tilde{S}^{12}$ which are zero the corresponding values of $\tilde{X}^{12}$ are again arbitrary. Explicitly

Proposition 1: The most general solution of (II. 1), (II. 2), (II. 9b), (II. 12), and (II. 14) for X^{12} and S^{12} is

$S^{12}(n_1, n_2, n_2, n'_2, n_3, \ldots; n-m)$

$$= U^{11} \begin{array}{c} \begin{matrix} n_1 & n_2 & n_2 & n'_2 & n_3 & \cdots & n-m \end{matrix} \\ \begin{pmatrix} s_1 & & & & & & \\ & s_2 & & & & & \\ & & s_2 & & & & \\ & & & s_2 & & & \\ & & & & 0 & & \end{pmatrix} \end{array} U^{22\dagger}, \tag{III. 9a}$$

$X^{12}(n_1, n_2, n_2, n'_2, n_3, \ldots; n-m)$

$$= U^{11} \begin{array}{c} \begin{matrix} n_1 & n_2 & n_2 & n'_2 & n_3 & \cdots & n-m \end{matrix} \\ \begin{pmatrix} x_1 & & & & & & \\ & x_2 \neq o & & & & & \\ & & -x_2 & & & & \\ & & & 0 & & & \\ & & & & x_3 & & \end{pmatrix} \end{array} U^{22\dagger}, \tag{III. 9b}$$

where the blocks labelled 1, $1 > s_2 > 0$, $0, \ldots; x_1$, $x_2 \neq 0$, ... are proportional to the unit matrix of the corresponding dimension (n_i or n'_i) with coefficient $1, s_2, 0, \ldots; x_1, x_2, \ldots$ The unitary matrices U^{11} $(m \times m)$ and $U^{22}(n \times n)$ are arbitrary.

C. A set of solutions of the second-class conditions

1. Generalities

We now turn to the problem of imposing the restrictions (II. 3) on the moduli of the elements of S^{12} and the phase conditions (II. 9a) between S^{12} and X^{12}. A matrix which satisfies these conditions will be called extremal. We have succeeded in solving this problem in a few cases only, namely when only one x_i of $\tilde{X}^{12}$ is different from zero. In Sec. IV we will show that these cases exhaust all the possible boundaries when $m = 2$, which is a physically interesting situation.

Let us first imagine that we take for the matrix elements of S^{12}

$$s^{12}_{ij} = a_{ij} \exp(i\varphi_{ij}) \quad [(i,j) \in \mathcal{J}], \tag{III. 10}$$

where φ_{ij} is a set of arbitrary phases. It is clear that the eigenvalues of $S^{12} S^{12+}$ (i.e., s^2_j) will in general be all different from each other, zero, and one. Our general Proposition 1 of Sec. III. B2 would then imply that the x_i be all equal to zero $(X = 0)$. This means that this S^{12} does not belong to an extremal matrix except if $s^{12}_{pq} = 0$. Indeed y_{pq} has to be equal to one, and

$$0 = x_{pq} = s^{12}_{pq} y_{pq} = s^{12}_{pq} = 0. \tag{III. 11}$$

This latter case $(s^{12}_{pq} = 0)$ corresponds to the minimal value possible for the modulus of s^{12}_{pq}. This extremum can be reached provided that the phases in (III. 10) can be chosen in such a way that the eigenvalues of $S^{12} S^{12+}$ are all smaller than one (and $s^{12}_{pq} = 0$). When the a_{ij} are sufficiently small such a type of solution always exists. This result can then be summarized in

Proposition 2: The hyperplanes $a_{ij} = 0$ are pieces of the $(nm - 1)$ dimensional boundary $\mathcal{B}$ of volume $\mathcal{V}$. There exists a neighborhood of the origin which is entirely contained in $\mathcal{V}$.

According to the characterization of volume $\mathcal{V}$ discussed in Sec. IIIB2, all other types of boundaries are related to the maximal possible value for the s_i, namely one. Following Proposition 1, X can then indeed be different from zero. Thus

Proposition 3: Except the case of Proposition 2, a necessary condition for S^{12} to be extremal is that $S^{12} S^{12+}$ has at least one eigenvalue equal to one.

As we have already said at the beginning of this section, we have been able to obtain the explicit extrema of $|s_{pq}|$ only when one x_i is different from zero and when the corresponding value of s is equal to one. Unfortunately, since, as is well known, the Lagrange parameter method may provide saddle points, and since we have not been able to solve the phase conditions when more than one x_i is different from zero, we cannot guarantee that we have obtained the true bounds. When $m = 2$ however, we will show that the discussion which follows exhausts all possible types of bounds.

2. x_1 only is different from zero. $U^{11}_{i1} \neq 0$ and $U^{22}_{i1} \neq 0$

Let $\tilde{X}^{12}_{11} = x$, the only non-zero element of $\tilde{X}^{12}$, and let the corresponding $\tilde{S}^{12}_{11}$ be equal to one.

According to (II. 9a) and (II. 14), X^{12} and S^{12}

$$S^{12} = U^{11} \tilde{S}^{12} U^{22\dagger}, \quad X^{12} = U^{11} \tilde{X}^{12} U^{22\dagger} \tag{III. 12}$$

must have the "same phase."

At this point it is useful to exploit the freedom contained in equation (II. 15) in order to choose U^{11}_{i1} and U^{22}_{i1} real. Since then

$$x^{12}_{ij} = x U^{11}_{i1} U^{22}_{j2} \tag{III. 13}$$

is purely real, the phase condition implies that S^{12} is purely real. One obtains then the matrix elements s^{12}_{ij} of $S^{12}(\epsilon)$ by

$$s^{12}_{ij} = \epsilon_{ij} a_{ij}, \tag{III. 14}$$

where $\epsilon=\{\epsilon_{ij}\}$ is an arbitrary set of signs. The extremal values of $|s_{pq}|$ are given by the extremal $S^{12}(\epsilon)$, i.e., those for which $S^{12}S^{12+}=S^{12}S^{12t}$ has one eigenvalue equal to one

$$\det\,(S^{12}(\epsilon)S^{12t}(\epsilon)-1)=0 \qquad \text{(III. 15)}$$

and the other eigenvalues smaller or equal to one. One then may choose U^{11} and U^{22} to be orthogonal matrices. In short

Proposition 4: Certain pieces of the boundary β of $\not\vee$ are of the form (III. 15).

3. x_1 only is different from zero. General case

The conclusion of the previous section does not hold if some elements of X^{12} are identically zero. Indeed, no phase condition then exists on the corresponding element of S^{12}. Remembering (III. 13) this situation implies that some U^{11}_{i1} and/or some U^{2*}_{i1} are zero. By re-labelling the lines and columns of S^{12}, it can be assumed that the $(m-k)$ last elements of U^{11}_{i1} are zero as well as the $(n-l)$ last elements of U^{11}_{i2}. The matrix X^{12} can then be written

$$X^{12}\begin{pmatrix}k, m-k\\ 1, n-1\end{pmatrix}=\begin{pmatrix}{}^{11}X^{12} & 0\\ 0 & 0\end{pmatrix}, \qquad \text{(III. 16)}$$

where the amputated ${}^{11}X^{12}$ is a $(k\times l)$ nonzero matrix. In a similar way the amputated ${}^{11}S^{12}$ is obtained from S^{12} by suppressing the $(n-l)$ last columns and the $(m-k)$ last lines.

It is quite obvious in view of the discussion of Sec. IIIC1 that $|s_{pq}|$ has to be chosen inside ${}^{11}S^{12}$. Otherwise $|s_{pq}|$ would be zero. On the other hand, the general theory of Lagrange multipliers tells us that when one of the multipliers y_{ij} is zero, the extremal value of $|s_{pq}|$ does not depend in general on the precise value of the corresponding a_{ij}. Indeed the bound of $|s_{pq}|$ depends only on the values of ${}^{11}S^{12}$ when this type of solution is realized. In Appendix C, we show that ${}^{11}S^{12}\,{}^{11}S^{12t}$ which by (III. 13) is real has an eigenvalue one and that the equation

$$\det({}^{11}S^{12}(\epsilon)\,{}^{11}S^{12t}(\epsilon)-1)=0 \qquad \text{(III. 17)}$$

represents a set of possible boundary values for $\not\vee$. Since this type of boundary does not exist for $m=n=2$, but is present in many other configurations, we conclude

Proposition 5: In almost all cases, certain pieces of the boundary of $\not\vee$ are of the cylindrical form (III. 17), where ${}^{11}S^{12}(\epsilon)$ is an amputated part of $S^{12}(\epsilon)$.

IV. APPLICATION OF THE FIRST CASE: $m=1, 2$

As an illustration of the results of the preceding section, we here present the general solution of our problem when S^{12} consists only of two lines ($m=2$). Indeed the case $m=1$

$$S^{12}=(s^{12}_{12}, s^{12}_{13}, s^{12}_{14}, \ldots, s^{12}_{1,n+1})$$

is completely trivial since our Proposition 4 implies that any one of these elements is maximal if

$$\sum_{i=2}^{n+1}a^2_{1i}=1, \qquad \text{(IV. 1)}$$

i.e., when this line saturates unitarity. Inside the case $m=2$ it is useful to distinguish two subcases $n=2$ and $n\geq 3$.

A. m = n = 2

When $m=n=2$ S^{12} has four elements

$$S^{12}=\begin{pmatrix}s_{13} & s_{14}\\ s_{23} & s_{24}\end{pmatrix}. \qquad \text{(IV. 2)}$$

The extremal S^{12} can be classified according to the eigenvalues of $S^{12}\,S^{12\dagger}$ or equivalently $\tilde{S}^{12}$.

When one of the values of $\tilde{S}^{12}$ is one, the boundary of Proposition 4 is given by

$$\det(S^{12}(\epsilon)S^{12t}(\epsilon)-1)=0. \qquad \text{(IV. 3)}$$

As can be seen directly, the only relevant sign among the ϵ_{ij} can be chosen to be ϵ_{24}, and $\epsilon_{13}=\epsilon_{14}=\epsilon_{23}=1$. Hence (IV. 3) describes two surfaces only. The surface $\epsilon_{24}=-1$ is the desired boundary of volume $\not\vee$ since it is always exterior to the surface $\epsilon_{24}=+1$ (application of proposition 0).

When the two values of $\tilde{S}^{12}$ are equal to one S^{12} simply is a unitary matrix. Using the freedom of phases (II. 15), S^{12} can be chosen real and is then a surface of lower dimension (2 instead of 3) entirely contained in boundary (IV. 3).

When one of the values of $\tilde{S}^{12}$ is zero, the two lines of S^{12} are proportional and

$$a_{13}/a_{23}=a_{14}/a_{24}. \qquad \text{(IV. 4)}$$

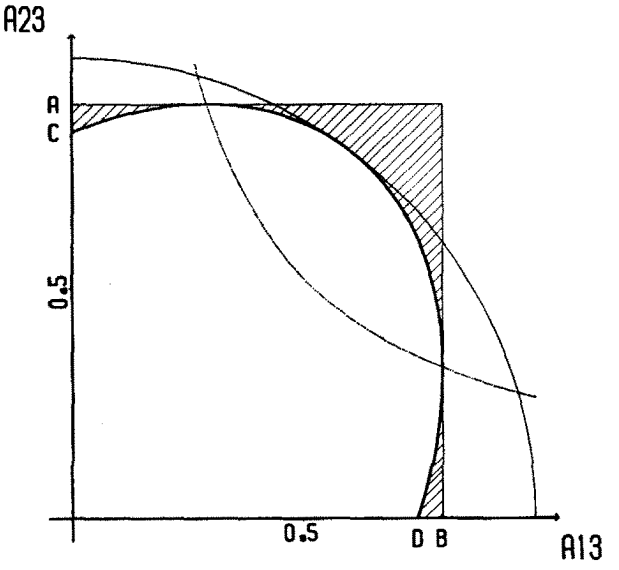


FIG. 1. Bounds on a_{23} as a function of a_{13} when a_{14} and a_{24} are fixed (here $a^c_{13}=\sqrt{1-a^2_{14}}=0.8$, $a^c_{23}=\sqrt{1-a^2_{24}}=0.9$). The shaded regions are those which are excluded by the full requirements of unitarity compared with the trivial bounds $a_{13}=a^c_{13}$ (vertical line from B) and $a_{23}=a^c_{23}$ (horizontal line from A). The ellipse starting from C and ending in D (the true bound) has equation $det(S^{12}S^{12\dagger}-1)=0$, i.e., $(a^2_{12}+a^2_{14}-1)(a^2_{23}+a^2_{24}-1)-(a_{13}a_{23}-a_{14}a_{24})^2=0$. The dotted line is the hyperbola $a_{13}a_{23}=a_{14}a_{24}\equiv\pi$, while the circle $a^2_{13}+a^2_{23}=1$ is another trivial bound.

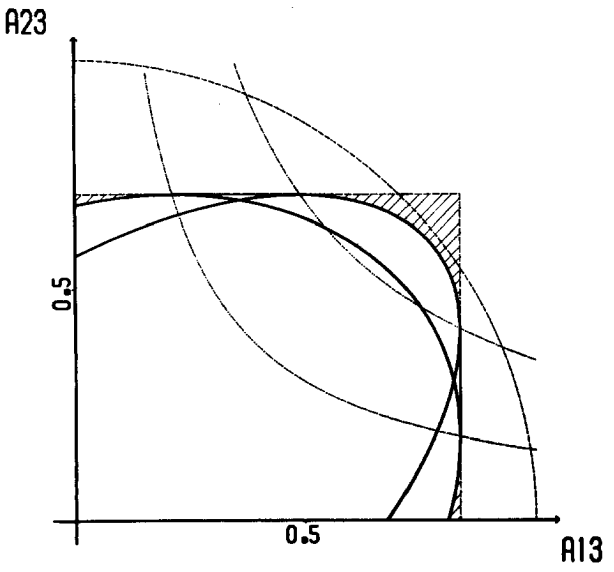


FIG. 2. Bounds on a_{23} as a function of a_{13} when $a_{1i}, a_{2i} (i = 4, \ldots, N)$ are fixed. Here $\pi^+ = 0.346$, $\pi^- = 0.15$, $a_{13}^c = 0.835$, and $a_{23}^c = 0.707$, we are in the case $0 < \pi^- < \pi^+ < a_{13}^c a_{23}^c$. The shaded regions are those excluded compared with the trivial bounds $a_{13} = a_{13}^c$ (vertical dotted line) and $a_{23} = a_{23}^c$ (horizontal dotted line). Remark that the parts of these straight lines between the two hyperbolas $a_{13}a_{23} = \pi^-$ and $a_{13}a_{23} = \pi^+$ are true bounds of the domain. The ellipses do not touch the circle $a_{13}^2 + a_{23}^2 = 1$, the other trivial bound any more.

The maximum occurs when the other value of $\tilde{S}^{12}$ is one, i.e., when

$$a_{13}^2 + a_{14}^2 + a_{23}^2 + a_{24}^2 = 1. \qquad \text{(IV. 5)}$$

Equation (IV. 4) together with Eq. (IV. 5) is of dimension 2, and is again entirely contained in boundary (IV. 3).

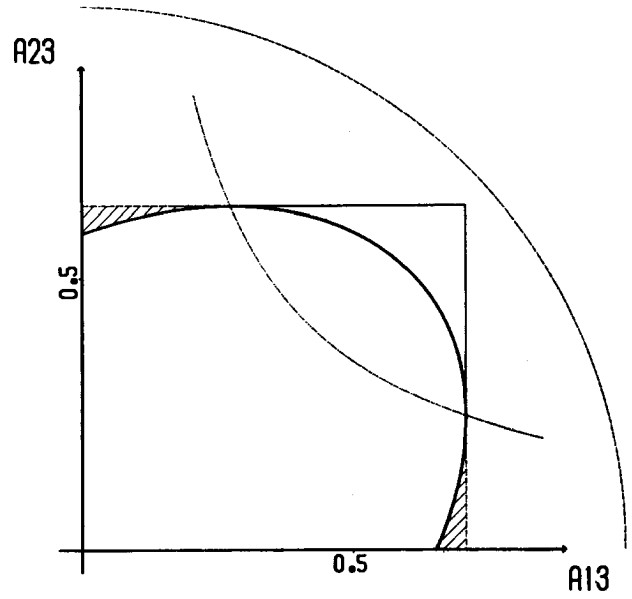


FIG. 3. Bounds on a_{23} as a function of a_{13} when $a_{1i}, a_{2i} (i = 4, \ldots, N)$ are fixed. Here $\pi^+ = 0.49$, $\pi^- = 0.173$, $a_{13}^c \cong 0.707$, and $a_{23}^c = 0.632$, we are in the case $0 < \pi^- < a_{13}^c a_{23}^c < \pi^+$. The shaded regions are those excluded compared with the trivial bounds $a_{13} = a_{13}^c$ and $a_{23} = a_{23}^c$. Remark that the parts of these straight lines between the hyperbola $a_{13}a_{23} = \pi^-$ and the point (a_{13}^c, a_{23}^c) are true bounds of the domain.

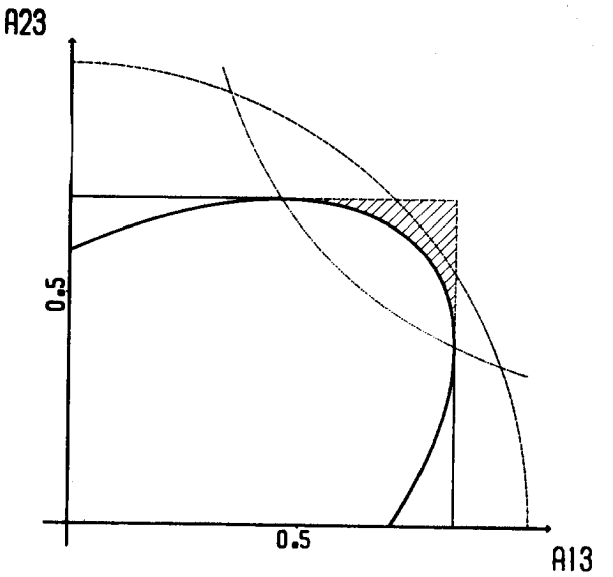


FIG. 4. Bounds on a_{23} as a function of a_{13} when $a_{1i}, a_{2i} (i = 4, \ldots, N; N > 5)$ are fixed. Here $\pi^+ = 0.324$, $\pi^- = -0.173$, $a_{13}^c = 0.835$, and $a_{23}^c = 0.707$, we are in the case $\pi^- < 0 < \pi^+ < a_{13}^c a_{23}^c$. The shaded regions are those excluded compared with the trivial bounds $a_{13} = a_{13}^c$ and $a_{23} = a_{23}^c$. Remark that the parts of these straight lines between the hyperbola $(a_{13}a_{23} = \pi^+)$ and the axis are true bounds of the domain.

The surface (IV. 3) ($\epsilon_{24} = -1$) is conveniently represented by an ellipse in the a_{13} a_{23} plane. (See Fig. 1). This corresponds to the intersection of the four-dimensional space of the a_{ij} by the hyperplanes $|s_{14}| = a_{14}$ and $|s_{24}| = a_{24}$, two given constants such that $a_{14}^2 + a_{24}^2 \leq 1$. All points inside the ellipse are allowed points of volume V. The ellipse is tangent to the straight lines

$$a_{13} = (1 - a_{14}^2)^{1/2} = a_{13}^c, \quad a_{23} = (1 - a_{24}^2)^{1/2} = a_{23}^c \qquad \text{(IV. 6)}$$

at the points where

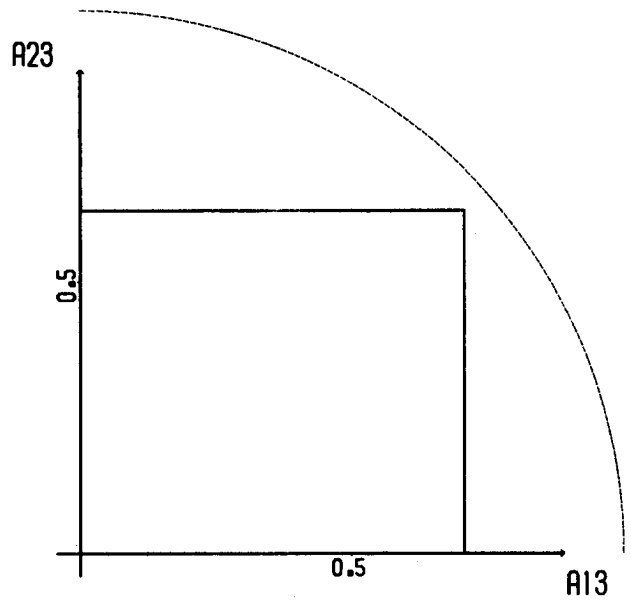


FIG. 5. Bounds on a_{23} as a function of a_{13} when a_{1i} and $a_{2i} (i = 4, \ldots, N; N > 5)$ are fixed. Here $\pi^+ = 0.52$, $\pi^- = -0.245$, $a_{13}^c = 0.707$, and $a_{23}^c = 0.632$, we are in the case $\pi^- < 0 < a_{13}^c a_{23}^c < \pi^+$. The bounds on the domain are the trivial straight lines $a_{13} = a_{13}^c$ and $a_{23} = a_{23}^c$.

$$a_{13}\,a_{23}=a_{14}\,a_{24}\equiv\pi, \tag{IV.7}$$

where π is introduced for later convenience. These points (corresponding to surfaces of dimension 2 and not 3 in $\mathcal{E}$) are the only remnants of the case of proposition 5 corresponding to S^{12} amputated of its first or second line. Finally the ellipse is also tangent to the circle

$$a_{13}^2+a_{23}^2=1 \tag{IV.8}$$

another remnant of S^{12} amputated of its second column.

B. $m=2, n>2$

When $m=2$ and $n>2$, S^{12} consists of two lines

$$S^{12}=\begin{pmatrix} s_{13} & s_{14} & s_{15} & \cdots & s_{1n+2} \\ s_{23} & s_{24} & s_{25} & \cdots & s_{2n+2} \end{pmatrix}. \tag{IV.9}$$

When the value of $\tilde{S}_{11}^{12}$ is one, the form of some boundaries is, as usual

$$\det(S^{12}(\epsilon)\,S^{12\,t}(\epsilon)-1)=0. \tag{IV.10}$$

In the plane a_{13}, a_{23}, once all the other a's have been fixed, (IV.10) is represented by ellipses or hyperbolas which are tangent to the critical straight lines

$$\begin{aligned} &\text{(a)}\ a_{13}=(1-\sum_{4}^{n+2} a_{1i}^2)^{1/2}\equiv a_{13}^c \\ &\text{(b)}\ a_{23}=(1-\sum_{4}^{n+2} a_{2i}^2)^{1/2}\equiv a_{23}^c. \end{aligned} \tag{IV.11}$$

At most two of these ellipses are relevant. Indeed (IV.11a) and (IV.11b) are of the form of equation (III.17) related to S^{12} amputated of one of its lines. So that part of the lines (IV.11a), (IV.11b) are bounds of domain $\not\mathcal{V}$. For these later bounds to appear, according to (C.13), one must be able to find phases such that the line which has been removed from S^{12} is unitary orthogonal to ${}^{11}U^{22}$. In this case ${}^{11}U^{22}$ is simply the remaining line in the amputated S^{12}. [Technically one uses (C.3) and (C.6) with ${}^{12}U^{11}=0$ being a consequence of unitarity for U^{11}].

Here we simply describe the final result. As suggested by (C.14), let π^+ and π^- be defined as follows:

$$\pi^+=\sum_{i=4}^{n+2} a_{1i}\,a_{2i}, \tag{IV.12}$$

$$\pi^-=a_{1l}\,a_{2l}-\sum_{\substack{i=4\\ i\neq l}}^{n+2} a_{1i}\,a_{2i}, \tag{IV.13}$$

where $a_{1l}\,a_{2l}$ is the largest of the products $a_{1i}\,a_{2i}$ ($i=4,\ldots,n+2$). Remark that, when $n=2$, $\pi^+=\pi^-=\pi$ (IV.7).

Four cases occur whether

$$\pi^+\gtrless a_{13}^c\,a_{23}^c, \quad \pi^-\gtrless 0. \tag{IV.14}$$

Specific examples are drawn in Figs. 2–5.

V. FURTHER STUDY OF DOMAIN $\not\mathcal{V}$ IN THE FIRST CASE

In this section, we study the relations between a matrix S^{12} and the matrices which are contained in it from the point of view of our problem.

Consider the decomposition

$$S^{12}=\begin{pmatrix} {}^{11}S^{12} & {}^{12}S^{12} \\ {}^{21}S^{12} & {}^{22}S^{12} \end{pmatrix}. \tag{V.1}$$

First let us remark that if a matrix S^{12} is unitarizable, then evidently any of its submatrices is unitarizable. This shows immediately that, if S^{12} has all its pseudo-eigenvalues smaller than or equal to one, i.e., if

$$1-S^{12}\,S^{12+}\geq 0 \tag{V.2}$$

is a semidefinite positive matrix, then for any amputated matrix ${}^{11}S^{12}$

$$1-{}^{11}S^{12}\,{}^{11}S^{12\,t}\geq 0. \tag{V.3}$$

Inversely, one may ask what are the conditions to be fulfilled by ${}^{11}S^{12}$, ${}^{12}S^{12}$, and ${}^{21}S^{12}$ for S^{12} to be unitarizable by the adjunction of a matrix ${}^{22}S^{12}$. When ${}^{21}S^{12}$ and ${}^{12}S^{12\,t}$ have one line only, the necessary and sufficient condition for the existence of a number ${}^{22}S^{12}$ is given by the theorem.

Theorem: Let $A=({}^{11}S^{12}, {}^{12}S^{12})$ and $B^t=({}^{11}S^{12\,t}, {}^{21}S^{12\,t})$ be given complex matrices (${}^{21}S^{12}$ and ${}^{12}S^{12\,t}$ one line only). If $\alpha_1=1-AA^+$ is a positive matrix, and if $\det(1-BB^+)$ is positive, then there exists a number ${}^{22}S^{12}$ such that S^{12} is unitarizable.

Proof: First we show that the conditions on A and B imply the existence of a number ${}^{22}S^{12}$ such that

$$\det(1-S^{12}\,S^{12+})>0. \tag{V.4}$$

Then we deduce that $1-S^{12}\,S^{12+}$ is positive, thus that S^{12} is unitarizable.

Writing $1-S^{12}\,S^{12+}$ in the form $(m-1, 1)$,

$$1-S^{12}\,S^{12+}=\begin{pmatrix} \alpha_1 & \alpha_2 \\ \alpha_2^+ & \alpha_4 \end{pmatrix}, \tag{V.5}$$

we find [see(D.9)]

$$\det(1-S^{12}\,S^{12+})=(\alpha_4-\alpha_2^+\,\alpha_1^{-1}\,\alpha_2)\det\alpha_1. \tag{V.6}$$

In this formula we are allowed to use α_1^{-1} since α_1 is positive. Consequently $\det\alpha_1$ is also positive, and (V.4) becomes

$$a\,|{}^{22}S^{12}|^2+b\,{}^{22}S^{12}+b^*\,{}^{22}S^{12*}+c>0, \tag{V.7}$$

where

$$\begin{aligned} a&=1+{}^{12}S^{12+}\,\alpha_1^{-1}\,{}^{12}S^{12}, \\ b&={}^{12}S^{12+}\,\alpha_1^{-1}\,{}^{11}S^{12}\,{}^{21}S^{12+}, \\ c&=-1+{}^{21}S^{12}\,{}^{21}S^{12+}+{}^{21}S^{12}\,{}^{11}S^{12+}\,\alpha_1^{-1}\,{}^{11}S^{12}\,{}^{21}S^{12+}. \end{aligned} \tag{V.8}$$

A solution ${}^{21}S^{12}$ of Eq. (V.7) exists if

$$|b|^2-ac>0. \tag{V.9}$$

As is shown in Appendix D, this condition can simply be written as

$$\det\alpha_1\det(1-BB^+)>0, \tag{V.10}$$

which is true by hypothesis.

Finally, to prove that $1-S^{12}\,S^{12+}$ is positive, let us consider an arbitrary vector V, and

$$M=V^+(1-S^{12}\,S^{12+})\,V. \tag{V.11}$$

Decomposing $V^t=(V_1^t, V_4^t)$, where V_4 is a number, one shows easily that M is always positive. Indeed

$$M=V_1^+\,\alpha_1 V_1+V_1^+\alpha_2 V_4+V_4^+\,\alpha_2^+\,V_1+V_4^+\,\alpha_4\,V_4. \tag{V.12}$$

Using the inequality $\alpha_4>\alpha_2^+\,\alpha_1^{-1}\,\alpha_2$ (V.6), we write

$$M > (V_1^+ \alpha_1^{1/2} + V_4^+ \alpha_2^+ \alpha_1^{-1/2})(\alpha_1^{1/2} V_1 + \alpha_1^{-1/2} \alpha_2 V_4) > 0 \tag{V. 13}$$

which is clearly positive for any V.

VI. SECOND CASE

A. Presentation of the problem

Let as usual all the matrices of the problem have the decomposition (m, n, p)

$$S(m,n,p) = \begin{pmatrix} S^{11} & S^{12} & S^{13} \\ S^{12t} & S^{22} & S^{23} \\ S^{13t} & S^{23t} & S^{33} \end{pmatrix}. \tag{VI. 1}$$

The problem we would like to solve is to obtain the bounds on the modulus of one element of S^{11} once the moduli of all the elements of S^{12} are given. We have succeeded in giving the general solution of this problem when $m = 2$ only.

According to (II. 6) and (II. 9) the general form of X is

$$X(m,n,p) = \begin{pmatrix} X^{11} & X^{12} & 0 \\ X^{12t} & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}, \tag{VI. 2}$$

where X^{11} is a $m \times m$ symmetric matrix whose only non-zero elements are those corresponding to the special element s_{pq} of S^{11}. The possible forms are either

$$X^{11}(1, m-1) = \begin{pmatrix} x_{11} & 0 \\ 0 & 0 \end{pmatrix}, \quad x_{11} = s_{11} \tag{VI. 3}$$

or

$$X^{11}(1,1,m-2) = \begin{pmatrix} 0 & x_{12} & 0 \\ x_{12} & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}, \quad x_{12} = s_{12}. \tag{VI. 4}$$

In Appendix E, we have solved with some details the problem when $m = 2$ and $p = 0$. We next present the results of this appendix and some indications of the more general problem.

B. Some solutions of the problem when $p = 0$

1. Real solution when m is arbitrary

Let a_{ij} $(i = (1, \ldots, m);\ j = (m+1, \ldots, m+n))$ be the given moduli of S^{12}. Let ϵ_{ij} be an arbitrary set of signs and let

$$s_{ij} = \epsilon_{ij} a_{ij} \quad (i = 1, \ldots, m;\ j = m+1, \ldots, m+n) \tag{VI. 5}$$

the elements of $S^{12}(\epsilon)$.

A system of extremal solutions is obtained by

Proposition 6: Let $A(\epsilon) = (S^{11}, S^{12})$ be a real matrix. It is extremal if

$$A(\epsilon) A^t(\epsilon) = 1_m. \tag{VI. 6}$$

This is not difficult to show by using the method outlined in Appendix E. In this case all matrices involved in the problem are real, and the phase conditions are satisfied trivially.

In order to obtain explicitly the extremal value of the element $|s_{pq}|$ belonging to S^{11}, one may use the following method.

(i) Choose a system of ϵ and compute the eigenvalues of $S^{12}(\epsilon) S^{12t}(\epsilon)$, namely s_j^2. The system (ϵ) is allowed if all s_j^2 are smaller than or equal to one.

(ii) Compute the orthogonal matrix $U^{11}(\epsilon)$ which diagonalizes $S^{12} S^{12t}$.

(iii) Choose ϵ_i $(i = 1, \ldots, m)$ and let the diagonal elements of the diagonal matrix $\tilde{S}^{11}(\epsilon)$ be

$$c_i = \epsilon_i \sqrt{1 - s_i^2}. \tag{VI. 7}$$

(iv) The modulus $a_{pq}(\epsilon)$ of the element s_{pq} of $S^{11}(\epsilon)$ computed from

$$S^{11} = U^{11} \tilde{S}^{11} U^{11t} \tag{VI. 8}$$

is an extremal value.

2. Other solutions when m = 2

When $m = 2$, we have shown in Appendix E that some trivial bounds can be reached which are not of the form of Proposition 6. In these cases the extremal matrices S are indeed complex.

The explicit forms of these bounds are of two types.

$$S^{11} = \begin{pmatrix} (1 - \sum_{i=3}^{n+2} a_{1i}^2)^{1/2} & 0 \\ 0 & (1 - \sum_{i=3}^{n+2} a_{2i}^2)^{1/2} \end{pmatrix} \tag{VI. 9}$$

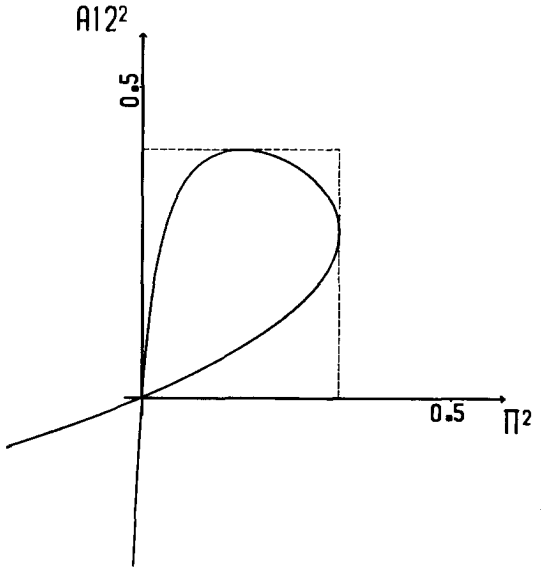


FIG. 6. The curve of a_{12}^2 as a function of π^2 once σ_1 and σ_2 have been fixed. Here $\sigma_1 = 0.6$ and $\sigma_2 = 0.2$. The vertical dotted line has equation $\pi^2 = (1-\sigma_1)(1-\sigma_2)$ and the horizontal one has $a_{12}^2 = \min(1-\sigma_1, 1-\sigma_2)$. Let $\pi^+ = \sum_{i=3}^{n+2} a_{1i} a_{2i}$ and $\pi^- = a_{1j} a_{2j} - \sum_{\substack{i=3 \\ i \neq j}}^{n+2} a_{1i} \cdot a_{2i}$ $(a_{1j} a_{2j} \geq a_{1i} a_{2i}$ for all $i)$. The intersection of the curve with $\pi = \pi^+$ or $\pi = \pi^-$ determines possible upper and lower bounds for a_{12}^2.

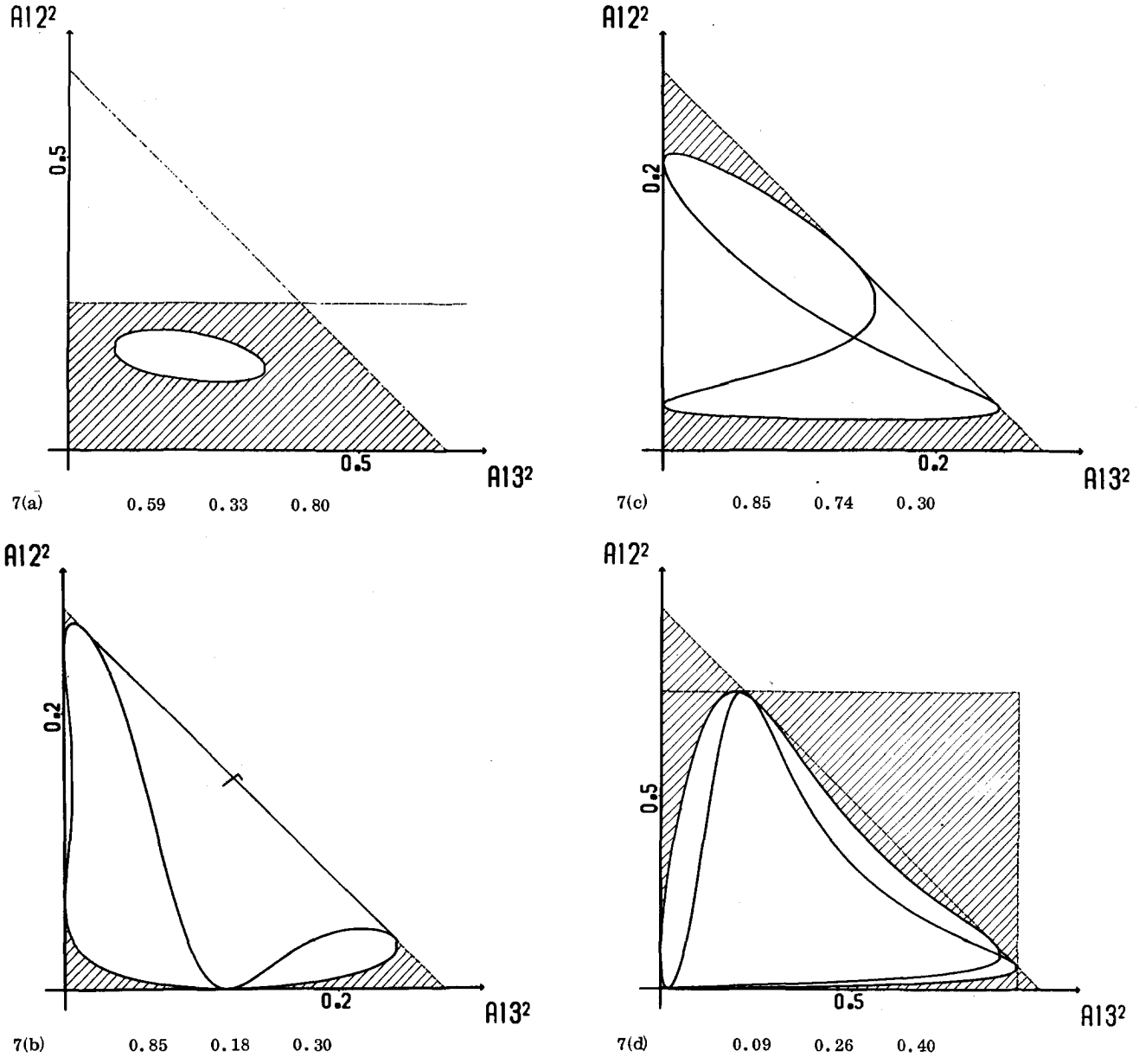


FIG. 7. Example of bounds of a_{12}^2 as a function of a_{13}^2 once a_{14}, a_{23} and a_{24} have been fixed. Remark that the straight lines $a_{12}^2 = 1 - a_{23} - a_{24}^2$ and $a_{12}^2 + a_{13}^2 = 1 - a_{14}^2$ may be parts of the true bound of the domain.

$$S^{11} = \begin{pmatrix} 0 & (1 - \sum_{i=3}^{n+2} a_{1i}^2)^{1/2} \\ (1 - \sum_{i=3}^{n+2} a_{1i}^2)^{1/2} & (\sum_{i=3}^{n+2} (a_{1i}^2 - a_{2i}^2)^{1/2} \end{pmatrix}, \quad \text{(VI. 10)}$$

where we have assumed that $\Sigma_{i=3}^{n+2} (a_{1i}^2 - a_{2i}^2)$ is positive. In both cases for this type of bound to appear there most exist a system of phases for the first two lines of S such that they are unitary orthogonal.

3. Drawing of the results

When $m = 2$, Eq. (VI. 6) may be represented by a surface in a $2n+1$ dimensional space. We here write the equation of the surface explicitly in the variables a_{12} and a_{ij} $(i = 1, 2;\ j = 3, \ldots n+2)$. Define the following combination:

$$\sigma_1 = \sum_{i=3}^{n+2} a_{1i}^2, \quad \text{(VI. 11)}$$

$$\sigma_2 = \sum_{i=3}^{n+2} a_{2i}^2, \quad \text{(VI. 12)}$$

$$\pi = \sum_{i=3}^{n+2} \epsilon_i a_{1i} a_{2i}. \quad \text{(VI. 13)}$$

Equation (VI. 6) implies that the scalar product of the first two lines of S is equal to zero, and reads

$$a_{12}^4((\sigma_1 - \sigma_2)^2 + 4\pi^2) - 2a_{12}^2(2 - \sigma_1 - \sigma_2)\pi^2 + \pi^4 = 0. \quad \text{(VI. 14)}$$

The curve of a_{12}^2 as a function of π^2 is given in Fig. 6 when σ_1 is larger than σ_2.

It may also be worthwhile to have a plot of a_{12}^2 as a function of one of the elements of S^{12} (say a_{13}^2). There is a great number of different configurations and we have chosen to restrict ourselves to four explicit curves [Figs. 7(a)–(d)].

APPENDIX A: GENERAL THEOREMS

We here recall some well-known theorems and give indications for their proofs.

(1) A unitary matrix U can always be written

$$U=\exp(iH), \tag{A1}$$

where H is a Hermitian matrix (Ref. 1, Vol. I, Chap IX, p. 278).

(2) A unitary symmetric matrix U can always be written

$$U=\exp(iR), \tag{A2}$$

where R is real symmetrical (Ref. 1, Vol. II, Chap. XI, p. 4).

(3) A Hermitian matrix H can be diagonalized by a unitary transformation U

$$H=U\tilde{H}U^{+}. \tag{A3}$$

The same is true for a unitary matrix [by (A1)]. $\tilde{H}$ is diagonal real (Ref. 1, Vol. I, Chap. IX p. 274).

(4) A real symmetrical matrix R can be diagonalized by a real orthogonal transformation O

$$R=O\tilde{R}O^{t}. \tag{A4}$$

The same is true for a unitary symmetric matrix [by (A. 2), (Ref. 1, Vol. I, Chap. IX, p. 285].

(5) Let A be an arbitrary $(m\times n)$ matrix. Then the semipositive Hermitian matrices AA^{+} and $A^{+}A$ have the same positive eigenvalues a_i^2 with the same multiplicities. For the eigenvalues zero, the multiplicities differ by $|n-m|$.

(6) There exists a unitary matrix U which diagonalizes AA^{+} and a unitary matrix V which diagonalizes $A^{+}A$, and such that if

$$A=U\tilde{A}V^{+}. \tag{A5}$$

$\tilde{A}$ is in a pseudo-diagonal form which can be chosen real positive. When $n\geq m$, this means that, if $\tilde{A}$ is decomposed in

$$\tilde{A}(m,\ n-m)=(\tilde{A}^{11},\ \tilde{A}^{12}). \tag{A6}$$

$\tilde{A}^{11}$ is a square $m\times m$ diagonal real positive matrix with diagonal elements a_j, and $\tilde{A}^{12}$ is identically zero.

Proof: In the proof of (5) and (6) one may use the lemma:

(7) If $AA^{+}=1_m$ and $A^{+}A=1_n$, where 1_k is the k-dimensional unit matrix, then $m=n$. (This follows trivially from rank considerations.) If $AA^{+}=0$ then $A=0$.

The existence of U's and V's which diagonalize, respectively, AA^{+} in $D^{(1)}$ and $A^{+}A$ in $D^{(2)}$ follows from (3). Using (A. 5) as a definition of $\tilde{A}$, one obtains

$$\tilde{A}\tilde{A}^{+}\tilde{A}=D_1\tilde{A}=\tilde{A}D_2. \tag{A7}$$

This implies that $\tilde{A}$ may be decomposed in blocks $\tilde{A}_{rs}$ corresponding to one given eigenvalue $d_r^{(1)}$ of D_1 and to $d_s^{(2)}$ of D_2, with

$$\tilde{A}_{rs}=0 \text{ if } d_r^{(1)}\neq d_s^{(2)}. \tag{A8}$$

By writing $\tilde{A}\tilde{A}^{+}=D_1$ and $\tilde{A}^{+}\tilde{A}=D_2$; one obtains

$$\begin{aligned} &A_{rr}A_{rr}^{+}=d_r 1_{(1)}, && \text{(1) is the multiplicity of } d_r^{(1)},\\ &&& d_r=d_r^{(1)}=d_r^{(2)}=a_r^2, \\ &A_{rr}^{+}A_{rr}=d_r 1_{(2)}, && \text{(2) is the multiplicity of } d_r^{(2)}. \end{aligned} \tag{A9}$$

The internal consistency of (A8) (A9) and our lemma imply (5). By ordering the eigenvalues d_r of D^1 and D^2 by decreasing order, the A_{rr} blocks appear on the main diagonal. Since the A_{rr} are unitary up to a factor they can be made proportional to the unit matrix by a suitable unitary transformation inside the subspace corresponding to one eigenvalue. A final phase transformation makes $\tilde{A}$ real positive. This completes the proof of (6).

(8) In complete analogy with (5) and (6), an arbitrary $(m\times n)$ real matrix A can be pseudo-diagonalized by real orthogonal transformations O and P

$$A=O\tilde{A}P. \tag{A10}$$

APPENDIX B: PROOF OF (III.9)

In this appendix we show that the general solution of Eqs. (II. 1), (II. 2), (II. 9b), (II. 12), invariant under Eq. (II. 14), (III. 4), is of the form (III. 9).

With the notation (III. 1), the Lagrange equations (II. 12) become

$$\begin{aligned} a-\tilde{S}^{12}\tilde{X}^{12+} &= \tilde{X}^{12}\tilde{S}^{12+},\\ b-\tilde{S}^{11}\tilde{X}^{12*} &= \tilde{X}^{12}\tilde{S}^{22+},\\ c-\ 0\ &= \tilde{X}^{12}\tilde{S}^{23*},\\ d-\tilde{S}^{12t}\tilde{X}^{12*} &= \tilde{X}^{12t}\tilde{S}^{12*},\\ e-\ 0\ &= \tilde{X}^{12t}\tilde{S}^{13*}, \end{aligned} \tag{B1}$$

while the unitarity conditions (II. 1), (II. 2) read

$$\begin{aligned} a-\tilde{S}^{11}\tilde{S}^{11+}+\tilde{S}^{12}\tilde{S}^{12+}+\tilde{S}^{13}\tilde{S}^{13+} &= 1,\\ b-\tilde{S}^{11}\tilde{S}^{12*}+\tilde{S}^{12}\tilde{S}^{22+}+\tilde{S}^{13}\tilde{S}^{23+} &= 0,\\ c-\tilde{S}^{11}\tilde{S}^{13*}+\tilde{S}^{12}\tilde{S}^{23+}+\tilde{S}^{13}\tilde{S}^{33+} &= 0,\\ d-\tilde{S}^{12t}\tilde{S}^{12*}+\tilde{S}^{22}\tilde{S}^{22+}+\tilde{S}^{23}\tilde{S}^{23+} &= 1,\\ e-\tilde{S}^{12t}\tilde{S}^{13*}+\tilde{S}^{22}\tilde{S}^{23*}+\tilde{S}^{23}\tilde{S}^{33+} &= 0,\\ f-\tilde{S}^{13t}\tilde{S}^{13*}+\tilde{S}^{23t}\tilde{S}^{23*}+\tilde{S}^{33}\tilde{S}^{33+} &= 1. \end{aligned} \tag{B2}$$

Since the matrices $S^{12}S^{12+}$, $X^{12}X^{12+}$, and $S^{12}X^{12+}$ are Hermitian and commute, due to Eq. (B1a), it can be shown (cf. Appendix A) that $\tilde{S}^{12}$ and $\tilde{X}^{12}$ can be "pseudo-diagonalized" in the canonical forms.

$$\tilde{S}^{12}(m,n-m)=\left(\begin{array}{ccc:c} s_1 & & 0 & \\ & s_2 & & 0 \\ & \ddots & & \\ 0 & & s_m & \end{array}\right), \tag{B3}$$

$$\tilde{X}^{12}(m,n-m)=\left(\begin{array}{ccc:c} x_1 & & 0 & \\ & x_2 & & 0 \\ & \ddots & & \\ 0 & & x_m & \end{array}\right), \tag{B4}$$

where we have assumed for definiteness that n is larger than or equal to m. Moreover the s_i can be taken to be real nonnegative while the x_i are real.

Since $\tilde{S}^{11}$ has to satisfy the equations

$$\tilde{S}^{11}\tilde{S}^{12*}\tilde{X}^{12t} = -\tilde{S}^{12}\tilde{X}^{12*}\tilde{S}^{11},$$
$$\tilde{S}^{11}\tilde{X}^{12*}\tilde{X}^{12t} = \tilde{X}^{12}\tilde{X}^{12*}\tilde{S}^{11}, \tag{B5}$$

which are consequences of (B2) and (B1), it assumes the form given by

$$\tilde{S}^{11} = \begin{array}{c} \begin{array}{cccccc} \text{given} & & s_i \neq 0 & & s=0 & s \\ +x_j & -x_j & +x_k & -x_k & x=0 & \end{array} \\ \begin{pmatrix} 1 & & & & & \\ & 1^t & & & & \\ & & 2 & 3 & & \\ & & 3^t & 2 & & \\ & & & & & 4 \end{pmatrix} \end{array}, \tag{B6}$$

where the lines and columns of $\tilde{S}^{11}$ have been labelled with the corresponding diagonal values of $\tilde{X}^{12}$ and $\tilde{S}^{12}$ and the shaded blocks or blocks labeled 1, 2, ...) are the only nonzero ones. More precisely $\tilde{s}^{11}_{ij}$ may be different from zero only if either

(1) $s_i = s_j \neq 0$ and $x_i = -x_j \neq 0$ (blocks 1, 1^t),

(2) $s_i = s_j = 0$ and $x_i = \pm x_j \neq 0$ (blocks 2, 3),

(3) s_i and s_j arbitrary and $x_i = x_j = 0$ (block 4).

An analogous reasoning applied to $\tilde{S}^{22}$ leads to

$$\tilde{S}^{22} = \begin{array}{c} \begin{array}{cccccc} \text{given} & & s_i \neq 0 & & s=0 & \text{any } s \\ +x_j & -x_j & +x_k & -x_k & x=0 & n-m \end{array} \\ \begin{pmatrix} 1 & & & & & \\ & 1^t & & & & \\ & & 2 & 3 & & \\ & & 3^t & 2 & & \\ & & & & & 4 \end{pmatrix} \end{array}. \tag{B7}$$

Compared with the $\tilde{S}^{11}$, the structure of $\tilde{S}^{22}$ is quite the same except that the block 4 in $\tilde{S}^{22}$ has been expanded to include the $n-m$ extra lines and columns. Equation (B. 1b) then implies the following relations between $\tilde{S}^{11}$ and $\tilde{S}^{22}$

$$\tilde{S}^{11}\ (\text{block } 1) = -\tilde{S}^{22}\ (\text{block } 1),$$
$$\tilde{S}^{11}\ (\text{block } 2) = +\tilde{S}^{22}\ (\text{block } 2),$$
$$\tilde{S}^{11}\ (\text{block } 3) = -\tilde{S}^{22}\ (\text{block } 3). \tag{B. 8}$$

Equations (B1c) and (B1d) then reduce $\tilde{S}^{13}$ and $\tilde{S}^{23}$ to the form

$$\tilde{S}^{13} = \begin{array}{c} x \neq 0 \\ x = 0 \end{array}\begin{pmatrix} \\ \\ \end{pmatrix}, \qquad \tilde{S}^{23} = \begin{array}{c} x \neq 0 \\ x = 0 \\ n-m \end{array}\begin{pmatrix} \\ \\ \\ \end{pmatrix}. \tag{B9}$$

Unitarity (B2) then implies that the blocks 1 in (B6) and (B7) are square matrices except when $s_i = 1$. When s is different from zero and one, this signifies that the nonzero x's appear in pairs of opposite sign ($x_j = -x_j$).

The fact that there are basically no other conditions on the possible values of x's and s's can be verified readily by checking that the matrix of Table I is unitary and symmetric, and satisfies the Lagrange equations. In this table specific values of $\tilde{S}^{13} = 0$, $\tilde{S}^{23} = 0$, $\tilde{S}^{11}$, $\tilde{S}^{22}$, and $\tilde{S}^{33} = 1$ have been chosen quite arbitrarily. We have the right to do so because we have only to show that there exists a solution for those submatrices which does not give further restrictions on the matrices $\tilde{X}^{12}$ and $\tilde{S}^{12}$.

APPENDIX C: PROOF OF (III.17)

When $\tilde{X}^{12}$ has one non-zero element ($\tilde{X}^{12}_{11} = x$) and U^{11} and U^{22} have respectively $(m-k)$ and $(n-l)$ zeros on their first columns, X^{12} is of the form

TABLE I.

$\tilde{S}$ =

		n_1	n_2	m n_2	n_2'	n_3	n_1	n_2	n_2	n_2'	n_3'	$n-m$	p
	any x						1						
	x			$\sqrt{1-s^2}$				s					
m	$-x$		$\sqrt{1-s^2}$						s				
	$x=0$				$\sqrt{1-s^2}$					s			
	any x					1					0		
		1											
			s						$-\sqrt{1-s^2}$				
n				s				$-\sqrt{1-s^2}$					
					s					$-\sqrt{1-s^2}$			
						0					1		
												1	
p													1

[a] All elements of $\tilde{S}$ are zero except where indicated.

$$X^{12}\begin{pmatrix}k, m-k\\ l, n-l\end{pmatrix}=\begin{pmatrix}{}^{11}X^{12} & 0\\ 0 & 0\end{pmatrix}, \tag{C1}$$

where ${}^{11}X^{12}$ is a $k\times 1$ nonzero real matrix. Analogously S^{12} will be written in rectangular blocks

$$S^{12}\begin{pmatrix}k, m-k\\ l, n-l\end{pmatrix}=\begin{pmatrix}{}^{11}S^{12} & {}^{12}S^{12}\\ {}^{21}S^{12} & {}^{22}S^{12}\end{pmatrix}, \tag{C2}$$

where ${}^{11}S^{12}$ is a real matrix.

Let U^{11} and U^{22} be the unitary transformations

$$U^{11}\begin{pmatrix}k, m-k\\ 1, m-1\end{pmatrix}=\begin{pmatrix}{}^{11}U^{11} & {}^{12}U^{11}\\ 0 & {}^{22}U^{11}\end{pmatrix} \tag{C3}$$

and

$$U^{22}\begin{pmatrix}l, n-l\\ 1, n-1\end{pmatrix}=\begin{pmatrix}{}^{11}U^{22} & {}^{12}U^{22}\\ 0 & {}^{22}U^{22}\end{pmatrix} \tag{C4}$$

which pseudo-diagonalize S^{12} in $\tilde{S}^{12}$

$$S^{12}=U^{11}\tilde{S}^{12}U^{22t}. \tag{C5}$$

The matrix S^{12} has $s_1=1$. Also remark that ${}^{11}U^{11}$ and ${}^{11}U^{22}$ are, respectively, $k\times 1$ and $l\times 1$ matrices.

We now show that ${}^{11}S^{12}\,{}^{11}S^{12t}={}^{11}S^{12}\,{}^{11}S^{12+}$ has ${}^{11}U^{11}$ as an eigenvector with eigenvalue 1. Indeed

$${}^{11}S^{12}=({}^{11}U^{11}, {}^{12}U^{11})\tilde{S}^{12}\begin{pmatrix}{}^{11}U^{22t}\\ {}^{12}U^{22t}\end{pmatrix}, \tag{C6}$$

and

$$\begin{aligned}{}^{11}S^{12}\,{}^{11}S^{12+}\,{}^{11}U^{11}&=({}^{11}U^{11}, {}^{12}U^{11})\tilde{S}^{12}\begin{pmatrix}1_1 & 0\\ 0 & {}^{12}U^{22t}\,{}^{12}U^{22*}\end{pmatrix}\\ &\quad\times\tilde{S}^{12}\begin{pmatrix}1\\ 0\end{pmatrix},\\ &=({}^{11}U^{11}, {}^{12}U^{11})\begin{pmatrix}1\\ 0\end{pmatrix},\\ &={}^{11}U^{11}.\end{aligned} \tag{C7}$$

Let $\tilde{\tilde{U}}^{11}$ and $\tilde{\tilde{U}}^{22}$ be the $k\times k$ and $l\times l$ orthogonal matrices which pseudo-diagonalize ${}^{11}S^{12}$. Then

$$\begin{aligned}S^{12}&=\begin{pmatrix}\tilde{\tilde{U}}^{11} & 0\\ 0 & 1_{m-k}\end{pmatrix}\begin{pmatrix}1_1 & 0 & 0\\ 0 & {}^{22}\tilde{\tilde{S}}^{12} & {}^{23}\tilde{\tilde{S}}^{12}\\ 0 & {}^{32}\tilde{\tilde{S}}^{12} & {}^{33}\tilde{\tilde{S}}^{12}\end{pmatrix}\begin{pmatrix}\tilde{\tilde{U}}^{22t} & 0\\ 0 & 1_{n-l}\end{pmatrix}\\ &=\begin{pmatrix}\tilde{\tilde{U}}^{11} & 0\\ 0 & 1_{m-k}\end{pmatrix}\tilde{\tilde{S}}^{12}\begin{pmatrix}\tilde{\tilde{U}}^{22t} & 0\\ 0 & 1_{n-l}\end{pmatrix}.\end{aligned} \tag{C8}$$

Let $\tilde{U}^{11}$ and $\tilde{U}^{22}$ be the $(m-1)\times(m-1)$ and $(n-1)\times(n-1)$ matrices which pseudo-diagonalize the submatrix

$$\begin{pmatrix}k-1, m-k\\ l-1, n-l\end{pmatrix}=\begin{pmatrix}{}^{22}\tilde{\tilde{S}}^{12} & {}^{23}\tilde{\tilde{S}}^{12}\\ {}^{32}\tilde{\tilde{S}} & {}^{33}\tilde{\tilde{S}}^{12}\end{pmatrix} \tag{C9}$$

of $\tilde{\tilde{S}}^{12}$. Thus

$$\tilde{\tilde{S}}^{12}=\begin{pmatrix}1_1 & 0\\ 0 & \tilde{U}^{11}\end{pmatrix}\quad\tilde{S}^{12}\begin{pmatrix}1_1 & 0\\ 0 & \tilde{U}^{22t}\end{pmatrix}. \tag{C10}$$

Combining (C. 10) with (C. 8), one sees that

$$U^{11}=\begin{pmatrix}\tilde{\tilde{U}}^{11} & 0\\ 0 & 1_{m-k}\end{pmatrix}\begin{pmatrix}1_1 & 0\\ 0 & \tilde{U}^{11}\end{pmatrix}, \tag{C11}$$

$$U^{22}=\begin{pmatrix}\tilde{\tilde{U}}^{22} & 0\\ 0 & 1_{n-l}\end{pmatrix}\begin{pmatrix}1_1 & 0\\ 0 & \tilde{U}^{22}\end{pmatrix}, \tag{C12}$$

which are of the desired form (C3), (C4). It is again clear that the first column of $\tilde{\tilde{U}}^{11}$ (resp. $\tilde{\tilde{U}}^{22}$) is precisely ${}^{11}U^{11}$ (resp. ${}^{11}U^{22}$).

The crucial property of this solution can be seen in Eq. (C8). When ${}^{11}S^{12}$ is diagonalized, the number ${}^{11}\tilde{\tilde{S}}^{12}$ is one. This implies immediately that ${}^{13}\tilde{\tilde{S}}^{12}$ and ${}^{31}\tilde{\tilde{S}}^{12}$ have to vanish because of unitarity. This last property is equivalent to

$$\begin{aligned}{}^{12}S^{12t}\,{}^{11}U^{11}&=0,\\ {}^{21}S^{12}\;{}^{11}U^{22}&=0,\end{aligned} \tag{C13}$$

since $(\tilde{\tilde{U}}^{11})^{-1}=\tilde{\tilde{U}}^{11t}$.

For this type of solution to exist, it is necessary, once a set of ϵ_{ij}'s $(=\pm 1)$ corresponding to the real part of ${}^{11}S^{12}$ has been chosen, and once one eigenvalue of ${}^{11}S^{12}\,{}^{11}S^{12t}$ has been set to one, to see whether there exists a set of phases for ${}^{12}S^{12}$, ${}^{21}S^{12}$, and ${}^{22}S^{12}$ such that (C13) is satisfied and such that the resulting S^{12} has all its corresponding s_j^2 between zero and one.

It should be remarked that for condition (C. 13) to hold, a sum of k or l complex numbers whose moduli R_i are known has to vanish. This is in general possible only when k or l is larger than or equal to three. The condition reads for all i

$$R_i\leq\sum_{j\neq i}R_j. \tag{C14}$$

It is not difficult to see that this type of solutions forms $(nm-1)$ dimensional pieces of boundary β by taking all moduli very small except along the pseudo-diagonal of S^{12}.

APPENDIX D: PROOF OF $\det\alpha_1\,(|b|^2-ac)=\det(1-BB^+)$

Let

$$A=({}^{11}S,\ {}^{12}S), \tag{D1}$$

$$B^t=({}^{11}S^t, {}^{21}S^t), \tag{D2}$$

$$\alpha_1=1-AA^+, \tag{D3}$$

$$\beta=1-{}^{11}S\,{}^{11}S^+, \tag{D4}$$

where ${}^{12}S$ and ${}^{21}S^t$ are one-column matrices, and α_1 and β are invertible. Let also

$$a=1+{}^{12}S^+\alpha_1^{-1}\,{}^{12}S, \tag{D5}$$

$$b={}^{12}S^+\alpha_1^{-1}\,{}^{11}S\,{}^{21}S^+, \tag{D6}$$

$$c=-1+{}^{21}S\,{}^{21}S^+ +{}^{21}S\,{}^{11}S^+\alpha_1^{-1}\,{}^{11}S\,{}^{21}S^+. \tag{D7}$$

We want to prove that

$$(|b|^2-ac)\det\alpha_1=\det(1-BB^+). \tag{D8}$$

For this we will use the two tricks

$$\begin{pmatrix} M_1 & M_2 \\ M_3 & M_4 \end{pmatrix}\begin{pmatrix} 1 & -M_1^{-1}M^2 \\ 0 & 1 \end{pmatrix} = \begin{pmatrix} M_1 & 0 \\ M_3 & M_4 - M_3 M_1^{-1} M_2 \end{pmatrix}, \tag{D9}$$

$$M_2^{-1} - M_1^{-1} = M_2^{-1}(M_1 - M_2)M_1^{-1}. \tag{D10}$$

First it easy to prove that

$$\det\beta = a \det\alpha_1 \tag{D11}$$

by taking the determinant of the equation

$$\begin{pmatrix} 1 & {}^{12}S \\ 0 & 1 \end{pmatrix}\begin{pmatrix} \beta & 0 \\ -{}^{12}S^{+} & 1 \end{pmatrix} = \begin{pmatrix} \alpha_1 & {}^{12}S \\ -{}^{12}S^{+} & 1 \end{pmatrix} \tag{D12}$$

$$= \begin{pmatrix} \alpha_1 & 0 \\ -{}^{12}S^{+} & 1 + {}^{12}S^{+}\alpha_1^{-1}\,{}^{12}S \end{pmatrix}\begin{pmatrix} 1 & \alpha_1^{-1}\,{}^{12}S \\ 0 & 1 \end{pmatrix}.$$

Using (D9) again for $M = 1 - BB^{+}$, taking its determinant

$$\det(1 - BB^{+}) = d \det\beta, \tag{D13}$$

$$d = 1 - {}^{21}S\,{}^{21}S^{+} - {}^{21}S\,{}^{11}S^{+}\beta^{-1}\,{}^{11}S\,{}^{21}S^{+} \tag{D14}$$

and replacing $\det\beta$ by (D11), it remains to prove that

$$|b|^2 - ac = ad, \tag{D15}$$

which follows simply by repeated application of (D10).

APPENDIX E

In this appendix, we present with some details the calculations leading to the bounds of s_{12} when $m = 2$ and $p = 0$, namely when

$$S(2,n) = \begin{pmatrix} S^{11} & S^{12} \\ S^{12t} & S^{22} \end{pmatrix}. \tag{E1}$$

The moduli of S^{12} are supposed to be given. Then

$$X(2,n) = \begin{pmatrix} X^{11} & X^{12} \\ X^{12t} & 0 \end{pmatrix} \tag{E2}$$

and

$$X^{11} = \begin{pmatrix} 0 & x_{12} \\ x_{12} & 0 \end{pmatrix}. \tag{E3}$$

1. Solution of the first-class condition

Following (II. 14), let us perform a transformation with the particular matrix

$$U(2,n) = \begin{pmatrix} U^{11} & 0 \\ 0 & U^{22} \end{pmatrix}. \tag{E4}$$

This transformation leaves the submatrices (E1), (E2) globally invariant, but does not respect the structure of X^{11} [(E3)]. Restoring this structure will be a supplementary second-class condition.

As is shown in Appendix A, $\tilde{S}^{12}$ can be pseudo-diagonalized

$$\tilde{S}^{12}(2, n-2) = ({}^{11}\tilde{S}^{12},\ {}^{12}\tilde{S}^{12}), \tag{E5}$$

with

$${}^{11}\tilde{S}^{12} = \begin{pmatrix} s_1 & 0 \\ 0 & s_2 \end{pmatrix}, \quad 0 \leq s_1 \leq 1, \quad 0 \leq s_2 \leq 1 \tag{E6}$$

$${}^{11}\tilde{S}^{12} = 0. \tag{E7}$$

We shall assume here, without loss of generality, that

$$1 \neq \det {}^{11}\tilde{S}^{12} \neq 0. \tag{E8}$$

At the same time, one may choose ϵ_i and

$$\tilde{S}^{11} = \begin{pmatrix} c_1 & 0 \\ 0 & c_2 \end{pmatrix}, \tag{E9a}$$

$$c_i = \epsilon_i(1 - s_i^2)^{1/2}, \tag{E9b}$$

$$\tilde{S}^{22} = \begin{pmatrix} {}^{11}\tilde{S}^{22} & {}^{12}\tilde{S}^{22} \\ {}^{12}\tilde{S}^{22t} & {}^{22}\tilde{S}^{22} \end{pmatrix}, \tag{E10a}$$

$${}^{11}\tilde{S}^{22} = -\tilde{S}^{11}, \tag{E10b}$$

$${}^{12}\tilde{S}^{22} = 0, \tag{E10c}$$

$${}^{22}\tilde{S}^{22} = 1_{m-2}. \tag{E10d}$$

The Lagrange equations

$$\tilde{S}^{+}\tilde{X} = \tilde{X}^{+}\tilde{S} \tag{E11}$$

then imply

$$\begin{aligned}
&(11)\ \tilde{S}^{11+}\tilde{X}^{11} + {}^{12}\tilde{S}^{12*}\tilde{X}^{12t} = \tilde{X}^{11+}\tilde{S}^{11} + {}^{11}\tilde{X}^{12*}\,{}^{11}\tilde{S}^{12t},\\
&(12)\ \tilde{S}^{11+}\,{}^{11}\tilde{X}^{12} = \tilde{X}^{11+}\,{}^{11}\tilde{S}^{12} + {}^{11}\tilde{X}^{12*}\,{}^{11}\tilde{S}^{22},\\
&(13)\ \tilde{S}^{11+}\,{}^{12}\tilde{X}^{12} = {}^{12}\tilde{X}^{12*}\,{}^{22}\tilde{S}^{22},\\
&(22)\ {}^{11}\tilde{S}^{12+}\,{}^{11}\tilde{X}^{12} = {}^{11}\tilde{X}^{12+}\,{}^{11}S^{12},\\
&(23)\ {}^{11}\tilde{S}^{12+}\,{}^{12}\tilde{X}^{12} = 0,
\end{aligned} \tag{E12}$$

where for X^{12} a decomposition $(2, m-2)$ has also been made. Equation (E12 (23)) together with (E6), (E8) imply

$${}^{12}\tilde{X}^{12} = 0 \tag{E13}$$

while (E12(22)) suggests the convenient parametrization

$${}^{11}\tilde{X}^{12} = Z\,{}^{11}\tilde{S}^{12}, \tag{E14}$$

$$Z = Z^{+}. \tag{E15}$$

Finally (E12(12)) leads, for $\tilde{X}^{11}$, to the value

$$\tilde{X}^{11} = Z\tilde{S}^{11} + \tilde{S}^{11}Z^{t}, \tag{E16}$$

since (E12(11)) and (E12(13)) are then identities.

2. Solutions of the second-class conditions

There are four types of second-class conditions:

(a) X^{11} must be of the form (E3),

(b) the phase of x_{12} must equal the phase of s_{12},

(c) the phase of the corresponding elements of S^{12} and X^{12} are equal,

(d) the moduli of the elements of S^{12} are fixed.

Let U^{11} be an arbitrary (2×2) unitary unimodular matrix

$$U^{11}=\begin{pmatrix} u & v \\ -v^* & u^* \end{pmatrix}, \quad |u|^2+|v|^2=1; \tag{E17}$$

the expression of S^{11} is

$$S^{11}=U^{11}\tilde{S}^{11}U^{11\dagger} =\begin{pmatrix} u^2c_1+v^2c_2 & -uv^*c_1+u^*vc_2 \\ -uv^*c_1+u^*vc_2 & v^{*2}c_1+u^{*2}c_2 \end{pmatrix} \tag{E18}$$

so that

$$x_{12}=-uv^*c_1+u^*vc_2. \tag{E19}$$

In turn

$$\tilde{X}^{11}=U^{11*}X^{11}U^{11*} =x_{12}\begin{pmatrix} -2u^*v & |u|^2-|v|^2 \\ |u|^2-|v|^2 & 2uv^* \end{pmatrix}. \tag{E20}$$

But, following (E16), the diagonal elements of X^{11} should be real; thus

$$u^*v=\epsilon' uv^* \tag{E21}$$

is purely real ($\epsilon'=+1$) or purely imaginary ($\epsilon'=-1$), and Z becomes ($s_1\neq s_2$)

$$Z=x_{12}\begin{pmatrix} -u^*v/c_1 & (|u|^2-|v|^2)/(c_2+\epsilon' c_1) \\ (|u|^2-|v|^2)/(c_1+\epsilon' c_2) & +uv^*/c_2 \end{pmatrix} \tag{E22}$$

The value of ${}^{11}\tilde{X}^{12}$ is then obtained from (E. 14). By modifying suitably the ϵ_i of (E. 9b) one may always choose

$$\epsilon'=+1 \tag{E23}$$

as it is not difficult to verify. The transformation U^{11} is then equivalent [by (II. 15)] to a purely real (thus orthogonal) one.

Accordingly there are three distinct possibilities:

(a) U^{22} is real. All phase conditions are trivially satisfied. The moduli of the elements of S^{12} can be given their pre-assigned value by using the freedom on U^{11}, s_i, and U^{22} as shown in Appendix A. At this point, it is clear that this type of solution exists for arbitrary values of m, since when all matrices are real the phase condition is trivially satisfied. One method to obtain the explicit value of that bound is explained in the main text.

(b) U^{22} is complex. As can be seen directly from

$$S^{12}=U^{11}\tilde{S}^{12}U^{22t}, \tag{E24}$$

$$X^{12}=U^{11}\tilde{X}^{12}U^{22t}. \tag{E25}$$

The condition for S^{12} and X^{12} to have the same phase is that each line of $U^{11}\tilde{S}^{12}$ is proportional to the corresponding line of $U^{11}\tilde{X}^{12}$. This leads to the relation

$$(u^2c_1+v^2c_2)(v^2c_1+u^2c_2)=0. \tag{E26}$$

Let us analyze one of these two analogous case, say

$$u^2c_1+v^2c_2=0 \tag{E27}$$

which implies in particular that $\epsilon_1\epsilon_2=-1$. Then, as is seen from (E18), s_{11} is equal to zero while X^{12} has its second line identical to zero. The maximal value of a_{12} is then simply

$$a_{12}=(1-\sum_{i=3}^{n+2}a_{1i}^2)^{1/2} \tag{E28}$$

and does not depend at all on the second line. The element (22) of S^{11} then becomes

$$a_{22}=(\sum_{i=3}^{n+2}a_{1i}^2-\sum_{i=3}^{n+2}a_{2i}^2)^{1/2}. \tag{E29}$$

This result is analogous to the amputated case of Proposition 5. For this maximum to be realized, it is evidently necessary that

$$\sum_{i=3}^{n+2}a_{1i}^2\geq\sum_{i=3}^{n+2}a_{2i}^2. \tag{E30}$$

[Otherwise the other solution of (E. 26) has to be considered.] Moreover, there should exist a system of phases for S^{11} and S^{12} such that the two lines of the $2\times(2+n)$ matrix

$$(S^{11},S^{12}) \tag{E31}$$

are unitary orthogonal. This means that a sum of $(n+1)$ complex numbers of known moduli $R_i=a_{1i}a_{2i}$ has to vanish. The condition is equivalent to (C14) and reads

$$R_i\leq\sum_{j\neq i}R_j. \tag{E32}$$

(c) $x_{12}=0$. This is the trivial case which corresponds to X identically zero, and consequently

$$s_{pq}=0 \tag{E33}$$

which is possible if, by a suitable choice of phases φ_{ij},

$$s_{ij}=\exp(i\varphi_{ij})a_{ij} \quad (i=1,2.\ j=3,\ldots,n+2), \tag{E34}$$

the two lines of S^{12} can be made unitary orthogonal. This leads again to conditions of the form (E32).

*Present address: University of Mons (Belgium).
†Laboratoire associé au Centre National de la Recherche Scientifique. Postal address: Laboratoire de Physique Théorique et Hautes Energies, Bâtiment 211, Université Paris Sud, 91-Orsay, France.

[1] F. R. Gantmacher, *The Theory of Matrices* (Chelsea, New York, 1959).

On the spontaneous breakdown of compound symmetries

G. Cicogna

Istituto di Fisica dell'Universita di Pisa and Dipartimento di Matematiche Applicate dell'Universita, Pisa, Italy

P. de Mottoni

Istituto per le Applicazioni del Calcolo del CNR, Roma, Italy
(Received 1 March 1973)

We consider the spontaneous breakdown of a symmetry group $\mathcal{G}$ which is the direct product of two groups: $\mathcal{G} = \mathcal{A} \times \mathcal{B}$. We study the conditions under which the breakdown of $\mathcal{G}$ entails that of both $\mathcal{A}$ and $\mathcal{B}$. Our results are corroborated by an explicit example, where $\mathcal{G}$ is $SU_2 \times L$, L being the Lorentz group, and such that any spontaneous breakdown of $\mathcal{G}$ entails that of both SU_2 and L.

I. INTRODUCTION

Spontaneously broken symmetries can be dealt with in different ways,[1-6] according to the general theory one is working in. Different approaches provide in general different characterizations for the breaking to occur, yet the group theoretical properties of the breakdown, as algebraic and therefore more abstract attributes are involved, should be independent of the approach, that is intrinsic to the system under consideration.[7]

In the present paper we shall consider a problem of algebraic type, namely concerning the breaking of a symmetry group $\mathcal{G}$ which is the direct product of two Lie groups, $\mathcal{A}$ and $\mathcal{B}$. We shall investigate this problem in the Wightman formulation, and, under some conditions on the representations of $\mathcal{A}$ and $\mathcal{B}$ involved, we shall show that the breaking of $\mathcal{G}$ entails that of both $\mathcal{A}$ and $\mathcal{B}$. Put in a more precise fashion, we shall show that if the physical quantities transform according to an irreducible nontrivial representation of $\mathcal{B}$, then, if $\mathcal{A}$ breaks down, $\mathcal{B}$ does the same.

Therefore, if $\mathcal{A}$ is broken, in order to avoid the breakdown of $\mathcal{B}$, the system should transform according to a representation of $\mathcal{B}$ containing the one-dimensional representation.

This situation will be illustrated by an example, where the symmetry group $\mathcal{G}$ is the direct product of the internal symmetry group SU_2 and of the Lorentz group, the physical quantities considered being self-interacting vector mesons. It will be shown that any spontaneous breakdown of $\mathcal{G}$ forces both the internal symmetry and the relativistic one to break down, and that this agrees with the general results mentioned above: in fact, in our example, both SU_2 and the Lorentz group act via irreducible nontrivial representations.

The following sections are organized as follows: in Sec. II we shall derive some general results within the Wightman formalism; in Sec. III we shall work out our example, which is described in terms of a Lagrangian; Sec IV will be devoted to some comments, regarding particularly the relationship between our results and other works on the breakdown of compound symmetries.

II. GENERAL RESULTS

We want to give some general results concerning the occurrence of the spontaneous breakdown of a symmetry, the latter being described by the direct product of two groups.

These results are quite general and can be conveniently stated in the Wightman formalism.

Let ϕ be a field (or a Wick polynomial in several fields) acting in the Hilbert space $\mathcal{H}$. Suppose ϕ to transform under the group $\mathcal{G} = \mathcal{A} \times \mathcal{B}$ ($\mathcal{A}$ and $\mathcal{B}$ being two groups) as follows:

$$\begin{aligned} &\phi^{a,b}(x) \to (\tau(A,B)\phi)^{a,b}(x), \\ &(\tau(A,B)\phi)^{a,b}(x) = D_{\mathcal{A}}^{a,a'}(A) D_{\mathcal{B}}^{b,b'}(B) \phi^{a',b'}(x), \\ &\forall A \in \mathcal{A},\ B \in \mathcal{B}, \end{aligned} \tag{1}$$

where $\mathcal{A} \ni A \to D_{\mathcal{A}}(A)$, $\mathcal{B} \ni B \to D_{\mathcal{B}}(B)$ are finite-dimensional representations of $\mathcal{A}$, $\mathcal{B}$, respectively.

We say that $\mathcal{A}$ is an exact symmetry if there exists a representation $A \to U_A$ of $\mathcal{A}$ into the unitary operators on $\mathcal{H}$ such that U_A leaves the vacuum invariant, for any A, and

$$(\tau(A, I_{\mathcal{B}})\phi)^{a,b}(x) = U_A \phi^{a,b}(x) U_A^{-1}, \forall A \in \mathcal{A}, \tag{2}$$

$I_{\mathcal{B}}$ being the identity operator of $\mathcal{B}$.
A quite similar definition holds of course for $\mathcal{B}$. Thus, if $\mathcal{A}$ is an exact symmetry, any Wightman function has the following property (covariance):

$$\begin{aligned} &(\Omega, \phi^{a_1 b_1}(x_1) \cdots \phi^{a_n b_n}(x_n)\Omega) \\ &= D_{\mathcal{A}}^{a_1 a'_1}(A) \otimes \cdots \otimes D_{\mathcal{A}}^{a_n\ a'_n}(A)(\Omega, \phi^{a'_1 b_1}(x_1) \cdots \phi^{a'_n b_n}(x_n)\Omega); \end{aligned} \tag{3}$$

in particular, for $n = 1$

$$(\Omega, \phi^{a,b}(x)\Omega) = D_{\mathcal{A}}^{a,a'}(A)(\Omega, \phi^{a',b}(x)\Omega). \tag{4}$$

But if we assume that $D_{\mathcal{A}}$ does not contain the identity representation, the latter equality entails

$$(\Omega, \phi^{a,b}(x)\Omega) = 0 \quad \forall a, \forall b$$

On the other hand, if $\mathcal{A}$ is not exact, we have a broken symmetry. This means that at least one Wightman function does not enjoy the covariance property (3). If we assume—as usual—that a symmetry proves broken through the noncovariance of the one-point Wightman function, and furtherly that $D_{\mathcal{A}}$ does not contain the identity representation, then the breakdown is characterized by the following statement:

(*S*) If $\mathcal{A}$ is broken, then there is at least one index $\tilde{a}$ such that

$$(\Omega, \phi^{\tilde{a},b}(x)\Omega) \neq 0, \ \forall b.$$

After this preparation, it is easy to prove the following result:

Theorem 1: Assume that $D_{\mathcal{A}}$ and $D_{\mathcal{B}}$ do not contain the identity representation; then if $\mathcal{A}$ is broken, $\mathcal{B}$ is broken as well.

Proof: If $\mathcal{A}$ is broken, by (S), $\exists\, \tilde{a}$ such that

$$(\Omega, \phi^{\tilde{a},b}(x)\Omega) \neq 0 \quad \forall b$$

but if $\mathcal{B}$ were an exact symmetry, the latter quantity $(\Omega, \phi^{a,b}(x)\Omega)$ would vanish for any a, b. Thus $\mathcal{B}$ must be broken as well.

A similar situation prevails if $\mathcal{A}$ proves broken through the noncovariance of another. Wightman function say, the nth. First of all, if $\mathcal{A}$ is exact, and we assume that $\overset{n}{\otimes} D_{\mathcal{A}}$ does not contain the identity representation, we have

$$(\Omega, \phi^{a_1 b_1}(x_1) \cdots \phi^{a_n b_n}(x_n)\Omega) = 0 \quad \forall a_1, \ldots, a_n;\ b_1, \ldots, b_n.$$

Thus if $\mathcal{A}$ is broken, there exists a n-tuple $\tilde{a}_1, \ldots, \tilde{a}_n$ such that

$$(\Omega, \phi^{\tilde{a}_1 b_1}(x_1) \cdots \phi^{\tilde{a}_n b_n}(x_n)\Omega) \neq 0 \quad \forall b_1, \ldots, b_n.$$

Within this framework, the following theorem holds true:

Theorem $1_{(n)}$: Assume that $\overset{n}{\otimes} D_{\mathcal{A}}$, $\overset{n}{\otimes} D_{\mathcal{B}}$ do not contain the identity representation. Then if $\mathcal{A}$ does not leave the n-point function invariant, $\mathcal{B}$ is broken as well.

The proof mimics that of the preceding Theorem 1.

In the following, we shall speak of a broken symmetry only in the sense of statement (S). It can be noted that this is the usual assumption in several field-theoretical models (σ models, tadpole techniques, variational approaches, etc.). Even with this assumption there are some relevant generalizations of Theorem 1, namely:

Theorem 2: Let $\mathcal{G}$ be any Lie group and $D_{\mathcal{G}}$ any representation of it acting on the fields ϕ. In order that a subgroup $\mathcal{G}' \subset \mathcal{G}$ is conserved (i.e., is contained in the "exact" residual symmetry), a necessary condition is that $D_{\mathcal{G}}$ contains, when reduced under $\mathcal{G}'$, the identity representation $I_{\mathcal{G}'}$ of $\mathcal{G}'$.

The proof of this statement is similar to that of Theorem 1.

Another generalization of Theorem 1 deals with groups of transformations acting not only on the indices a, b, of $\phi^{a,b}(x)$ (corresponding to the internal degrees of freedom), but on the coordinates x as well. Let us for instance suppose that $\mathcal{B}$ is the homogeneous Lorentz group. Then $(\tau(A,B)\phi)^{a,b}(x)$ reads:

$$(\tau(A,B)\phi)^{a,b}(x) = D_{\mathcal{A}}^{a,a'}(A)\, D_{\mathcal{B}}^{b,b'}(B)\, \phi^{a',b'}(\Lambda_B x)$$

and, if $\mathcal{B}$ is an exact symmetry, we have in particular:

$$(\Omega, \phi^{a,b}(x)\Omega) = D_{\mathcal{B}}{}^{bb'}(B)\, (\Omega, \phi^{a,b'}(\Lambda_B x)\Omega).$$

But if we assume translation invariance of the theory, $(\Omega, \phi(y)\Omega)$ does not depend on the argument of ϕ, so the last equality reads

$$(\Omega, \phi^{a,b}\Omega) = D_{\mathcal{B}}{}^{b,b'}(B)(\Omega, \phi^{a,b'}\Omega)$$

and, if, as above, $D_{\mathcal{B}}$ does not contain the identity representation,

$$(\Omega, \phi^{a,b}\Omega) = 0 \quad \forall a, b.$$

Thus:

Theorem 3: Theorem 1 still holds when $\mathcal{B}$ is the Lorentz group.[8]

The relevance of this theorem is evident in the elementary particle physics, when describing for instance an isospin multiplet of vector particles (e.g., the ρ mesons), that is, particles transforming in a definite way both under an internal symmetry group $\mathcal{A}$, and under the Lorentz group. Then, under the hypothesis of theorem 1, if $\mathcal{A}$ is broken, the Lorentz group is broken as well.

Our basic results (Theorems 1 and 3) might appear surprising at first sight, since one knows that for nonspontaneous symmetry breaking the internal symmetry is broken but Lorentz invariance is not; to clarify this point, we want to emphasize that what is assumed here is that the direct-product structure of the two groups is preserved. In other words, we have the situation:

	Internal Symmetry	Lorentz-Invariance	Direct-Product Structure
Nonspontaneous breaking:	Broken	Preserved	Broken
Spontaneous breaking:	Broken	Broken	Preserved.

III. A CONCRETE MODEL

In the present section we shall consider a concrete model which behaves according to Theorem 3: more specifically, the breakdown of the Lorentz invariance will arise in any spontaneous breakdown of the whole symmetry group. This will be shown, for simplicity sake, in the framework of the "semiclassical" Lagrangian formalism. It is known, actually, that the equations obtained by means of this formalism, if suitably interpreted, remain essentially valid in the quantum-field-theoretical approach.[6,9] To our purpose, therefore, we may use equally well the Lagrangian formalism, which provides in fact the most simple and convenient scheme for the analysis of the abstract situation, because it retains all the group-theoretical structure underlying the problem. More specifically, we want to obtain:

(i) The directions in the representation space along which the symmetry may break down, together with the residual symmetry group.[5-7,10] These directions correspond to the stationary points of any quantity $Q = Q(\Phi)$ which is left invariant by the symmetry group (the static part $\mathcal{L}_{st}$ of the Lagrangian in our case), namely the points $\bar{\Phi}$ where

$$\frac{\delta \mathcal{L}_{st}}{\delta \Phi} = 0.$$

These directions are only determined by the group-theoretical structure of the representation space, and in particular are independent of the very source of the

breaking (e.g., tadpole terms, bootstraps, external driving forces.[11,12])

(ii) The "mass spectrum" of the system under consideration. In the Lagrangian formalism, after the introduction of the "shifted" fields

$$\varphi = \Phi - \tilde{\Phi}$$

with vanishing vacuum expectation values (as in the σ model or in the tadpole formalism[13,14]), the mass matrix is given by the second-order coefficient in the expansion of $\mathcal{L}$ in terms of the "true" fields φ

$$\left(\frac{\delta^2 \mathcal{L}}{\delta\varphi_i \delta\varphi_j}\right)_{\tilde{\Phi}}$$

around the stationary point $\tilde{\Phi}$. Of course, this definition is meaningful only for c-number valued Φ's; it is known, however, that the latter quantity is the analog of the quantum-field-theoretical mass tensor, which is given in fact by the second (functional) derivative of the Legendre transform of the functional generating the T-ordered connected Green functions.[9,6]

After these preliminaries, let us define the model, which describes a system of selfinteracting vector mesons.[15]

The symmetry group is $\mathcal{G} = L \times SU(2)$, L being the Lorentz group and $SU(2)$ the isospin group. We denote the fields by $V^{l\lambda}(\lambda = 0,1,2,3;\ l = 1,2,3)$, with the 'metric' g:

$$g^{\mu\gamma} = \begin{pmatrix} 1 & & & \\ & -1 & & \\ & & 0 & \\ & & & -1 \end{pmatrix}.$$

We define the Lagrangian

$$\mathcal{L} = -\tfrac{1}{4} F^l_{\lambda\mu} F^{l\lambda\mu} + \tfrac{1}{2} m^2 V^l_\lambda V^{l\lambda} + g\epsilon^{ijl}(\partial^\mu V^{i\lambda}) V^j_\lambda V^l_\mu + \tfrac{1}{4} g^2 (V^l_\lambda V^l_\mu V^{j\lambda} V^{j\mu} - V^i_\lambda V^{i\lambda} V^j_\mu V^{j\mu}), \tag{5}$$

where

$$F^l_{\lambda\mu} = \partial_\lambda V^l_\mu - \partial_\mu V^l_\lambda;\ \partial_\lambda = \frac{\partial}{\partial x^\lambda};\ m, g \in \mathbf{R}.$$

The spontaneously broken solutions are the nonzero solutions of

$$\frac{\delta \mathcal{L}_{st}}{\delta V^l_\lambda} = 0, \tag{6}$$

where $\mathcal{L}_{st}$ denotes the part of $\mathcal{L}$ which does not contain any derivative of the V's. After some manipulations, (see the Appendix), one sees that the solutions are two:

(A) $V^{l0} = 0, \quad l = 1, 2, 3.$

$V^{li} = \delta^{li}\mu(g\sqrt{2})^{-1}, \quad l, i = 1, 2, 3$

where $\mu^2 = -m^2 > 0,$

(B) $V^{l0} = V^{l3} = 0, \quad l = 1, 2, 3$

$V^{li} = \delta^{li}\mu g^{-1}, \quad l, i = 1, 2,$

with $\mu^2 = -m^2 > 0.$

These solutions exhaust all the solutions of (6) in the sense that all the solutions (apart from the trivial one $V^{l\lambda} = 0$) can be reduced to either of them by means of a transformation of the symmetry group $\mathcal{G}$, and therefore fall into two inequivalent orbits of equivalent (i.e., $\mathcal{G}$-conjugate) points.

Note that in order to get real symmetry breaking solutions and at the same time a Hermitian Lagrangian we are forced to take $m^2 < 0$; hence our Lagrangian (5) cannot be viewed as the sum of a free and an interacting term.[16]

In the following, we shall consider the residual invariance and the mass spectrum associated with the broken solutions given above. As to the solution (A), the residual symmetry group is an $SU(2)$ group with generators

$$X_i = \epsilon_{ijk} M^{jk} + T^i, \quad (i,j,k = 1,2,3),$$

where M and T are the generators of the Lorentz and of the isospin group, respectively: This group acts on both the isospin and space variables, rotating them simultaneously, so to speak, through the same angle. This exhibits the occurrence of the combined breakdown of the Lorentz and of the internal symmetry: the breaking of the Lorentz invariance cannot be avoided if the internal symmetry breaks spontaneously, which agrees with the general argument of Sec. II. In addition, the model predicts some "mixing" between the spatial and the internal variables. As to the mass spectrum, this is given by the eigenvalues of the matrix

$$M^{l,\lambda;n,\nu} = \left[\frac{\delta^2 \mathcal{L}_{st}}{\delta V^l_\lambda \delta V^n_\nu}\right]_{V \text{ stationary}},$$

which, after some computations, are

(a) $2\mu^2$ (simple),

(b) $-\mu^2$ (degeneracy 5),

(c) zero (degeneracy 6).

Here, the eigenvectors associated with (c) play the role of Goldstone particles,[7] as expected.

As to the solution B, the breaking of the symmetry is here even more drastic, as the residual symmetry group is abelian, namely the direct product of $O(1,1)$ with $SO(2)$, with generators M^{03} and $T^3 + M^{12}$, respectively. The same considerations as before apply, and, regarding the mass matrix, the eigenvalues are

(a) $2\mu^2$ (simple),
(b) $-2\mu^2$ (degeneracy 2),
(c) μ^2 (degeneracy 2),
(d) zero (degeneracy 7).

IV. FINAL REMARKS

As we have already mentioned, our results show two relevant features, namely, the large breaking of the Lorentz invariance induced by the spontaneous breakdown of the internal symmetry, and the mixing of spatial and internal transformations in the residual invariance group.

There are some other points which deserve mention. We have found, for instance, that the masses of our multiplet turn out to be independent of the "coupling constant" g. This amounts to say that the size of the mass splitting does not depend on the size of the interaction term (if only not zero) in the Lagrangian (5). Therefore, the constant g plays a very peculiar role: it is a sort of "catalyst" for the occurrence of the

spontaneous breakdown. The existence of such quantities seems to be a quite general feature of the spontaneously broken symmetries.

Another interesting point is that both solutions of Eq. (6) are actually unstable, due to the occurrence of negative eigenvalues of the squared mass operator. A similar situation has been already observed in Ref. 7, this result being possibly related to the dynamical approximations involved with the Lagrangian formalism.

To conclude, a few words on the breakdown of the Lorentz invariance. The occurrence of such breakdown is of course a rather disturbing feature, yet, the spontaneous breakdown of the internal symmetry when vector or spinor particles are involved is generally assumed in the literature (see, for instance, the case of the $\omega - \phi$ mixing,[17,18,12] or the chiral symmetries.[14,19]) For spinor particles, this situation has been investigated by Swieca.[20] In the general case, our results (Sec. II) indicate that a possible way to avoid breakdowns of the Lorentz group is to start with a reducible representation of this group, containing in particular some one-dimensional representations. In terms of the Lagrangian formalism, this amounts to say that one should consider Lagrangian functions depending not only on nonscalar fields (e.g., vector fields V^l, as in our model), but also on some scalar fields S, both V and S transforming under some representation of the internal symmetry group. In this way, spontaneous breakdown of the internal symmetry group is introduced by assuming that some of the fields S have non vanishing vacuum expectation values

$$(\Omega, S\,\Omega) \neq 0$$

and that all the vector (and in general nonscalar) fields have VEV's equal to zero:

$$(\Omega, V\Omega) = 0.$$

The latter condition prevents the breakdown of the Lorentz invariance, whereas the former ensures the breakdown of the internal symmetry. The effects of this appear not only in the multiplet of the S fields, but also in the other subspaces, thus giving rise to the mass splittings, etc., of all multiplets as imposed by the smaller residual (internal) symmetry. Roughly speaking, we may say that the breakdown is spontaneous in the subspace of the S fields, and is induced in the others; in fact, we can note that in this way the Goldstone particles belong exactly to the subspace of the S fields. An approach of this type has been already used in the literature, with remarkable physical results.[14,19]

ACKNOWLEDGMENTS

The analysis of the model considered in Sec. III was suggested to us by Professor E. Fabri. We would like to thank him for many stimulating discussions. One of us (P.d.M.) expresses his gratitude to Professor G. Jona-Lasinio for useful comments.

APPENDIX

We give a brief outline of the proofs needed in order to establish the results announced in Sec. III. To start with, some general results:

Proposition A1: Let (E, g) an m-dimensional real Euclidean space E provided with a pseudometric g (real symmetric bilinear form on E, not necessarily positive definite). Let the product of two matrices T, S over E be defined by

$$(T, S)^\mu_\nu = T^\mu_\lambda S^\lambda_\nu = T^{\mu\rho} g_{\rho\lambda} S^\lambda_\nu.$$

Then if a matrix T satisfies the equation $P(T) = 0$ (P being an arbitrary polynomial) with simple roots, T has a basis of eigenvectors.

Proof: Standard, apart from the occurrence of g: the decomposition of any vector $v \in E$ into eigenvectors of T: $v = \Sigma v_{(a)}$ is done by putting

$$v_{(a)} = [Q_{(a)}(T)/Q_{(a)}(\lambda_{(a)})]v,$$

where $\lambda_{(a)}$ are the roots of $P(z) = 0$, and $Q_{(a)}$ is such that $P(T) = (T - \lambda_{(a)} g)\, Q_{(a)}(T) = 0$.

Proposition A2: If (E, g) is as above, and T is a real symmetric matrix with real eigenvalues, then the eigenvectors of T can be chosen to be real, and eigenvectors belonging to different eigenvalues are orthogonal with respect to g. The proof is trivial, as it mimics the standard one.

An easy consequence of the above propositions[21] is:

Proposition A3: Let (E, g) be as above. Then a matrix T which satisfies both the hypothesis of Propositions A1 and A2 can be diagonalized by means of a g-preserving matrix.

Let us now turn back to Eq. (6)

$$\frac{\delta \mathfrak{L}_{st}}{\delta V},$$

which, when written explicitly, reads

$$m^2 V^{l\lambda} + g^2 (V^l_\mu V^{j\mu} V^{j\lambda} - V^{l\lambda} V^{j\mu} V^j_\mu) = 0.$$

Multiplying by V^l_ν and summing over l we get

$$m^2 T^\lambda_\nu + g^2 (T_{\nu\mu} T^{\mu\lambda} - T^\lambda_\nu T^\mu_\mu) = 0,$$

where

$$T^{\lambda\nu} = V^{l\lambda} V^{l\nu}.$$

Using the above Propositions we may diagonalize T (its eigenvalues, t_0, t_1, t_2, t_3 are real and positive in view of the reality of the V's) and obtain the following equations:

$$t_0(m^2 + g^2(t_1 + t_2 + t_3)) = 0,$$

$$t_i(-m^2 + g^2(t_i + t_0 - t_1 - t_2 - t_3)) = 0, \quad i = 1, 2, 3.$$

Apart from the trivial (zero) solution, there are only two solutions of the above equation, namely

(a) $t_0 = 0$, $t_1 = t_2 = t_3 = -(m^2/2g^2)$,

(b) $t_0 = 0$, only one of the t_i's equal to zero, the remaining two equal to $-(m^2/g^2)$.

These solutions correspond precisely to the solutions (A), (B), given in Sec. III.

As to the mass spectrum, we give the explicit form for the mass matrix and for the corresponding eigenvectors (the entries of the matrix are labeled in the order V^{10}, V^{20}, V^{30}, V^{11}, V^{21}, ...)

Case (A) $-m^{-2}M$

$$= \frac{1}{2} \begin{pmatrix} 0 & & 0 & & 0 & & 0 \\ & 0\,0\,0 & & 0\,-2\,0 & & 0\,0\,-2 & \\ 0 & 0\,1\,0 & & 1\,0\,0 & & 0\,0\,0 & \\ & 0\,0\,1 & & 0\,0\,0 & & 1\,0\,0 & \\ & 0\,1\,0 & & 1\,0\,0 & & 0\,0\,0 & \\ 0 & -2\,0\,0 & & 0\,0\,0 & & 0\,0\,-2 & \\ & 0\,0\,0 & & 0\,0\,1 & & 0\,1\,0 & \\ & 0\,0\,1 & & 0\,0\,0 & & 1\,0\,0 & \\ 0 & 0\,0\,0 & & 0\,0\,1 & & 0\,1\,0 & \\ & -2\,0\,0 & & 0\,-2\,0 & & 0\,0\,0 & \end{pmatrix}.$$

The eigenvectors are

(a) $V^{11} + V^{22} + V^{33}$ with eigenvalue $-2m^2$.

(b) $V^{12} + V^{21}$; $V^{13} + V^{31}$; $V^{23} + V^{32}$; $V^{11} - V^{22}$; $V^{11} - V^{33}$, with eigenvalue $+m^2$.

(c) V^{10}; V^{20}; V^{30}; $V^{12} - V^{21}$; $V^{23} - V^{32}$, $V^{31} - V^{13}$ with eigenvalue 0.

Case (B) $-m^{-2}M$

$$= \begin{pmatrix} 0\,0\,0 & & & & & & \\ 0\,0\,0 & & 0 & & 0 & & 0 \\ 0\,0\,1 & & & & & & \\ & & 0\,0\,0 & & 0\,-2\,0 & & \\ & 0 & 0\,1\,0 & & 1\,0\,0 & & 0 \\ & & 0\,0\,0 & & 0\,0\,0 & & \\ & & 0\,1\,0 & & 1\,0\,0 & & \\ & 0 & -2\,0\,0 & & 0\,0\,0 & & 0 \\ & & 0\,0\,0 & & 0\,0\,0 & & \\ & & & & & & 0\,0\,0 \\ & 0 & & 0 & & 0 & 0\,0\,0 \\ & & & & & & 0\,0\,-1 \end{pmatrix}.$$

The eigenvectors are

(a) $V^{11} + V^{22}$ with eigenvalue $-2m^2$.

(b) $V^{12} + V^{21}$, $V^{11} - V^{22}$ with eigenvalue $+2m^2$.

(c) V^{30}, V^{33} with eigenvalue $-m^2$.

(d) V^{10}; V^{20}; V^{13}; V^{23}; V^{31}; V^{32}; $V^{12} - V^{21}$ with eigenvalue 0.

[1] J. Goldstone, Nuovo Cimento **19**, 154 (1961).
[2] R. F. Streater, Proc. R. Soc. (Lond.) A **287**, 510 (1965).
[3] D. Kastler, D. W. Robinson, and A. Swieca, Comm. Math. Phys. **2**, 108 (1966).
[4] R. Ferrari and L. E. Picasso, Nucl. Phys. B **20**, 553 (1970).
[5] G. Cicogna, F. Strocchi, and R. Vergara Caffarelli, Phys. Rev. D **1**, 1197 (1970).
[6] G. Cicogna, F. Strocchi, and R. Vergara Caffarelli, Phys. Rev. D **6**, 301 (1972).
[7] P. de Mottoni and E. Fabri, Nuovo Cimento A **54**, 42 (1968).
[8] Note, of course, that the same is not true for Theorem $1_{(m}$.
[9] G. Jona-Lasinio, Acta Phys. Hung. **19**, 139 (1965).
[10] L. Michel and L. A. Radicati, in *Proceedings of the Fifth Coral Gables Conference on Symmetry Principles at High Energies* (Benjamin, New York, 1968); and in "Atti Convegno Mendeleviano," Rome, 1969.
[11] A. Pais. Phys. Rev. **173**. 1587 (1968).
[12] C.-L. Chan, J. Math. Phys. **11**, 1815 (1970).
[13] M. Gell-Mann and M. Lévy, Nuovo Cimento **16**, 705 (1960); B. W. Lee, Nucl. Phys. B **9**, 649 (1969).
[14] M. Lévy, Nuovo Cimento A **52**, 23 (1967).
[15] V. I. Ogievetskij and I. V. Polubarinov, Ann. Phys. (N. Y.) **25**, 358 (1963).
[16] A similar situation prevails in the $SO(n)$ breaking.[7]
[17] G. Cicogna and R. Vergara Caffarelli, Nuovo Cimento A **65**, 89 (1970).
[18] C.-L. Chan, Phys. Rev. D **1**, 2421 (1970).
[19] C. Cecchini and G. Cicogna, Nuovo Cimento A **5**, 527 (1971).
[20] J. A. Swieca, Nuovo Cimento A **57**, 573 (1968).
[21] In proving these propositions, a little care is needed only in presence of nonzero vectors with zero "norm."

Symmetries of extended $3j$ coefficients

A. D'Adda

Institut di Fisica dell'Università, Torino, Italy

R. D'Auria and G. Ponzano

Istituto di Fisica dell'Università, Torino, Italy

Istituto Nazionale di Fisica Nucleare, sezione di Torino, Italy

(Received 23 October 1972)

This paper studies the symmetries of a function which generalizes the $3j$ coefficients of $SU(2)$ and of $SU(1,1)$ involving discrete unitary representations. As a by-product of the analysis, the symmetries of the $SU(1,1)$ coefficient are obtained.

INTRODUCTION

Recent works[1-4] have pointed out a deep connection between $SU(1,1)$ and $SU(2)$ unitary representations, and between corresponding Clebsch—Gordan coefficients. This connection can be best visualized through analytic continuation in the representation parameters in such a way that discrete and continuous representations appear essentially on the same footing.[1,3] In this respect one might say that the $SU(1,1)$ case can be regarded as a suitable extension of $SU(2)$.

On the other hand, in the copious literature about quantum angular momenta there exists an extension[5-7] of $SU(2)$ representations, coupling and recoupling coefficients whose connection with the aforementioned unifying treatment of $SU(2)$ and $SU(1,1)$ has not yet been clarified. We mean the generalization to values $\bar{j}$ of the representation parameters of $SU(2)$ related to the usual ones by $\bar{j} = -j - 1$.

In our work we shall show that there exists a highly symmetrical structure which contains all these extensions in so far as discrete representations are considered. In so doing, the relations between $SU(2)$ and $SU(1,1)$ coupling and recoupling coefficients will appear as particular instances of the symmetries of the structure. Moreover, we shall obtain as a byproduct of our analysis a complete characterization of the symmetries of $SU(1,1)$ coefficients with discrete representations.

In the particular case of $6j$-coefficients, our work will contribute to the solution of a problem which arose recently. In fact new symmetries of the $6j$ of $SU(2)$ were claimed to hold[8]; however, they are not actual symmetries, as they violate triangular conditions.[9] In a forthcoming article we shall show that they are elements of the invariance group of the general structure.

The present paper is entirely devoted to the $3j$-coefficient. In Secs. 1 and 2 we introduce a set of real variables which are particularly suited to express the $3j$ of $SU(2)$ in terms of entire functions proportional to hypergeometric series ${}_3F_2(u,v,w;y,z;1)$. The restriction of these variables to a discrete set of values imposed by $SU(2)$ triangular conditions, suggests how to extend the $SU(2)$ $3j$-coefficient to a larger discrete domain $R \subset \mathbb{R}^{(5)}$. This is achieved in Sec. 3 where we consider a set of functions defined over R which satisfy a system of relations whose role is essentially equivalent to the recurrence relations of the aforementioned ${}_3F_2$ series. These functions provide the extension of the $3j$ of $SU(2)$ to the whole R and coincide—apart from a well-defined phase factor—with the $SU(1,1)$ Clebsch—Gordan coefficient in a suitable subregion of R; moreover, in the remaining portion of R, they coincide with the extension of $SU(2)$ and $SU(1,1)$ $3j$-coefficients to values $-j-1$ of at least one angular momentum parameter j labelling $SU(2)$, respectively $SU(1,1)$ discrete unitary representations. We stress that, especially when it is not explicitly stated, we consider only this class of $SU(1,1)$ representations.

In Sec. 4 we study the symmetries of the generalized functions and find $72 \cdot 20$ $(=1440)$ points in R where they differ by a phase factor which is explicitly determined. Through these symmetries the values of our generalized functions are known in the whole R once we know the $SU(2)$ $3j$-coefficient in the subregion of R characterized by $SU(2)$ constraints. Furthermore, by restricting the analysis to $SU(1,1)$ subregions of R, we find 6.18 $(=108)$ symmetries of $SU(1,1)$ $3j$-coefficients with generic discrete unitary representations; they are studied in detail in Appendix C. All symmetries of the generalized functions can be thought as consisting of the $72+108$ $SU(2)$ and $SU(1,1)$ symmetries enlarged by the transformations which change at least one j into $-j-1$; in fact, $(72+108)\cdot 8 = 1440$.

Finally we show in Sec. 5 how the Regge square-symbol for the $3j$ of $SU(2)$ can be generalized to our extended structure; we are led to a geometric configuration of lines, points, and planes in $\mathbb{R}^{(3)}$.

1. BASIC DEFINITIONS

According to Whipple notations,[10] we shall use[11] the variables $r_a \in \mathbb{R}$, $a \in \omega \equiv \{0,1,2,3,4,5\}$ or, in short, $\mathbf{r} = (r_0, r_1, r_2, r_3, r_4, r_5)$, restricted to a space $\mathbb{R}^{(5)}$ by

$$\sum_{a\in\omega} r_a = 0. \tag{1.1}$$

It is also convenient to introduce the auxiliary variables

$$\beta_{ab}(\mathbf{r}) = r_a - r_b + 1, \quad \forall\{a,b\} \subset \omega, \tag{1.2}$$

$$\alpha_{abc}(\mathbf{r}) = r_a + r_b + r_c + \tfrac{1}{2}, \quad \forall\{a,b,c\} \subset \omega; \tag{1.3}$$

lower case italic letters a, b, c, d, e, etc. will denote elements of ω, while the lower case Greek letters σ, τ, χ will denote subsets of ω with $|\sigma| = |\tau| = |\chi| = 3$ and $\sigma' = \omega - \sigma$, etc. Therefore, we shall write also $\alpha_\sigma(\mathbf{r}) = \sum_{a\in\sigma} r_a + \frac{1}{2}$ and, if no confusion arises, we shall abbreviate $\alpha_\sigma(\mathbf{r}) = \alpha_\sigma$.

Equation (1.1) entails many linear relations among the auxiliary variables as there are $\binom{6}{3} = 20$ different α's

and $\binom{6}{2}=15$ different β's; we quote only the basic identities

$$\alpha_\sigma+\alpha_{\sigma'}=1, \quad \forall\sigma, \tag{1.4}$$

$$\alpha_{abc}+\alpha_{ade}+\alpha_{bef}+\alpha_{cdf}=2, \tag{1.5}$$

where a,b,c,d,e,f is any permutation of $0,1,2,3,4,5$.

In $\mathbb{R}^{(5)}$ we consider the function

$$F(-\beta_{20},-\beta_{40},-\beta_{30},-\beta_{10},-\beta_{50}) = \frac{{}_3F_2(\alpha_{234},\alpha_{124},\alpha_{245};\beta_{20},\beta_{40};1)}{\Gamma(\alpha_{135})\Gamma(\beta_{20})\Gamma(\beta_{40})} \tag{1.6}$$

where, of course, $\beta_{a0}=\beta_{a0}(\mathbf{r})$, $a=1,2,3,4,5$, $\alpha_{234}=\alpha_{234}(\mathbf{r})$, etc. F has the following fundamental properties:

1.7 *Theorem*: $F(-\beta_{20},-\beta_{40},-\beta_{30},-\beta_{10},-\beta_{50})$ is symmetrical with respect to any permutation of its variables $\beta_{0a}(\mathbf{r})$, $a\in\omega-0$, namely, of r_1,r_2,r_3,r_4,r_5; hence we may write $F(-\beta_{a0},\forall a\in\omega-0)$.

The proof is due to A. C. Dixon.[12]

1.8 *Theorem*: F is an entire function of each variable.

Proof: See Appendix A.

By exchanging in Eq. (1.6) r_0 with any $r_a, a\in\omega-0$, one can define $F(-\beta_{ab}(\mathbf{r}),\forall a\in\omega-b)$, $\forall b\in\omega$; owing to Theorem 1.7 it is customary[13] to adopt the notation

$$\begin{aligned} F_p(b;\mathbf{r}) &= F(-\beta_{ab}(\mathbf{r}),\forall a\in\omega-b) \\ F_n(b;\mathbf{r}) &= F(-\beta_{ab}(-\mathbf{r}),\forall a\in\omega-b) \end{aligned}, \quad \forall b\in\omega \tag{1.10}$$

and to write, if no confusion arises, $F_p(b)$, $F_n(b)$. These 12 functions are linearly dependent; they satisfy[13] the following relations:

$$\frac{\sin(\pi\beta_{bc})F_p(a)}{\pi\Gamma(\alpha_{abc})} = \frac{F_n(b)}{\Gamma(\alpha_{cde},\alpha_{cef},\alpha_{cdf})} - \frac{F_n(c)}{\Gamma(\alpha_{bde},\alpha_{bef},\alpha_{bdf})}, \tag{1.11}$$

$$\frac{\sin(\pi\beta_{cb})F_n(a)}{\pi\Gamma(\alpha_{def})} = \frac{F_p(b)}{\Gamma(\alpha_{abd},\alpha_{abe},\alpha_{abf})} - \frac{F_p(c)}{\Gamma(\alpha_{acd},\alpha_{ace},\alpha_{acf})} \tag{1.12}$$

where a,b,c,d,e,f are any permutations of $0,1,2,3,4,5$ and $\Gamma(x_1,x_2,\cdots)=\prod_h\Gamma(x_h)$. If one or more α's assume integer values, part of Eqs. (1.11), (1.12) simplify remarkably; in particular, if $\alpha_\sigma(\mathbf{r})$ is a positive integer, these equations imply[13]:

$$\prod_{c\in\sigma'}\Gamma(\alpha_{\{\sigma-a\}\cup c}(\mathbf{r}))\ F_p(a;\mathbf{r}) = (-1)^{\alpha_{\sigma'}(\mathbf{r})}\prod_{c\in\sigma}\Gamma(\alpha_{\{\sigma-c\}\cup b}(\mathbf{r}))\ F_n(b;\mathbf{r}), \tag{1.13}$$

$\forall a\in\sigma,\ \forall b\in\sigma'$,

so that the value of both sides of Eq. (1.13) depend on neither $a\in\sigma$, nor $b\in\sigma'$. These two-term relations will play a basic role in what follows.

2. CONNECTION WITH THE 3j COEFFICIENT OF SU(2)

As the $3j$-coefficient of $SU(2)$ depends upon five independent parameters, j_ξ,μ_ξ, $\xi=1,2,3$, $\sum_{\xi=1}^3\mu_\xi=0$, the r variables are certainly suitable to express this coefficient as shown in Ref. 3 in a similar context. It will become apparent in Sec. 4 that they are the right variables to study the symmetries of our generalized functions. We make the following identification:

$$\begin{aligned} \begin{pmatrix} j_1 & j_2 & j_3 \\ \mu_1 & \mu_2 & \mu_3 \end{pmatrix} &= \begin{pmatrix} (r_1-r_0-1)/2 & (r_3-r_2-1)/2 & (r_5-r_4-1)/2 \\ (r_2+r_3-r_4-r_5)/2 & (r_4+r_5-r_0-r_1)/2 & (r_0+r_1-r_2-r_3)/2 \end{pmatrix} \\ &= \boxed{\begin{matrix} j_2+j_3-j_1 & j_3+j_1-j_2 & j_1+j_2-j_3 \\ j_1-\mu_1 & j_2-\mu_2 & j_3-\mu_3 \\ j_1+\mu_1 & j_2+\mu_2 & j_3+\mu_3 \end{matrix}} \\ &= \boxed{\begin{matrix} \alpha_{035}-1 & \alpha_{125}-1 & \alpha_{134}-1 \\ \alpha_{145}-1 & \alpha_{013}-1 & \alpha_{235}-1 \\ \alpha_{123}-1 & \alpha_{345}-1 & \alpha_{015}-1 \end{matrix}}. \end{aligned} \tag{2.1}$$

For future convenience we remark that Eq. (2.1) entails

$$\begin{aligned} &3r_0=-3j_1+\mu_3-\mu_2-\tfrac{3}{2}, \quad 3r_1=3j_1+\mu_3-\mu_2+\tfrac{3}{2}, \\ &3r_2=-3j_2+\mu_1-\mu_3-\tfrac{3}{2}, \quad 3r_3=3j_2+\mu_1-\mu_3+\tfrac{3}{2}, \\ &3r_4=-3j_3+\mu_2-\mu_1-\tfrac{3}{2}, \quad 3r_5=3j_3+\mu_2-\mu_1+\tfrac{3}{2}, \end{aligned} \tag{2.1'}$$

The row- and column-sum in the Regge square-symbol[14] in Eq. (2.1) is $\sum_{\xi=1}^3 j_\xi=(r_1+r_3+r_5-r_0-r_2-r_4-3)/2=\alpha_{135}-2$ by virtue of Eq. (1.3). As the arguments of the square-symbol have to be nonnegative integers to fulfill $SU(2)$ constraints, we see that α_χ, $|\chi\cap\{1,3,5,\}\geq 2$, are restricted to assume positive integer values (actually $\alpha_{135}>1$).

Now we recall[15] the identity

$$\begin{aligned} \begin{pmatrix} j_1 & j_2 & j_3 \\ \mu_1 & \mu_2 & \mu_3 \end{pmatrix} &= \delta_{\mu_1+\mu_2+\mu_3=0}\frac{(-1)^{j_1-j_2-\mu_3}}{(j_3-j_2+\mu_1)!\,(j_3-j_1-\mu_2)!} \\ &\times\left\{\frac{(j_2+j_3-j_1)!\,(j_1+\mu_1)!\,(j_2-\mu_2)!\,(j_3+\mu_3)!\,(j_3-\mu_3)!}{(j_1+j_2-j_3)!\,(j_1+j_2+j_3+1)!\,(j_1-\mu_1)!\,(j_2+\mu_2)!}\right\}^{1/2} \\ &\times{}_3F_2(j_3-j_1-j_2,\mu_1-j_1,-j_2-\mu_2;j_3-j_2+\mu_1+1,j_3-j_1-\mu_2+1;1); \end{aligned} \tag{2.2}$$

if in Eq. (2.2) we express j's and μ's in terms of α's according to Eq. (2.1), and make use of definition (1.10) with $a=4$, we obtain

$$\begin{pmatrix} j_1 & j_2 & j_3 \\ \mu_1 & \mu_2 & \mu_3 \end{pmatrix} = (-1)^{r_0-r_2}\left\{\frac{\Gamma(\alpha_{135},\alpha_{013},\alpha_{035},\alpha_{051},\alpha_{213},\alpha_{235},\alpha_{251})}{\Gamma(\alpha_{413},\alpha_{435},\alpha_{451})}\right\}^{1/2}F_p(4), \tag{2.3}$$

where r_0-r_2 is integer, as $r_0-r_2=\alpha_{035}-\alpha_{235}$. Equation (2.3) is the starting point of our generalization.

2.4. *Definition*: For any $\sigma=\{a,b,c\}\subset\omega$ let us denote with $R_\sigma\equiv R_{abc}$ the discrete set of points $\mathbf{r}\in\mathbb{R}^{(5)}$ such that $\alpha_\sigma(\mathbf{r})-1$ and $\alpha_\tau(\mathbf{r})$, $\forall\tau:|\sigma\cap\tau|=2$ are positive integers; by virtue of Eq. (1.1), the remaining $\alpha_\chi: 0\leq|\chi\cap\sigma|\leq 1$, assume nonpositive integer values in R_σ. Clearly there are 20 different R_σ's and we also define their joining $R=\cup_\sigma R_\sigma$.

2.5. *Lemma*: $R_\sigma \neq \emptyset, \forall\sigma$; $R_\sigma \cap R_\tau = \emptyset$, $\forall\sigma \neq \tau$. The first statement is true because it holds when $\sigma = \{1,3,5\}$ by virtue of Eq. (2.3). As for the second statement, let $\mathbf{r} \in R_\sigma$: (1) $\sigma \cap \tau \neq \emptyset$; then $\sigma \cap \tau' \neq \emptyset$, $\sigma' \cap \tau' \neq \emptyset$. Let $a \in \sigma \cap \tau'$, $b \in \sigma \cap \tau$, $c \in \sigma' \cap \tau'$; from Definition 2.4 it follows that $\alpha_{abc}(\mathbf{r}) > 0$, so that $\mathbf{r} \notin R_\tau$ because $\{a,b,c\} \cap \tau = b$. (2) $\sigma \cap \tau = \emptyset$; then $\alpha_\chi(\mathbf{r}) > 0\ \forall\chi : |\chi \cap \sigma| = 2$ and $\mathbf{r} \notin R_\tau$ because $|\chi \cap \tau| = 1$.[16]

2.6. *Remark*: In R every β_{ab} is integer; therefore $\alpha_{abc} + \beta_{ab} + 2\beta_{ca} - 3 = 3r_c + \frac{1}{2}$ is integer, so that $6r_c + 1 = 0 \bmod 2$, $\forall c \in \omega$. This yields $2\sum_{a\in\sigma} r_a + 1 = 0 \bmod 2$, $\forall\sigma$ and $r_a - r_b = 3r_a + 3r_b + 1 \bmod 2$, $\forall a \neq b$. These relations will prove useful in what follows to compute phase factors.

2.7. *Lemma*: Let $a \in \sigma$, $b \in \sigma'$; in $\mathbf{r} \in R_\sigma$, $F_p(a)$ and $F_n(b)$ have a zero of the second order; moreover,

$$(-1)^{3r_b - 1/2}\Big\{\prod_{c,d\in\sigma}\Gamma(\alpha_{bcd})\Big\}^{-1}F_p(b),$$

$$(-1)^{3r_a + 1/2}\Big\{\prod_{c,d\in\sigma'}\Gamma(1-\alpha_{acd})\Big\}^{-1}F_n(a)$$

do not depend on b, respectively a, and $F_n(a)$, $F_p(b)$ are related by

$$\Big\{\prod_{c\in\sigma-a}\Gamma(\alpha_{abc})\Big\}F_n(a) = (-1)^{\alpha\{\sigma'-b\}\cup a}\Big\{\prod_{c\in\sigma'-b}\Gamma(1-\alpha_{abc})\Big\}F_p(b). \tag{2.8}$$

Proof: Set $\sigma = \{a,e,f\}$, $\sigma' = \{b,g,h\}$ and let α_{efb} assume a positive integer value; Eq. (1.13) yields

$$\begin{aligned}\Gamma(\alpha_{fba}, \alpha_{fbg}, \alpha_{fbh})F_p(e) &= \Gamma(\alpha_{eba}, \alpha_{ebg}, \alpha_{ebh})F_p(f)\\ &= \Gamma(\alpha_{efa}, \alpha_{efg}, \alpha_{efh})F_p(b)\\ &= (-1)^{\alpha_{agh}}\Gamma(\alpha_{aef}, \alpha_{afb}, \alpha_{aeb})F_n(a)\\ &= (-1)^{\alpha_{agh}}\Gamma(\alpha_{gef}, \alpha_{gfb}, \alpha_{geb})F(g)\\ &= (-1)^{\alpha_{agh}}\Gamma(\alpha_{hef}, \alpha_{hfb}, \alpha_{heb})F(h).\end{aligned} \tag{2.9}$$

As $\mathbf{r} \in R_{aef}$ all Γ's which multiply $F_p(b)$, $F_n(a)$ are finite; on the contrary, in front of $F_p(e)$, $F_p(f)$, $F_n(g)$, $F_n(h)$ there are two exploding Γ's which must be matched by a second order zero of these functions. As F_n, F_p are entire functions of each variable, the relation $\Gamma(\alpha_{efg}, \alpha_{efh})F_p(b) = (-1)^{\alpha_{agh}}\Gamma(\alpha_{abf}, \alpha_{abe})F_n(a)$ does not depend upon along which direction the point $\mathbf{r} \in R_\sigma$ is approached in $\mathbf{R}^{(5)}$. The proof is completed by making α_{efg} positive integer and by comparing $F_p(b)$ to $F_p(g)$ through their relation with $F_n(a)$.

2.10. *Corollary*: For any given σ, the functions

$$\varphi_p(b,\mathbf{r}) = (-1)^{3r_b - 1/2 + \alpha_\sigma}\{P_\sigma(\mathbf{r})\}^{1/2}\Big\{\prod_{c,d\in\sigma}\Gamma(d_{bcd}(\mathbf{r}))\Big\}^{-1}F_p(b;\mathbf{r}),$$

$$b \in \sigma',$$

$$\varphi_n(a;\mathbf{r}) = (-1)^{3r_a + 1/2}\{P_\sigma(\mathbf{r})\}^{1/2}\Big\{\prod_{c,d\in\sigma'}\Gamma(1-\alpha_{acd}(\mathbf{r}))\Big\}^{-1}F_n(a;\mathbf{r}),$$

$$a \in \sigma,$$

where $P_\sigma = \prod_{\forall\chi:|\chi\cap\sigma|\geq 2}\Gamma(\alpha_\chi)$, coincide if $\mathbf{r} \in R_\sigma$: $\varphi_n(a;\mathbf{r}) = \varphi_p(b;\mathbf{r}) \equiv \varphi(\sigma;\mathbf{r})$, $\forall a \in \sigma$, $b \in \sigma'$.

2.11. *Remark*: We notice that in the particular case $\sigma = \{1,3,5\}$, recalling Eq. (2.3), we have from 2.10

$$\begin{pmatrix} j_1 & j_2 & j_3 \\ \mu_1 & \mu_2 & \mu_3 \end{pmatrix} = \varphi(\{1,3,5\};\mathbf{r}). \tag{2.12}$$

It is straightforward to deduce from 2.10 and from Eq. (2.5) the symmetries of the $SU(2)$ $3j$-coefficient. In fact, it is clearly invariant under (1) permutations of r_0, r_2, r_4; (2) permutations of r_1, r_3, r_5; (3) exchange of r_0, r_2, r_4 with r_1, r_3, r_5 and sign inversion of all $r_a, a \in \omega$. This transformation multiplies the coefficient by $(-1)^{\alpha_{135}}$ because $\varphi_p(b;\mathbf{r})$, $b \in \{0,2,4\}$, transforms into

$$(-1)^{-3r_a - 1/2 + \alpha_{135}}\{P_{024}(-\mathbf{r})\}^{1/2}\Big\{\prod_{c,d\in\sigma'}\Gamma(\alpha_{acd}(-\mathbf{r}))\Big\}^{-1}F_p(a;-\mathbf{r})$$

for some $a \in \{1,3,5\}$, and $P_{024}(-\mathbf{r}) = P_{135}(\mathbf{r})$, $\alpha_{acd}(-\mathbf{r}) = 1 - \alpha_{acd}(\mathbf{r})$, $F_p(a;-\mathbf{r}) = F_n(a;\mathbf{r})$ according to definition (1.10) These symmetries correspond to permutations of rows and (or) columns of the Regge symbol, and to reflection with respect to the main diagonal.

One way to generalize Eq. (2.3) is to continue analytically $\varphi_p(b;\mathbf{r})$, $\varphi_n(a;\mathbf{r})$, $b \in \{0,2,4\}$, $a \in \{1,3,5\}$, from $\mathbf{r} \in R_{135}$ to $\mathbf{r}' \in R - R_{135}$. In so doing one should take into account that $F_p(b;\mathbf{r}')$, $F_n(a;\mathbf{r}')$ develop a zero of the second order, according to 2.7; but it is easy to check that a pole of the second order is provided at $\mathbf{r}'$ by the factors which multiply F_p, F_n in φ_p, φ_n. This suggests that the analytic extension of φ_p, φ_n to the entire R can be devised and in Ref. 2 a Pochhammer double-integral representation of ${}_3F_2$ has been used for this purpose. However, we feel that this procedure—which is powerful enough to cover even the continuous representations of $SU(1,1)$—in so far as the discrete ones are considered, does not clarify the symmetry of the underlying algebraic structure; for this reason we shall develop a different method.

Beforehand we want to analyze the meaning of each R_σ from the point of view of the angular momentum parameters. More precisely, we associate to R_{abc} the square-symbol

$$[a,b,c] \equiv \begin{vmatrix} \alpha_{dbc} - 1 & \alpha_{eac} - 1 & \alpha_{fab} - 1 \\ \alpha_{fca} - 1 & \alpha_{dab} - 1 & \alpha_{ebc} - 1 \\ \alpha_{eab} - 1 & \alpha_{fbc} - 1 & \alpha_{dca} - 1 \end{vmatrix} \tag{2.13}$$

where $\{d,e,f\} = \omega - \{a,b,c\}$; the row and column sum is $\alpha_{abc} - 2$ and, from Definition 2.4, in R_{abc} all variables which appear in Eq. (2.13) in addition to $\alpha_{abc} - 1$ are nonnegative integers. Sticking to the identification specified by Eq. (2.1) among j's, μ's, and $\mathbf{r}$, we want to see what inequalities hold true for j's, μ's in R_{abc}.

2.14. If $R_{abc} = R_{135}$ we obviously find the $SU(2)$ constraints:

$$j_\xi \geq 0 \tag{2.14a}$$

$$j_\xi \geq |\mu_\xi| \qquad \xi = 1,2,3, \tag{2.14b}$$

$$j_1 + j_2 + j_3 \geq 0, \tag{2.14c}$$

$$j_1\frac{\mu_2\mu_3}{|\mu_2\mu_3|} + j_2\frac{\mu_3\mu_1}{|\mu_3\mu_1|} + j_3\frac{\mu_1\mu_2}{|\mu_1\mu_2|} \leq 0, \tag{2.14d}$$

where Eq. (2.14d), which we have borrowed from Ref. 17, implies the well-known triangular relations, because $\mu_1 + \mu_2 + \mu_3 = 0$ provides the only constraint over the sign of μ's.

2.15. Let $\mathbf{r}$ belong to any one of R_{035}, R_{125}, R_{134}; R_{124}, R_{034}, R_{025}; R_{024}. In each case we find the same inequalities of Eqs. (2.14) provided j_ξ is replaced by $-j_\xi - 1$, $\forall\xi \in \Xi$, where

$$\Xi = \begin{array}{ccccccc} R_{035} & R_{125} & R_{134} & R_{124} & R_{034} & R_{025} & R_{024} \\ \hline \{1\} & \{2\} & \{3\} & \{2,3\} & \{3,1\} & \{1,2\} & \{1,2,3\} \end{array}$$

For instance, if $\mathbf{r} \in R_{024}$, $\alpha_{024} = -j_1 - j_2 - j_3 - 1 \geq 2$ implies $(-j_1 - 1) + (-j_2 - 1) + (-j_3 - 1) \geq 0$, etc. Therefore, the extension of the $3j$ of $SU(2)$ to these other regions of R, where the $SU(2)$ inequalities are satisfied but at least one j assumes a negative value, is just the extension considered in Refs. 5, 6, 7.

2.16. Finally, if $\mathbf{r}$ belongs to one of the remaining R_σ's, setting for convenience $\hat{j}_\xi = j_\xi + \frac{1}{2}$, $\xi = 1, 2, 3$, we find

$$|\hat{j}_\xi| + \tfrac{1}{2} \leq |\mu_\xi|, \quad \xi = 1, 2, 3, \tag{2.16a}$$

$$\frac{\mu_2\mu_3}{|\mu_2\mu_3|}(|\hat{j}_1| + \tfrac{1}{2}) + \frac{\mu_3\mu_1}{|\mu_3\mu_1|}(|\hat{j}_2| + \tfrac{1}{2}) + \frac{\mu_1\mu_2}{|\mu_1\mu_2|}(|\hat{j}_3| + \tfrac{1}{2}) \geq 0, \tag{2.16b}$$

and in each region the sign of only one $\hat{j}$ and of all μ's is specified as follows:

$\hat{\xi}=1$	2	3			
$\mathbf{r} \in R_{145}$	R_{013}	R_{235}:	$\hat{j}_\xi \geq \frac{1}{2}$,	$\mu_\xi \leq -1$,	$\mu_\eta \geq +\frac{1}{2}$,
R_{123}	R_{345}	R_{015}:	$\hat{j}_\xi \geq \frac{1}{2}$,	$\mu_\xi \geq +1$,	$\mu_\eta \leq -\frac{1}{2}$,
R_{023}	R_{245}	R_{014}:	$\hat{j}_\xi \leq -\frac{1}{2}$,	$\mu_\xi \geq +1$,	$\mu_\eta \leq -\frac{1}{2}$,
R_{045}	R_{012}	R_{234}:	$\hat{j}_\xi \leq -\frac{1}{2}$,	$\mu_\xi \leq -1$,	$\mu_\eta \geq +\frac{1}{2}$,

(2.16c)

$\eta \in \{1, 2, 3\} - \xi$. Here we have quoted the stronger inequalities; notice that Eq. (2.16b) provides only one relation for each R_σ. It is easily checked (see Ref. 2) that inside the subregions of $R_{023}, R_{245}, R_{014}, R_{045}, R_{012}, R_{234}$ where $\hat{j}_\xi \leq \frac{1}{2}$, $\xi = 1, 2, 3$, the parameters j's, μ's satisfy the inequalitites which characterize $SU(1,1)$ unitary discrete representations having a nonvanishing $3j$-coefficient. The remaining subregions of these R_σ's, as well as the whole of $R_{145}, R_{013}, R_{235}, R_{123}, R_{345}, R_{015}$, give rise to the same $SU(1,1)$ inequalities with sign inversion of at least one $\hat{j}$, namely they correspond to the same extension[18] as the one considered in 2.15 for the $SU(2)$ case.

For future convenience we denote with $\hat{R}_\sigma$, $\sigma = \{0, 1, 2\}$, $\{0, 1, 4\}, \{2, 3, 4\}, \{2, 3, 0\}, \{4, 5, 0\}, \{4, 5, 2\}$ the subset of R_σ where $\hat{j}_\xi \leq +\frac{1}{2}$, $\xi = 1, 2, 3$, and write $R(SU(1,1)) = \cup_\sigma R_\sigma$.

3. THE GENERALIZED FUNCTIONS

In this section we shall construct a set of functions in R which coincide with the $3j$-coefficient of $SU(2)$ in R_{135}; it will turn out that in $R(SU(1,1))$ these functions coincide within a phase with the $3j$ of $SU(1,1)$ which couples discrete unitary representations.

3.1. *Definition*: We introduce the step function $\vartheta(\sigma;\mathbf{r})$, $\forall \mathbf{r} \in R$, such that $\vartheta(\sigma, \mathbf{r}) = 1$ if $\alpha_\sigma(\mathbf{r}) > 0$, while $\vartheta(\sigma;\mathbf{r}) = 0$ if $\alpha_\sigma(\mathbf{r}) \leq 0$; notice that $\vartheta(\sigma;\mathbf{r}) + \vartheta(\sigma';\mathbf{r}) = 1\ \forall\sigma$.

To simplify our notations, it is very convenient to use the function

$$\tilde{\Gamma}(\alpha_\sigma(\mathbf{r})) = \frac{\{\Gamma(\alpha_\sigma(\mathbf{r}))\}^{\vartheta(\sigma;\mathbf{r})}}{\{\exp[i\pi\alpha_{\sigma'}(\mathbf{r})]\Gamma(\alpha_{\sigma'}(\mathbf{r}))\}^{\vartheta(\sigma';\mathbf{r})}}. \tag{3.2a}$$

As we are concerned with integer values of α_σ's, Eq. (3.2a) amounts to modify the Γ-function by replacing its infinite values at the poles with the corresponding residues; this procedure has been used in Refs. 5, 6, 7 to extend the $3j$ of $SU(2)$ to negative values of j's. Eq. (3.2a) must be implemented with the following phase prescription:

$$\{\tilde{\Gamma}(\alpha_\sigma(\mathbf{r}))\}^{1/2} = \exp\{-\tfrac{1}{2}i\pi\vartheta(\sigma';\mathbf{r})\alpha_{\sigma'}(r)\}\,|\,\tilde{\Gamma}(\alpha_\sigma(\mathbf{r}))\,|^{1/2}.$$

This is enough for our purposes as we shall not run into noninteger powers of $\tilde{\Gamma}$'s different from $\frac{1}{2}$.

We notice that

$$\tilde{\Gamma}(\alpha_\sigma(\mathbf{r}))\tilde{\Gamma}(\alpha_{\sigma'}(\mathbf{r})) = \exp\{-i\pi[\vartheta(\sigma;\mathbf{r})\alpha_\sigma(\mathbf{r}) + \vartheta(\sigma';\mathbf{r})\alpha_{\sigma'}(\mathbf{r})]\}. \tag{3.3}$$

Moreover, the following useful identity holds $\forall a \in \omega$:

$$(1/2)\sum_{\tau \subset \omega - a}\{\vartheta(\tau;\mathbf{r})\alpha_\tau(\mathbf{r}) + \vartheta(\tau';\mathbf{r})\alpha_{\tau'}(\mathbf{r})\} = -\tfrac{1}{2} \bmod 2. \tag{3.4}$$

By means of definition (3.2) we may cast the functions $\varphi_p(b;\mathbf{r})$, $\varphi_n(a;\mathbf{r})$, $\mathbf{r} \in R_\sigma$, $a \in \sigma$, $b \in \sigma'$, defined in 2.10, into a more symmetrical form; recalling 2.6, we obtain easily

$$\varphi_p(b;\mathbf{r}) = \exp\{i\pi(-\tfrac{3}{2}\gamma_b + \tfrac{3}{4})\}\{Q_p(b;\mathbf{r})\}^{1/2}F_p(b;\mathbf{r}), \quad b \in \sigma', \tag{3.5a}$$

$$\varphi_n(a;\mathbf{r}) = \exp\{i\pi[\tfrac{3}{2}\gamma_a - \tfrac{1}{4} + \alpha_{\sigma'}(\mathbf{r})]\}\{Q_n(a;\mathbf{r})\}^{1/2}F_n(a;\mathbf{r}), \quad a \in \sigma, \tag{2.5b}$$

where

$$Q_p(a;\mathbf{r}) = \prod_{\tau \subset \omega - a}\tilde{\Gamma}(\alpha_\tau(\mathbf{r})) \tag{3.5c}$$

$$Q_n(a;\mathbf{r}) = \prod_{\tau \subset \omega - a}\tilde{\Gamma}(1 - \alpha_\tau(\mathbf{r})). \tag{3.5d}$$

We postulate now a set of equations whose solution will provide the basic ingredients of our construction. In fact, the functions $\tilde{F}_p(a;\mathbf{r})$, $\tilde{F}_n(a;\mathbf{r})$, $a \in \omega$, which appear as unknown in these equations, can be chosen in such a way that—upon multiplication by suitable factors as specified in Definition 3.11—they yield the desired generalized functions. The equations read, if $\mathbf{r} \in R_\sigma$,

$$\prod_{c \in \tau'}\tilde{\Gamma}(\alpha_{\{\tau - b\}\cup c}(\mathbf{r}))\ \tilde{F}_p(b;\mathbf{r}) = (-1)^{|\tau\cap\sigma| + \alpha_{\tau'}(\mathbf{r})} \times \prod_{c \in \tau}\tilde{\Gamma}(\alpha_{\{\tau - c\}\cup a}(\mathbf{r}))\ F_n(a;\mathbf{r}), \tag{3.6}$$

$\forall \tau: 2 \leq |\tau \cap \sigma| \leq 3$, and $\forall a \in \tau'$, $b \in \tau$. These relations—apart from the phase $(-1)^{|\tau\cap\sigma|}$—are obtained from Eq. (1.13) by considering in turn each $\alpha_\tau(\mathbf{r}) > 0$ and by replacing $\Gamma(\alpha)$ with $\tilde{\Gamma}(\alpha)$.

3.7. *Lemma*: Equations (3.6) are compatible and admit the following solutions at $\mathbf{r} \in R_\sigma$:

$$\tilde{F}_p(a;\mathbf{r}) = F_p(a;\mathbf{r}), \quad \forall a \in \sigma', \tag{3.8a}$$

$$\tilde{F}_n(a;\mathbf{r}) = F_n(a;\mathbf{r}), \quad \forall a \in \sigma, \tag{3.8b}$$

$$\tilde{F}_p(b;\mathbf{r}) = (-1)^{3r_a + 1/2}\prod_{c \in \omega - \{a, b\}}\Gamma(\alpha_{abc}(\mathbf{r}))\ F_n(a;\mathbf{r}), \quad \forall b \in \sigma,\ a \in \sigma - b; \tag{3.8c}$$

$$\tilde{F}_n(b;\mathbf{r}) = (-1)^{3r_a-1/2} \prod_{c\in\omega-\{a,b\}} \Gamma(1-\alpha_{abc}(\mathbf{r}))\ F_p(a;\mathbf{r}),$$

$$\forall b\in\sigma',\quad a\in\sigma'-b. \qquad (3.8d)$$

Proof: We delete the proof of consistency to Appendix B. Here we show that there are equations of the system (3.6) which are satisfied by the choice (3.8). Let $a\in\sigma\cap\tau'$, $b\in\sigma'\cap\tau$ in (3.6); this implies $|\sigma\cap\tau|=2$ and $\sigma=(\sigma\cap\tau)\cup a$, $\tau=(\sigma\cap\tau)\cup b$; therefore, if $c\in\tau'$, we have $\alpha_{\{\tau-b\}\cup c}=\alpha_{(\sigma\cap\tau)\cup c}>0$ and if $c\in\tau:|(\{\tau-c\}\cup a)\cap\sigma|=2$, so that $\alpha_{\{\tau-c\}\cup a}>0$. According to definition (3.2), it follows that $\tilde{\Gamma}(\alpha_\chi)=\Gamma(\alpha_\chi)$ for every χ involved in (3.6); moreover, $\Gamma(\alpha_{(\sigma\cap\tau)\cup a})$ appears on both sides of (3.6) and $\alpha_{\tau'}=\alpha_{\{\sigma'-b\}\cup a}$. Therefore, $\tilde{F}_p(b;\mathbf{r})$, $b\in\sigma'$, and $\tilde{F}_n(a;\mathbf{r})$, $a\in\sigma$, satisfy Eq. (2.8). As for Eqs. (3.8.c,d), they follow from (3.8.a,b) and from Eqs. (B1b), (B1c) of Appendix B.

Remark: The consistency of Eqs. (3.6) shows up also from the fact that the rhs of (3.8c), (3.8d) is actually independent of $a\in\sigma$, respectively $a\in\sigma'$. In fact, we may write these equations as follows:

$$\tilde{F}_p(b;\mathbf{r}) = \Gamma(\alpha_\sigma(\mathbf{r}))\left\{\prod_{\substack{c\in\sigma'\\ d\in\sigma-b}}\Gamma(\alpha_{bcd}(\mathbf{r}))\right\}\frac{(-1)^{3r_a+1/2}F_n(a;\mathbf{r})}{\prod_{c,d\in\sigma'}\Gamma(1-\alpha_{acd}(\mathbf{r}))},$$

$$\forall b\in\sigma,\quad a\in\sigma-b; \qquad (3.9a)$$

$$\tilde{F}_n(g;\mathbf{r}) = \Gamma(1-\alpha_{\sigma'}(\mathbf{r}))\left\{\prod_{\substack{c\in\sigma\\ d\in\sigma'-g}}\Gamma(1-\alpha_{gcd}(\mathbf{r}))\right\}\frac{(-1)^{3r_e-1/2}F_p(e;\mathbf{r})}{\prod_{c,d\in\sigma}\Gamma(\alpha_{ecd}(\mathbf{r}))},$$

$$\forall g\in\sigma',\quad e\in\sigma'-g; \qquad (3.9b)$$

then we recall 2.7 and realize that Eqs. (3.9a), (3.9b) hold also when $a=b$, $e=g$. In this particular case we know from Appendix B that, if $\mathbf{r}\in R_\sigma$, then

$$\tilde{F}_p(b;\mathbf{r}) = -Q_n(b;\mathbf{r})\tilde{F}_n(b;\mathbf{r}),\quad \forall b\in\sigma, \qquad (3.10a)$$

$$\tilde{F}_n(g;\mathbf{r}) = -Q_p(g;\mathbf{r})\tilde{F}_p(g;\mathbf{r}),\quad \forall g\in\sigma'. \qquad (3.10b)$$

3.11. *Definition*: $\mathbf{r}\in R$, $a\in\omega$, we set

$$\Phi_p(a;\mathbf{r}) = \exp\{i\pi[\psi_a(\mathbf{r})+A(a;\mathbf{r})]\}\{Q_p(a;\mathbf{r})\}^{1/2}\tilde{F}_p(a;\mathbf{r}), \qquad (3.11a)$$

$$\Phi_n(a;\mathbf{r}) = \exp\{i\pi[\nu_a(\mathbf{r})+A(a;\mathbf{r})]\}\{Q_n(a;\mathbf{r})\}^{1/2}\tilde{F}_n(a;\mathbf{r}), \qquad (3.11b)$$

where

$$\begin{aligned}\psi_a(\mathbf{r}) &= 3r_a-\tfrac{1}{2}+\alpha_{135}(\mathbf{r})\\ \nu_a(\mathbf{r}) &= 3r_a+\alpha_{135}(\mathbf{r})\end{aligned}\quad \text{if } a\in\{0,2,4\}, \qquad (3.11c)$$

$$\begin{aligned}\psi_a(\mathbf{r}) &= 3r_a+1\\ \nu_a(\mathbf{r}) &= 3r_a+\tfrac{1}{2}\end{aligned}\quad \text{if } a\in\{1,3,5\}, \qquad (3.11d)$$

$$A(a;\mathbf{r}) = \tfrac{1}{2}\sum_{b,c\in\chi}\{\vartheta(\{a,b,c\};\mathbf{r})\alpha_{abc}(\mathbf{r})+\vartheta(\omega-\{a,b,c\};\mathbf{r})\alpha_{\omega-\{a,b,c\}}(\mathbf{r})\},$$

$$\chi=\{1,3,5\}\ \text{if } a\in\{0,2,4\};\quad \chi=\{0,2,4\}\ \text{if } a\in\{1,3,5\}. \qquad (3.11e)$$

This choice of phase factors will be justified *a posteriori* as we shall show in 3.13 that it forces all Φ_p's, Φ_n's to coincide in R_{135} with the $SU(2)$ $3j$-coefficient; this clarifies also the privileged role of R_{135}.

3.12. *Theorem*: For any σ, recalling the definition of $\varphi(\sigma;\mathbf{r})$ given in 2.10, if $r\in R_\sigma$, we have

$$\Phi_{p,n}(a;\mathbf{r}) = \exp\{i\pi\Psi_{p,n}(a;\mathbf{r})\}\varphi(\sigma;\mathbf{r}),$$

where

$$\Psi_n(a;\mathbf{r}) = \nu_a(\mathbf{r})-\tfrac{3}{2}r_a+\tfrac{1}{4}+\alpha_{\sigma'}(\mathbf{r})+A(a;\mathbf{r}),\quad \text{mod 2, if } a\in\sigma$$
$$= \nu_a(\mathbf{r})+\tfrac{3}{2}r_a+\tfrac{3}{4}+A(a;\mathbf{r}),\quad \text{mod 2, if } a\in\sigma', \qquad (3.12a)$$

$$\Psi_p(a;\mathbf{r})-\Psi_n(a;\mathbf{r}) = \delta_{a\in(\sigma\cap\{0,2,4\})\cup(\sigma'\cap\{1,3,5\})},\quad \text{mod 2}. \qquad (3.12b)$$

Here, $\delta_{a\in\tau}=1$ if $a\in\tau$, $\delta_{a\in\tau}=0$ if $a\notin\tau$.

Proof: Let $a\in\sigma$; then, from Eq. (3.8a), $\tilde{F}_n(a;\mathbf{r})=\tilde{F}_n(a;\mathbf{r})$. Therefore, comparing Eqs. (3.11b), (3.5b), we obtain (3.12a) for the case $a\in\sigma$. On the other hand, if $a\in\sigma'$, we may use Eq. (3.10b) so that (3.11b) becomes

$$\Phi_n(a;\mathbf{r}) = \exp\{i\pi[\nu_a(\mathbf{r})+A(a;\mathbf{r})+1]\}\{Q_n(a;\mathbf{r})\}^{1/2}Q_p(a;\mathbf{r})\tilde{F}(a;\mathbf{r}).$$

Now notice that

$$Q_p(a;\mathbf{r})Q_n(a;\mathbf{r}) = \prod_{\tau\subset\omega-a}\Gamma(\alpha_\tau(\mathbf{r}))\tilde{\Gamma}(\alpha_{\tau'}(\mathbf{r}));$$

therefore, from Eqs. (3.3), (3.4), (3.8a) we obtain

$$\Phi_n(a;\mathbf{r}) = \exp\{i\pi[\nu_a(\mathbf{r})+A(a;\mathbf{r})-\tfrac{1}{2}]\}\{Q_p(a;\mathbf{r})\}^{1/2}F_p(a;\mathbf{r})$$

and by comparison with Eq. (3.5a), we get (3.12a) in the case $a\in\sigma'$. Then consider the ratio $\Phi_p(a;\mathbf{r})/\Phi_n(a;\mathbf{r})$; from 3.11 we have

$$\frac{\Phi_p(a;\mathbf{r})}{\Phi_n(a;\mathbf{r})} = \exp\{i\pi[\psi_a(\mathbf{r})-\nu_a(\mathbf{r})]\}\left\{\frac{Q_p(a;\mathbf{r})}{Q_n(a;\mathbf{r})}\right\}^{1/2}\frac{\tilde{F}_p(a;\mathbf{r})}{\tilde{F}_n(a;\mathbf{r})}.$$

Using Eqs. (3.10), (3.3), (3.4) we deduce

$$\left\{\frac{Q_p(a;\mathbf{r})}{Q_n(a;\mathbf{r})}\right\}^{1/2}\frac{\tilde{F}_p(a;\mathbf{r})}{\tilde{F}_n(a;\mathbf{r})} = -\{Q_p(a;\mathbf{r})Q_n(a;\mathbf{r})\}^{1/2-\delta_{a\in\sigma'}}$$
$$= \exp\ i\frac{\pi}{2}(\delta_{a\in\sigma'}-\delta_{a\in\sigma}),$$

which leads to (3.12b) if we recall (3.11c,d).

3.13. *Corollary*: If $\mathbf{r}\in R_{135}$, then: $\Phi_p(a;\mathbf{r})=\Phi_n(a;\mathbf{r})=\begin{pmatrix}j_1 & j_2 & j_3\\ \mu_1 & \mu_2 & \mu_3\end{pmatrix}$, $\forall a\in\omega$, provided we make the identifications specified by Eq. (2.1).

In fact, by direct computation of $A(a;\mathbf{r})$, Eq. (3.12a) yields for any $a\in\omega$

$$\Psi_n(a;\mathbf{r}) = 3\sum_{b\in\{1,3,5\}}r_b+1,\quad \text{mod 2, if } |\sigma\cap\{1,3,5\}|=0, \qquad (3.14a)$$

$$= 3r_{\sigma'\cap\{0,2,4\}}+\delta_{a\in\sigma\cap\{0,2,4\}},\quad \text{mod 2, if } |\sigma\cap\{1,3,5\}|=1, \qquad (3.14b)$$

$$= 3\sum_{b\in\sigma\cap\{1,3,5\}}r_b+\delta_{a\in\sigma\cap\{1,3,5\}},\quad \text{mod 2, if } |\sigma\cap\{1,3,5\}|=2, \qquad (3.14c)$$

$$= 0,\quad \text{mod 2, if } |\sigma\cap\{1,3,5\}|=3. \qquad (3.14d)$$

Therefore, if $\sigma=\{1,3,5\}$, we get from Eq. (3.12b): $\Psi_p(a;\mathbf{r})=\Psi_n(a;\mathbf{r})$, $\forall a\in\omega$, and we have just to recall 2.12.

3.15. *Corollary*: If $\mathbf{r}\in R(SU(1,1))$, then $\Phi_p(a;\mathbf{r})$, $\Phi_n(a;\mathbf{r})$, $\forall a\in\omega$, coincide—apart from a phase factor—

with the $SU(1,1)$ $3j$-coefficient involving the discrete unitary representations j_ξ, $\xi=1,2,3$, specified by Eq. (2.1).

In fact, by means of Dougall's formula, it is straightforward to check that the Clebsch–Gordan coefficient of $SU(1,1)$ with discrete unitary representations $C^{j_1\ j_2\ j_3}_{\mu_1\ \mu_2\ -\mu_3}$, divided by $\sqrt{-2j_3-1}$, as computed in Ref. 2, is related to $\varphi(\sigma;\mathbf{r})$ as follows:

$$\frac{C^{j_1\ j_2\ j_3}_{\mu_1\ \mu_2\ -\mu_3}}{\sqrt{-2j_3-1}}$$

$$=(-1)^{3r_2+1/2}\varphi(\{2,3,4\};\mathbf{r}) \quad \text{[Eq. (3.11), Ref. 2]}$$

$$=(-1)^{3r_0+1/2}\varphi(\{0,1,4\};\mathbf{r}) \quad \text{[Eq. (3.15), Ref. 2]}$$

$$=(-1)^{r_4-r_5}\varphi(\{0,1,2\};\mathbf{r}) \quad \text{[Eqs. (3.18), (3.20), Ref. 2]}$$

$$=(-1)^{r_4-r_5}\varphi(\{0,2,3\};\mathbf{r}) \quad \text{[Eqs. (3.17), (3.19), Ref. 2],}$$

and we guess

$$=(-1)^{3r_2+1/2}\varphi(\{2,4,5\};\mathbf{r})$$
$$=(-1)^{3r_0+1/2}\varphi(\{0,4,5\};\mathbf{r}).$$

To complete the argument, we recall 3.12.

Remarks: At this stage it is clear, in the light of definition (3.2a), that within the regions specified in 2.15 all Φ_n's, Φ_p's coincide—apart from a possible phase—with one of the extensions considered in Refs. 5, 6, 7 of the $SU(2)$ $3j$-coefficient. Similarly, inside the regions considered in 2.16 which are not proper $SU(1,1)$ regions, these functions coincide—always within a phase—with $SU(1,1)$ $3j$-coefficients in which at least one j is replaced by $-j-1$.

All Φ_p's, Φ_n's are acceptable generalizations; they coincide essentially with the GWC of Ref. 2 and with the $3j$-symbol of Ref. 3 provided the corresponding variables are restricted to R through a suitable limiting procedure.

4. SYMMETRIES

To study the symmetries of Φ_p's, Φ_n's it is convenient to consider a suitable linear combination $\Phi(\mathbf{r})$ of these functions. We choose

$$\Phi(\mathbf{r})=\tfrac{1}{6}\sum_{a\in\{1,3,5\}}[\Phi_p(a;\mathbf{r})+\Phi_n(a;\mathbf{r})+\sum_{a\in\{0,2,4\}}[\Phi_p(a;\mathbf{r})-\Phi_n(a;\mathbf{r})]\ . \tag{4.1}$$

As will soon be clear, $|\Phi(\mathbf{r})|=|\Phi_p(a;\mathbf{r})|=|\Phi_n(a;\mathbf{r})|$, $\forall a\in\omega$, $\forall\mathbf{r}\in R$; moreover, the phase of $\Phi(\mathbf{r})$ keeps unchanged the features which are shared by the phases of all Φ_p's, Φ_n's, as will be specified in 4.2. These conditions do not determine $\Phi(\mathbf{r})$ uniquely; the additional degree of freedom might be exploited to produce a function which enjoys other properties shared by $SU(2)$, $SU(1,1)$ $3j$-coefficients. Here we shall not enter into these questions as our purpose is merely to unify the analysis of the symmetries of $\Phi_{p,n}$'s.

4.2. *Lemma*: $\Phi(\mathbf{r})=\exp\{i\pi\Psi(\mathbf{r})\}\varphi(\sigma;\mathbf{r})$ if $\mathbf{r}\in R_\sigma$, where

$$\Psi(\mathbf{r})=3\sum_{a\in\{1,3,5\}}r_a,\ \text{mod 2, if } |\sigma\cap\{1,3,5\}|=0,$$
$$=3r_{\sigma\cap\{0,2,4\}},\ \text{mod 2, if } |\sigma\cap\{1,3,5\}|=1,$$
$$=3\sum_{a\in\sigma\cap\{1,3,5\}}r_a+1,\ \text{mod 2, if } |\sigma\cap\{1,3,5\}|=2,$$
$$=0,\ \text{mod 2. if } |\sigma\cap\{1,3,5\}|=3.$$

These formulas can be checked by direct computation of $\Phi(\mathbf{r})$ by means of Eq. (3.14) and 3.12.

4.3. *Remark*: It is clear that through Eq. (4.1) we made a sort of average over the phases of Φ_p's, Φ_n's, without changing the phase when $\mathbf{r}\in R_{135}$; therefore, $\Phi(\mathbf{r})$ coincides with the $3j$ of $SU(2)$ in R_{135}.

4.4. *Definition*: We consider now a group $G(3j)$ of automorphisms P_{12} of R onto itself, provided by the following transformation: (1) permutations P^+_{12} of components of $\mathbf{r}^{(1)}\in R$ which send $\mathbf{r}^{(1)}$ into a different point $\mathbf{r}^{(2)}_+\in R$; (2) permutations of components of $\mathbf{r}^{(1)}$ and sign inversion of all $r^{(1)}_a$, $\forall a\in\omega$; these transformations, which we denote with P^-_{12}, send $\mathbf{r}^{(1)}$ into a point $\mathbf{r}^{(2)}_-\in R$. Clearly, $G(3j)\sim S_6\otimes S_2$ and its order is $6!\cdot 2$ $(=1440)$.

4.5. *Remark*: It is worth noticing that the elements of $G(3j)$ can be partitioned into 20 subsets of $3!\cdot 3!\cdot 2$ $(=72)$ elements each. In fact, if $\mathbf{r}^{(1)}\in R_{\sigma_1}$, there are 72 different $P_{12}\in G(3j)$ which send $\mathbf{r}^{(1)}$ into correspondingly different points $\mathbf{r}^{(2)}\in R_{\sigma_2}$, for any given σ_2. For instance, if $\sigma_1=\{1,3,5\}$, we have $\mathbf{r}^{(2)}=(r^{(2)}_a,r^{(2)}_b,r^{(2)}_c,r^{(2)}_d,r^{(2)}_e,r^{(2)}_f)\in R_{abc}$, provided we set either $(r^{(2)}_a,r^{(2)}_b,r^{(2)}_c)=S_3(r^{(1)}_1,r^{(1)}_3,r^{(1)}_5)$, $(r^{(2)}_d,r^{(2)}_e,r^{(2)}_f)=S_3(r^{(1)}_0,r^{(1)}_2,r^{(1)}_4)$, or $(r^{(2)}_a,r^{(2)}_b,r^{(2)}_c)=S_3(-r^{(1)}_0,-r^{(1)}_2,-r^{(1)}_4)$, $(r^{(2)}_d,r^{(2)}_e,r^{(2)}_f)=S_3(-r^{(1)}_1,r^{(1)}_3,-r^{(1)}_5)$, where S_3 denotes any element of the permutation group over 3 objects and $(u,v,w,\cdots)$ denotes an ordered set. In the particular case $\sigma_1=\sigma_2=\{1,3,5\}$, recalling 2.11, we see that the corresponding 72 elements of $G(3j)$ coincide with the symmetries of the $SU(2)$ $3j$-coefficient.

4.6. *Remark*: If $\mathbf{r}^{(1)}\in R_{\sigma_1}$ and $\mathbf{r}^{(2)}=P_{12}\mathbf{r}^{(1)}\in R_{\sigma_2}$, we want to point out that (1) $P_{12}=P^+_{12}$ implies $r^{(2)}_a=r^{(1)}_b$, so that if $a\in\sigma_2$, then $b\in\sigma_1$ and if $a\in\sigma_2'$ then $b\in\sigma_1'$; (2) $P_{12}=P^-_{12}$ implies $r^{(2)}_a=-r^{(1)}_b$, so that, if $a\in\sigma_2$, then $b\in\sigma_1'$, while if $a\in\sigma_2'$, then $b\in\sigma_1$. These remarks stem directly from Definition 2.4.

4.7. *Theorem*: $\forall\mathbf{r}^{(1)}\in R_{\sigma_1}$, $\forall P^\pm_{12}\in G(3j)$, setting $\mathbf{r}^{(2)}_\pm=P^\pm_{12}\mathbf{r}^{(1)}$, we have $\Phi(\mathbf{r}^{(2)}_\pm)=\exp\{i\pi\Psi^\pm(\mathbf{r}^{(1)},\mathbf{r}^{(2)}_\pm)\}\Phi(\mathbf{r}^{(1)})$, where $\Psi^+(\mathbf{r}^{(1)},\mathbf{r}^{(2)}_+)=\Psi(\mathbf{r}^{(2)}_+)-\Psi(\mathbf{r}^{(1)})$, and $\Psi^-(\mathbf{r}^{(1)},\mathbf{r}^{(2)}_-)=\Psi(\mathbf{r}^{(2)})-\Psi(\mathbf{r}^{(1)})+\alpha_{\sigma_1}(\mathbf{r}^{(1)})$.

Proof: (1) P^+_{12}: According to 4.6, choose any $a\in\sigma_2$ and $b\in\sigma_1$ such that $(\mathbf{r}^{(2)}_+)_a=r^{(1)}_b$; from 4.2 we have $\Phi(\mathbf{r}^{(2)}_+)=\exp\{i\pi\Psi(\mathbf{r}^{(2)}_+)\}\varphi(\sigma_2;\mathbf{r}^{(2)}_+)$. Recalling 2.10 and Eq. (3.5b), we deduce

$$\varphi(\sigma_2;\mathbf{r}^{(2)}_+)=\varphi_n(a;\mathbf{r}^{(2)}_+)$$
$$=\exp\{i\pi[\tfrac{3}{2}(\mathbf{r}^{(2)}_+)_a-\tfrac{1}{4}+\alpha_{\sigma_2'}(\mathbf{r}^{(2)}_+)]\}Q_n(a;\mathbf{r}^{(2)}_+)F_n(a;\mathbf{r}^{(2)}_+);$$

then it is easily checked that $\alpha_{\sigma_2'}(\mathbf{r}^{(2)}_+)=\alpha_{\sigma_1'}(\mathbf{r}^{(1)})$, $Q_n(a;\mathbf{r}^{(2)}_+)=Q_n(b;\mathbf{r}^{(1)})$, $F_n(a;\mathbf{r}^{(2)}_+)=F_n(b;\mathbf{r}^{(1)}$, where the last identity stems from Eqs. (1.10). Therefore $\varphi(\sigma_2;\mathbf{r}^{(2)}_+)=\varphi(\sigma_1;\mathbf{r}^{(1)})=\exp\{-i\pi\Psi(\mathbf{r}^{(1)})\}\Phi(\mathbf{r}^{(1)})$.

(2) P^-_{12}: Choose any $a\in\sigma_2$, $b\in\sigma_1'$ such that $(\mathbf{r}^{(2)}_-)_a$

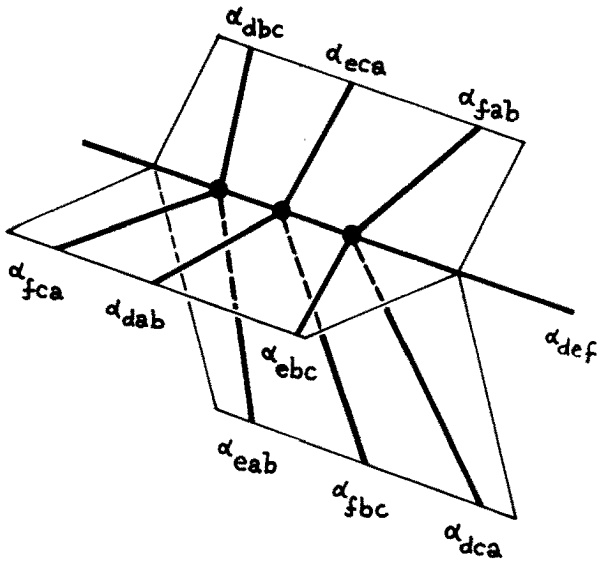


FIG. 1. Geometric representation of Regge symbol $[a, b, c]$.

$= -r_b^{(1)}$; arguing as in (1), we have now $Q_n(a;\mathbf{r}_-^{(2)})$ $= Q_p(b;\mathbf{r}^{(1)})$ and, recalling Eqs. (1.10), $F_n(a;\mathbf{r}_-^{(2)})$ $= F(-\beta_{cq}(-\mathbf{r}_-^{(2)}), \forall c \in \omega - a) = F(-\beta_{cb}(\mathbf{r}^{(1)}), \forall c \in \omega - b)$ $= F_p(b;\mathbf{r}^{(1)})$. Therefore, taking into account Eq. (3.5a), and noting that by virtue of 4.6 $\alpha_{\sigma_2'}(\mathbf{r}_-^{(2)}) = \frac{1}{2} + \sum_{c\in\sigma_2'}(\mathbf{r}_-^{(2)})_c$ $= \frac{1}{2} - \sum_{d\in\sigma_1} r_d^{(1)} = 1 - \alpha_{\sigma_1}(\mathbf{r}^{(1)})$, we get $\varphi(\sigma_2;\mathbf{r}_-^{(2)}) = \varphi_n(a;\mathbf{r}_-^{(2)})$ $= \exp[i\pi\alpha_{\sigma_1}(\mathbf{r}^{(1)})]\varphi_p(b;\mathbf{r}^{(1)}) = \exp[i\pi[\alpha_{\sigma_1}(\mathbf{r}^{(1)}) - \Psi(\mathbf{r}^{(1)})]]$ $\times\Phi(\mathbf{r}^{(1)})$.

4.8. *Corollary*: For any given σ, the numerical tabulation of $\Phi(\mathbf{r})$, $\mathbf{r}\in R_\sigma$, provides the complete tabulation of $\Phi(\mathbf{r})$, $\forall\mathbf{r}\in R$.

This follows from 4.7 and from the remark that $\forall\mathbf{r}^{(2)}$ $\in R_{\sigma_2}$, there are $\mathbf{r}^{(1)}\in R_\sigma$, $P_{12}\in G(3j)$ such that $\mathbf{r}^{(2)}$ $= P_{12}\mathbf{r}^{(1)}$.

In particular, the existing tables of $SU(2)$ $3j$-coefficients cover a portion not only of R_{135}, but of every R_σ, as, taking any $\mathbf{r}^{(1)}\in R_{135}$, we have $\forall P_{12}^{\pm}\in G(3j)$:

$$\Phi(\mathbf{r}_\pm^{(2)} = P_{12}^{\pm}\mathbf{r}^{(1)}) = \exp\{i\pi\Psi^{\pm}(\mathbf{r}^{(1)}, \mathbf{r}_\pm^{(2)})\}\begin{pmatrix} j_1 & j_2 & j_3 \\ \mu_1 & \mu_2 & \mu_3 \end{pmatrix},$$

where j's, μ's are given in terms of $r^{(1)}$ by Eqs. (2.1′). For the sake of possible convenience, we quote the phases $\Psi^{\pm}$(mod 2) in the following table for each possible σ_2, $r_\pm^{(2)}\in R_{\sigma_2}$:

σ_2	$\Psi^+(\mathbf{r}^{(1)}, \mathbf{r}_+^{(2)})$	$\Psi^-(\mathbf{r}^{(1)}, \mathbf{r}_-^{(2)})$
$\{1,3,5\}$	0	$j_1+j_2+j_3$
$\{a,3,5\}$	$j_2-j_3+\mu_1$	$j_1+\mu_2-\mu_3$
$\{a,5,1\}$, $a\in\{0,2,4\}$	$j_3-j_1+\mu_2$	$j_2+\mu_3-\mu_1$
$\{a,1,3\}$	$j_1-j_2+\mu_3$	$j_3+\mu_1-\mu_2$
$\{0,2,a\}$	$\mu_1-\mu_2-j_3-\frac{1}{2}$	$j_1-j_2-\mu_3+\frac{1}{2}$
$\{2,4,a\}$, $a\in\{1,3,5\}$	$\mu_2-\mu_3-j_1-\frac{1}{2}$	$j_2-j_3-\mu_1+\frac{1}{2}$
$\{4,0,a\}$	$\mu_3-\mu_1-j_2-\frac{1}{2}$	$j_3-j_1-\mu_2+\frac{1}{2}$
$\{0,2,4\}$	$-j_1-j_2-j_3-\frac{3}{2}$	$\frac{1}{2}$

We notice that the phases Ψ^+ which belong to $\sigma_2=\{0,2,4\}$, $\{0,3,4\}$, $\{0,3,5\}$ coincide respectively with those of formulas (18.4), (18.5), (18.6) of Ref. 6, apart from a possible ±1 discrepancy whose origin can be traced back to our particular definition (4.1).

By restricting the results of 4.7 to $R(SU(1,1))$, it is possible to classify the symmetries of $SU(1,1)$ $3j$-coefficients involving discrete unitary representations. Here we count them, while in Appendix C a particular phase definition of the $3j$ of $SU(1,1)$ is proposed and its symmetries are characterized explicitly.

4.9. *Corollary*: $\forall\mathbf{r}^{(1)}\in R(SU(1,1))$ there are in general 108 distinct points ($\mathbf{r}^{(1)}$ included) $\mathbf{r}\in R(SU(1,1))$ such that $|\Phi(\mathbf{r})| = |\Phi(\mathbf{r}^{(1)})|$.

In fact, by virtue of 4.5 and 4.7, there are $72\cdot12$ distinct points in the collection of the 12 regions R_σ considered in 2.16 where Φ has the same value within a phase; for every one of these points, which does fulfill $SU(1,1)$ constraints, there exist in general seven others which do not, because they are obtained from the given one through $G(3j)$ operations which change the sign of at least one $\hat{j}$ parameter, as remarked in 2.15. Therefore, the symmetries of the $3j$ of $SU(1,1)$ with generic discrete unitary representations are in number of $72\cdot12/8$ $= 108$; some of them have been pointed out in Ref. 3.

5. A GEOMETRICAL REALIZATION OF REGGE SYMBOL AND ITS EXTENSION

We conclude our discussion by presenting a possible generalization of Regge square-symbol. As explained in Sec. 2, to each $R_\sigma\equiv R_{abc}$ we associate the square-symbol of Eq. (2.13) whose structure is characterized by the property that the sum of each column- or row-elements with $\alpha_{\{a,b,c\}'} - 1$ equals -2; these linear relations are the only ones which involve $\alpha_{\{a,b,c\}'}$ inside the class $\mathcal{Q}_1$ of all possible relations defined by Eq. (1.5). Our aim is to find a structure which takes into account all relations of $\mathcal{Q}_1$ in a way similar to that of Regge symbol, so that the elements of $G(3j)$ correspond to those permutations of elements of the structure which map $\mathcal{Q}_1$ onto itself. Clearly in $\mathcal{Q}_1$ there are altogether $20\cdot6/4$ $= 30$ elements.

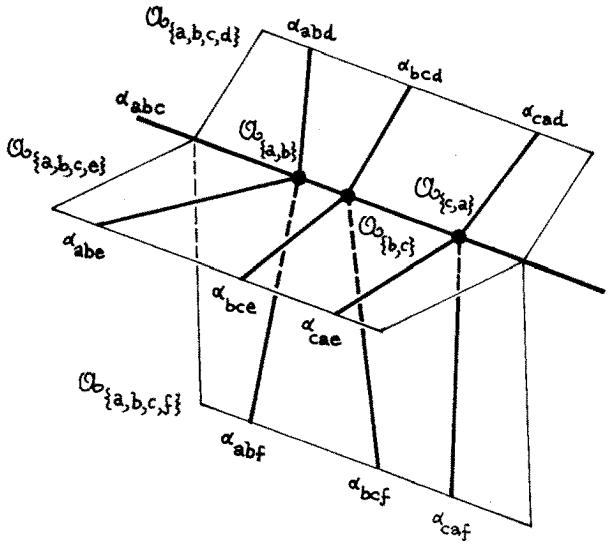


FIG. 2. Geometric representation of the modified Regge square-symbol $\langle d, e, f\rangle$.

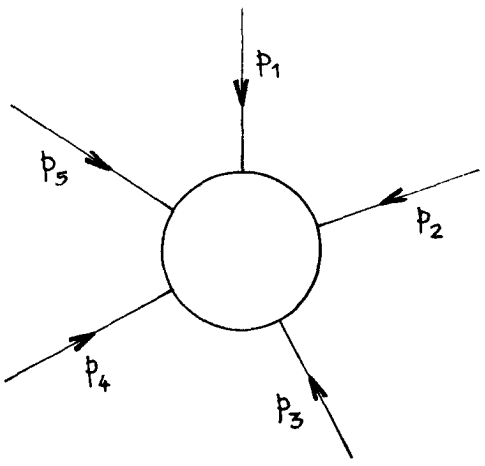


FIG. 3. Kinematic variables of 5-point dual amplitude.

The looked for structure turns out to be a configuration Ω of 15 points, 15 planes, and 20 straight lines in a three-dimensional Euclidean space, characterized by:

(1) a one-to-one correspondence λ_1 exists between straight lines and the 20 α_σ's, and between points and planes of the configuration and elements of $\mathcal{Q}_1$.

(2) If an element of $\mathcal{Q}_1$ corresponds to a point (plane), the latter belongs to (contains) the four straight lines associated to the α's which are involved in the given relation.

(3) Each line contains three points and belongs to three planes.

Clearly, the Regge square-symbol $[a,b,c]$ corresponds under λ_1 to the portion of Ω shown in Fig. 1 where points, planes correspond to the rows, respectively columns of $[a,b,c]$, and $\omega=\{a,b,c,d,e,f\}$.

To characterize the whole configuration and prove that it does exist, let us define a modified square-symbol $\langle d,e,f\rangle$ associated with each $[a,b,c]$:

$$\langle d,e,f\rangle=\begin{Vmatrix} \alpha_{abd}-1 & \alpha_{abe}-1 & \alpha_{abf}-1 \\ \alpha_{bcd}-1 & \alpha_{bce}-1 & \alpha_{bcf}-1 \\ \alpha_{cad}-1 & \alpha_{cae}-1 & \alpha_{caf}-1 \end{Vmatrix}. \tag{5.1}$$

It is easily checked that the 72 symmetries of $[a,b,c]$ are preserved by $\langle d,e,f\rangle$; namely, induced permutations of rows, columns and exchange rows with columns of $\langle d,e,f\rangle$. Furthermore, the structure of $\langle d,e,f\rangle$ suggests we define the collection $\mathcal{Q}_2$ of the 15 sets $\mathcal{A}_{\{a,b\}}=\{\alpha_\chi:\{a,b\}\subset\chi\}$ and of the other 15 sets $\mathcal{A}_{\omega-\{a,b\}}=\{\alpha_\chi:\chi\subset\omega-\{a,b\}\}$, $\forall a\neq b\in\omega$. In this way we see that, while we had to add $\alpha_{\{a,b,c\}'}-1$ to each row- or column-sum of $[a,b,c]$ to characterize elements of $\mathcal{Q}_1$, in the case of any $\langle d,e,f\rangle$ we have to associate $\alpha_{\{d,e,f\}'}$ with each row or column of $\langle d,e,f\rangle$ to find elements of $\mathcal{Q}_2$; for instance, the rows of $\langle d,e,f\rangle$ characterize respectively $\mathcal{A}_{\{a,b\}}$, $\mathcal{A}_{\{b,c\}}$, $\mathcal{A}_{\{c,a\}}$ and its columns: $\mathcal{A}_{\omega-\{e,f\}}$, $\mathcal{A}_{\omega-\{f,d\}}$, $\mathcal{A}_{\omega-\{d,e\}}$. Therefore, in the same spirit as we represented $[a,b,c]$ geometrically in Fig. 1, we may represent $\langle d,e,f\rangle$ in Fig. 2 where points correspond to rows, and planes to columns of $\langle d,e,f\rangle$.

This geometrical representation can be generalized as follows. In the five-dimensional space X of the homogeneous real variables $(x_0,x_1,x_2,x_3,x_4,x_5)$, let us define the linear varieties

$$\bar{V}_{\{a,b,c\}}=\{x_a=x_b=x_c\}, \tag{5.2a}$$

$$\bar{V}_{\{a,b\}}=\{x_c=x_d=x_e=x_f\}, \tag{5.2b}$$

$$\bar{V}_{\omega-\{a,b\}}=\{x_a=x_b\}, \tag{5.2c}$$

$\forall\, a,b,c\in\omega=\{0,1,2,3,4,5\}\equiv\{a,b,c,d,e,f\}$. Then consider any two-dimensional linear variety T in generic position with respect to all $\bar{V}$'s and define

$$V=\bar{V}\cap T,\quad \forall\bar{V}. \tag{5.3}$$

Of course, $V_{\{a,b,c\}}$ are straight lines, while $V_{\{a,b\}}$ and $V_{\omega-\{a,b\}}$ are respectively points and planes in $X\cap T$. These elements form the configuration Ω whose inclusion properties[19] are

$$V_{\{a,b\}}\subset V_{\{a,b,s\}},\ V_{\{a,b,s,t\}}\quad \forall s,t\in\omega-\{a,b\}, \tag{5.4a}$$

$$V_{\{p,q\}}\subset V_{\{a,b,c\}}\subset V_{\omega-\{p,q\}}\quad \forall p,q\in\{a,b,c\}, \tag{5.4b}$$

$$V_{\omega-\{a,b\}}\supset V_{\{s,t\}},\ V_{\{a,b,s\}}\quad \forall s,t\in\omega-\{a,b\}. \tag{5.4c}$$

Therefore, the one-to-one map $\lambda_2:\Omega\to\{\alpha_\sigma\},\mathcal{Q}_2$ defined by

$$\begin{aligned} &V_{\{a,b,c\}}\to\alpha_{abc},\quad \forall a,b,c\in\omega,\\ &V_{\{a,b\}}\to\mathcal{A}_{\{a,b\}},\\ &V_{\omega-\{a,b\}}\to\mathcal{A}_{\omega-\{a,b\}} \end{aligned} \tag{5.5}$$

provides the generalization of the modified square-symbol, as represented in Fig. 2. On the other hand, if $\lambda:\{\alpha_\sigma\}\to\{\alpha_\sigma\}$ is the map defined by

$$\begin{aligned} &\langle 0,2,4\rangle\to[1,3,5],\\ &\alpha_{135}\to\alpha_{024}, \end{aligned} \tag{5.6a}$$

and by the constraint

$$\lambda(\alpha_{\sigma_1})=\alpha_{\sigma_2}\iff\lambda(\alpha_{\sigma_1'})=\alpha_{\sigma_2'},$$

then $\lambda_1=\lambda\lambda_2:\Omega\to\{\alpha_\sigma\},\mathcal{Q}_1$ is the one-to-one correspondence introduced in (1), (2), (3) which yields a generalization of the Regge square-symbol as represented in Fig. 1.

From the previous discussion it follows that the elements P_{12}^+ of $G(3j)$ permute lines, points, planes of Ω respectively among themselves, while the elements P_{12}^- also exchange points with planes.

ACKNOWLEDGMENTS

We like to thank Professor T. Regge for his kind interest, and Dr. F. Gliozzi and Dr. S. Sciuto for useful discussions.

APPENDIX A: CONNECTION WITH VENEZIANO FIVE-POINT AMPLITUDE AND PROOF OF (1.8)

To describe the scattering of 5 spinless neutral particles (Fig. 3) with 4-momenta p_i, $i=1,2,\ldots,5$, the following amplitude has been proposed recently[20]:

$$\begin{aligned} &B(\gamma_{12},\gamma_{23},\gamma_{34},\gamma_{45},\gamma_{51})\\ &\quad=\int_0^1 dx\int_0^1 dy\, x^{-\gamma_{12}-1}y^{-\gamma_{45}-1}(1-x)^{-\gamma_{23}-1}(1-y)^{-\gamma_{34}-1}\\ &\qquad\times(1-xy)^{-\gamma_{15}+\gamma_{23}+\gamma_{34}}, \end{aligned} \tag{A1}$$

where

$$\gamma_{ij}=\gamma_{ji}=1+(p_i+p_j)^2=-1+2p_i\cdot p_j\ \forall i\neq j=1,\dots,5, \tag{A2}$$

having chosen the metric $p_i^2=-1$. As energy-momentum conservation yields

$$\sum_{i=1}^{5}p_i=0, \tag{A3}$$

only 5 γ_{ij} are linearly independent and $\gamma_{i,i+1}$ ($\gamma_{5,5+1}=\gamma_{51}$) have been selected in (A1). The main properties of B can be summarized as follows:

A4. $B(\gamma_{12},\gamma_{23},\gamma_{34},\gamma_{45},\gamma_{51})$ is invariant under any cyclical and anticyclical permutations of labels $1,\dots,5$; this property reflects the symmetry of Fig. 3.

A5. The only singularities of B are simple poles at nonnegative integer values of each $\gamma_{i,i+1}$; moreover, B has poles in $\gamma_{i,i+1}$ and $\gamma_{j,j+1}$ if and only if $\{i,i+1\}\cap\{j,j+1\}=\emptyset$. This property is known as "duality."

A6. The integral (A1) can be written in terms of the F function introduced in Eq. (1.6) as follows:

$$\begin{aligned}&B(\gamma_{12},\gamma_{23},\gamma_{34},\gamma_{45},\gamma_{51})\\&\quad=\Gamma(-\gamma_{12})\Gamma(-\gamma_{23})\Gamma(-\gamma_{34})\Gamma(-\gamma_{45})\Gamma(-\gamma_{51})\\&\quad\times F(\gamma_{51}+\gamma_{12},\gamma_{12}+\gamma_{23},\gamma_{23}+\gamma_{34},\gamma_{34}+\gamma_{45},\gamma_{45}+\gamma_{51}).\end{aligned}$$

Therefore, from A5 it follows that F is an entire function of each variable $\gamma_{i-1,i}+\gamma_{i,i+1}$.

APPENDIX B: PROOF OF CONSISTENCY OF EQS. (3.6)

B1. We show that, given any $a,b\in\omega$, $a\neq b$, $\mathbf{r}\in R_\sigma$, then $\tilde F_p(b;\mathbf{r})/\tilde F_n(a;\mathbf{r})$ computed from Eq. (3.6) does not depend upon any τ such that $b\in\tau$, $a\in\tau'$, $|\tau\cap\sigma|=2,3$. In fact, we have

$$\frac{\prod_{c\in\tau}\tilde\Gamma(\alpha_{\{\tau-c\}\cup a}(\mathbf{r}))}{\prod_{c\in\tau'}\tilde\Gamma(\alpha_{\{\tau-b\}\cup c}(\mathbf{r}))}=\frac{\prod_{c\in\tau-b}\tilde\Gamma(\alpha_{abc}(\mathbf{r}))}{\prod_{c\in\tau'-a}\tilde\Gamma(\alpha_{\{\tau-b\}\cup c}(\mathbf{r}))},$$

and recalling Eq. (3.3)

$$\begin{aligned}\Big\{\prod_{c\in\tau'-a}\tilde\Gamma(\alpha_{\{\tau-b\}\cup c}(\mathbf{r}))\Big\}^{-1}&=\prod_{c\in\tau'-a}\tilde\Gamma(\alpha_{abc}(\mathbf{r}))\\&\times\exp\{i\pi[\vartheta(\{a,b,c\};\mathbf{r})\alpha_{abc}(\mathbf{r})+\vartheta(\{a,b,c\}';\mathbf{r})\alpha_{\{a,b,c\}'}(\mathbf{r})\}\\&=\Big\{\prod_{c\in\tau'-a}\tilde\Gamma(\alpha_{abc}(\mathbf{r}))\Big\}\\&\times\exp\Big\{i\pi\prod_{c\in\tau'-a}[\vartheta(\{a,b,c\};\mathbf{r})+\alpha_{\{a,b,c\}'}(\mathbf{r})]\Big\}\end{aligned}$$

because the following identity holds mod 2:

$$\vartheta(\chi;\mathbf{r})\alpha_\chi(\mathbf{r})+\vartheta(\chi';\mathbf{r})\alpha_{\chi'}(\mathbf{r})=\vartheta(\chi;\mathbf{r})+\alpha_{\chi'}(\mathbf{r}). \tag{B1a}$$

Then by direct computation we obtain

$$\begin{aligned}&\alpha_{\tau'}+\prod_{c\in\tau'-a}\alpha_{\{a,b,c\}'}=3r_a+\tfrac12\ \text{mod 2},\\&\sum_{c\in\tau'-a}\vartheta(\{a,b,c\})=|\tau\cap\sigma|+\delta_{a\in\sigma'}\delta_{b\in\sigma}\ \text{mod 2}.\end{aligned} \tag{B1b}$$

Therefore

$$\frac{\tilde F_p(b;\mathbf{r})}{\tilde F_n(a;\mathbf{r})}=\exp\{i\pi(3r_a+\tfrac12+\delta_{a\in\sigma'}\delta_{b\in\sigma})\}\prod_{\chi\supset\{a,b\}}\tilde\Gamma(\alpha_\chi(\mathbf{r})),$$

which is independent of τ. Similarly, we may prove the inverse relation

$$\frac{\tilde F_n(a;\mathbf{r})}{\tilde F_p(c;\mathbf{r})}=\exp\{i\pi(3r_c-\tfrac12+\delta_{c\in\sigma}\delta_{a\in\sigma'})\}\prod_{\chi\subset\omega-\{a,c\}}\tilde\Gamma(\alpha_\chi(\mathbf{r})). \tag{B1c}$$

B2. We show that, $\forall a,b,c\in\omega$, $a\neq b\neq c\neq a$, the ratio $\tilde F_p(b;\mathbf{r})/\tilde F_p(c;\mathbf{r})|_a=\{\tilde F_p(b;\mathbf{r})/\tilde F_n(a;\mathbf{r})\}\cdot\{\tilde F_n(a;\mathbf{r})/\tilde F_p(c;\mathbf{r})\}$ does not depend on a. In fact using Eqs. (3.3), (B1a), we get

$$\begin{aligned}&\tilde F_p(b;\mathbf{r})/\tilde F_p(c;\mathbf{r})|_a\\&\quad=\exp\{i\pi[3r_a+3r_c+\delta_{a\in\sigma'}(\delta_{b\in\sigma}+\delta_{c\in\sigma})]\}\\&\quad\times\exp\{i\pi[\vartheta(\{a,b,c\};\mathbf{r})+\alpha_{\omega-\{a,b,c\}}(\mathbf{r})\}\prod_{b\in\chi\subset\omega-c}\tilde\Gamma(\alpha_\chi(\mathbf{r})).\end{aligned}$$

We notice that $\vartheta(\{a,b,c\})=\delta_{b\in\sigma}\delta_{c\in\sigma}+\delta_{a\in\sigma}(\delta_{b\in\sigma}\delta_{c\in\sigma'}+\delta_{b\in\sigma'}\delta_{c\in\sigma})$, mod 2, so that $\delta_{a\in\sigma'}(\delta_{b\in\sigma}+\delta_{c\in\sigma'})+\vartheta(\{a,b,c\})=1+\delta_{b\in\sigma'}\delta_{c\in\sigma'}$, mod 2. Therefore,

$$\frac{\tilde F_p(b;\mathbf{r})}{\tilde F_p(c;\mathbf{r})}=\exp\{i\pi(3r_b-\tfrac12+\delta_{b\in\sigma'}\delta_{c\in\sigma'})\}\prod_{b\in\chi\subset\omega-c}\tilde\Gamma(\alpha_\chi(\mathbf{r})). \tag{B2a}$$

Similarly, we obtain

$$\frac{\tilde F_n(b;\mathbf{r})}{\tilde F_n(c;\mathbf{r})}=\exp\{i\pi(3r_b+\tfrac12+\delta_{b\in\sigma}\delta_{c\in\sigma})\}\prod_{c\in\chi\subset\omega-b}\tilde\Gamma(\alpha_\chi(\mathbf{r})) \tag{B2b}$$

B3. Finally, we show that $\forall b,c\in\omega$, $b\neq c$, $\{\tilde F_n(c;\mathbf{r})/\tilde F_p(b;\mathbf{r})\}\cdot\{\tilde F_p(b;\mathbf{r})/\tilde F_p(c;\mathbf{r})\}$ is independent of b. In fact, by means of Eqs. (B1c), (B2a), (3.5c), this expression takes the form

$$\exp\{i\pi[(3r_b-\tfrac12+\delta_{b\in\sigma}\delta_{c\in\sigma'})+(3r_b-\tfrac12+\delta_{b\in\sigma'}\delta_{c\in\sigma'})]\}Q_p(c;\mathbf{r});$$

therefore

$$\frac{\tilde F_n(c;\mathbf{r})}{\tilde F_p(c;\mathbf{r})}=\exp(i\pi\delta_{c\in\sigma'})Q_p(c;\mathbf{r}), \tag{B3a}$$

and, similarly

$$\frac{\tilde F_p(c;\mathbf{r})}{\tilde F_n(c;\mathbf{r})}=\exp(i\pi\delta_{c\in\sigma})Q_n(c;\mathbf{r}). \tag{B3b}$$

APPENDIX C: SYMMETRIES OF THE $SU(1,1)$ $3j$ COEFFICIENT WITH DISCRETE UNITARY REPRESENTATIONS

To catalog the symmetries of the $SU(1,1)$ $3j$-coefficient with unitary discrete representations, it is convenient to use the cycle notation to represent permutations of point-vectors $\mathbf{r}$, $\mathbf{r}^{(1)}$, etc. in R; for instance, the operators $P_{12}^{\pm}=\pm(052)(34)\in G(3j)$ send $\mathbf{r}^{(1)}\in R$ into $\mathbf{r}_\pm^{(2)}\in R$, where $(r_0^{(2)})_\pm=\pm r_5^{(1)}$, $(r_1^{(2)})_\pm=\pm r_1^{(1)}$, $(r_2^{(2)})_\pm=\pm r_0^{(1)}$, $(r_3^{(2)})_\pm=\pm r_4^{(1)}$, $(r_4^{(2)})_\pm=\pm r_3^{(1)}$, $(r_5^{(2)})_\pm=\pm r_2^{(1)}$.

We define the $SU(1,1)$ $3j$-coefficient as follows:

$$\begin{pmatrix}j_1&j_2&j_3\\\mu_1&\mu_2&\mu_3\end{pmatrix}_{SU(1,1)}=i\Phi(\mathbf{r}),\quad\forall\mathbf{r}\in R(SU(1,1)). \tag{C1}$$

Recalling 3.15, 4.2, we obtain the phase relative to the corresponding Clebsch–Gordan coefficients of Ref. 2:

$$\begin{pmatrix}j_1&j_2&j_3\\\mu_1&\mu_2&\mu_3\end{pmatrix}_{SU(1,1)}=\exp(i\pi\rho)\frac{C^{j_1\ j_2\ j_3}_{\mu_1\ \mu_2-\mu_3}}{\sqrt{-2j_3-1}}, \tag{C2}$$

where

$$\rho=j_3+\mu_1-\mu_2\quad\text{if }\mathbf{r}\in\hat R_{012}\cup\hat R_{023}, \tag{C3}$$

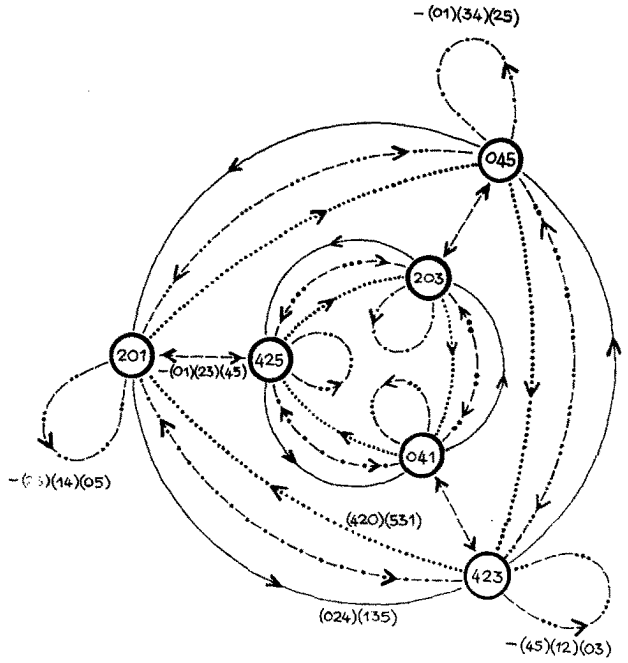


FIG. 4. Usual (non-Regge) symmetries of the 3j of $SU(1,1)$.

$$\rho = j_1 - j_2 - \mu_3 \quad \text{if } \mathbf{r} \in \hat{R}_{234} \cup \hat{R}_{245} \cup \hat{R}_{045} \cup \hat{R}_{014}.$$

We like definition (C1) because the usual (non-Regge) symmetries of the $SU(2)$ 3j-coefficient hold also in phase in the $SU(1,1)$ case.

In fact, these symmetries read

$$\begin{pmatrix} j_1 & j_2 & j_3 \\ \mu_1 & \mu_2 & \mu_3 \end{pmatrix}_{SU(1,1)} = (-1)^J \begin{pmatrix} j_{i1} & j_{i2} & j_{i3} \\ \mu_{i1} & \mu_{i2} & \mu_{i3} \end{pmatrix}_{SU(1,1)}$$

$$= (-1)^{j_1+j_2+j_3} \begin{pmatrix} j_1 & j_2 & j_3 \\ -\mu_1 & -\mu_2 & -\mu_3 \end{pmatrix}_{SU(1,1)}$$

where $J=0$, $J=j_1+j_2+j_3$ if (i_1,i_2,i_3) is an even, respectively odd, permutation of $(1,2,3)$; from Eq. (2.1) it is clear that the even permutations of j's are realized by $(024)(135)$, $(420)(531)$, while the odd ones are realized by $-(45)(12)(03)$, $-(01)(25)(34)$, $-(23)(14)(05)$ and the sign inversion of μ's by $-(01)(23)(45)$. All these operators are in $G(3j)$ and send any $\mathbf{r}^{(1)} \in R(SU(1,1))$; into some $\mathbf{r}^{(2)} \in R(SU(1,1))$ the phases are easily checked by means of 4.2, 4.7. It is worth analyzing into which region $\hat{R}_{\sigma_2}$ is sent a point $\mathbf{r}^{(1)} \in \hat{R}_{\sigma_1}$ by these operators. All cases are summarized directed toward vertex σ_2 if, $\forall \mathbf{r}^{(1)} \in \hat{R}_{\sigma_1} : \mathbf{r}^{(2)} = P_{12}\mathbf{r}^{(1)} \in \hat{R}_{\sigma_2}$; the edge is obviously labelled P_{12}. Notice that, for each $\hat{R}_\sigma$, there is a particular P_{12}^- such that $\mathbf{r}^{(2)} = P_{12}^- \mathbf{r}^{(1)} \in \hat{R}_\sigma$ if $\mathbf{r}^{(1)} \in \hat{R}_\sigma$; for instance, $P_{12}^- = -(23)(05)(14)$ if $\sigma = \{0,1,2\}$.

From 4.9 we know that the symmetries (C4) do not exhaust all the invariance properties of the $SU(1,1)$ 3j-coefficient. To complete the analysis, we may search, for any given $\hat{R}_\sigma$, all those elements P_{12} of $G(3j)$ such that, if $\mathbf{r}^{(1)} \in \hat{R}_\sigma$, then $\mathbf{r}^{(2)} = P_{12}\mathbf{r}^{(1)} \in \hat{R}_\sigma$. To this end we notice that, setting $\{a_0, a_2, a_4\} = \{0,2,4\}$, $a_i = a_{i-1} + 1$, $i \in \{1,3,5\}$, the regions $\hat{R}_\sigma$'s are of the type $\hat{R}_{a_0,a_2,a_1}$.

If $\mathbf{r} \in R_{a_0,a_2,a_1}$, according to 2.4, the following pattern p of inequalities holds:

$$p = \begin{matrix} r_{a_0} - r_{a_4} > 0, & r_{a_2} - r_{a_4} > 0, & r_{a_1} - r_{a_4} > 0 \\ r_{a_0} - r_{a_3} > 0, & r_{a_2} - r_{a_3} > 0, & r_{a_1} - r_{a_3} > 0 \\ r_{a_0} - r_{a_5} > 0, & r_{a_2} - r_{a_5} > 0, & r_{a_1} - r_{a_5} > 0 \end{matrix}, \qquad \text{(C5)}$$

and if $\mathbf{r}$ is further restricted to $\hat{R}_{a_0,a_2,a_1}$, we also have

$$r_{a_0} - r_{a_1} \geq -1, \qquad \text{(C6)}$$

$$r_{a_4} - r_{a_5} \geq -1. \qquad \text{(C7)}$$

Clearly the 72 elements of $G(3j)$ which send $\mathbf{r}^{(1)} \in \hat{R}_{a_0,a_2,a_1}$ into $\mathbf{r}^{(2)} \in \hat{R}_{a_0,a_2,a_1}$ permute rows, columns, and rows with columns of p; however, only part of them yield $\mathbf{r}^{(2)}$ which fulfills (C6), (C7) as well. In fact, if $\mathbf{r}^{(2)} = (a_0 a_1)\mathbf{r}^{(1)}$, then $r^{(2)}_{a_0} - r^{(2)}_{a_1} = r^{(1)}_{a_1} - r^{(1)}_{a_0} \leq 1$, so that (C6) is satisfied only in the particular cases $r^{(1)}_{a_0} - r^{(1)}_{a_1} = \pm 1, 0$; similarly, the permutation $(a_4 a_5)$ leads to $\mathbf{r}^{(2)} \in \hat{R}_{a_0,a_2,a_1}$ only if $r^{(1)}_{a_5} - r^{(1)}_{a_4} = \pm 1, 0$.

For the time being, let us restrict ourselves to considering $\mathbf{r}^{(1)}$ satisfying (C5) and

$$r_{a_0} - r_{a_1} \geq 0, \qquad \text{(C6')}$$

$$r_{a_4} - r_{a_5} \geq 0, \qquad \text{(C7')}$$

so that (C6') in general is violated by $\mathbf{r}^{(2)} = (a_0 a_1)\mathbf{r}^{(1)}$, and (C7') by $\mathbf{r}^{(2)} = (a_4 a_5)\mathbf{r}^{(1)}$. We notice that $\mathbf{r}^{(2)} = (a_0 a_2)\mathbf{r}^{(1)}$ satisfies (C6') if $r^{(2)}_{a_0} - r^{(2)}_{a_1} = r^{(1)}_{a_2} - r^{(1)}_{a_1} \geq 0$, while the analogous conditions for the operators $(a_1 a_2)$, $(a_0 a_1 a_2)$, $(a_0 a_2 a_1)$, are, respectively

$$r^{(1)}_{a_0} - r^{(1)}_{a_2} \geq 0, \quad r^{(1)}_{a_1} - r^{(1)}_{a_2} \geq 0, \quad r^{(1)}_{a_2} - r^{(1)}_{a_0} \geq 0.$$

Similar conditions are obtained for the permutations $(a_3 a_4)$, $(a_3 a_5)$, $(a_3 a_5 a_4)$, $(a_3 a_4 a_5)$; they read, respectively

$$r^{(1)}_{a_3} - r^{(1)}_{a_5} \geq 0, \quad r^{(1)}_{a_4} - r^{(1)}_{a_3} \geq 0, \quad r^{(1)}_{a_3} - r^{(1)}_{a_4} \geq 0,$$
$$r^{(1)}_{a_5} - r^{(1)}_{a_3} \geq 0.$$

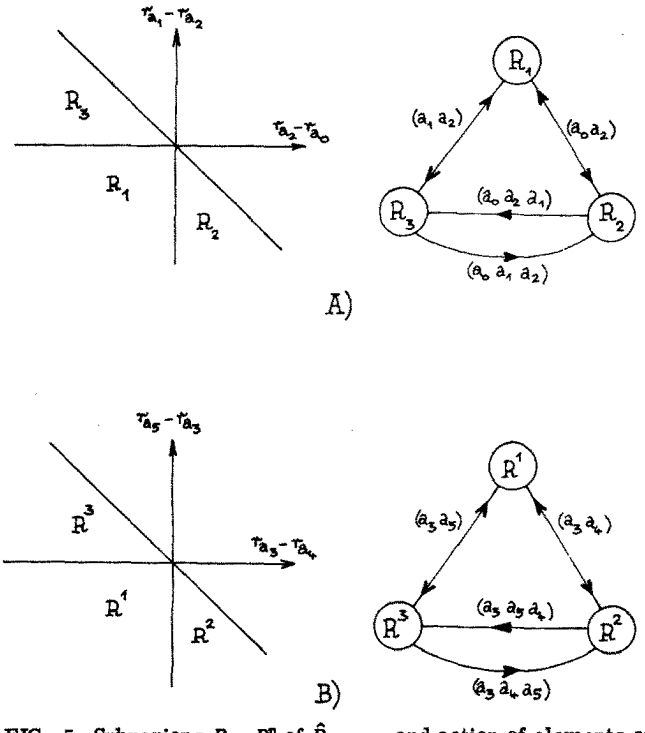


FIG. 5. Subregions R_η, R^η of $\hat{R}_{a_0,a_1,a_2}$ and action of elements of p_η, p^η, $\eta = 1,2,3$.

Therefore, let us call R_η, R^η, $\eta=1,2,3$, the subregions of $\hat{R}_{a_0,a_2,a_1}$ where (C6′), (C7′) are satisfied in addition to the inequalities (Fig. 5):

$$\begin{aligned} &R_1\colon\ r_{a_0}-r_{a_2}\geq 0,\quad r_{a_2}-r_{a_1}\geq 0;\\ &R_2\colon\ r_{a_0}-r_{a_2}\leq 0;\\ &R_3\colon\ r_{a_2}-r_{a_1}\leq 0; \end{aligned} \tag{C8a}$$

$$\begin{aligned} &R^1\colon\ r_{a_4}-r_{a_3}\geq 0,\quad r_{a_3}-r_{a_5}\geq 0;\\ &R^2\colon\ r_{a_4}-r_{a_3}\leq 0;\\ &R^3\colon\ r_{a_3}-r_{a_5}\leq 0. \end{aligned} \tag{C8b}$$

Then we call ρ_η, ρ^η the sets of those permutations over $(r^{(1)}_{a_0}, r^{(1)}_{a_2}, r^{(1)}_{a_1})$, respectively $(r^{(1)}_{a_3}, r^{(1)}_{a_5}, r^{(1)}_{a_4})$, which send $\mathbf{r}^{(1)}\in R_\eta$, respectively $\mathbf{r}^{(1)}\in R^\eta$, $\eta=1,2,3$, into $\mathbf{r}^{(2)}$ which satisfies (C5), (C6′), (C7′). It is easily checked that

$$\begin{aligned} &\rho_1=\{\mathbb{1},(a_0a_2),(a_1a_2)\},\\ &\rho_2=\{\mathbb{1},(a_0a_2),(a_0a_2a_1)\},\\ &\rho_3=\{\mathbb{1},(a_1a_2),(a_0a_1a_2)\}, \end{aligned} \tag{C9a}$$

$$\begin{aligned} &\rho^1=\{\mathbb{1},(a_3a_4),(a_3a_5)\},\\ &\rho^2=\{\mathbb{1},(a_3a_4),(a_3a_5a_4)\},\\ &\rho^3=\{\mathbb{1},(a_3a_5),(a_3a_4a_5)\}, \end{aligned} \tag{C9b}$$

where $\mathbb{1}$ denotes the identical operator. Notice that the elements of ρ_{η_1} commute with those of ρ^{η_2}, so that in $\rho_{\eta_1}\otimes\rho^{\eta_2}$ there are 9 distinct elements and each of them maps any $\mathbf{r}^{(1)}\in R_{\eta_1}\cap R^{\eta_2}$ into distinct $\mathbf{r}^{(2)}\in\hat{R}_{a_0,a_2,a_1}$. From Fig. 4 we realize that, if $\mathbf{r}^{(1)}\in\hat{R}_{a_0,a_2,a_1}$, then

$$\mathbf{r}^{(2)}=-(a_2a_3)(a_0a_5)(a_1a_4)\mathbf{r}^{(1)}\in R_{a_0,a_2,a_1};$$

the 18 operators which map $\mathbf{r}^{(1)}\in R_{\eta_1}\cap R^{\eta_2}$ into 18 different points of $\hat{R}_{a_0,a_2,a_1}$ ($\mathbf{r}^{(1)}$ included) are

$$\{\mathbb{1},-(a_2a_3)(a_0a_5)(a_1a_4)\}\otimes\rho_{\eta_1}\otimes\rho^{\eta_2},\quad \eta_1\eta_2=1,2,3. \tag{C10}$$

We shall not write these symmetries explicitly in terms of j's, μ's, because for this purpose we should consider separately each one of the nine subregions of $\hat{R}_{a_0,a_2,a_1}$; for the same reason we shall not writte the corresponding phases, which—at any rate—can be easily computed by means of 4.2, 4.7.

Finally, we comment briefly on the case in which $\mathbf{r}^{(1)}$ satisfies (C6), (C7); this implies the possibilities $r_{a_0}-r_{a_1}=\pm1,0$, and $r_{a_4}-r_{a_5}=\pm1,0$, so that the operators (a_0a_1), (a_4a_5) may become acceptable symmetries and consequently may modify the picture we have just drawn.[21] However, this remark applies only to a very particular class of $3j$-coefficients which lay on the boundary of $R(SU(1,1))$ and involve the lowest representations; for this reason we feel justified in skipping the corresponding details.

[1] A. Giovannini and M. Verde, Nuovo Cimento 34, 1936 (1964).
[2] W. J. Holman and L. C. Biedenharn, Ann. Phys. 39, 1 (1966).
[3] I. Ferretti and M. Verde, Nuovo Cimento 55A, 110 (1968); M. Verde, *The Coupling Problem for Infinite-Multiplets* (Atti Convegno Mendeleeviano, Torino-Roma, 1969).
[4] W. J. Holman and L. C. Biedenharn, Ann. Phys. 47, 205 (1968).
[5] A. A. Bandzaitis, A. V. Karosene, A. Yu. Savukinas, and A. P. Yucys, Physics-Doklady 9, 139 (1964).
[6] A. P. Yucys and A. A. Bandzaitis, *The Theory of Angular Momentum in Quantum Mechanics* (Mintis, Vilnius, Lithuanian SSR, 1965) (in Russian).
[7] A. P. Yucys, A. Yu. Savukinas, and A. A. Bandzaitis, Lietuvos Fiz. Rinkinys, 5, 171 (1965).
[8] B. M. Minton, J. Math. Phys. 11, 3061 (1970).
[9] E. Yakimiw, J. Math. Phys. 12, 1134 (1971); V. Joshi, J. Math. Phys. 12, 1134 (1971).
[10] F. J. Whipple, Proc. Lond. Math. Soc. (2) 23, 104 (1925).
[11] They have been employed in a similar context in Ref. 3.
[12] A. C. Dixon, Proc. Lond. Math. Soc (2) 2, 8 (1905).
[13] See, for instance, L. J. Slater, *Generalized Hypergeometric Functions* (Cambridge U. P., Cambridge, 1966).
[14] T. Regge, Nuovo Cimento 10, 296 (1958).
[15] M. E. Rose, *Multipole Fields* (Wiley, New York, 1955); see also Ref. 6, formula (19.25).
[16] It is also easily checked that R is strictly contained into the subset of $\mathbb{R}^{(5)}$ where $|r_a+r_b+r_c+\frac{1}{2}|$, $a,b,c\in\omega$, is integer.
[17] S. Sciuto, On the Clebsch—Gordan Theorems for the 3-dimensional Lorentz Group, internal report, Università di Torino, di Torino, 1965 (unpublished).
[18] From this point of view, the particular values $\hat{j}_\xi=0,\ \pm\frac{1}{2}$ are peculiar because they are unaffected by the transformation $\hat{j}_\xi\to-\hat{j}_\xi$, namely $j_\xi\to-j_\xi-1$.
[19] It is interesting to remark that Ω contains several different Desargues configurations [D. Hilbert and S. Cohn-Vossen, *Geometry and the Imagination* (Chelsea, New York, 1952)]. In fact, the elements of Ω whose defining set contains a given index $a\in\omega$ are 10 lines, 5 points, and 5 planes which form a space Desargues configuration; on the other hand, the elements of Ω which do not contain a form the dual Desargues configuration obtained from the first one by exchanging the words "plane," "point."
[20] See, for instance, K. Bardakçi and H. Ruegg, Phys. Lett. 28B, 342 (1968).
[21] Compare also with footnote 18.

Can outside fields destroy black holes?

Petr Hájíček

Institute for Theoretical Physics, University of Berne, Switzerland
(Received 21 March 1974)

Stationary, axisymmetric, asymptotically flat space-times with a black hole surrounded by matter rings, disks, or shells are considered. A certain set of invariant functions, called local invariants, is defined which contains full information about the metric and electromagnetic field in a small electrovacuum neighborhood of the horizon. The local invariants are shown to satisfy an inequality, which is a generalization of the well-known Kerr–Newman inequality $m^2 > a^2 + e^2$, and which places an upper bound on the gravimagnetic, electric, and magnetic field strengths as measured at the surface of the black hole, independently of whether the fields are produced by the black hole itself or originate in outside sources.

1. INTRODUCTION

An excellent tool to study the properties of black holes—at least theoretically—is the simple model of an equilibrium system consisting of a black hole surrounded by some matter configuration and otherwise perfectly isolated from any other influence. The corresponding space–time is, therefore, stationary, axisymmetric, and asymptotically flat. For example, by using this model, Bardeen, Carter, and Hawking were able to extract an important equation governing the energy balance of a black hole.[1] Among other factors the striking resemblance of this equation to the second law of thermodynamics has led to very interesting speculations.[2,3]

On the other hand, the full content of the equilibrium model does not seem to be exhausted. It still offers many problems which are of importance for black hole physics though not directly related to such fascinating fields as "black hole thermodynamics" or even "black hole quantum mechanics." With this in mind, the model was more thoroughly studied in,[4] in particular concerning the structure of the horizon. A promising set of invariant functions, called "local invariants," from which the metric and the electromagnetic field in a small electrovacuum neighbourhood of the horizon can in principle be reconstructed, has been found. We hope to gain new insights into black hole physics by studing properties of these invariants. Some support for these hopes was made in Ref. 5, where simple and symmetric relations have been written between the local invariants and the black hole degrees of freedom (surface area, angular momentum, and charges).

In the present paper, we develop the idea a step further. In Sec. 2, a concise definition of the local invariants is given together with a review of their properties to the extent necessary to understand the new results. These consist in part of a great number of small improvements, interrelations and interpretations and, in part, of an interesting inequality which must be obeyed by the local invariants. The discussion of some implications of the inequality closes the section. All more difficult proofs of the new statements are then presented in Sec. 3.

We use the dimension and sign conventions following Newman and Penrose,[6] because many spin coefficient equations are used. Thus, the dielectric constant ϵ_0 of vacuum, the light velocity c in vacuum, and the gravitation constant G all equal to 1. The signature of the space—time metric is -2 and the definitions of the curvature and Ricci tensors are

$$V_{i;jk} - V_{i;kj} = V_l R^l_{ijk},$$
$$R_{ij} = R^l_{ijl},$$

the semicolon denoting the space—time covariant derivative.

2. LOCAL INVARIANTS

In Ref. 7, the structure of all null hypersurfaces whose generators (= rays) have zero convergence and shear

$$\rho = \sigma = 0 \tag{1}$$

and which possess compact two-dimensional spacelike sections has been investigated. All Killing horizons representing black holes[8] have these properties, but, e.g., the null hyperplanes in Minkowski space—time are excluded, because none of their sections is compact. We call such hypersurfaces "perfect horizons."[9]

Choosing a pseudo-orthonormal tetrad $l^i, n^i, m^i, \bar{m}^i$ in a neighbourhood of a perfect horizon $\mathcal{H}$ such that l^i and m^i are tangential to $\mathcal{H}$ and parallelly propagated along the rays of $\mathcal{H}$, we can write at $\mathcal{H}$

$$\begin{aligned} &l^i{}_{;j} l^j = m^i{}_{;j} l^j = 0, \\ &l^i{}_{;j} m^j = (\bar{\alpha} + \beta) l^i, \\ &m^i{}_{;j} m^j = \bar{\lambda} l^i - (\bar{\alpha} - \beta) m^i, \\ &\bar{m}^i{}_{;j} m^j = \mu l^i + (\bar{\alpha} - \beta) \bar{m}^i. \end{aligned} \tag{2}$$

Here, we use the spin coefficients $\rho, \sigma, \alpha, \beta, \lambda, \mu$ as introduced in Ref. 6 and the relations (1). The following definition shortens many equations

$$\Omega = \bar{\alpha} + \beta, \quad \Gamma = \bar{\alpha} - \beta.$$

The relations (2) imply that the vector $a^i{}_{;j} b^j$ is tangential to $\mathcal{H}$, if a^i and b^i are tangential to $\mathcal{H}$, because all such vector fields are combinations of just l^i, m^i, and $\bar{m}^i$. Thus, (2) defines an intrinsic affine connection on $\mathcal{H}$. In addition, the well-known degenerate metric on $\mathcal{H}$ is defined by

$$\begin{aligned} &(l^i, m^i) = (l^i, l^i) = (m^i, m^i) = 0, \\ &(m^i, \bar{m}^i) = -1. \end{aligned} \tag{3}$$

Considering $\mathcal{H}$ as a three-dimensional manifold with regular affine connection (2) and degenerate metric (3), we can answer many questions by working within $\mathcal{H}$ and

forgetting about the outside space–time $\mathcal{M}$. In such a way, a simple classification of perfect horizons according to their symmetry has been worked out in Ref. 7.

We restrict our attention to axisymmetric perfect horizons whose spacelike sections have spherical topology. Then, the set of invariants[10]

$$R,\ A(\vartheta),\ B(\vartheta),\ E(\vartheta),\ H(\vartheta), \qquad 0 \leq \vartheta \leq \pi, \tag{4a}$$

can be defined as follows (cf. Ref. 4). Let S be a compact intersection of $\mathcal{H}$ with an axisymmetric spacelike hypersurface Σ in $\mathcal{M}$. Then, S is topologically a sphere and the induced metric on S can be written as

$$ds^2 = -R^2(d\vartheta^2 + A^2(\vartheta)\, d\varphi^2) \tag{5}$$

with unique $R > 0$, and $A(\vartheta) \geq 0$ given up to the transformation

$$\vartheta \to \pi - \vartheta \tag{6}$$

reversing the poles $\vartheta = 0$ (north) and $\vartheta = \pi$ (south). φ corresponds to the parameter of the $SO(2)$-group with the period 2π, and the direction of its increase can be fixed by a convention—say, from the west to the east. R can be called the "radius" and $A(\vartheta)$ the "shape function" of $\mathcal{H}$.

The axisymmetry Killing field ξ^i corresponding to φ is, therefore, uniquely determined up to its sign, which changes under (6). Let l'^i, now, denote an arbitrary vector field tangential to the rays. It is fixed up to a scaling

$$l'^i \to \eta l'^i \tag{7}$$

where η is any function on $\mathcal{H}$. It follows that $B(\vartheta)$ defined by the equation

$$\xi^i{}_{;j}\, l'^j = B(\vartheta) l'^i, \tag{8}$$

(proved in Sec. 3) is an invariant with one exception. Under (6), B transforms as

$$B(\vartheta) \to -B(\pi - \vartheta).$$

It is also shown in Sec. 3, that the following equation holds

$$l'^i{}_{;j}\, \xi^j = B(\vartheta) l'^i, \tag{9}$$

if we restrict l'^i to be axially symmetric, and that $B = 0$, if the outside space–time $\mathcal{M}$ is static. Let us call $B(\vartheta)$ "gravimagnetic field."

Finally, let n_t^i be the unit future directed normal vector to Σ and n_r^i the unit outwards pointing normal vector to S tangential to Σ. Then, $E(\vartheta)$ is the component of the electric field and $H(\vartheta)$ that of the magnetic field in the direction of n_r^i as measured by an observer with the 4-velocity n_t^i at any point of S with the coordinate ϑ. $E(\vartheta)$ and $H(\vartheta)$ do not depend on the choice of Σ. We have, namely,

$$\Phi_1 = \tfrac{1}{2}(E - iH), \tag{10}$$

where

$$\Phi_1 = \tfrac{1}{2} F_{ij}(l^i n^j + \bar{m}^i m^j)$$

is a component of the Maxwell spinor Φ as defined in Ref. 6. Φ_1 is invariant against boosts in the plane defined by l^i and n^i, rotations in the orthogonal plane as well as null rotations about l^i [because $\Phi_0 = 0$, see Ref. 7; and $(\Phi_1)_{,i} l^i = 0$, see Ref. 7]. l^i and n^i can be chosen as

$$l^i = (1/\sqrt{2})(n_t^i + n_r^i), \qquad n^i = (1/\sqrt{2})(n_t^i - n_r^i),$$

which implies (10) and the required invariance. $E(\vartheta)$ and $H(\vartheta)$ transform under (6):

$$E(\vartheta) \to E(\pi - \vartheta), \qquad H(\vartheta) \to H(\pi - \vartheta).$$

The functions $A(\vartheta)$, $B(\vartheta)$, $E(\vartheta)$, and $H(\vartheta)$ must satisfy the obvious regularity and smoothness conditions:

$$\begin{aligned} &\vartheta = 0: && A = 0, \quad A' = 1, \quad A^{-1}A'', \ A^{-1}B \text{ regular;} \\ &0 < \vartheta < \pi: && A > 0, \\ &\vartheta = \pi: && A = 0, \quad A' = -1, \ A^{-1}A'', \ A^{-1}B \text{ regular.} \end{aligned} \tag{4b}$$

The invariants (4) are well defined for all axisymmetric perfect horizons. Their role is particularly interesting, if the horizon admits an additional symmetry—the collineation group C. This is the case for any of the following symmetry types: $SO(3)\times C$, $SO(3)\times CT$, $SO(2)\times C$, $SO(2)\times CT$ (Ref. 11)—which were all called AC-horizons in Ref. 4. For example, the Kerr–Newman horizon with $m^2 > a^2 + e^2$ belongs to this class, and the horizon of any generic black hole in equilibrium with surrounding matter is generally considered as necessarily being of this AC type.[1] But it is important to keep in mind that all results given from now on apply only to axisymmetric stationary space–times with bifurcate Killing horizon. Nonaxisymmetric static space–times with bifurcate horizon or axisymmetric stationary space–times with horizon of the symmetry type $SO(3)\times T$, $SO(2)\times T$[7] (such as extreme Kerr–Newman) need *essentially different* sets of invariants to describe their structure,[4] even if the extreme types seem to be limit cases of the AC types.

Two facts have essentially been shown for the AC horizons in Ref. 4.

Lemma 1: Let $\mathcal{M}$ be an axisymmetric stationary space–time which contains an AC-horizon $\mathcal{H}^+$ as a Killing horizon and which satisfies Einstein–Maxwell electrovacuum equations in a neighbourhood of $\mathcal{H}^+$. Then, there is another AC-horizon $\mathcal{H}^-$ in $\mathcal{M}$, intersecting $\mathcal{H}^+$ in a compact spacelike surface S. If the pole convention for $\mathcal{H}^+$ and $\mathcal{H}^-$ coincides at S, then the corresponding invariants (4) satisfy

$$R^+ = R^-, \quad A^+(\vartheta) = A^-(\vartheta), \quad B^+(\vartheta) = -B^-(\vartheta),$$

$$E^+(\vartheta) = E^-(\vartheta), \quad H^+(\vartheta) = H^-(\vartheta),$$

and $\mathcal{H}^+ \cup \mathcal{H}^-$ is a bifurcate Killing horizon symmetric under the (φ, t) – transform.[12]

Lemma 2: Let the conditions of Lemma 1 be satisfied. Then, the characteristic initial data for Einstein–Maxwell electrovacuum equations satisfying the corresponding constraints along the two intersecting null hypersurfaces $\mathcal{H}^+$ and $\mathcal{H}^-$ (this all abbreviated by CID) determines uniquely a value of the invariants (4), and any C^3-value of the invariant (4) determines a unique CID.

From these two Lemmas, and from the well-known uniqueness theorem for the space–time development of

a given CID,[13] we infer that a value of the set (4) contains complete information about the metric tensor g_{ij} and the electromagnetic tensor F_{ij} in an electrovacuum neighbourhood of the corresponding horizon. We call, therefore, the quantities (4) "local invariants" (to distinguish them from other quantities in black hole physics such as the electric potential difference between the black hole surface and infinity,[14] rotation of the black hole with respect to infinity,[1] surface gravity,[1] and other possible potential-difference-like properties, which need not be determined completely by the local space–time structure at the surface of the hole). We have shown

Theorem 1: The most general stationary axisymmetric solution of Einstein–Maxwell electrovacuum equations that contains a regular bifurcate Killing horizon depends on arbitrary parameters and functions which can be chosen as in (4).[15]

An interesting problem is the following. It is a well-known fact that there are Killing horizons which cannot form an absolute event horizon[16] in any space–time. For example, the Killing horizon in the Ehlers–Kundt C-metric space–time,[17] or the inner horizons in the Kerr–Newman space–times with $m^2 > a^2 + e^2$ (Ref. 18) (they all violate the condition that there should not be trapped surfaces outside of them[16]). Can one distinguish the Killing horizons which may be absolute event horizons from those which may not just by looking at the values of their local invariants? A partial answer is given by the following:

Theorem 2: A necessary condition for a Killing horizon $\mathcal{H}$ of AC type to form an absolute event horizon in an axisymmetric stationary space–time is that the corresponding local invariants (4) satisfy

$$\int_0^{\pi}\left(\frac{B^2}{A^2}+R^2E^2+R^2H^2\right)A\,d\vartheta<2. \tag{11}$$

The proof is given in Sec. 3. In order to see implications of Theorem 2, recall the following properties of the local invariants (shown essentially in Ref. 5):

Theorem 3: Let the value (4) of the local invariants correspond to the horizon $\mathcal{H}$ of a black hole in an axisymmetric stationary space–time $\mathcal{M}$ which is electrovacuum in a neighbourhood of $\mathcal{H}$. Let A, J, Q_e, and Q_h be the net surface area, net angular momentum, net electric, and magnetic charge of the black hole as defined in Refs. 1 and 14. Then

$$A=\int_S \sqrt{g}\,d\vartheta\,d\varphi, \tag{12}$$

$$Y=-\frac{1}{8\pi}\int_S B(\vartheta)\sqrt{g}\;d\vartheta\,d\varphi, \tag{13}$$

$$Q_e=\frac{1}{4\pi}\int_S E(\vartheta)\sqrt{g}\;d\vartheta\,d\varphi, \tag{14}$$

$$Q_h=-\frac{1}{4\pi}\int_S H(\vartheta)\sqrt{g}\;d\vartheta\,d\varphi, \tag{15}$$

where S is any compact spacelike section of $\mathcal{H}$ and g is the determinant of the metric (5) on S,

$$\sqrt{g}=R^2A(\vartheta). \tag{16}$$

Wheeler's well-known magic formula "black holes have no hair" could, now, take on the following form:

Conjecture: Any equilibrium state of the system consisting of a black hole surrounded by matter and charge configurations is fixed, if one prescribes the values of A, J, Q_e, and Q_h of the black hole, and specifies the mass and charge currents of the matter configuration.

In Ref. 5, this conjecture is illustrated and made plausible in the neighbourhood of the Schwarzschild solution by considering small axisymmetric, stationary gravitational and electromagnetic perturbations of the background containing a thin charged spherical matter shell (see, e.g., Ref. 19). In particular, the fields of outside sources can deform the black hole surface and change R and $A(\vartheta)$ (see, also Ref. 20), produce a non-vanishing gravimagnetic field B preserving $Y=0$ (because the space–time in which a rotating matter ring is present cannot be static even if the black hole in the middle does not rotate), and the radial component of the electric and the magnetic field originating in the charge and current of the shell can be nonzero at the horizon without contributing to the integrals (14) and (15).

Now, we can discuss the meaning of (11). It is clearly a generalization and sharpening of the well-known inequality

$$m^2 > a^2 + e^2 + h^2 \tag{17}$$

holding for the bifurcate horizon in the Kerr–Newman space–time with the total mass m, angular momentum am, electric charge e, and magnetic charge h. The main similarities and differences between (11) and (17) can be summarized as follows:

(a) On setting the Kerr–Newman values of the local invariants into (11), we obtain not only (17), but also the condition

$$r_H=m+(m^2-a^2-e^2-h^2)^{1/2}.$$

Thus, (11) is stronger than (17), excluding the inner horizon. Indeed, all known Killing horizons which cannot play the role of absolute event horizon (C-metric horizon, Kerr–Newman inner horizons, Robinson–Bertotti horizon as well as other CT-horizons[4,5]) are excluded by (11), as one can immediately verify.

(*b*) Just as the inequality (17) does, (11) places limitations on the active degrees of freedom of a black hole, if its irreducible mass is fixed (angular momentum and charges are "active," because their presence allows the extraction of energy from the hole). Comparing (11) with (12)–(16) shows indeed that J, Q_e, and Q_h cannot be arbitrarily large for a hole whose A is kept constant.

(c) (11) implies limitations on the gravimagnetic field B, the electric field E and the magnetic field H at the surface of a black hole *independently of what the source of these fields is*. They can originate from the black hole itself as under (b) [which is the only posibility with (17)], but they can have their source outside the hole as well. One integrates squares in (11) so that the fields, which need not contribute to (13), (14), and (15), do contribute to (11), if they are nonzero.

We can speculate about what these limitations on the outside fields actually mean:

(i) Outside fields which were stronger than the allowed upper bound would destroy the black hole, either by removing horizon (in an analogy to the well-known phenomenon observed in the Kerr family—as a grows over m, the horizon disappears), or by producing singularities and/or new horizons outside it.

(ii) Outside fields can never be produced and brought down to the surface of a black hole so strong and so directed as to violate (11).

How large is the upper bound? For a spherical "dead" black hole with radius R, we have $A=\sin\vartheta$, $J=Q_e=Q_h=0$. Apply a "homogeneous" electric field $E=E_0\cos\vartheta$. (11) takes the form $E_0<\sqrt{3}R^{-1}$. Using $1\ \sec^{-1}=(2/\sqrt{3})\times10^{14}\ CGS$[21], we obtain

$$E_0<6\times10^{24}\,R^{-1}\ [CGS].$$

Take, for example, $R=2\ km$, corresponding to a solar mass black hole. The upper bound is, then, $3\times10^{19}\ CGS$. Thus, if (11) is everything black holes must satisfy, they are robust rather than fragile.

3. PROOFS

Equations (8) *and* (9): The left-hand side of Eq. (8) is a real vector tangential to $\mathcal{H}$, hence

$$\xi^i{}_{;j}\,l'^j=Bl'^i+Cm^i+\bar{C}\bar{m}^i, \tag{18}$$

where B is a real and C a complex function on $\mathcal{H}$. Multiply (18) by a covenient function η so that $l''^i=\eta l'^i$ is axially symmetric, i.e.,

$$\pounds_\xi l''^i=l''^i{}_{;j}\,\xi^j-\xi^i{}_{;j}\,l''^j=0.$$

Hence, the left-hand side of (18) is equal to $l''^i{}_{;j}\,\xi^j$, and this must be a multiple of l''^i according to (2). Thus, $C\eta=0$, implying $C=0$, because $\eta\neq0$. As a by-product, we obtain (9).

It remains to be proved that B depends only on ϑ, or $B_{,i}l^i=0$. Observe that $(\xi^i,\xi^i)=-R^2A^2$; we can, therefore, choose m^i such that the relation

$$\xi^j=\frac{i}{\sqrt{2}}RA\,(m^j-\bar{m}^j)$$

holds at $\mathcal{S}$; otherwise let l^i and m^i satisfy (2). Then, it follows from (8) that

$$\xi^j=\frac{i}{\sqrt{2}}RA(m^j-\bar{m}^j)+yl^j,$$

where y is a function on $\mathcal{H}$ determined by $y_{,i}l^i=B$, $y|_{\mathcal{S}}=0$. Now, (9) and (2) imply

$$B=\frac{i}{\sqrt{2}}RA(\bar{\alpha}-\alpha+\beta-\bar{\beta}).$$

On the other hand, the relevant Newman–Penrose equations as given in Ref. 6 and adapted to our tetrad read

$$\alpha_{,i}l^i=2\Phi_1\bar{\Phi}_0, \tag{4.2d}$$

$$\beta_{,i}l^i=\Psi_1, \tag{4.2e}$$

$$0=2\Phi_0\bar{\Phi}_0, \tag{4.2a}$$

$$0=-\Psi_1+2\Phi_0\bar{\Phi}_1, \tag{4.2k}$$

where we have used the Einstein's equations in the form

$$\Phi_{ij}=2\Phi_i\bar{\Phi}_j$$

corresponding to our choice $G=c=1$. The required equation $B_{,i}l^i=0$ follows immediately.

$B=0$ *in static space–times*. Let the spacetime $\mathcal{M}$ outside $\mathcal{H}$ be static and axisymmetric. Let Σ and $\mathcal{S}$ be chosen as in Sec. 2. Through any point p of Σ outside $\mathcal{S}$, there passes a totally geodesic axisymmetric spacelike hypersurface Σ_p which intersects Σ in a topologically spherical surface $\mathcal{S}_p$. Let n_p^i be the unit future oriented normal vector field to Σ_p at $\mathcal{S}_p$. Clearly

$$n_p^i{}_{;j}\,\xi^j=0,$$

because ξ^i is tangential to Σ_p. On the other hand, if p approaches a point p_0 at $\mathcal{S}$, Σ_p converges to $\mathcal{H}$ and the direction of its normal n_p^i must, therefore, approach that of l^i. Multiplying the vector n_p^i by an axially symmetric factor η_p such that

$$n_p^i\eta_p\to l^i,$$

we have

$$(\eta_p n_p^i)_{ij}\,\xi^j=0,$$

which implies the statement.

Theorem 2: The strategy of the proof will be to find outer trapped surfaces (see Ref. 16, p. 319) outside of $\mathcal{H}$ in the cases when (11) is violated, and then to use Proposition 9.2.8 of Ref. 16 which forbids such a situation. The conditions under which the proposition holds are supposed to be satisfied in $\mathcal{M}$. This does not mean any essential restriction of generality at least as concerns physically interesting space–times.

$\mathcal{H}$ is an event horizon, i.e., the outside of $\mathcal{H}$ coincides with its past in $\mathcal{M}$. Let us indicate this by the superscript "+" at $\mathcal{H}$. As $\mathcal{H}^+$ is a Killing horizon of AC type in $\mathcal{M}$, there must be another perfect horizon in $\mathcal{M}$, $\mathcal{H}^-$, say, whose properties are described by Lemma 1. $\mathcal{H}^-$ is a particle horizon, i.e., its future coincides with its outside in $\mathcal{M}$. As $\mathcal{H}^-$ is again of AC type, we can choose the coordinates α, ϑ, φ and the triad l^i, m^i along $\mathcal{H}^-$ in the canonical way described by Theorem 5 in Ref. 4. Then, using Lemma 1, we express all relevant quantities of $\mathcal{H}^-$ by means of the local invariants R, A, B, E, H of $\mathcal{H}^+$, e.g.,

$$l=\frac{\partial}{\partial\alpha},\quad \mathrm{m}=\frac{1}{\sqrt{2}}\frac{1}{R}\left(\frac{\partial}{\partial\vartheta}+\frac{i}{A}\frac{\partial}{\partial\varphi}-i\alpha\frac{B}{A}\frac{\partial}{\partial\alpha}\right),$$

$$\Omega=\frac{i}{\sqrt{2}}\frac{1}{R}\frac{B}{A},\quad \Gamma=-\frac{1}{\sqrt{2}}\frac{1}{R}\frac{A'}{A}, \tag{19}$$

$$\Psi_2=\tfrac{1}{2}E^2+\tfrac{1}{2}H^2+\frac{1}{2}\frac{1}{R^2}\frac{A''}{A}+\frac{i}{2}\frac{1}{R^2}\frac{B'}{A}.$$

$\alpha=0$ at the intersection $\mathcal{H}^+\cap\mathcal{H}^-$ and $\alpha<0$ outside of $\mathcal{H}^+$.

Let us choose an arbitrary axisymmetric compact spacelike section S of the lower half ($\alpha<0$) of $\mathcal{H}^-$, propagate it by the C-group along $\mathcal{H}^-$, and compute the convergence of the outgoing null geodesics orthogonal to the resulting family of sections. To this end, we perform the transformation

$$l_1^i=\alpha l^i,\quad m_1^i=m^i+\alpha\Omega l^i. \tag{20}$$

The new vectors l_1^i, m_1^i satisfy

$$m_{1\ ;j}^i l_1^j = \Omega l_1^i, \quad l_{1\ ;j}^i m_1^j = \Omega l_1^i, \tag{21}$$

$$\bar{m}_{1\ ;j}^i m_1^j = (\Psi_2 + \bar{\Omega}_{,j} m^j - \Gamma\bar{\Omega} + \Omega\bar{\Omega}) l_1^i + \Gamma m_1^i, \tag{22}$$

where Γ, Ω, Ψ_2 and m^i are given by (19). The pair of vector fields m_1^i, $\bar{m}_1^i$ is (1) surface-tangential, because the imaginary part of the corresponding μ [the coefficient at l_1^i on the right of (22)] is zero [see Ref. 6 and Eq. (15) of Ref. 7], (2) C-group-propagated, for (21) implies

$$[m_1^i, l_1^i] = 0$$

and l_1^i is a generator of the C-group (cf. Theorem 5 in Ref. 4), and (3) axisymmetric. Any other axially and C-symmetric pair $m_2^i, \bar{m}_2^i$ is of the form (up to a rotation $m_2^k \to \exp(i\psi) m_2^k$)

$$m_2^i = m_1^i + \xi l_1^i,$$

where ξ is a complex function on the lower half of $\mathcal{H}^-$ satisfying $\partial\xi/\partial\alpha = \partial\xi/\partial\varphi = 0$ everywhere and $\xi = 0$ at the pole rays $\vartheta = 0, \pi$. The corresponding μ is given by

$$\mu = (\Psi_2 + \bar{\Omega}_{,j} m^j - \Gamma\bar{\Omega}) + (\bar{\xi}_{,j} m^j - \Gamma\bar{\xi}) + (\Omega + \xi)(\bar{\Omega} + \bar{\xi}). \tag{23}$$

m_2^i, $\bar{m}_2^i$ are surface-tangential, iff

$$\bar{\xi}_{,j} m^j - \xi_{,j} \bar{m}^j - \Gamma\bar{\xi} + \bar{\Gamma}\xi = 0,$$

and this is equivalent to[7]

$$\xi = \xi_{,j} m^j, \tag{24}$$

where ξ is a *real* function on the lower half of $\mathcal{H}^-$, which must satisfy

$$\frac{\partial\xi}{\partial\alpha} = \frac{\partial\xi}{\partial\varphi} = 0 \quad \text{at all points of the lower half of } \mathcal{H}^-, \tag{25}$$

$$\frac{\partial\xi}{\partial\vartheta} = 0 \quad \text{at the pole rays } \vartheta = 0, \pi. \tag{26}$$

Replacing the first term on the right of (23) by its real part, substituting for Γ, Ω, Ψ_2, and m^i from (19), (24) for ξ, and using (25), we find

$$\mu = \frac{1}{2}\frac{1}{R^2}\left(\frac{B^2}{A^2} + R^2E^2 + R^2H^2 + \frac{A''}{A} + \xi'' + \frac{A'\xi'}{A} + (\xi')^2\right). \tag{27}$$

μ is the desired convergence, all possible sections S being represented by all possible functions ξ [indeed, μ is a convergence, not an expansion,[6] because it is defined as a coefficient at l_1^i in an equation of the type of (22) and l_1^i is oriented in the *past* as it follows from Eq. (20) and the fact that $\alpha < 0$].

The next step is to choose ξ in such a way that the sum of the first six terms in the parentheses on the right-hand side of (27) is equal to a constant real number k. If the conditions (26) and (4b) are to be satisfied, k is uniquely determined, namely

$$k = \left(\int_0^\pi A d\vartheta\right)^{-1} \cdot \int_0^\pi \left(\frac{B^2}{A^2} + R^2E^2 + R^2H^2 + \frac{A''}{A}\right) A d\vartheta.$$

Using (4b) again we obtain

$$k = \left(\int_0^\pi A d\vartheta\right)^{-1} \cdot \left[\int_0^\pi \left(\frac{B^2}{A^2} + R^2E^2 + R^2H^2\right) A\, d\vartheta - 2\right].$$

Thus, if $k < 0$, (11) follows immediately. Let $k \geq 0$. Then $\mu \geq 0$ at all points of S, and μ remains nonnegative and regular near $\mathcal{H}^-$ in the future of $\mathcal{H}^-$, because S is compact and weak energy inequality holds. But, then, we have outer trapped surfaces outside of $\mathcal{H}^+ \cup \mathcal{H}^-$. QED

[1] J.M. Bardeen, B. Carter, and S.W. Hawking, Commun. Math. Phys. **31**, 161 (1973).
[2] J.D. Bekenstein, "The Generalized Second Law of Thermodynamics in Black Hole Physics," Preprint, Austin, 1973.
[3] S.W. Hawking, "Black Hole Explosions?" Preprint, Cambridge, January 1974.
[4] P. Hájíček, Commun. Math. Phys. **34**, 53 (1973).
[5] P. Hájíček, Commun. Math. Phys. (to be published).
[6] E. Newman and R. Penrose, J. Math. Phys. **3**, 566 (1962).
[7] P. Hájíček, Commun. Math. Phys. **34**, 37 1973).
[8] B. Carter, *Lecturers at the Summer School*, Les Houches, August 1972.
[9] Null hypersurfaces exist which are perfect horizons but not Killing horizons at the same time, so the difference between the two notions goes beyond the condition of compact sections.
[10] In fact, the functions of (4a) are not invariant under the pole exchange (defined later).
[11] The structure and action of the collineation group C, translation group T, as well as the two-dimensional group CT is explained in Ref 7.
[12] Under the (φ, t)-transform, $t \to -t$ and $\varphi \to -\varphi$; therefore, $\mathcal{H}^+ \to \mathcal{H}^-$.
[13] R.K. Sachs, J. Math. Phys. **3**, 908 (1962).
[14] B. Carter, "The Electrical Equilibrium of a Black Hole." Preprint, Cambridge, 1972.
[15] The maximal analytic extension of such a general solution cannot be always regular outside the horizon: either the singularity must appear where the outside sources —material rings, disks, and shells—should be, or, no asymptotically flat space—time exists at all containing a part of the analytic extension as a subspace (see Theorem 2).
[16] S.W. Hawking and G.F.R. Ellis, *The Large Scale Structure of Spacetime* (Cambridge U. P., Cambridge, 1973).
[17] M. Walker, J. Math. Phys. **11**, 2280 (1970).
[18] B. Carter, Phys. Rev. **174**, 1559 (1968).
[19] K. Kuchar, Czech. J. Phys. **B18**, 435 (1968).
[20] W. Israel, Lett. Nuovo Cim. **6**, 267 (1973), G.W. Gibbons, Commun. Math. Phys. **35**, 13 (1974).

Complex structures and representations of the Einstein equations*

Carl H. Brans†

Joseph Henry Laboratories, Physics Department, Princeton University, Princeton, New Jersey 08540
(Received 23 April 1974)

The space–time model of general relativity is that of a four-dimensional manifold M, with a metric of Minkowski signature. The space of two-forms on M is endowed with a natural complex structure, **J**, generated by the star duality operator. The existence of such a structure is an accidental characteristic of the dimension four and of the metric signature. The full differential geometric structure equations are expressed in this two-form language and it is pointed out that the weakened Einstein empty space equations, i.e., $R_{ab}-(1/4)g_{ab}R=0$, reduce to the condition that the curvature form commute with **J**. This fact, together with the isomorphism of the two-form space with the Lorentz Lie algebra, l_0, are then shown to provide the basis for the importance of the various complex representations of l_0, such as $SO(3,C)$ and the spinor $SL(2,C)$, in understanding the real geometry of Einstein spaces. In fact, the complexified Einstein structure equations naturally divide into two sets, each the complex conjugate of the other, each involving only one-half of the basis.

INTRODUCTION

The study of Einstein's general relativistic field equations has had a profound impact on differential geometry. These equations have been expressed in many different forms, each showing a different facet of Einstein space–time.[1] In this paper we will concentrate on the representations of the Einstein equations generated by representations of the Lorentz Lie algebra, l_0, and try to understand why the *complex* representations of l_0, such as $SO(3,C)$,[2] and the spinor one, $SL(2,C)$,[3] provide such natural and effective framework for the study of the *real* differential geometry of Einstein spaces. Our approach will be to start with an expression of the Einstein equations in terms of conditions on the curvature forms. Next, the curvature forms will be related to linear maps of the space of two-forms, F_p^2, at each point into itself. For the special case of four-dimensional manifolds with indefinite metrics F_p^2 is endowed with another linear map onto itself, **J**, which has the characteristics of what mathematicians call a "complex structure."[4] The existence of **J** then gives rise to a natural isomorphism of the real six-dimensional F_p^2 with the complex three-dimensional vector space. We then note that the Einstein equations can be expressed simply by saying that the *curvature commutes with* **J**. Thus, in explicit complex representations the equations naturally split into two sets, each of which involves only one-half of the basis and is the complex conjugate of the other, so that the *complexification has essentially halved the number of equations to be solved.*

In order to carry out this program, we must introduce a formalism in which the basis for vectors, forms, and other geometric quantities can be freely chosen at each point to conform to the structure of the metric, curvature, etc. The language of modern differential geometry, especially exterior differential forms, is well adapted for this program. Those readers familiar with the use of bundle theory will find it a natural framework for the expression of these ideas. However, for simplicity of exposition, we will not explicitly use bundle techniques in this paper.

FORMS, VECTORS, AND GEOMETRY

This section contains a brief review of the formalism to be used in this paper. This approach involves the use of differential forms, and, in particular, takes advantage of the availability of arbitrary bases at each point to describe the forms, or, equivalently the dual space of tangent vectors. These arbitrary bases are sometimes referred to as "tetrads" in the physics literature. For background information on the use of differential forms presented intuitively see Misner, Thorne and Wheeler.[5] For a more complete and mathematical treatment of these and other tools for differential geometry, including complex structures, see Kobayashi and Nomizu.[4]

For our purposes, space–time can be represented as a four-dimensional, orientable, differentiable manifold M. At each point $p\in M$ let T_p be the space of tangent vectors at p. An element of T_p can be defined intrinsically as a "differentiation operator", or more concretely represented in terms of components, v^a $(a=0,1,2,3)$, with respect to some local coordinates, x^a, near p.[6] Thus, if $v\in T_p$, we will write

$$v=v^a\partial/\partial x^a, \tag{1}$$

where the symbols $\partial/\partial x^a$ stand for differentiation along the corresponding coordinate lines. Equation (1) shows clearly the usual contravariant transformation properties of the list of components, v^a, under a coordinate transformation. The vector space dual to T_p, i.e., the space of all linear functions on T_p is denoted by T_p^*. An element of T_p^* will be represented by a Greek letter, say ρ, and will be called a differential form of degree one, or simply a one-form. Let the symbols dx^a be used to represent the dual basis to $\partial/\partial x^a$, so that as linear functions on T_p the dx^a satisfy

$$dx^a(\partial/\partial x^b)=\delta^a{}_b. \tag{2}$$

An element of T_p^* can be written in terms of its components, ρ_a, with respect to dx^a as

$$\rho=\rho_a dx^a,$$

showing the covariant transformation law for ρ_a.

Tensors are then defined as usual as linear combinations of ordered products of vectors and forms. One particularly important class of tensors for our purposes will be totally antisymmetirc tensors built from one-forms. Those of degree r are called r-forms and a special symbol, $\wedge$, is used to represent the antisymmetric

product of forms. Thus, an r-form ω can be written with respect to the coordinates x^a as

$$\omega = (1/r!)\omega_{a_1 \cdots a_r} dx^{a_1} \wedge \cdots \wedge dx^{a_r}, \tag{3}$$

where the components $\omega_{a_1 \cdots a_r}$ are totally antisymmetric and the normalization by $r!$ is standard convention. We will denote the space of r-forms at p by $F^r{}_p$ so that $T^*_p \equiv F^1{}_p$.

The symbol "d" used in the initial coordinate representation of one-forms can be used in an extended sense as an operator taking r-forms into $(r+1)$-forms defined directly by

$$d\omega = [1/(r+1)!]\omega_{[a_1 \cdots a_r, a_{r+1}]} dx^{a_{r+1}} \wedge dx^{a_1} \cdots \wedge dx^{a_r}, \tag{4}$$

where the comma denotes ordinary differentiation and the brackets indicate antisymmetrization of enclosed indices.

Now consider the notion of generalized bases for T_p, that is, bases which need not correspond to differentiation along coordinates in any neighborhood. Thus if u_a are four linearly independent vectors at p, then any vector v can be written

$$v = v^a u_a, \tag{5}$$

in terms of components v^a. Let us now introduce the notation of using boldface letters to represent bases, so that u stands for the matrix of four independent vectors, $\{u_a\}$. Let us call such a u a "vector frame" and denote the set of all vector frames at p by B_p.[7] Similarly, let B^*_p be the dual space of "form frames" so that if $\rho \in B^*_p$, then ρ stands for a matrix of four independent one-forms, $\{\rho^a\}$. The duality relationship between T_p and T^*_p provides a natural, unique, isomorphism between B_p and B^*_p defined by $\rho \leftrightarrow \mathrm{u}$ iff

$$\rho^a(u_b) = \delta^a{}_b, \tag{6}$$

or, in more compact notation,

$$\rho(\mathrm{u}) = 1. \tag{7}$$

As noted above, the introduction of arbitrary frames for vectors includes of necessity a wider class of frames than those which can be represented by sets of vectors each of which is differentiation along some coordinate line. It can be shown that a necessary and sufficient condition for the existence of a local coordinate system such that

$$u_a = \partial/\partial x^a, \tag{8}$$

or, equivalently, for the dual basis $\{\rho^a\}$,

$$\rho^a = dx^a, \tag{9}$$

is that

$$d\rho = 0. \tag{10}$$

A frame for which (10) is satisfied is sometimes said to be *holonomic*. In general, however, there will be a nonzero matrix $A^a{}_{bc} = -A^a{}_{cb}$ associated with each frame ρ, such that

$$d\rho^a = \tfrac{1}{2} A^a{}_{bc} \rho^b \wedge \rho^c. \tag{11}$$

The family of frames, B^*_p, is tied together by the action of the general linear group, $GL(4,R)$. Thus if $\rho \in B^*_p$ and $g \in GL(4,R)$, with g represented by the matrix $\{g^a{}_b\}$, then g can be regarded as acting on B^*_p taking ρ into $\rho' = g\rho$,

$$\rho'^a = g^a{}_b \rho^b. \tag{12}$$

The dual, contragradient, action of $GL(4, R)$ on B_p is obviously defined by means of the relationship (6). Thus, if u is dual to ρ, then $\mathrm{u}' = \mathrm{u}g^{-1}$ will be dual to ρ', where

$$u'_a = u_b (g^{-1})^b{}_a. \tag{13}$$

A metric on M can be defined in terms of an inner product on each T_p, denoted by (,). This corresponds to a nonsingular, symmetric matrix γ_{ab} for each frame $\mathrm{u} \in B_p$ such that

$$\gamma_{ab}(\mathrm{u}) = (u_a, u_b). \tag{14}$$

For physical manifolds each γ_{ab} is required to be of signature $(-,+,+,+)$. Equivalently, using the duality isomorphism we can write the metric in terms of an invariant expression, ds^2,

$$ds^2 = \gamma_{ab}(\boldsymbol{\rho}) \rho^a \rho^b. \tag{15}$$

The significance of the product of forms on the right side of (15) can be seen by relating the ρ^a to the dx^a, so that (15) becomes a bilinear quadratic expression in the dx^a, reducing to the definition of metric as a rule for deriving infinitesimal distance ds from the dx^a regarded as coordinate differentials.

The introduction of a metric permits a specialization of the frames to those for which the metric matrix assumes some standard form, for example, $\eta_{ab} \equiv \mathrm{diag}(-1, +1, +1, +1)$. Let B^0_p and B^{0*}_p be those subsets of B_p and B^*_p respectively for which the metric matrix $\{\gamma_{ab}\}$ assumes the value $\{\eta_{ab}\}$. These will be called the families of *Lorentz frames*. Further, for simplicity, we will assume that B^0_p and B^{0*}_p have been restricted by the preferred orientation on M. The reduction of B^*_p to B^{0*}_p thus means that the allowed group of transformations has been reduced from $GL(4, R)$ to the proper Lorentz group, $L_0 \equiv SO(3, 1)$. The value of this approach lies in the fact that the metric assumes a constant standard value at each point and the group connecting all admissible frames has been reduced to the well-studied L_0. The price paid, of course, is that the frames will in general no longer be holonomic, that is, composed simply of coordinate differentials.

Another approach to the geometry of M is by means of the definition of covariant differentiation and the associated connection forms. By using appropriate linearity properties for derivatives, the value of the covariant derivative $\nabla_v w$ of any vector w along a given vector v can be obtained from the action of ∇_v on the members of a basis, which in turn can be written

$$\nabla_v u_a = -\omega^b{}_a(v) u_b, \tag{16}$$

or

$$\nabla_v \mathrm{u} = -\omega(v) \cdot \mathrm{u}. \tag{17}$$

The matrix $\omega(v)$ depends linearly on the vector v and so defines a matrix of one-forms, ω, called the *connection forms*.

The relationship between covariant differentiation and the metric is established by requiring that the parallel displacement naturally defined by the covariant deriva-

tive preserve the metric inner product. Regarding $\{-\omega^a{}_b(v)u_a\}$ as the infinitesimal displacement of the frame $\{u_a\}$, the metric condition can be expressed by requiring that for any $v \in T_p$, the matrix $\omega^a{}_b(v)$ be an "infinitesimal" Lorentz transformation, or, more accurately, that it belong to the Lie algebra, l_0, of L_0. Thus, the metric condition for the connection is

$$\omega: T_p \to l_0, \tag{18}$$

or,

$$\eta_{ab}\omega^b{}_c + \eta_{cb}\omega^b{}_a = 0. \tag{19}$$

Finally, we add the usual condition that the connection be torsion free so that for the basis $\rho \in B_p^{0*}$ dual to u,

$$d\rho^a = \omega^a{}_b \wedge \rho^b, \tag{20}$$

or simply

$$d\rho = \omega \wedge \rho. \tag{21}$$

It should be noted that (20) and (19), or, equivalently, (21) and (18), serve to uniquely define the connection forms ω in terms of the basis ρ and the standard metric η.

The curvature tensor can now be defined by

$$d\omega^a{}_b + \omega^a{}_c \wedge \omega^c{}_b = \Omega^a{}_b. \tag{22}$$

or in terms of the full list of Riemann curvature components $R^a{}_{bcd}$,

$$\Omega^a{}_b = \tfrac{1}{2}R^a{}_{bcd}\rho^c \wedge \rho^d. \tag{23}$$

Again, we can make use of the matrix notation to write (22) simply as

$$d\omega + \omega \wedge \omega = \Omega, \tag{24}$$

and express a result of the metric condition in terms of an equation that will be of importance later.

$$\Omega: T^{[2]}{}_p \to l_0, \tag{25}$$

where $T^{[2]}{}_p$ is the linear space formed from antisymmetric pairs of vectors. The fact that at each point the metric provides a natural isomorphism between $T^{[2]}{}_p$ and l_0 means that the *curvature tensor translates to an endomorphism of l_0 with itself*. In the same way, the metric induces an isomorphism of $F^2{}_p$ with l_0.[8] The approach of this paper will be to emphasize this central role of l_0, and thus its representations, in studying general relativity.

In summary, the differential geometric structure of M can be expressed by a choice of a frame of forms ρ. The metric is then obtained from this choice according to (15) while the connection and curvature properties are described by (21), (18), and (24). The geometry itself (in the sense of the metric) is unchanged by Lorentz transformation on ρ, so in this sense the appropriate description of the geometry is carried by the full set of Lorentz frames, B_p^{0*}, at each point, each acted on transitively by L_0.

STRUCTURE EQUATIONS FOR $F^2{}_p$

In this section we will translate the above geometirc formalism into one based on $F^2{}_p$. First, we must establish a relationship between B_p^* and the bases for $F^2{}_p$ denoted by β_p with $\phi \in \beta_p$ representing a matrix of six linearly independent two-forms, $\{\phi^i\}$, $i = 1, \ldots, 6$. Clearly,

$$\phi^i = \tfrac{1}{2}f^i{}_{ab}\rho^a \wedge \rho^b, \tag{26}$$

for some quantities $f^i{}_{ab}$ which, for the moment, are to be regarded as fixed. Thus different ϕ's are generated by different ρ's. How then is the choice of ϕ related to the choice of ρ and thus the metric? Assume that the metric has been defined by choosing the ρ used in (26) as a Lorentz frame, so that ds^2 is given by (15) and $\rho \in B_p^{0*}$ by definition. Consider the 6×6 symmetric matrix Δ^{ij} defined by

$$\phi^i \wedge \phi^j = \Delta^{ij} I, \tag{27}$$

where

$$I \equiv \rho^0 \wedge \rho^1 \wedge \rho^2 \wedge \rho^3. \tag{28}$$

From the linear independence of the ϕ^i, it follows that Δ^{ij} is nonsingular, with inverse denoted as usual by Δ_{ij}. This matrix plays the role of a metric on $F^2{}_p$.

We can now state the important converse to this relationship: *Any six independent two-forms satisfying* (27) *define uniquely* (*up to inversions, with which we are not here concerned*) *a basis ρ for which* (26) *is true*. In order to prove this statement, it is sufficient to consider one particular choice for $f^i{}_{ab}$ and associated ϕ^i since, because of the assumed completeness of the ϕ^i, any other set of ϕ^i can be obtained as linear combinations of this particular one. The entire proof can then be translated into the other basis by a linear transformation. Let us choose $f^i{}_{ab}$ so that

$$\phi^i = \rho^0 \wedge \rho^i, \quad i \leq 3, \tag{29}$$

and

$$\phi^{i+3} = \tfrac{1}{2}\epsilon_{ijk}\rho^j \wedge \rho^k, \quad i \leq 3, \tag{30}$$

where ϵ_{ijk} is the three-dimensional alternating symbol. The conditions (27) become in this case

$$\phi^{i+3} \wedge \phi^j = \delta^{ij} I, \quad i \leq 3, \tag{31}$$

and

$$\phi^i \wedge \phi^j = \delta^{(i-3)j} I, \quad i \geq 3. \tag{32}$$

Assume now that some other linearly independent set is given, say ϕ'^i, satisfying (31) and (32). More precisely, with respect to some bases, say ρ', the ϕ'^i can be expressed

$$\phi'^i = \tfrac{1}{2}g^i{}_{ab}\rho'^a \wedge \rho'^b, \tag{33}$$

and (31) and (32) are satisfied with I replaced by $I' \equiv \rho'^0 \wedge \rho'^1 \wedge \rho'^2 \wedge \rho'^3$. We must now show that ρ' can be adjusted so that $g^i{}_{ab} = f^i{}_{ab}$. To begin, consider the set of equations

$$\phi'^1 \wedge \phi'^1 = \phi'^4 \wedge \phi'^4 = 0, \tag{34}$$

and

$$\phi'^1 \wedge \phi'^4 = I'. \tag{35}$$

From (34) it follows that ϕ'^1 and ϕ'^4 are each simple two-forms, i.e., hook products of pairs of one-forms, while (35) implies that this set of four one-forms is independent, and thus, by redefining the basis ρ', we can set

$$\phi'^1=\rho'^0\wedge\rho'^1 \tag{36}$$

and

$$\phi'^4=\rho'^2\wedge\rho'^3. \tag{37}$$

The proof then proceeds by filling out the remaining ϕ'^i similarly, with the only ambiguities arising being dealt with by reordering the index i.

Further, we note that the action of the Lorentz group L_0 on the basis ρ gives rise to a corresponding group action on the basis ϕ which preserves the "metric" Δ^{ij} as defined by (27). This is, of course, equivalent to the antisymmetric two-tensor representation of L_0. In a manner analogous to the restriction of the one-form bases B^*_p to B^{0*}_p we can also define β^0_p as the subset of β_p generated by the group preserving Δ^{ij}, denoted by l_0.

For a fixed representation $f^i{}_{ab}$, we can summarize these results: A *metric can be specified in terms of the two-form* $F^2{}_p$ *structure by the conditions* (27) *corresponding to the choice of Lorentz basis* ρ *defined by* (26). *The equivalence class* $\beta^0{}_p$ *of such bases producing this metric,* Δ_{ij} *in* (27), *corresponds to* B^{0*}_p *for the one-forms.*

We now proceed to construct the structure equations with respect to $F^2{}_p$, defining the analogs of the connection and curvature forms along the way. The metric induced isomorphism between l_0 and $F^2{}_p$ discussed in Footnote 8 will play an important role. First, we note that, by lowering an index on the matrix of connection forms $\{\omega^a{}_b\}$, we obtain a matrix of forms $\{\Gamma^i\}$, representable by the $f^i{}_{ab}$. Thus there is a one-to-one relationship between possible matrices of connection forms and matrices $\{\Gamma^i\}$ of one-forms defined by

$$\omega_{ab}=\Gamma_i f^i{}_{ab} \quad (\Gamma_i\equiv\Delta_{ij}\Gamma^j). \tag{38}$$

In fact, we can use (27) to explicitly evaluate the Γ^i

$$\Gamma^i\equiv-\tfrac{1}{4}\epsilon^{abcd}\omega_{ab}f^i{}_{cd}, \tag{39}$$

where now ϵ^{abcd} is the totally antisymmetric symbol with $\epsilon^{0123}=-1$. On the other hand, by using the metric to raise the first index of the components of elements of $F^2{}_\rho$, we can introduce the Lie algebra bracket operation on $F^2{}_\rho$

$$[f^i,f^j]_{ab}\equiv f^i{}_a{}^c f^j{}_{cb}-f^j{}_a{}^c f^i{}_{cb}. \tag{40}$$

Because of the assumed completeness of the set $f^i{}_{ab}$, we can find a set of constants (we are always assuming the $f^i{}_{ab}$ to be fixed and constant from point to point in M) $S^{ij}{}_k$ such that

$$[f^i,\ f^j]=S^{ij}{}_k f^k. \tag{41}$$

The $S^{ij}{}_k$ will of course be recognized as the structure constants of l_0 with respect to the basis defined by the $f^i{}_{ab}$.

With this background it is now a straightforward matter to write the differential structure equations. Using (26), (20), (38), and (41), we get

$$d\phi=[\Gamma,\phi] \tag{42}$$

or, in terms of components,

$$d\phi^i=S^{ij}{}_k\Gamma_j\phi^k. \tag{43}$$

Finally, the curvature equations (22) become

$$d\Gamma_i+\tfrac{1}{2}S^{jk}{}_i\Gamma_j\wedge\Gamma_k=R_{ij}\phi^j, \tag{44}$$

in which the curvature is represented by the 6×6 matrix R_{ij}.

Thus, we have a complete $F^2{}_\rho$ *representation of the metric, connection and curvature, described by* (27), (43), *and* (44). *Once a set of six independent two-forms* ϕ^i *is given satisfying* (27) *for the* Δ^{ij} *defined by the* $f^i{}_{ab}$, *we can decompose them uniquely into a basis* ρ *satisfying* (26). *The differentials of the* ϕ^i *then give the connection forms* Γ^i *by way of* (43), *in which the* $S^{ij}{}_k$ *are also determined by the* $f^i{}_{ab}$ *as in* (41). As they stand, these equations do not appear to offer any advantage over the one-form equations (15), (20), (19), and (22). However, in the next section, we will find that the introduction of the complex structure J on $F^2{}_\rho$, in terms of the star duality operator *which can be done only for Minkowski signature metrices and dimension four,* enables us to re-express these equations in such a way as to produce a significant simplification of general relativistic field equations, especially the vacuum Einstein equations, in various representations.

COMPLEX STRUCTURES

The important role of $F^2{}_\rho$, or equivalently its isomorphic image l_0, in the geometric structure equations given above leads us to investigate the structure of these spaces more closely. We will find a natural negative-square endomorphism on these spaces, J, which satisfies the condition of being a complex structure in the mathematical sense. Using this, we will study the form of the complexifications of $F^2{}_\rho$ and l_0, leading naturally to complex representations of the geometry associated with representations of the Lorentz Lie algebra. The most important of these representations correspond to the group $SO(3,\ C)$, $SL(2,\ C)$, the latter being the spinor representation.

Now, let us define an operator J on $F^2{}_\rho$ in terms of the Minkowski-signature metric and the alternating, totally antisymmetric symbol, ϵ_{abcd} for which

$$\epsilon_{0123}=-\epsilon^{0123}=+1. \tag{45}$$

Let Ψ be a two-form, having components Ψ_{ab} with respect to the basis ρ for which the metric is γ_{ab}. Define $J\Psi$

$$J\Psi=-\tfrac{1}{4}\gamma^{ab}\gamma^{cd}\Psi_{ac}\epsilon_{bdnm}\rho^n\wedge\rho^m. \tag{46}$$

Because we are dealing with a space—time of dimension four, J is a linear map of $F^2{}_\rho$ onto itself, and, because of the indefinite signature of the metric,

$$J^2=-1, \tag{47}$$

where 1 is the identity map on $F^2{}_\rho$. Such a linear map of a vector space onto itself is called a "complex structure." It will also be recognized, perhaps in more familiar form, as the "star operator," generally represented by the symbol $*$. Here we have chosen the symbol J for two reasons: first, to emphasize its role in the complexification of the geometric structures in which it will be replaced by the usual imaginary number $i\equiv\sqrt{-1}$, and, second, to stimulate investigation into the possibilities of regarding J as a variable operator in

terms of which general relativistic geometries can be described. This second approach will be developed in other papers.

Now let us recall that the curvature can also be expressed in this formalism as an operator of $F^2{}_p$ onto itself as described in Eq. (44). Thus define the operator $\mathbf{R}$ on $F^2{}_p$ in terms of the 6×6 matrix $\mathcal{R}_{ij}$, and the two-form metric Δ_{ij},

$$\mathbf{R}(\phi^i)\equiv\Delta^{ij}\mathcal{R}_{jk}\phi^k. \tag{48}$$

Now decompose the curvature operator $\mathbf{R}$ with respect to the complex structure J,

$$\mathbf{R}=\mathbf{P}+\mathbf{Q}, \tag{49}$$

where P and Q are the parts of $\mathbf{R}$ that commute and anticommute with J respectively. Therefore,

$$\mathbf{P}=(\mathbf{R}-\mathbf{JRJ})/2, \tag{50}$$

$$\mathbf{Q}=(\mathbf{R}+\mathbf{JRJ})/2. \tag{51}$$

It is now a straightforward algebraic matter to show that this decomposition corresponds to the more familiar decomposition of the components of the Riemann curvature tensor R_{abcd} into Weyl conformal part, traceless Ricci part, and curvature scalar part. In fact, by using

$$\mathcal{R}_{ij}f^i{}_{ab}f^j{}_{cd}=R_{abcd}, \tag{52}$$

and the definition of J in (46), it follows that P depends on the Weyl plus curvature scalar part and Q on the traceless Ricci part, $R_{ab}-\frac{1}{4}\gamma_{ab}R$. Thus *the weakend Einstein equations can be expressed by saying that the curvature operator on $F^2{}_p$ commutes with the complex structure* J. See Chap. 13 of Misner, Thorne, and Wheeler[5] for details of these calculations in terms of one-form components, R_{abcd}.

In order to take advantage of this result, we note that the condition (47) satisfied by J leads naturally to a complexification of $F^2{}_p$ in which the $F^2{}_p$ regarded as a *real* vector space of six dimensions can be replaced by a *complex* vector space of three dimensions with the imaginary scalar multiplication being associated with J. Thus, suppose that ϕ is some basis for $F^2{}_p$, and let capital Latin letters, A, B, $C,\ldots$, assume only the values 1, 2, 3. Consider then the three complex dimensional vector space, $F^{2c}{}_p$ spanned by the three complex two-forms σ^A,

$$\sigma^A\equiv\phi^A+i\mathrm{J}\phi^A,\quad i^2=-1. \tag{53}$$

Thus, $F^{2c}{}_p$ is the set of complex two-forms σ that can be written

$$\sigma=z_A\sigma^A, \tag{54}$$

where the z_A are three complex numbers. Similarly, the conjugate space, $\bar{F}^{2c}{}_p$ is defined in terms of the basis $\bar{\sigma}^A$,

$$\bar{\sigma}^A\equiv\phi^A-i\mathrm{J}\phi^A. \tag{55}$$

It is easy to see that there is a one-to-one relationship between $F^2{}_p$ and $F^{2c}{}_p$ defined as follows. If $\Psi\in F^2{}_p$, let the components of Ψ with respect to the basis $(\phi^A,\mathrm{J}\phi^A)$ be the six real numbers $(x_A,\ y_A)$ so that Ψ can be uniquely written

$$\Psi=x_A\phi^A+y_A\mathrm{J}\phi^A. \tag{56}$$

Defining the three complex numbers z_A by

$$z_A=x_A-iy_A, \tag{57}$$

we can then define the image of Ψ in $F^{2c}{}_p$ as Ψ^c,

$$\Psi^c\equiv z_A\sigma^A. \tag{58}$$

Conversely, the real Ψ can be obtained from Ψ^c by

$$\Psi=(\Psi^c+\bar{\Psi}^c)/2=(z_A\sigma^A+\bar{z}_A\bar{\sigma}^A)/2. \tag{59}$$

This relationship is clearly one-to-one and its consistency depends on the condition (47). It is this reduction of the dimension of the vector space from six to three by complexification here, and in the structure equations below, that makes this approach helpful.

Next, we note that J commutes with the $\wedge$ operator,

$$(\mathrm{J}\Phi)\wedge\Psi=\Phi\wedge\mathrm{J}\Psi, \tag{60}$$

and

$$(\mathrm{J}\Phi)\wedge(\mathrm{J}\Psi)=-\Phi\wedge\Psi. \tag{61}$$

Thus, regarding as an inner product on $F^2{}_p$, with metric Δ^{ij}, we see that this six-dimensional array commutes with J and respects the division between σ^A and $\bar{\sigma}^A$. That is,

$$\sigma^A\wedge\sigma^B=\Delta^{AB}I, \tag{62}$$

$$\sigma^A\wedge\bar{\sigma}^B=0, \tag{63}$$

$$\bar{\sigma}^A\wedge\bar{\sigma}^B=\bar{\Delta}^{AB}I. \tag{64}$$

Note that (64) is merely the complex conjugate of (62) so that when working over the complex field, it is necessary only to satisfy (62) and (63).

Similarly, it is easy to see that J commutes with the Lie bracket operation for two-forms. Thus, from the definitions (40) and (46) it follows directly that for any two-forms Φ and Ψ

$$[\mathrm{J}\Phi,\ \Psi]=-[\Phi,\ \mathrm{J}\Psi], \tag{65}$$

so that

$$[\mathrm{J}\Phi,\ \mathrm{J}\Psi]=[\Phi,\ \Psi]. \tag{66}$$

When these results are applied to the basis σ^A defined in (53) and (55), it follows immediately that the real structure constants $S^{ij}{}_k$ for which the indices run from one to six can be replaced by complex $S^{AB}{}_C$ with indices running only from one to three,

$$[\sigma^A,\ \sigma^B]=S^{AB}{}_C\sigma^C, \tag{67}$$

$$[\sigma^A,\ \bar{\sigma}^B]=0, \tag{68}$$

$$[\bar{\sigma}^A,\ \bar{\sigma}^B]=\bar{S}^{AB}{}_C\bar{\sigma}^C. \tag{69}$$

Finally, the differential structure equations (43) and (44) become

$$d\sigma^A=S^{AB}{}_C\Gamma_B\wedge\sigma^C, \tag{70}$$

$$d\Gamma_C+\tfrac{1}{2}S^{AB}{}_C\Gamma_A\wedge\Gamma_B=P_{CA}\sigma^A+Q_{CA}\bar{\sigma}^A. \tag{71}$$

The matrices P and Q correspond to those defined in (50) and (51) so that the Einstein equations correspond to the condition

$$Q_{AB}=0. \tag{72}$$

Further, P must be symmetric and, for the stronger

Einstein condition, $R=0$, traceless. Clearly then, the Einstein equations have a formal three-dimensional structure over the complex field. Of course, we must recall that the complex basis σ^A, while containing only three elements, consists of two-forms over a four-dimensional M.[9] This apparent three-dimensional structure will be analyzed in more detail in the following section.

REPRESENTATIONS OF l_0 AND THE EINSTEIN EQUATIONS

Now consider how various representations of the Einstein equations correspond to representations of l_0 in the two-form structure equations.

First, we note that we can freely choose the metric Δ_{AB} to be any nonsingular complex symmetric matrix by appropriate choice of basis, $(\sigma^A, \bar{\sigma}^A)$. In particular, we can set

$$\Delta_{AB}=i\delta_{AB}, \tag{73}$$

corresponding to

$$\sigma^A=\rho^0\wedge\rho^A+i\rho^B\wedge\rho^C \quad [A,B,C=\text{cyclic }(1,2,3)]. \tag{74}$$

Thus, when the basis σ^A is changed by an element of L_0, the σ^A are transformed in such a way as to leave the inner product form δ_{AB} unchanged. Such a transformation belongs, of course, to the complex group $SO(3,C)$ and this procedure can be described as the $SO(3,C)$ representation of l_0.

For this basis, the structure constants have the form

$$S^{AB}{}_C=i\epsilon^{ABC}, \tag{75}$$

since lowering and raising indices is accomplished by multiplying by i or $-i$, respectively, for the metric in (73). The Einstein structure equations become

$$d\sigma^A=i\epsilon^{ABC}\Gamma_B\wedge\sigma^C, \tag{76}$$

$$d\Gamma_C+(i/2)\epsilon^{ABC}\Gamma_A\wedge\Gamma_C=P_{CA}\sigma^A. \tag{77}$$

The essential three-dimensional form of these equations can be more clearly displayed by using a three-vector notation, $\sigma=(\sigma^A)$, $\Gamma=(\Gamma^A)$, $\mathrm{P}=(P^A{}_B)$, with the special symbol $\otimes$ representing the combination of hook form-product with three-dimensional vector "cross product." Thus (76) and (77) become[2]

$$d\sigma=-\Gamma\otimes\sigma, \tag{78}$$

$$d\Gamma-\Gamma\otimes\Gamma=i\mathrm{P}\cdot\sigma. \tag{79}$$

This approach is obviously well suited for the Petrov analysis of the algebraic structure of the Riemann tensors of Einstein spaces. In fact, the basis (σ^A) defining the geometry is arbitary up to a complex orthogonal transformation. Thus, the matrix P which carries the content of the curvature tensor can be freely transformed

$$\mathrm{P}\to\mathrm{P}'=\mathrm{S}^{-1}\mathrm{PS}, \quad \mathrm{S}\in SO(3,C). \tag{80}$$

Petrov[2] found that most general P could be transformed as in (80) to one of essentially three different canonical forms,

$$\text{type I: } \mathrm{P}=\begin{pmatrix}\alpha&0&0\\0&\beta&0\\0&0&\gamma\end{pmatrix}, \tag{81}$$

$$\text{type II: } \mathrm{P}=\begin{pmatrix}\alpha&0&0\\0&\beta-i&1\\0&1&\beta+i\end{pmatrix}, \tag{82}$$

$$\text{type III: } \mathrm{P}=\begin{pmatrix}\alpha&1&0\\1&\alpha&i\\0&i&\alpha\end{pmatrix}, \tag{83}$$

where α, β, γ are arbitrary complex numbers, which can vary from point to point, of course. Further, the stronger Einstein condition, $R=0$, requires $\alpha+\beta+\gamma=0$ for type I, $\alpha+2\beta=0$ for type II, and $\alpha=0$ for type III.

Hence, one natural approach is to choose for P one of these types, insert it into (78) and (79), thus reducing the freedom of choice for the basis σ^A, at least in the nondegenerate cases. This technique has been used successfully to reduce the type III problem to one equation for one-function, for which a one-function family of solutions has been obtained.[10] Similar results can be obtained for the type N case (type II with $\alpha=\beta=0$).

However, the most intriguing possibility seems to lie in the general type I case in which no two of the α, β, γ are equal. Further, adding the $\alpha+\beta+\gamma=0$ condition reduces the triple (α, β, γ) to two independent complex functions, precisely the number (if they are functionally independent) to uniquely determine a canonical set of coordinates for the base manifold M, as well as a fully determined basis for two-forms and thus one-forms. In other words, the "most general" Petrov case is the one in which the coordinate and frame degeneracy vanishes, suggesting that it might also be the one for which solutions could be easily obtained. Alternatively, it might be profitable to postpone the $\alpha+\beta+\gamma=0$ condition and regard the α, β, γ as a set of complex three-dimensional coordinates, solve the resulting three-dimensional structure equations and then search for ways to map the forms σ^A into two-forms over a four-manifold, satisfying also (62) and (63).

A closely related approach is based on the choice

$$\Delta_{AB}=i\begin{pmatrix}0&1&0\\1&0&0\\0&0&1\end{pmatrix}. \tag{84}$$

The structure equations can still be written as (78) and (79) but where the $\otimes$ symbol means

$$(\mathrm{u}\otimes\mathrm{v})^1=u^1v^3-u^3v^1, \tag{85}$$

$$(\mathrm{u}\otimes\mathrm{v})^2=u^3v^2-u^2v^3, \tag{86}$$

$$(\mathrm{u}\otimes\mathrm{v})^3=u^2v^1-u^1v^2, \tag{87}$$

for a pair of complex three-vectors u and v. The algebraic conditions (62) and (63) for the σ^A imply

$$\sigma^1\wedge\sigma^1=\sigma^2\wedge\sigma^2=0, \tag{88}$$

$$\sigma^1\wedge\sigma^2=\sigma^3\wedge\sigma^3=iI. \tag{89}$$

From these it follows that the σ^A can be decomposed in terms of two real, null, one-forms κ, λ and a complex-complex conjugate pair μ, $\bar{\mu}$,

$$\sigma^1=\sqrt{2}\,\kappa\wedge\mu, \tag{90}$$

$$\sigma^2=\sqrt{2}\,\lambda\wedge\bar{\mu}, \tag{91}$$

$$\sigma^3=i(\kappa\wedge\lambda+\mu\wedge\bar{\mu}). \tag{92}$$

The metric is

$$ds^2=2\kappa\lambda+2\mu\bar{\mu}. \tag{93}$$

This approach is well suited for the study of gravitational radiation problems. In fact, the null vector dual to κ can be chosen to be a "principal null direction,"[11] tangent to null geodesics representing gravitational rays by requiring $P^1{}_2=0$. This last condition can be met for any but the algebraically nondegenerate type I case.

Let us now fix Δ_{AB} say to have the value given in (73), and consider the possibility of representing two-form geometric quantities in terms of various representations of l_0. Such an n-dimensional representation can be specified in terms of matrices $T^{A\alpha}{}_\beta$, $\alpha,\beta=1,\ldots,n$, whose matrix commutators have the $S^{AB}{}_C$ for structure constants. In this case,

$$T^{A\alpha}{}_\mu T^{B\mu}{}_\beta-T^{B\alpha}{}_\mu T^{A\mu}{}_\beta=i\epsilon^{ABC}T^{C\alpha}{}_\beta. \tag{94}$$

We can now use these matrices to represent the basis σ^A. First, let $\tau^\alpha{}_\beta$ denote a matrix of two-forms lying in the linear matrix space spanned by the T^A, that is,

$$\tau^\alpha{}_\beta=\phi_A T^{A\alpha}{}_\beta \tag{95}$$

for some (complex) two-forms ϕ_A. It is now easy to see that a necessary and sufficient condition that the matrix $\tau^\alpha{}_\beta$ provide a basis satisfying (62), (63), and (73) is that

$$\tau^\alpha{}_\beta\wedge\tau^\mu{}_\nu=t^\alpha{}_\beta{}^\mu{}_\nu I, \tag{96}$$

$$\tau^\alpha{}_\beta\wedge\bar{\tau}^\mu{}_\nu=0, \tag{97}$$

where

$$t^\alpha{}_\beta{}^\mu{}_\nu\equiv\Delta_{AB}T^{A\alpha}{}_\beta T^{B\mu}{}_\nu. \tag{98}$$

The quantities $t^\alpha{}_\beta{}^\mu{}_\nu$ will be recognized as the components in this representation of the metric associated with the Killing inner product defined on semisimple Lie algebras.[12] If we further restrict ourselves to faithful representations, it is clear then that (96), (97), and (98) are the conditions that $\tau^\alpha{}_\beta$ be decomposable according to (95) into a basis for $F^2{}_p$ satisfying (62), (63), and (73).

The differential structure equations are now easy to write,

$$d\tau=-[\Gamma,\tau], \tag{99}$$

which defines the connection forms $\Gamma=\{\Gamma^\alpha{}_\beta\}$ in this representation and

$$d\Gamma-\tfrac{1}{2}[\Gamma,\Gamma]=i\mathbf{P}\cdot\tau, \tag{100}$$

which defines the Einstein curvature forms, $\mathbf{P}=\{P^\alpha{}_\beta{}^\mu{}_\nu\}$. Again the satisfaction of the weakened Einstein equations is equivalent to the absence of the complex conjugate terms $\bar{\tau}$ on the right-hand side of (100).

For example, consider the $SO(3,C)$ representation. In this formalism, we would represent the basis by 3×3 antisymmetric matrices of complex two-forms τ. Because of the three-dimensional characteristic and the antisymmetry, however, we can go over to a three-vector formalism, $\sigma^A=\frac{1}{2}\epsilon^{ABC}\tau^B{}_C$. This, of course, leads directly to (78) and (79) above.

The basic representation of l_0 is, of course, the spinor one. Here $n=2$ and the bases are represented in terms of 2×2 matrices $\tau^\alpha{}_\beta$ of complex two-forms belonging to the Lie algebra of $SL(2,C)$ and hence required to be traceless,

$$\tau^\alpha{}_\alpha=0. \tag{101}$$

This implies that there are only three independent components, say $\tau^1{}_1$, $\tau^1{}_2$, $\tau^2{}_1$ with $\tau^2{}_2=-\tau^1{}_1$. By using Pauli matrices for the $T^{A\alpha}{}_\beta$, it is seen that the conditions (96), (97), and (98) reduce to

$$\tau^1{}_1\wedge\tau^1{}_1=(-i/4)I, \tag{102}$$

$$\tau^1{}_2\wedge\tau^2{}_1=(-i/2)I, \tag{103}$$

and all other products zero. Thus, these conditions are equivalent to (88) and (89), so that, again, the basis matrix $\tau^\alpha{}_\beta$ can be decomposed into the radiation adapted one-form frame $(\kappa,\lambda,\mu,\bar{\mu})$, as in (90), (91), and (92).

The differential structure equations are given by (99) and (100) in which the curvature is represented by the matrix $P^\alpha{}_\beta{}^\mu{}_\nu$. If we add the $R=0$ condition, the full set of algebraic conditions on this spinor representation of the Einstein curvature tensor becomes

$$P^\alpha{}_\beta{}^\mu{}_\nu=P^\mu{}_\nu{}^\alpha{}_\beta, \tag{104}$$

$$P^\alpha{}_\alpha{}^\mu{}_\nu=P^\alpha{}_\beta{}^\beta{}_\alpha=0. \tag{105}$$

By noting that the indices assume only the values 1, 2, it is easy to see that there are only five independent components for a $P^\alpha{}_\beta{}^\mu{}_\nu$ satisfying (104) and (105). Penrose[3] has fully analyzed the properties of such a matrix in a manner analogous to that of Petrov for the $SO(3,C)$ representation, resulting in a "principal spinor" decomposition.

Finally, it should be noted that this formalism is well adapted to an extension in which the $F^2{}_p$, or l_0, structure at each point is regarded as the geometric part of a larger structure. For example, l_0 can be imbedded in the conformal Lie algebra referring perhaps to electromagnetic or quantum mechanical symmetries. If the complex structure can similarly be imbedded, the purely differential structure equations, (70) and (71), can be naturally extended to a set involving a mixing of the geometry and the other symmetries. This problem will be considered in another paper.

ACKNOWLEDGMENTS

The author would like to express his gratitude to the Physics Department of Princeton University, and especially to R.H. Dicke's Gravity Research Group, for their hospitality during this academic year.

*This research was supported in part by the National Science Foundations.

†On leave from Loyola University, New Orleans, Louisiana.

[1] In this paper the weakened form of the Einstein equations, that is, $R_{ab}-\frac{1}{4}g_{ab}R=0$ will be studied for the most part. The additional requirement $R=0$ can be imposed by demanding that the traces of the quantities on the right-hand side of the various forms of the two-form curvature structure equations vanish. Thus, for example, $R=0$ is equivalent to $P_{AB}\Delta^{AB}=0$, for P_{AB} given in (71).

[2]R. Debever, Cahiers Phys. 168, 303 (1964); A. Z. Petrov, Sci. Trans. Kazan State Univ. 114, 55 (1954); or, A. Z. Petrov, *Einstein Spaces* (Pergamon, New York, 1969).
[3]L. Witten, Phys. Rev. 113, 357 (1959); R. Penrose, Ann. Phys. (N.Y.) 10, 171 (1960).
[4]S. Kobayashi and K. Nomizu, *Foundations of Differential Geometry* (Interscience, New York, Vol. 1, 1963, Vol. 2, 1969).
[5]C. Misner, K. Thorne, and J. Wheeler, *Gravitation* (Freeman, San Francisco, 1973).
[6]When mention is made of some property "at p," we will understand that the discussion refers not only to p itself, but differentiably in some neighborhood of p. Thus, for example, when we discuss a vector v, at p, we assume that v is actually a differentiable vector field in some neighborhood of p.
[7]The union of the collection of bases at different points of M, suitably topologized, is just the principal bundle of frames over M.
[8]Specifically, with repsect to a given basis Lorentz basis, $\rho \epsilon B_p^0$, the set $T^2{}_p$ and $F^2{}_p$ can be described as sets of antisymmetric matrices t^{ab} and f_{ab}, respectively, while l_0 corresponds to the set of matrices λ^a_b such that $\eta_{ac}\lambda^c{}_b = -\eta_{bc}\lambda^c{}_b$. Thus, η_{ac} as a lowering operator provides the isomorphism between $T^2{}_p$ and l_0 and then between l_0 and $F^2{}_p$. This procedure is clearly independent of the original choice of Lorentz basis.
[9]An interesting question arises here concerning the possibility of regarding **J** as an "almost complex structure" on M in some sense analogous to that used for complex structures on tangent vector spaces. Thus, we could ask whether or not **J** can be defined in terms of some complex coordinate operations on M. Such a condition is probably equivalent to flatness.
[10]C. Brans, J. Math. Phys. 12, 1616 (1971).
[11]R. Debever, Bull. Soc. Math. Belg. 10, 112 (1959); R. Sachs, Proc. Roy. Soc. (London) A 264, 309 (1961).
[12]See, for example, N. Jacobson, *Lie Algebras* (Wiley, New York, 1962).

Absence of trivial subrepresentations from tensor products of unitary representations of pseudo-orthogonal groups*

S. A. Fulling†

Physics Department, University of Wisconsin-Milwaukee, Milwaukee, Wisconsin 53201
(Received 1 March 1973; revised manuscript received 11 October 1973)

A unitary tensor product representation of the group $SO_0(1,n)$, $n=2,3,4$, does not contain the trivial representation as a discrete direct summand unless each of the factors does.

INTRODUCTION

The following question arose in the course of an investigation[1] in quantum field theory: Can a tensor product representation[2] $D_1 \otimes D_2 \otimes \cdots \otimes D_l$, where each D_j is a (continuous) unitary representation of the group $SO_0(1,n)$, contain a vector which is invariant under the action of all the elements of the group? In other words, can the decomposition of such a representation into irreducible representations include the trivial (identity) representation? The answer is negative, at least for $n=2, 3$, or 4, unless each of the factor representations contains such an invariant vector (in which case the product of those vectors is invariant under the product representation).

This theorem can be applied to prove uniqueness of the vacuum[3] in a quantum field theory invariant under the de Sitter group $SO_0(1,4)$, or in an analogous model of space–time dimension 2, invariant under $SO_0(1,2)$. There[4] one is led to construct a Fock space

$$\mathcal{F} = \sum_{l=0}^{\infty}{}^{\oplus} \mathcal{H}_l,$$

$$\mathcal{H}_0 = \mathbb{C}, \quad \mathcal{H}_1 = D, \quad \mathcal{H}_2 = D \otimes D, \cdots,$$

where D is a Hilbert space supporting an irreducible unitary representation (or ray representation) of $SO_0(1,n)$. D is interpreted as the space of possible quantum states of a single particle. Since each $\mathcal{H}_n$ is invariant under the natural action of $SO_0(1,n)$, the theorem implies that the only invariant vectors in $\mathcal{F}$ are those in the one-dimensional subspace $\mathcal{H}_0$, representing the state with no particles present.

The theorem will be proved first for $SO_0(1,2)$ and then extended to the two next higher dimensions.

TWO IRREDUCIBLE REPRESENTATIONS OF $SO_0(1,2)$

The irreducible unitary representations of $SO_0(1,2)$ can be described in infinitesimal terms as follows.[5] Let J_0, J_1, J_2 be the standard choice of basis for the Lie algebra of the group, and $Q = J_1^2 + J_2^2 - J_0^2$ be the Casimir invariant operator. Each representation possesses a basis consisting of vectors $|q;p\rangle$, where

$$Q|q;p\rangle = q|q;p\rangle, \quad J_0|q;p\rangle = p|q;p\rangle,$$
$$(J_1 \pm iJ_2)|q;p\rangle = [q + p(p \pm 1)]^{1/2}|q;\, p \pm 1\rangle.$$

The fixed number q parametrizes the representation (but does not determine it uniquely[6]), and the index p labels the basis vectors within the representation. Besides the trivial representation ($q=0$, $p=0$) there are representations corresponding to all positive values of q (the continuous series) and to certain other values (the discrete series). In each nontrivial representation the range of p is a discrete set of points which extends by unit steps to infinity in at least one direction. These results are derived by observing that in a unitary representation the factors $[q + p(p \pm 1)]^{1/2}$ must be real numbers, and that the sequence of p's in a representation can terminate only when one of these factors vanishes.

Theorem 1: Let D_1 and D_2 be any two irreducible unitary representations of $SO_0(1,2)$, not both trivial. Then $D_1 \otimes D_2$ does not contain the trivial representation as a discrete direct summand.

Proof[7]: The argument is analogous to Pukánszky's determination of the discrete representations which appear in the tensor product of two continuous representations.[8] Let the values of the Casimir operator for D_1 and D_2 be q_1 and q_2, respectively. We must show that no vector

$$\Psi = \sum_{p_1,p_2} a_{p_1 p_2} |q_1;p_1\rangle \otimes |q_2;p_2\rangle \quad \text{(infinite sum)}$$

in the tensor product space is annihilated by all the basis elements of the Lie algebra of the tensor product representation. (The latter have the form

$$J_i = J_i^{(D_1)} \otimes 1 + 1 \otimes J_i^{(D_2)}$$

in an obvious notation.) The condition $J_0\Psi = 0$ implies that Ψ is of the form

$$\Psi = \sum_p a_p |q_1;p\rangle \otimes |q_2;-p\rangle.$$

Requiring that $(J_1 \pm iJ_2)\Psi = 0$ leads (after taking a scalar product with each of the basis vectors) to the equations

$$a_p[q_1 + p(p+1)]^{1/2} + a_{p+1}[q_2 + p(p+1)]^{1/2} = 0,$$
$$a_p[q_2 + p(p+1)]^{1/2} + a_{p+1}[q_1 + p(p+1)]^{1/2} = 0.$$

These are consistent only if $q_1 = q_2$, and one then has $a_{p+1} = -a_p$ for all p in the range of p in the representation D_1. [One must check points where the coefficient $q_1 + p(p+1)$ vanishes; but these simply mark the boundary of the representation D_1.] It follows that nonvanishing a's satisfying $a_{p+1} = -a_p$ must extend to infinity in at least one direction. Consequently, the sequence $\{a_p\}$ is not square-summable; no normalizable invariant vector can exist.

Remark: The presence of an irreducible representation D in a tensor product $D_1 \otimes D_2$ is often detected by coupling the basis vectors of D_1, D_2, and D^{-1T} to form an invariant object,[9] where D^{-1T} is the representation contragredient to D. In the case of $SO_0(1,2)$ this technique has been used to calculate Clebsch–Gordan coefficients and matrix elements of tensorial operators.[10] The principle involved is demonstrated in the proof of Theorem 1, where one succeeded in constructing an in-

variant from (and only from) two irreducible representations with the same value of q and values of p of opposite signs[11]; such representations are mutually contragredient. For the infinite-dimensional representations of noncompact groups the invariant is not normalizable, in general, and hence does not constitute a trivial subrepresentation of $D_1 \otimes D_2 \otimes D^{-1T}$.

MANY REDUCIBLE REPRESENTATIONS OF SO_0 (1,2)

Theorem 2: Let D_j $(j=1,\ldots,l)$ be unitary representations of $SO_0(1,2)$, not all of which contain the trivial representation discretely. Then $D_1 \otimes D_2 \otimes \cdots \otimes D_l$ does not contain the trivial representation discretely.

Proof: Any representation, for instance D_l, is a direct integral[12] of irreducible representations:

$$D_l = \int^{\oplus} d\nu(s)[j_s^{(\nu)} \cdot D(s)].$$

Here $D(s)$ ranges over the inequivalent irreducible representations, which are labeled by a parameter s; ν is a measure on the space of allowed values of s; $j_s^{(\nu)}$ is the multiplicity (which may be countable infinity) with which $D(s)$ occurs in D_l; and $[j \cdot D]$ denotes the direct sum of j copies of the representation D:

$$[j \cdot D] = D \oplus \cdots \oplus D \quad (j \text{ terms}).$$

We use such notations as D_l and $[j \cdot D]$ to stand both for a representation in the abstract sense and for the Hilbert space in which the representation acts. In the latter sense an element of a direct integral $\int^{\oplus} d\nu(s)\mathcal{H}(s)$ is a function $f(s)$ taking values in $\mathcal{H}(s)$ which is square-integrable with respect to ν:

$$\int d\nu(s)\|f(s)\|^2_{\mathcal{H}(s)} < \infty.$$

(We shall always realize the Hilbert spaces as generalized L^2-spaces in this way.) In particular, a member of D_l can be thought of as a function $f(s, i_s^{(\nu)}, p^{(s)})$, where $i_s^{(\nu)}$ is an integer in the range $1 \leq i_s^{(\nu)} \leq j_s^{(\nu)}$ and, for fixed s and $i_s^{(\nu)}$, $f(s, i_s^{(\nu)}, p^{(s)})$ is the coefficient of an element of the representation space $D(s)$ with respect to a basis of eigenvectors of J_0. (Such coefficients were abbreviated as a_p above.) If the support of $\nu(s)$ consists of discrete points, the direct integral is an ordinary direct sum. All the discussion below may be rephrased in terms of the spectral analysis of the operators Q and J_0 in the various representation spaces.

We now prove the theorem by induction. (The major step is from irreducible to reducible representations, after which the extension to more than two factors is immediate.) Theorem 1 says that the product of two irreducible representations, $D(r)$ and $D(s)$, has a direct integral decomposition

$$D(r) \otimes D(s) = \int^{\oplus} d\omega_{r,s}(t)\,[j_t^{(\omega)} \cdot D(t)] \tag{1}$$

in which t_0, the value of t corresponding to $q=0$ and $p=0$ (the trivial representation), does not appear as a discrete point [i.e., $\omega_{rs}(\{t_0\}) > 0$ is not true, where $\{t_0\}$ is the set whose only member is t_0]. We assume that this statement has been extended to a tensor product of $l-1$ factors ($l \geq 2$), which may themselves be reducible (direct integrals). Thus

$$D_1 \otimes \cdots \otimes D_{l-1} = \int^{\oplus} d\mu(r)\,[j_r^{(\mu)} \cdot D(r)],$$

where the same restriction with regard to the trivial representation holds. Then we have

$$\begin{aligned} D_1 \otimes \cdots \otimes D_l &= \left\{\int^{\oplus} d\mu(r)\,[j_r^{(\mu)} \cdot D(r)]\right\} \otimes \left\{\int^{\oplus} d\nu(s)\,[j_s^{(\nu)} \cdot D(s)]\right\} \\ &= \int^{\oplus} d\mu(r)\,d\nu(s)\,[(j_r^{(\mu)} j_s^{(\nu)}) \cdot (D(r) \otimes D(s))]. \end{aligned} \tag{2}$$

The associative law used here is obvious if the direct integrals are direct sums; that it holds in general is most easily seen by observing[13] that the elements of the space on either side of the equality sign are functions of the type $f(r, s, i_r^{(\mu)}, i_s^{(\nu)}, p^{(r)}, p^{(s)})$.

Combining Eqs. (1) and (2), we have

$$\begin{aligned} D_1 \otimes \cdots \otimes D_l = \int^{\oplus} d\mu(r)\,d\nu(s)\,d\omega_{r,s}(t) \\ \times [(j_r^{(\mu)} j_s^{(\nu)} j_t^{(\omega)}) \cdot D(t)]. \end{aligned} \tag{3}$$

An element of this space is of the form

$$f(r, s, t, i_r^{(\mu)}, i_s^{(\nu)}, i_t^{(\omega_{r,s})}, p^{(t)}).$$

If this vector is different from zero in the L^2-space sense, then, for some set of values (r, s) with positive measure, the functions $f(r, s, t, \cdots)$ with fixed r and s are nonzero on sets of positive measure in t. Hence it is impossible that f could have its support concentrated at $t = t_0$—which was to be proved. An intuition for this part of the proof can be acquired by imagining the integrals in Eq. (3) replaced by direct sums; then one would just "collect terms" corresponding to each value of t and observe that t_0 is never present.

The proof of Theorem 2 involves only the general properties of direct integrals. Consequently, it applies, with appropriate changes in notation, to any Type I group (see Ref. 12) for which the analog of Theorem 1 can be proved.

REPRESENTATIONS OF SO_0 (1,n)

By arguments closely analogous to those just given for $SO_0(1,2)$, the corresponding theorems can be proved for $SO_0(1,3)$ and $SO_0(1,4)$, on the basis of explicit formulas for the irreducible representations of their Lie algebras. A rather obvious conjecture is that the conclusions are valid for all $SO_0(1,n)$.

Theorem 3: Let D_j $(j=1,\ldots,l)$ be unitary representations of $SO_0(1,n)$, $n=2, 3$, or 4, not all of which contain the trivial representation discretely. Then $D_1 \otimes D_2 \otimes \cdots \otimes D_l$ does not contain the trivial representation discretely.

Proof for $SO_0(1,3)$: The representations of the Lie algebra corresponding to irreducible unitary group representations are given by Naimark.[14] For brevity only the information directly relevant to the proof will be cited here. There are six independent generators, H_j and F_j, $j=1,2,3$; the H_j generate a subgroup isomorphic to $SO(3)$. [Under the usual physical interpretation of $SO_0(1,3)$ the H_j generate rotations and the F_j generate Lorentz boosts.] The irreducible unitary representations are labeled by pairs of numbers (c, k_0), where (among other restrictions) k_0 is a nonnegative

integer. The most general vector in a representation $D \equiv (c, k_0)$ has the form

$$\sum_{k=k_0}^{\infty} \sum_{m=-k}^{k} a_{km} |D; k, m\rangle,$$

where

$$H_3 |D; k, m\rangle = m |D; k, m\rangle,$$
$$(H_1 \pm iH_2) |D; k, m\rangle = [k(k+1) - m(m \pm 1)]^{1/2} |D; k, m \pm 1\rangle,$$

and

$$F_3 |D; k, m\rangle = [k^2 - m^2]^{1/2} C(D; k) |D; k-1, m\rangle - mA(D; k) |D; k, m\rangle - [(k+1)^2 - m^2]^{1/2} C(D; k+1) |D; k+1, m\rangle,$$

with similar formulas for $F_1 \pm iF_2$. Here

$$A(c, k_0; k) = ick_0/[k(k+1)],$$
$$C(c, k_0; k) = ik^{-1}[(k^2 - k_0^2)(k^2 - c^2)/(4k^2 - 1)]^{1/2}.$$

Following the proof of Theorem 1, we consider the general vector in the tensor product of two irreducible representations, $D_1 \otimes D_2$:

$$\Psi = \sum_{k_1=k_{01}}^{\infty} \sum_{m_1=-k_1}^{k_1} \sum_{k_2=k_{02}}^{\infty} \sum_{m_2=-k_2}^{k_2} a_{k_1 m_1 k_2 m_2} |D_1; k_1, m_1\rangle \otimes |D_2; k_2, m_2\rangle.$$

If Ψ is invariant, it must be annihilated, in particular, by H_3 and $H_\pm$. Hence one sees, by direct verification or by knowledge of the $SO(3)$ Clebsch–Gordan coefficient

$$\langle k_1 k_2 m_1 m_2 | 0\,0\rangle = (-1)^{k_1 + m_1} (2k_1 + 1)^{-1/2} \delta_{k_1 k_2} \delta_{m_1, -m_2},$$

that

$$\Psi = \sum_{k=k_0}^{\infty} a_k (2k+1)^{1/2} |D_1, D_2; k, 0\rangle,$$

where

$$|D_1, D_2; k, 0\rangle = (2k+1)^{-1/2} \sum_{m=-k}^{k} (-1)^m |D_1; k, m\rangle \otimes |D_2; k, -m\rangle.$$

We operate upon this vector with

$$F_3 = F_3^{(D_1)} \otimes 1 + 1 \otimes F_3^{(D_2)},$$

obtaining, after some redefinition of summation indices,

$$F_3 \Psi = \sum_{k=k_0}^{\infty} \sum_{m=-k}^{k} (-1)^m \times \{[k^2 - m^2]^{1/2} [C(D_1; k) a_k - C(D_2; k) a_{k-1}] |D_1; k-1, m\rangle \otimes |D_2; k, -m\rangle - m a_k [A(D_1; k) - A(D_2; k)] |D_1; k, m\rangle \otimes |D_2; k, -m\rangle - [k^2 - m^2]^{1/2} [C(D_1; k) a_{k-1} - C(D_2; k) a_k] |D_1; k, m\rangle \otimes |D_2; k-1, -m\rangle\}.$$

Every coefficient in this expansion must vanish if Ψ is invariant; hence

$$A(D_1; k) = A(D_2; k),$$
$$C(D_1; k) a_k = C(D_2; k) a_{k-1},$$
$$C(D_2; k) a_k = C(D_1; k) a_{k-1}.$$

It follows that $C(D_1; k) = C(D_2; k)$ and, most importantly, that $a_k = a_{k-1}$ unless $C(D_1; k) = 0$. (The formulas for A and C then imply that $k_{01} = k_{02} \leqslant k_0$ and $c_1 = c_2$; i.e., $D_1 = D_2$.)

For a nontrivial unitary representation (c_1 imaginary or $|c_1| < 1$), we have $C(c_1, k_{01}; k) = 0$ only when $k = k_{01}$. Thus the coefficient a_k has the same value for all k in the representation ($k \geqslant k_0 = k_{01}$). Since the vectors $|D_1, D_2; k, 0\rangle$ are normalized, the "norm" (squared) of Ψ is

$$\sum_{k=k_0}^{\infty} |a_k|^2 (2k+1) = \infty,$$

and Ψ is not in the Hilbert space. This establishes the analog of Theorem 1.

The argument of Theorem 2 now applies to complete the proof.

Proof for $SO_0(1, 4)$: Since the ideas are the same as before but the explicit formulas become voluminous, we simply outline the argument. The representations are given in a convenient form by Ström.[15] There are ten generators: the M_j ($j = 1, 2, 3$) generate $SO(3)$, the P_j and M_j generate $SO(4)$, the N_j and M_j generate $SO_0(1, 3)$, and P_0 (which commutes with the M_j) completes the Lie algebra. The basis vectors of an irreducible representation D of $SO_0(1, 4)$ are $|D; n, l; j, m\rangle$, where j and m are $SO(3)$ indices, and n and l label representations of $SO(4)$.[16]

The normalized $SO(4)$-invariant vectors in a tensor product of two $SO_0(1, 4)$ representations are of the form

$$(l^2 - n^2)^{-1/2} \sum_{j=|n|}^{l-1} \sum_{m=-j}^{j} (-1)^{j+m} |D_1; n, l; j, m\rangle \otimes |D_2; n, l; j, -m\rangle.$$

[Since the structure of the representation formulas for $SO(4)$ is very similar to that for $SO_0(1, 3)$, this result can almost be read off from our considerations above on the construction of $SO_0(1, 3)$ invariants.] The most general vector Ψ in the tensor product is a linear combination of such vectors (summed over l and n), with coefficients which we may denote $a_{nl}(l^2 - n^2)^{1/2}$.

One operates on Ψ with P_0 [as given by Ref. 15, Eq. (3.3), with identities among the coefficients stated on p. 458], and requires in the result that the coefficients of linearly independent basis vectors vanish. Just as in the lower-dimensional cases, one concludes that $|a_{nl}| = |a_{n,l+1}| = \cdots$ for infinitely many values of the l index. Thus an invariant vector cannot exist with a finite norm,

$$\|\Psi\|^2 = \sum_{l,n} |a_{nl}|^2 (l^2 - n^2).$$

ACKNOWLEDGMENTS

I am grateful to Professors M.C. Reed, Y. Chow, and W. Miller for helpful conversations. Theorems 1 and 2 appeared in my thesis (Ref. 1), which research was supported in part by AFOSR Contract F44620-71-C-0108 with Princeton University.

*Research supported in part by University of Wisconsin-Milwaukee Graduate School.

†Present address: Mathematics Department, King's College, University of London, Strand, London, WC2R 2LS, U.K.

[1]S.A. Fulling, "Scalar Quantum Field Theory in a Closed Universe of Constant Curvature," Ph. D. dissertation, Princeton University, 1972 (unpublished); paper in preparation.

[2]Also known as a Kronecker product or direct product representation.

[3]Cf. R.F. Streater and A.S. Wightman, *PCT, Spin and Statistics, and All That* (Benjamin, New York, 1964), p. 97.

[4]E.g., O. Nachtmann, Sitzber. Österr. Akad. Wiss., Math.-naturw. Klasse **167**, 363 (1968); N.A. Chernikov and E.A. Tagirov, Ann. Inst. Henri Poincaré A **9**, 109 (1968).

[5]V. Bargmann, Ann. Math. **48**, 568 (1947). Our remarks also apply with little change to the representations of the covering groups of $SO_0(1,2)$ [notably the two-sheeted covering group $SL(2,\mathbb{R})$ and the infinite-sheeted universal covering group)].

[6]In particular, there are two representations with $q=0$ besides the trivial one.

[7]The statement can also be extracted from the known Clebsch–Gordan series of $SO_0(1,2)$ as cataloged by K.-H. Wang, J. Math. Phys. **11**, 2077 (1970). Our direct method of proof is applicable to other groups for which the complete Clebsch–Gordan series are not known.

[8]L. Pukánszky, Trans. Am. Math. Soc. **100**, 116 (1961).

[9]For the case of $SO(3)$ see, e.g., V. Bargmann, Rev. Mod. Phys. **34**, 829 (1962), Sec. 3. For unitary representations the contragredient is just the complex conjugate.

[10]E.g., A.O. Barut and C. Fronsdal, Proc. Roy. Soc. (London) A **287**, 532 (1965), Sec. 6; M.J. Cunningham, J. Math. Phys. **13**, 33 (1972).

[11]See also W.J. Holman, III, and L.C. Biedenharn, Jr., Ann. Phys. (N.Y.) **39**, 1 (1966). J. Patera and P. Winternitz, J. Math. Phys. **14**, 1977 (1973), study a generalization of this construction, in which the invariant vector is replaced by a nontrivial unitary representation of $SU(2)$.

[12]K. Maurin, *General Eigenfunction Expansions and Unitary Representations of Topological Groups* (Polish Scientific Publishers, Warsaw, 1968), Chap. V; A.J. Coleman, in *Group Theory and Its Applications*, edited by E.M. Loebl (Academic, New York, 1968), Sec. IV. Note that a representation of $SO_0(m,n)$ has a unique direct integral decomposition, because connected semisimple groups are Type I groups [Maurin, *loc. cit.*; G.W. Mackey, *Induced Representations of Groups and Quantum Mechanics* (Benjamin, New York, 1968), Sec. 1.4].

[13]If $\mathcal{H}_i$ $(i=1,2)$ are Hilbert spaces of functions $f(s_i)$, possibly vector-valued, square-integrable with respect to μ_i, then the elements of $\mathcal{H}_1 \otimes \mathcal{H}_2$ are the functions $f(s_1,s_2)$ which are square-integrable with respect to the product measure [that is, $\int d\mu_1(s_1)\,d\mu_2(s_2)\,||f(s_1,s_2)\,||^2 < \infty$].

[14]M.A. Naimark, *Linear Representations of the Lorentz Group* (Pergamon, Oxford, 1964), Sec. 8.

[15]S. Ström, Arkiv Fysik **30**, 455 (1965). The second form of the representations, given in Sec. III, will be used here. In Eq. (3.1) the factor $l^2-(j+1)^2$ should be l^2-j^2 (S. Ström, private communication).

[16]The Lie algebra of $SO(4)$ is the direct sum of two copies of the algebra of $SU(2)$. An irreducible representation of $SO(4)$, therefore, corresponds to two representations of $SU(2)$, labeled by two integers or half-integers, k and k'. One then has $l=k+k'+1$, $n=k'-k$.

Statistical mechanical approaches to fluid turbulence

Jon Lee

Aerospace Research Laboratories, Wright-Patterson AFB, Ohio 45433
(Received 9 May 1973; revised manuscript received 21 March 1974)

In the moment formulation, the direct-interaction approximation equations have been derived by developing perturbation about three different states: the laminar flow (Wyld), a turbulent flow (Kraichnan), and the Gaussian random process (Phythian). Along the parallel line, the pertubation theories in the distribution function formulation have also been developed about the same three states: the laminar flow (Balescu–Senatorski), a turbulent flow (Herring), and the Gaussian random process (Edwards). Herring's modal energy and averaged Green's equations are the basic turbulence equations of the distribution function formalism. The modal energy equation, however, represents the simultaneous-time limit of Kraichnan's covariance equation. This paper provides a unified statistical mechanical framework for the three turbulence theories of the distribution function formalism. We have first revised the derivation of Balescu and Senatorski and then presented an alternate method for Herring's self-consistent-field approximation in terms of the action-angle variables. Finally, we have shown that Edwards' theory cannot give the totally correct stationary dynamics because it is not possible to uniquely determine the dynamic friction and diffusion coefficient in the distribution function formulation.

1. INTRODUCTION

Shortly after Kraichnan[1,2] presented the direct-interaction approximation (DIA) equations, Wyld[3] has demonstrated an alternate derivation by first developing perturbation about the laminar flow and then consolidating certain classes of the expansion terms of all orders. The DIA is the moment formulation, thereby addressing itself to the covariance evolution. Instead, it is also possible, and equivalent in a limited sense, to evolve the distribution function of *fictitious* eddy particles as in statistical mechanics. For the Liouville equation for the eddy motion, Herring[4,5] has devised the self-consistent-field approximation (SCFA), thereby obtaining the modal energy and averaged Green's equations. The SCFA equations represent the simultaneous-time limit of the DIA equations. The main point of this paper is to show that the theory of Balescu and Senatorski[6] rederives the SCFA equations by the laminar perturbation and renormalization. Hence, it bears the same relation to Herring's SCFA as does Wyld's theory to Kraichnan's DIA in the moment formulation.

Balescu and Senatorski begin with the triad-interaction representation of homogeneous turbulence and resolve the velocity field variables into action-angle coordinates (Sec. 2). The advantages of the action-angle representation are twofold: Only the reduced distribution function averaged over all angle variables enters into the isotropic turbulence theory (Sec. 3), and the nonlinear interaction is elegantly expressed by the transition matrix (Sec. 4). The disadvantage, however, is that it cannot represent the Gaussian random process.[7] At this point, we introduce the important ingredient of nonequilibrium statistical mechanics developed by the Brussels[7] school. It is the general kinetic equation. We shall show by a simple derivation that it represents an elaborate rearrangement of the Liouville equation under iterative perturbation (Sec. 5).

The kinetic equation is derived formally and hence applies in principle to the turbulent eddy motion as well as the statistical mechanical systems. We must, however, recognize the divergence of fluid turbulence from the classical many-body problems (Sec. 6). (i) In the homogeneous turbulence, we have artificially created the concept of discrete, denumerable degrees of freedom corresponding to the Fourier modes. Since the number of Fourier modes in a box volume L^3 is also proportional to L^3, the number of eddy particles and the box size are not independent. Hence, the thermodynamic limit [i.e., the concentration (N/L^3) being finite as the number of particles $N \to \infty$ and the volume $L^3 \to \infty$] has no important part in the turbulence theory. (ii) In the classical statistical mechanics, the distribution function may be factorized as a consequence of molecular chaos. In the homogeneous turbulence, however, the factorization is implied by Kraichnan's weak-dependence hypothesis.[2] Kinematically, in the limit as $L \to \infty$, homogeneity requires the Fourier modes to be statistically independent just as stationarity demands the random process to have orthogonal increments. (iii) For the fully developed turbulence theory, we can suppress the destruction fragments which represent relaxation of the initial correlation by means other than the triad interactions. This is indeed in accordance with Kraichnan's maximal-randomness hypothesis.[2] (iv) The renormalization requires a new ingredient called the Green's operator. Since the theory of the Brussels school does not involve the explicit formulation for the Green's operator equation, Balescu and Senatorski have guessed at it from the nonlinear interaction diagram. Here we derive the Green's operator equation from the kinetic equation.

The kinetic equation together with the Green's operator equation provide a sufficiently general basis for deriving the modal energy and averaged Green's equations. Upon renormalization, they agree with the SCFA equations (Sec. 7). Renormalization calls for establishing one-to-one correspondence between the consolidated term and a class of perturbation terms. This tedious renormalization may be avoided by developing perturbation about a turbulent flow which has the same Green's operator as the actual turbulent flow (Sec. 8). Herring's SCFA also relies on this sort of perturbation scheme (Sec. 9).

The covariance of the DIA involves the two independent time arguments, say t and t', so that we can evolve the covariance in the entire t–t' plane. In the

distribution function formalism, the covariance is defined in terms of the reference and difference times. Since the reference time runs along $t=t'$, the evolution of the modal energy is restricted to the diagonal of the $t-t'$ plane. More specifically, the modal energy equation represents the simultaneous-time limit of Kraichnan's covariance equation. We cannot directly recover the DIA equations from the Liouville equation. This is the inherent limitation of the distribution function formalism (Sec. 8B). A further limitation shows up in Edwards' theory[8] of stationary turbulence dynamics. There we cannot simultaneously determine the dynamic friction and diffusion coefficient. This is why Edwards' theory yields the dynamic friction with a relaxation factor different from Herring's SCFA (Appendix B).

Diagrams are used in this paper only as a shorthand notation for bulky mathematical expressions. Hence, we do not execute any algebraic operation with the diagrams.

2. EQUATIONS OF EDDY MOTION

The Navier–Stokes equations describe the motion of fluids as continuum. In the homogeneous flow, we can formally decompose the continuum dynamics into an equivalent many-degrees-of-freedom problem by Fourier analyzing the velocity field in a box of side L:

$$U_i(\mathbf{x},t)=\left(\frac{2\pi}{L}\right)^{3/2}\sum_{\mathbf{k}} U_i(\mathbf{k},t)\exp(i\mathbf{k}\cdot\mathbf{x}), \tag{2.1}$$

where

$$\mathbf{k}=\frac{2\pi}{L}\mathbf{n},\quad \mathbf{n}=\begin{pmatrix}n_x\\ n_y\\ n_z\end{pmatrix},\quad n_x,n_y,n_z=\text{all integers}\ .$$

Then, the incompressible Navier–Stokes equations will give rise to the Fourier-amplitude equations which are infinitely coupled through the convolution sum

$$\left(\frac{\partial}{\partial t}+\nu k^2\right)U_i(\mathbf{k},t)$$
$$=-i\left(\frac{2\pi}{L}\right)^{3/2}k_j P_{in}(\mathbf{k})\sum_{\mathbf{k}=\mathbf{p}+\mathbf{q}} U_n(\mathbf{p},t)U_j(\mathbf{q},t), \tag{2.2}$$

where ν is the kinematic viscosity and $P_{ij}(\mathbf{k})=\delta_{ij}-k_ik_j/k^2$. Since the incompressibility $k_iU_i(\mathbf{k},t)=0$ restricts the motion to a plane perpendicular to the wave vector $\mathbf{k}$, we may span $U_i(\mathbf{k})$ by polarization vectors[6]

$$U_i(\mathbf{k},t)=\sum_{\mu=1,2}\epsilon_i^\mu(\mathbf{k})u^\mu(\mathbf{k},t). \tag{2.3}$$

By definition, the polarization vectors $\epsilon^\mu(\mathbf{k})$ are perpendicular to $\mathbf{k}$, $k_i\epsilon_i^\mu(\mathbf{k})=0$, and orthonormal $\epsilon_i^\mu(\mathbf{k})\epsilon_i^\nu(\mathbf{k})=\delta_{\mu\nu}$. Further, the orthonormal vectors $(\epsilon^1(\mathbf{k}),\ \epsilon^2(\mathbf{k}),\ \mathbf{k}/k)$ satisfy the identity $\sum_\mu\epsilon_i^\mu(\mathbf{k})\epsilon_j^\mu(\mathbf{k})=P_{ij}(\mathbf{k})$. Upon introducing (2.3) into (2.2), we obtain the equations of eddy motion involving the three-Fourier-mode interactions

$$\left(\frac{\partial}{\partial t}+\nu k^2\right)u^\mu(\mathbf{k},t)$$
$$=-i\left(\frac{2\pi}{L}\right)^{3/2}\sum_{\lambda,\rho}\sum_{\mathbf{k}=\mathbf{p}+\mathbf{q}}\phi^{\mu|\lambda,\rho}_{\mathbf{k}|\mathbf{p},\mathbf{q}}u^\lambda(\mathbf{p},t)u^\rho(\mathbf{q},t), \tag{2.4}$$

where the coupling coefficient is

$$\phi^{\mu|\lambda,\rho}_{\mathbf{k}|\mathbf{p},\mathbf{q}}=[\mathbf{k}\cdot\epsilon^\rho(\mathbf{q})][\epsilon^\mu(\mathbf{k})\cdot\epsilon^\lambda(\mathbf{p})].$$

We have documented[9,10] the advantages of the triad-interaction representation (2.4) over the Fourier-amplitude form (2.2); hence it needs no further elaboration here.

In the remainder of this section, we put (2.4) in an alternate form convenient for the statistical mechanical formulation. Because of the reality requirement $u^\mu{}^*(\mathbf{k})=u^\mu(-\mathbf{k})$, we may rewrite the convolution sum as

$$\sum_{\mathbf{k}=\mathbf{p}+\mathbf{q}}u^\lambda(\mathbf{p})u^\rho(\mathbf{q})$$
$$=\sum_{\mathbf{p}}^{\mathbf{p}<\mathbf{k}}u^\lambda(\mathbf{p})u^\rho(\mathbf{k}-\mathbf{p})+\sum_{\mathbf{p}>\mathbf{k}}^{+}u^\lambda(\mathbf{p})u^{\rho*}(\mathbf{p}-\mathbf{k})$$
$$+\sum_{\mathbf{p}}^{+}u^{\lambda*}(\mathbf{p})u^\rho(\mathbf{k}+\mathbf{p}), \tag{2.5}$$

where $\sum_{\mathbf{p}}^{+}$ implies the sum over all $\mathbf{p}>0$. Splitting (2.4) into the real and imaginary parts by $u^\mu(\mathbf{k})=v^\mu(\mathbf{k})+iw^\mu(\mathbf{k})$, we obtain the equations with the positive wave vector domain ($\mathbf{k}>0$)

$$\left(\frac{\partial}{\partial t}+\nu k^2\right)\begin{pmatrix}v^\mu(\mathbf{k})\\ w^\mu(\mathbf{k})\end{pmatrix}=\left(\frac{2\pi}{L}\right)^{3/2}\sum_{\lambda,\rho}\sum_{\mathbf{p}}^{+}$$
$$\left\{\phi^{\mu|\lambda,\rho}_{\mathbf{k}|\mathbf{p},s(\mathbf{k}-\mathbf{p})}\begin{pmatrix}v^\lambda(\mathbf{p})\,sw^\rho(s(\mathbf{k}-\mathbf{p}))+w^\lambda(\mathbf{p})\,v^\rho(s(\mathbf{k}-\mathbf{p}))\\ -v^\lambda(\mathbf{p})\,v^\rho(s(\mathbf{k}-\mathbf{p}))+w^\lambda(\mathbf{p})sw^\rho(s(\mathbf{k}-\mathbf{p}))\end{pmatrix}\right.$$
$$\left.+\phi^{\mu|\lambda,\rho}_{\mathbf{k}|\mathbf{p},(\mathbf{k}+\mathbf{p})}\begin{pmatrix}v^\lambda(\mathbf{p})\,w^\rho(\mathbf{k}+\mathbf{p})-w^\lambda(\mathbf{p})\,v^\rho(\mathbf{k}+\mathbf{p})\\ -v^\lambda(\mathbf{p})\,v^\rho(\mathbf{k}+\mathbf{p})-w^\lambda(\mathbf{p})\,w^\rho(\mathbf{k}+\mathbf{p})\end{pmatrix}\right\}, \tag{2.6}$$

where

$$s(\mathbf{k}-\mathbf{p})=\begin{cases}\mathbf{k}-\mathbf{p}, & \mathbf{k}>\mathbf{p},\\ \mathbf{p}-\mathbf{k}, & \mathbf{k}<\mathbf{p},\end{cases}$$

and

$$sw^\rho(s(\mathbf{k}-\mathbf{p}))=\begin{cases}w^\rho(\mathbf{k}-\mathbf{p}), & \mathbf{k}>\mathbf{p},\\ -w^\rho(\mathbf{p}-\mathbf{k}), & \mathbf{k}<\mathbf{p}.\end{cases}$$

Since v^μ and w^μ are the 2D Cartesian coordinates, we may transform them to a sort of polar coordinates known as action-angle variables

$$v^\mu(\mathbf{k})=\eta^{1/2}_{\mathbf{k},\mu}\cos2\pi\xi_{\mathbf{k},\mu},$$
$$w^\mu(\mathbf{k})=\eta^{1/2}_{\mathbf{k},\mu}\sin2\pi\xi_{\mathbf{k},\mu}. \tag{2.7}$$

Therefore, the triad-interaction representation in action-angle variables becomes

$$\dot\eta_{\mathbf{k},\mu}+2\nu k^2\eta_{\mathbf{k},\mu}$$
$$=-2(2\pi/L)^{3/2}\sum_{\lambda,\rho}\sum_{\mathbf{p}}^{+}\left\{\phi^{\mu|\lambda,\ \rho}_{\mathbf{k}|\mathbf{p},s(\mathbf{k}-\mathbf{p}),\rho}\eta^{1/2}_{s(\mathbf{k}-\mathbf{p}),\rho}\sin2\pi A\right.$$
$$\left.+\phi^{\mu|\lambda,\rho}_{\mathbf{k}|\mathbf{p},(\mathbf{k}+\mathbf{p})}\eta^{1/2}_{(\mathbf{k}+\mathbf{p}),\rho}\sin2\pi B\right\}\eta^{1/2}_{\mathbf{p},\lambda}\ \eta^{1/2}_{\mathbf{k},\mu},$$
$$\dot\xi_{\mathbf{k},\mu}=-\frac{1}{2\pi}\left(\frac{2\pi}{L}\right)^{3/2}\sum_{\lambda,\rho}\sum_{\mathbf{p}}^{+}\left\{\phi^{\mu|\lambda,\ \rho}_{\mathbf{k}|\mathbf{p},s(\mathbf{k}-\mathbf{p})}\eta^{1/2}_{s(\mathbf{k}-\mathbf{p}),\rho}\cos2\pi A\right.$$
$$\left.+\phi^{\mu|\lambda,\rho}_{\mathbf{k}|\mathbf{p},(\mathbf{k}+\mathbf{p})}\eta^{1/2}_{(\mathbf{k}+\mathbf{p}),\rho}\cos2\pi B\right\}\eta^{1/2}_{\mathbf{p},\lambda}\ \eta^{-1/2}_{\mathbf{k},\mu}, \tag{2.8}$$

where the dot denotes $\partial/\partial t$ and $A=\xi_{\mathbf{k},\mu}-\xi_{\mathbf{p},\lambda}-s\,\xi_{s(\mathbf{k}-\mathbf{p}),\rho}$ and $B=\xi_{\mathbf{k},\mu}+\xi_{\mathbf{p},\lambda}-\xi_{\mathbf{k}+\mathbf{p},\rho}$. Note that the above differs from Eq. (2.10) of Ref. 6 by the factor $2(2\pi/L)^{3/2}$. Be-

cause of the square root, the action variable itself represents the modal energy (divided by density)

$$\eta_{\mathbf{k},\mu} = v^{\mu}(\mathbf{k})^2 + w^{\mu}(\mathbf{k})^2. \tag{2.9}$$

The action-angle variables are canonical in classical mechanics. In the inviscid limit, (2.8) obeys the Liouville theorem

$$\frac{\partial \dot{\eta}_{\mathbf{k},\mu}}{\partial \eta_{\mathbf{k},\mu}} + \frac{\partial \dot{\xi}_{\mathbf{k},\mu}}{\partial \xi_{\mathbf{k},\mu}} = 0, \tag{2.10}$$

implying invariance of the measure of a point set in the phase space.

3. THE LIOUVILLE EQUATION

In the phase space spanned by $\eta = \{\eta_{\mathbf{k},\mu}\}$ and $\xi = \{\xi_{\mathbf{k},\mu}\}$, the evolution of the distribution function $F(\eta, \xi, t)$ is governed by the continuity equation

$$\frac{\partial F}{\partial t} = \mathcal{L} F, \tag{3.1}$$

where the Liouville operator is

$$\mathcal{L} = -\sum_{\mathbf{k}}^{+} \sum_{\mu} \left\{ \frac{\partial}{\partial \eta_{\mathbf{k},\mu}} \dot{\eta}_{\mathbf{k},\mu} + \frac{\partial}{\partial \xi_{\mathbf{k},\mu}} \dot{\xi}_{\mathbf{k},\mu} \right\}.$$

Because of (2.10), we can therefore reduce (3.1) to the so-called Liouville equation

$$\frac{\partial F}{\partial t} = -\sum_{\mathbf{k}}^{+} \sum_{\mu} \left\{ \dot{\eta}_{\mathbf{k},\mu} \frac{\partial}{\partial \eta_{\mathbf{k},\mu}} + \dot{\xi}_{\mathbf{k},\mu} \frac{\partial}{\partial \xi_{\mathbf{k},\mu}} \right\} F. \tag{3.2}$$

Since it is not any more advantageous to use (3.2) than (3.1), we shall directly work with (3.1) and hereafter call it the Liouville equation. The flow system may be rendered conservative by suppressing the viscous effect: Either drop the viscous term altogether or counteract the dissipation by external forces. The measure of F is time-invariant for conservative systems; we may then normalize F for all t

$$\int d\eta \int d\xi\, F(\eta, \xi, t) = 1. \tag{3.3}$$

Let us expand the distribution function with respect to the periodic angle variables:

$$F(\eta, \xi, t) = \sum_{\{m\}} f_{\{m\}}(\eta, t) \exp(i2\pi m\, \xi), \tag{3.4}$$

where $\{m\}$ represents the set of all integers $\{\ldots, m_{\mathbf{k},\mu}, \ldots\}$ and $m\xi = \sum_{\mathbf{k}}^{+} \sum_{\mu} m_{\mathbf{k},\mu}\, \xi_{\mathbf{k},\mu}$. Multiplying (3.4) by $\exp(-2i\pi m' \xi)$ and integrating over all ξ, we find that $f_{\{m\}}(\eta, t)$ are the Fourier coefficients. In particular, we have

$$f_{\{0\}}(\eta, t) = \int d\xi\, F(\eta, \xi, t). \tag{3.5}$$

That is, $f_{\{0\}}(\eta, t)$ is a reduced distribution. We shall show presently that it plays the central role in the isotropic turbulence theory.

Isotropic energy spectrum

In the homogeneous field, the spectral tensor is defined by

$$\Phi_{ij}(\mathbf{k}, t) = \overline{U_i^*(\mathbf{k}, t) U_j(\mathbf{k}, t)}, \tag{3.6}$$

where the overbar denotes ensemble average. In terms of the action-angle variables, we find that

$$U_i^*(\mathbf{k}) U_j(\mathbf{k}) = \sum_{\mu,\nu} \epsilon_i^{\mu}(\mathbf{k}) \epsilon_j^{\nu}(\mathbf{k})\, \eta_{\mathbf{k},\mu}^{1/2} \eta_{\mathbf{k},\nu}^{1/2} \exp[-i2\pi(\xi_{\mathbf{k},\mu} - \xi_{\mathbf{k},\nu})]. \tag{3.7}$$

Therefore, by averaging over the distribution function (3.4), the spectral tensor takes the form

$$\Phi_{ij}(\mathbf{k}) = \sum_{\mu} \epsilon_i^{\mu}(\mathbf{k})\, \epsilon_j^{\mu}(\mathbf{k}) \int d\eta\, \eta_{\mathbf{k},\mu} f_{\{0\}}(\eta, t) + \sum_{\mu \neq \nu} \epsilon_i^{\mu}(\mathbf{k})\, \epsilon_j^{\nu}(\mathbf{k}) \int d\eta\, \eta_{\mathbf{k},\mu}^{1/2} \eta_{\mathbf{k},\nu}^{1/2} \times f_{\{1_{\mathbf{k},\mu}, -1_{\mathbf{k},\nu}\}}(\eta, t), \tag{3.8}$$

where $\{1_{\mathbf{k},\mu}, -1_{\mathbf{k},\nu}\} = \{\ldots, 0, m_{\mathbf{k},\mu} = 1, m_{\mathbf{k},\nu} = -1, 0, \ldots\}$. Denote the first integral of (3.8) by

$$I_{\mathbf{k},\mu}(t) = \int d\eta\, \eta_{\mathbf{k},\mu} f_{\{0\}}(\eta, t). \tag{3.9}$$

Let us now suppose that

(i) $\int d\eta\, \eta_{\mathbf{k},\mu}^{1/2} \eta_{\mathbf{k},\nu}^{1/2} f_{\{1_{\mathbf{k},\mu}, -1_{\mathbf{k},\nu}\}}(\eta, t) = 0 \quad (\mu \neq \nu)$, (3.10a)

(ii) $I_{\mathbf{k},\mu}(t)$ are independent of μ, (3.10b)

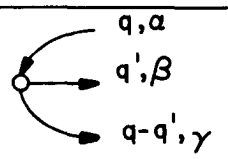
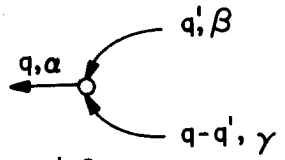
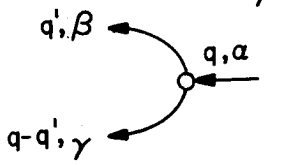
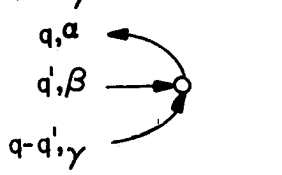
(iii) $I_{\mathbf{k},\mu}(t)$ are functions of k. (3.10c)

Then, by identifying

$$I_{\mathbf{k},\mu}(t) \to I_k(t) = E(k)/4\pi k^2,$$

where $E(k)$ is the energy spectrum, the spectral tensor reduces to the well-known isotropic form[11] $\Phi_{ij}(\mathbf{k}) = P_{ij}(\mathbf{k}) E(k)/4\pi k^2$. Clearly, (3.10) states the isotropic requirements: (i) imposes the reflexional symmetry, (ii) guarantees the rotational symmetry, and (iii) demands the spherical symmetry of isotropic scalar functions. For the isotropic theory, therefore, $f_{\{0\}}$ is all that we need to compute the energy spectrum.

TABLE I. Typical simple transitions.

Type	Mathematical expression	Vertex diagram
I	$\langle 0 \mid \angle'^{\alpha,\beta,\gamma}_{q,q',q\pm q'} \mid m'_{q,\alpha}=1,\ m'_{q',\beta}=-1,\ m'_{q-q',\gamma}=-1\rangle$ $=\frac{i}{\sqrt{2}}\left(\frac{2\pi}{L}\right)^{3/2}\left(\bar{\phi}^{\alpha\mid\beta,\gamma}_{q\mid q',q-q'}\frac{\partial}{\partial\eta_{q,\alpha}}-\bar{\phi}^{\beta\mid\gamma,\alpha}_{q'\mid q-q',q}\frac{\partial}{\partial\eta_{q',\beta}}-\bar{\phi}^{\gamma\mid\alpha,\beta}_{q-q'\mid q,q'}\frac{\partial}{\partial\eta_{q-q',\gamma}}\right)\eta^{1/2}_{q,\alpha}\eta^{1/2}_{q',\beta}\eta^{1/2}_{q-q',\gamma}$	
II	$\langle m_{q,\alpha}=1 \mid \angle'^{\alpha,\beta,\gamma}_{q,q',q\pm q'} \mid m'_{q',\beta}=1, m'_{q-q',\gamma}=1\rangle$ $=\frac{i}{\sqrt{2}}\left(\frac{2\pi}{L}\right)^{3/2}\eta^{1/2}_{q,\alpha}\left(-\bar{\phi}^{\alpha\mid\beta,\gamma}_{q\mid q',q-q'}\frac{\partial}{\partial\eta_{q,\alpha}}+\bar{\phi}^{\beta\mid\gamma,\alpha}_{q'\mid q-q',q}\frac{\partial}{\partial\eta_{q',\beta}}+\bar{\phi}^{\gamma\mid\alpha,\beta}_{q-q'\mid q,q'}\frac{\partial}{\partial\eta_{q-q',\gamma}}\right)\eta^{1/2}_{q',\beta}\eta^{1/2}_{q-q',\gamma}$	
III	$\langle m_{q',\beta}=1,\ m_{q-q',\gamma}=1 \mid \angle'^{\alpha,\beta,\gamma}_{q,q',q\pm q'} \mid m'_{q,\alpha}=1\rangle$ $=\frac{i}{\sqrt{2}}\left(\frac{2\pi}{L}\right)^{3/2}\eta^{1/2}_{q',\beta}\eta^{1/2}_{q-q',\gamma}\left(\bar{\phi}^{\alpha\mid\beta,\gamma}_{q\mid q',q-q'}\frac{\partial}{\partial\eta_{q,\alpha}}-\bar{\phi}^{\beta\mid\gamma,\alpha}_{q'\mid q-q',q}\frac{\partial}{\partial\eta_{q',\beta}}-\bar{\phi}^{\gamma\mid\alpha,\beta}_{q-q'\mid q,q'}\frac{\partial}{\partial\eta_{q-q',\gamma}}\right)\eta^{1/2}_{q,\alpha}$	
IV	$\langle m_{q,\alpha}=1,\ m_{q',\beta}=-1,\ m_{q-q',\gamma}=-1 \mid \angle'^{\alpha,\beta,\gamma}_{q,q',q\pm q'} \mid 0\rangle$ $=\frac{i}{\sqrt{2}}\left(\frac{2\pi}{L}\right)^{3/2}\eta^{1/2}_{q,\alpha}\eta^{1/2}_{q',\beta}\eta^{1/2}_{q-q',\gamma}\left(-\bar{\phi}^{\alpha\mid\beta,\gamma}_{q\mid q',q-q'}\frac{\partial}{\partial\eta_{q,\alpha}}+\bar{\phi}^{\beta\mid\gamma,\alpha}_{q'\mid q-q',q}\frac{\partial}{\partial\eta_{q',\beta}}+\bar{\phi}^{\gamma\mid\alpha,\beta}_{q-q'\mid q,q'}\frac{\partial}{\partial\eta_{q-q',\gamma}}\right)$	

4. TRANSITION MATRIX

Let us introduce (3.4) into (3.1). By separating out the Fourier components $f_{\{m\}}(\eta, t)$, we obtain an equivalent infinite set of Liouville equations

$$\frac{\partial f_{\{m\}}(\eta,t)}{\partial t}=\sum_{\{m'\}}\langle m\mid\angle\mid m'\rangle f_{\{m'\}}(\eta,t), \tag{4.1}$$

where

$$\langle m\mid\angle\mid m'\rangle=\int d\xi\exp(-i2\pi m\xi)\angle\exp(i2\pi m'\xi). \tag{4.2}$$

We shall compute the elements of (4.2) by introducing the Liouville operator

$$\langle m\mid\angle\mid m'\rangle=\sum_{\mathbf{k}}^{+}\sum_{\mu}\langle m\mid\angle^{\circ}_{\mathbf{k},\mu}\mid m'\rangle + \sum_{\mathbf{k},\mathbf{p}}^{+}\sum_{\mu,\lambda,\rho}\langle m\mid\angle'^{\mu,\lambda,\rho}_{\mathbf{k},\mathbf{p},\mathbf{k}\pm\mathbf{p}}\mid m'\rangle. \tag{4.3}$$

Here the first term reflects the angle-independent viscous effect

$$\langle m\mid\angle^{\circ}_{\mathbf{k},\mu}\mid m'\rangle=2\nu k^2\frac{\partial}{\partial\eta_{\mathbf{k},\mu}}\eta_{\mathbf{k},\mu}\,\delta(m-m'), \tag{4.4}$$

and the second term is derived from the nonlinear interaction

$$\begin{aligned}\langle m\mid\angle'^{\mu,\lambda,\rho}_{\mathbf{k},\mathbf{p},\mathbf{k}\pm\mathbf{p}}\mid m'\rangle &=i\left(\frac{2\pi}{L}\right)^{3/2}\sum_{a=\pm1}\left(-a\frac{\partial}{\partial\eta_{\mathbf{k},\mu}}+\frac{m_{\mathbf{k},\mu}}{2\eta_{\mathbf{k},\mu}}\right)\eta^{1/2}_{\mathbf{k},\mu}\eta^{1/2}_{\mathbf{p},\lambda}\\ &\times\{\phi^{\mu\mid\lambda,\rho}_{\mathbf{k}\mid\mathbf{p},s(\mathbf{k}-\mathbf{p})}\eta^{1/2}_{s(\mathbf{k}-\mathbf{p}),\rho}\,\delta(m'_{\mathbf{k},\mu}+a-m_{\mathbf{k},\mu})\\ &\times\delta(m'_{\mathbf{p},\lambda}-a-m_{\mathbf{p},\lambda})\delta(m'_{s(\mathbf{k}-\mathbf{p}),\rho}-as-m_{s(\mathbf{k}-\mathbf{p}),\rho})\\ &+\phi^{\mu\mid\lambda,\rho}_{\mathbf{k}\mid\mathbf{p},\mathbf{k}+\mathbf{p}}\eta^{1/2}_{\mathbf{k}+\mathbf{p},\rho}\,\delta(m'_{\mathbf{k},\mu}+a-m_{\mathbf{k},\mu})\\ &\times\delta(m'_{\mathbf{p},\lambda}+a-m_{\mathbf{p},\lambda})\delta(m'_{\mathbf{k}+\mathbf{p},\rho}-a-m_{\mathbf{k}+\mathbf{p},\rho})\}\\ &\times\prod_{\substack{\mathbf{k}'\neq\mathbf{k},\mathbf{p},\mathbf{k}\pm\mathbf{p}\\ \alpha\neq\mu,\lambda,\rho}}\delta(m'_{\mathbf{k}',\alpha}-m_{\mathbf{k}',\alpha}).\end{aligned} \tag{4.5}$$

Again, compare $i(2\pi/L)^{3/2}$ with the factor $-(i/2)(2\pi/L)^3$ in Eq. (2.18) of Ref. 6.

The simple transitions

Following the Brussels school, we shall call

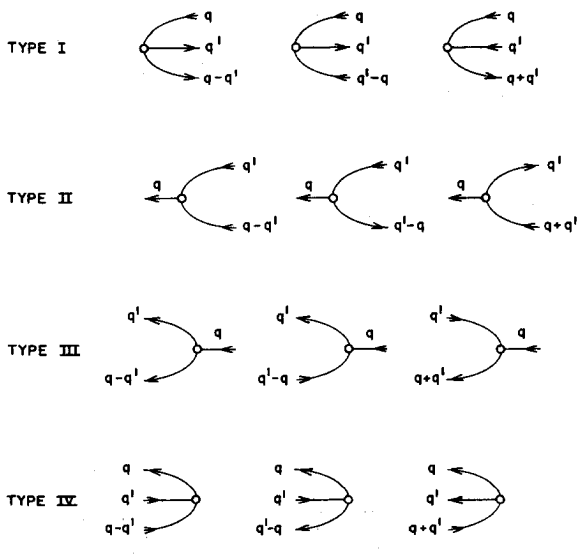


FIG. 1. Simple transitions (α, β, and γ deleted).

$\langle m|\mathcal{L}|m'\rangle$ the transition matrix from state $\{m'\}$ to $\{m\}$ (always read from the right to left). We then see that the viscous term induces no state transition, whereas the nonlinear interaction brings about transitions between the states which are constrained by the δ functions in (4.5). Since the isotropic turbulence theory evolves around the state $\{m\}=\{0\}$, we shall consider a class of transitions, called the simple transitions, between the states represented by $m=0$ and ± 1. By the complete enumeration of (4.5), we find 12 distinct simple transitions. They can, however, be divided into four types, each having the typical member as shown in Table I.

In the table, there appears the symmetrized coupling coefficient

$$\overline{\phi}^{\mu|\lambda,\rho}_{\mathbf{k}|\mathbf{p},\mathbf{q}} = \phi^{\mu|\lambda,\rho}_{\mathbf{k}|\mathbf{p},\mathbf{q}} + \phi^{\mu|\rho,\lambda}_{\mathbf{k}|\mathbf{q},\mathbf{p}} \tag{4.6}$$

Note that we have introduced the factor $2^{-1/2}$ into the simple transitions, so as to avoid double counting the elementary interactions composed of two simple transitions. The first simple transition (Type I) of Table I is obtained as follows: Consider $\{m\}=\{0\}$ and $\{m'\}=\{\ldots,0,m'_{\mathbf{q},\alpha}=1,m'_{\mathbf{q}',\beta}=-1,m'_{\mathbf{q}-\mathbf{q}',\gamma}=-1,\ldots\}$. There are six ways that transition from $\{m'\}$ to $\{m\}$ can be realized:

(i) $(\mathbf{q},\alpha)=(\mathbf{k},\mu),(\mathbf{q}',\beta)=(\mathbf{p},\lambda),(\mathbf{q}-\mathbf{q}',\gamma)=(\mathbf{k}-\mathbf{p},\rho)$,

(ii) $(\mathbf{q},\alpha)=(\mathbf{k},\mu),(\mathbf{q}',\beta)=(\mathbf{k}-\mathbf{p},\rho),(\mathbf{q}-\mathbf{q}',\gamma)=(\mathbf{p},\lambda)$,

(iii) $(\mathbf{q},\alpha)=(\mathbf{p},\lambda),(\mathbf{q}',\beta)=(\mathbf{k},\mu),(\mathbf{q}-\mathbf{q}',\gamma)=(\mathbf{p}-\mathbf{k},\rho)$,

(iv) $(\mathbf{q},\alpha)=(\mathbf{p},\lambda),(\mathbf{q}',\beta)=(\mathbf{p}-\mathbf{k},\rho),(\mathbf{q}-\mathbf{q}',\gamma)=(\mathbf{k},\mu)$,

(v) $(\mathbf{q},\alpha)=(\mathbf{k}+\mathbf{p},\rho),(\mathbf{q}',\beta)=(\mathbf{p},\lambda),(\mathbf{q}-\mathbf{q}',\gamma)=(\mathbf{k},\mu)$,

(vi) $(\mathbf{q},\alpha)=(\mathbf{k}+\mathbf{p},\rho),(\mathbf{q}',\beta)=(\mathbf{k},\mu),(\mathbf{q}-\mathbf{q}',\gamma)=(\mathbf{p},\lambda)$.

The transitions (i) and (ii) give the term multiplied by $\overline{\phi}^{\alpha|\beta,\gamma}_{\mathbf{q}|\mathbf{q}',\mathbf{q}-\mathbf{q}'}$. Similarly, (iii) and (vi) give the term multiplied by $\overline{\phi}^{\beta|\gamma,\alpha}_{\mathbf{q}'|\mathbf{q}-\mathbf{q}',\mathbf{q}}$, and (iv) and (v) the term multiplied by $\overline{\phi}^{\gamma|\alpha,\beta}_{\mathbf{q}-\mathbf{q}'|\mathbf{q},\mathbf{q}'}$. Upon adding these, we obtain the simple transition of Type I; the remaining three types can be verified similarly.

Since the simple transitions have bulky mathematical expressions, it is convenient to have a shorthand notation for them. To this end, we adopt the vertex diagram constructed by the following rules: (i) Represent $\langle m|\mathcal{L}'|m'\rangle$ by a small circle to which we attach lines for the nonzero elements of $\{m\}$ and $\{m'\}$. (ii) Place the lines (if any) for $\{m'\}$ to the right of the circle and the lines (if any) for $\{m\}$ to the left of the circle. And, (iii) attach an arrowhead pointing to the left for the lines corresponding to the positive elements of $\{m\}$ and $\{m'\}$, an arrowhead pointing to the right for lines corresponding to the negative elements. By using this diagrammatics, the simple transitions can be represented by the vertex diagrams of Table I. The two other transitions of each type are shown in Fig. 1. For each diagram of the figure, the conjugate pair is obtained by simply reversing the arrowheads. Of course, the particular configurations of the diagrams are immaterial because they are a topological representation.

Now we can convert the vertex diagrams into the corresponding mathematical expressions by the following recipe: (i) Assign a plus sign to the incoming arrows and a minus sign to the outgoing arrows (at each vertex, the incoming and outgoing wave vectors add up to zero). (ii) Interpret the open circle as the summation of three lines attached to it. (iii) Associate the line with the indices, say $\mathbf{q}$ and α, with

$$\overline{\phi}^{\alpha|\beta,\gamma}_{\mathbf{q}|\mathbf{q}',\mathbf{q}-\mathbf{q}'} \quad \frac{\partial}{\partial\eta_{\mathbf{q},\alpha}}\,\eta^{1/2}_{\mathbf{q},\alpha}\eta^{1/2}_{\mathbf{q}',\beta}\eta^{1/2}_{\mathbf{q}-\mathbf{q}',\gamma}$$

[where $(\mathbf{q}',\beta)$ and $(\mathbf{q}-\mathbf{q}',\gamma)$ are the indices of the two other lines] if it appears to the right of the circle, and

$$\eta^{1/2}_{\mathbf{q},\alpha}\,\overline{\phi}^{\alpha|\beta,\gamma}_{\mathbf{q}|\mathbf{q}',\mathbf{q}-\mathbf{q}'} \quad \frac{\partial}{\partial\eta_{\mathbf{q},\alpha}}\,\eta^{1/2}_{\mathbf{q}',\beta}\eta^{1/2}_{\mathbf{q}-\mathbf{q}',\gamma}$$

if it appears to the left of the circle. And, (iv) introduce the factor $(i/\sqrt{2})(2\pi/L)^{3/2}$ to complete the expression. With this convention, we can recover the mathematical expression for any of the vertex diagrams in Fig. 1.

5. THE GENERAL KINETIC EQUATION

After having defined the simple transitions, Balescu and Senatorski follow very faithfully the theoretical apparatus of nonequilibrium statistical mechanics developed by the Brussels school. The starting point, and by far the most essential step, is the general kinetic equation which was originally derived by Prigogine and Resibois.[12] In fact, it represents a rearrangement of the products of transition matrix elements which would appear under the iterative solution of Liouville equation. The derivation of Prigogine and Resibois involves decomposition and rearrangement of the iterated transition matrix elements with the aid of the diagrams; however, Zwanzig[13] has later presented an analytic derivation using the projection operator technique. For the purpose of demonstrating the gist of kinetic equation, we present here an elementary derivation for the leading terms of the kinetic equation. In turbulence work, however, the lowest-order terms are kinematically very important because they share the same structure with the renormalized lowest-order terms. In this respect, the present derivation is more than an illustration: It actually serves as a practical tool in the subsequent discussion.

For the notational compactness, rewrite (4.3) as

$$\langle m|\mathcal{L}|m'\rangle = \mathcal{L}^{\circ}\delta(m-m') + \langle m|\mathcal{L}'|m'\rangle, \tag{5.1}$$

where

$$\mathcal{L}^{\circ} = \sum_{\mathbf{k}}^{+}\sum_{\mu}\mathcal{L}^{\circ}_{\mathbf{k},\mu},$$

$$\mathcal{L}' = \sum_{\mathbf{k},\mathbf{p}}^{+}\sum_{\mu,\lambda,\rho}\mathcal{L}'^{\,\mu,\lambda,\rho}_{\mathbf{k},\mathbf{p},\mathbf{k}+\mathbf{p}}.$$

Then (4.1) takes the form

$$\frac{\partial f_{\{m\}}(\eta,t)}{\partial t} = \mathcal{L}^{\circ} f_{\{m\}}(\eta,t) + \sum_{\{m'\}} \langle m|\mathcal{L}'|m'\rangle f_{\{m'\}}(\eta,t). \quad (5.2)$$

By treating the second right-hand side as the inhomogeneous term, the formal solution is

$$f_{\{m\}}(\eta,t) = \exp(\mathcal{L}^{\circ}t) f_{\{m\}}(\eta,0) + \sum_{\{m'\}} \int_0^t dt' \exp[\mathcal{L}^{\circ}(t-t')] \times \langle m|\mathcal{L}'|m'\rangle f_{\{m'\}}(\eta,t'). \quad (5.3)$$

Now solve this by iteration. The iterative solution of the second-order is

$$f_{\{m\}}(\eta,t) = \exp(\mathcal{L}^{\circ}t) f_{\{m\}}(\eta,0) + \sum_{\{m'\}} \int_0^t dt' \exp[\mathcal{L}^{\circ}(t-t')] \times \langle m|\mathcal{L}'|m'\rangle \exp(\mathcal{L}^{\circ}t') f_{\{m'\}}(\eta,0) + \sum_{\{m'\}} \sum_{\{m''\}} \int_0^t dt' \times \int_0^{t'} dt'' \exp[\mathcal{L}^{\circ}(t-t')] \langle m|\mathcal{L}'|m'\rangle \exp[\mathcal{L}^{\circ}(t'-t'')] \times \langle m'|\mathcal{L}'|m''\rangle f_{\{m''\}}(\eta,t''). \quad (5.4)$$

Upon differentiating with respect to t, we obtain the following equation which is a simple variation of (5.2) without any approximation:

$$\frac{\partial f_{\{m\}}(\eta,t)}{\partial t} = \mathcal{L}^{\circ} f_{\{m\}}(\eta,t) + \sum_{\{m'\}} \langle m|\mathcal{L}'|m'\rangle \exp(\mathcal{L}^{\circ}t) f_{\{m'\}}(\eta,0) + \sum_{\{m'\}} \sum_{\{m''\}} \int_0^t dt' \langle m|\mathcal{L}'|m'\rangle \exp[\mathcal{L}^{\circ}(t-t')] \times \langle m'|\mathcal{L}'|m''\rangle f_{\{m''\}}(\eta,t'). \quad (5.5)$$

We have therefore decomposed the second term of (5.2) into the last two terms of (5.5), which may be considered as the equivalent interaction terms. By iterating (5.4) to higher orders, we can further decompose the second term of (5.2) into arbitrarily many equivalent interaction terms. Upon singling out the term for $\{m''\} = \{m\}$, we obtain at once the leading terms of the general kinetic equation

$$\frac{\partial f_{\{m\}}(\eta,t)}{\partial t} = \mathcal{L}^{\circ} f_{\{m\}}(\eta,t) + \sum_{\{m'\}} \langle m|\mathcal{L}'|m'\rangle \exp(\mathcal{L}^{\circ}t) f_{\{m'\}}(\eta,0) + \sum_{\{m'\}} \int_0^t dt' \langle m|\mathcal{L}'|m'\rangle \exp[\mathcal{L}^{\circ}(t-t')] \langle m'|\mathcal{L}'|m\rangle \times f_{\{m\}}(\eta,t') + \cdots, \quad (5.6)$$

where the three dots denote the double sum

$$\sum_{\{m'\}} \sum_{\{m''\}\neq\{m\}}.$$

We shall show that the second and third terms of (5.6) are the respective leading terms of the destruction and diagonal fragments. The general kinetic equation of the Brussels school has the form[7,14]

$$\frac{\partial f_{\{m\}}(\eta,t)}{\partial t} = \mathcal{L}^{\circ} f_{\{m\}}(\eta,t) + \mathcal{D}_{mm'}(t) + \int_0^t dt' \mathcal{E}_m(t-t') f_{\{m\}}(\eta,t'). \quad (5.7)$$

Here the destruction and diagonal fragments are given respectively by

$$\mathcal{D}_{mm'}(t) = (2\pi i)^{-1} \oint_c dz \exp(zt) \times \sum_{\{m'\}\neq\{m\}} \hat{D}_{mm'}(z) f_{\{m'\}}(\eta,0), \quad (5.8)$$

$$\mathcal{E}_m(t) = (2\pi i)^{-1} \oint_c dz \exp(zt) \hat{\mathcal{E}}_m(z), \quad (5.9)$$

where

$$\hat{D}_{mm'}(z) = \sum_{n=0}^{\infty} \langle m|\mathcal{L}'[(z-\mathcal{L}^{\circ})^{-1}\mathcal{L}']^n (z-\mathcal{L}^{\circ})^{-1}|m'\rangle_{\mathrm{irr}}, \quad (5.10)$$

$$\hat{\mathcal{E}}_m(z) = \sum_{n=0}^{\infty} \langle m|\mathcal{L}'[(z-\mathcal{L}^{\circ})^{-1}\mathcal{L}']^n|m\rangle_{\mathrm{irr}}. \quad (5.11)$$

The subscript irr implies that no intermediate state is identical to $\{m\}$. Note that (5.8) and (5.9) are the standard inverse Laplace transforms with the usual integration path denoted by $\oint_c$. Considering the leading terms of (5.10) and (5.11), we have

$$\mathcal{D}_{mm'}(t) = \sum_{\{m'\}\neq\{m\}} \langle m|\mathcal{L}'|m'\rangle \exp(\mathcal{L}^{\circ}t) f_{\{m'\}}(\eta,0) + \cdots, \quad (5.12)$$

$$\mathcal{E}_m(t) = \sum_{\{m'\}} \langle m|\mathcal{L}'|m'\rangle \exp(\mathcal{L}^{\circ}t) \langle m'|\mathcal{L}'|m\rangle + \cdots \quad (5.13)$$

In (5.13) we have used the decomposition formula

$$\langle m|\mathcal{A}\mathcal{B}|m'\rangle = \sum_{\{m''\}} \langle m|\mathcal{A}|m''\rangle\langle m''|\mathcal{B}|m'\rangle,$$

where $\mathcal{A}$ and $\mathcal{B}$ are operators involving $(z-\mathcal{L}^{\circ})^{-1}$ and $\mathcal{L}'$. The substitution of (5.12) and (5.13) into (5.7) verifies that the leading terms of the destruction and diagonal fragments are identical to those derived in (5.6).

6. TOWARDS THE TURBULENCE APPLICATION

The kinetic equation has been derived formally; hence it applies to both the turbulent eddy motion and statistical mechanical systems. Since the similarity between the eddy motion and classical many-body problem is superficial, the application of (5.7) to turbulence must take into account the divergence between them. The classical statistical mechanics deals with a system of

N particles contained in the volume L^3. Then, the concentration (N/L^3) must be finite in the limit as $N \to \infty$ and $L \to \infty$, thereby assuring finiteness of the intensive thermodynamic properties.[7] This thermodynamic limit must therefore be incorporated into the construction of a distribution function in order for it to be physically sensible. In turbulence, however, we have artifically created the concept of discrete, denumerable eddies corresponding to the Fourier modes. Since the number of Fourier modes in a box volume L^3 is also proportional to L^3, the number of eddies and the box volume are no longer independent parameters. Consequently, the thermodynamic limit plays no important role in the turbulence theory. Often the position coordinates are Fourier analyzed in statistical mechanics.[7] The use of Fourier representation there has the purpose of reducing the unperturbed Hamiltonian to a diagonal form, whereas the interaction Hamiltonian becomes off-diagonal. This therefore permits approximation of the interaction Hamiltonian in terms of the vacuum state.

Under the disguise of deceptive simplicity, (5.7) actually represents an infinite perturbation expansion about the laminar flow. To bring it to a form useful for turbulence work, we must therefore introduce such turbulence concepts as Kraichnan's weak-dependence and maximal-randomness hypotheses and Green's operator.

A. The product hypothesis

Let us assume factorization of the distribution function

$$f_{\{0\}}(\eta, t) = \prod_{\mathbf{k},\mu} \varphi(\eta_{\mathbf{k},\mu}, t). \tag{6.1}$$

Here, $\varphi(\eta_{\mathbf{k},\mu}, t)$ is the single mode distribution

$$\varphi(\eta_{\mathbf{k},\mu}, t) = \int_{(\mathbf{k},\mu)} d\eta\, f_{\{0\}}(\eta, t), \tag{6.2}$$

where $\int_{(\mathbf{k},\mu)}$ denotes the integration over all action variables except $\eta_{\mathbf{k},\mu}$. For a conservative system, (3.3) implies the normalization

$$\int d\eta_{\mathbf{k},\mu}\, \varphi(\eta_{\mathbf{k},\mu}, t) = 1. \tag{6.3}$$

In the classical mechanics, the factorization is a consequence of molecular chaos. In turbulence, (6.1) states the statistical independence of Fourier modes in the homogeneous field. In the limit as $L \to \infty$, homogeneity requires the Fourier modes to be statistically independent just as stationarity demands the random process to have orthogonal increments. Indeed, the product hypothesis embodies Kraichnan's weak-dependence hypothesis. In Appendix A, we shall briefly show that the cycle approximation (Sec. 6D) of kinetic equation cannot generate statistical dependence from the initial independence.

B. The destruction fragments

The destruction fragments represent relaxation of the initial correlation by means other than the triad interactions. For the fully developed turbulence theory, it is therefore natural to suppress the destruction fragments

$$\frac{\partial f_{\{m\}}(\eta, t)}{\partial t} = \mathcal{L}^{\circ} f_{\{m\}}(\eta, t) + \int_0^t dt'\, \mathcal{E}_m(t-t') f_{\{m\}}(\eta, t'). \tag{6.4}$$

We must, however, specify the initial condition $f_{\{m\}}(\eta, 0)$ which will then be relaxed through the diagonal fragments. The suppression of destruction fragments is consistent with Kraichnan's maximal-randomness hypothesis which postulates the fully developed turbulence to be as random as is possible consistent with the Navier-Stockes dynamics, but not at all dependent upon the initial and boundary conditions.

C. The Green's operator equation

The general kinetic equation is useless for strong turbulence, unless we can consolidate it by summing up certain classes of the expansion terms of all orders. This consolidation process is called the renormalization. To carry out renormalization, however, requires the Green's operator equation. Balescu and Senatorski suggested the Green's operator equation from the inspection of the interaction diagram (their propagator equation given by Fig. 8 of Ref. 6 is in error; no double line should appear along the cycle loop). Here we shall derive the Green's operator equation. This is not an unusual proposition because the Green's operator $\exp(\mathcal{L}t)$ also satisfies the Liouville equation.

Recall that $f_{\{0\}}$ is the average of F over all ξ, and $f_{\{m\}}$ are the fluctuation amplitudes with different periods. Let us consider the perturbation of $f_{\{0\}}$ induced by a disturbance to the mode $(\mathbf{k}, \mu)$ and denote it by $f_{\{1_{\mathbf{k},\mu}\}}$, where $\{1_{\mathbf{k},\mu}\} = \{\ldots, 0, m_{\mathbf{k},\mu} = 1, 0, \ldots\}$. Then, its equation can be written down from (6.4):

$$\partial f_{\{1_{\mathbf{k},\mu}\}}(\eta, t)/\partial t = \mathcal{L}^{\circ} f_{\{1_{\mathbf{k},\mu}\}}(\eta, t) + \int_0^t dt'\, \mathcal{E}_{1_{\mathbf{k},\mu}}(t-t') f_{\{1_{\mathbf{k},\mu}\}}(\eta, t'). \tag{6.5}$$

We now extend the product hypothesis

$$f_{\{1_{\mathbf{k},\mu}\}}(\eta, t) = \Big(\prod_{(\mathbf{q},\nu)\neq(\mathbf{k},\mu)} \varphi(\eta_{\mathbf{q},\nu}, t)\Big) \mathcal{G}(\eta_{\mathbf{k},\mu}, t). \tag{6.6}$$

In analogy to (6.1), this amounts to factoring out the averaged Green's function from the velocity covariances. After inserting (6.6) into (6.5), the integration over all η but $\eta_{\mathbf{k},\mu}$ yields the Green's operator equation

$$\partial \mathcal{G}(\eta_{\mathbf{k},\mu}, t)/\partial t = \mathcal{L}^{\circ}_{\mathbf{k},\mu} \mathcal{G}(\eta_{\mathbf{k},\mu}, t) + \int_0^t dt' \int_{(\mathbf{k},\mu)} d\eta\, \mathcal{E}_{1_{\mathbf{k},\mu}}(t-t') \times \Big(\prod_{(\mathbf{q},\nu)\neq(\mathbf{k},\mu)} \varphi(\eta_{\mathbf{q},\nu}, t')\Big) \mathcal{G}(\eta_{\mathbf{k},\mu}, t'), \tag{6.7}$$

with the initial condition $\mathcal{G}(\eta_{\mathbf{k},\mu}, 0) = 1$.

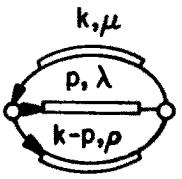
D. The cycle approximation

The leading term of the diagonal fragments for $\{m\} = \{0\}$ is

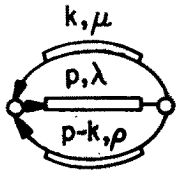
$$\mathcal{E}_0(t) = \sum_{\{m'\}} \langle 0 | \mathcal{L}' | m' \rangle \exp(\mathcal{L}^{\circ} t) \langle m' | \mathcal{L}' | 0 \rangle. \tag{6.8}$$

Recall that $\langle 0 | \mathcal{L}' | m' \rangle$ and $\langle m' | \mathcal{L}' | 0 \rangle$ have the three simple transitions of Types I and IV, respectively (Fig. 1). For the sum $\sum_{\{m'\}}$, we must therefore consider all possible combinations of these transitions. Using the vertex diagrams, we have

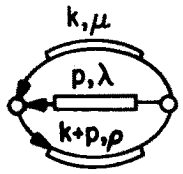
$$\mathcal{E}_0(t)=\sum_{\mathbf{k}}^{+}\sum_{\mu,\lambda,\rho}\Bigg\{\sum_{\mathbf{p}}^{\mathbf{p}\neq\mathbf{k}}$$



$$+\sum_{\mathbf{p}\neq\mathbf{k}}^{+}$$



$$+\sum_{\mathbf{p}}^{+}$$

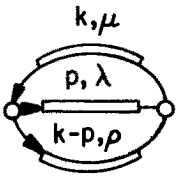


$$\Bigg\}$$

$$+\sum_{\mathbf{k}}^{+}\sum_{\mu,\lambda,\rho}\{\text{three diagrams as in the above with the arrows reversed}\},\tag{6.9}$$

where the double line with the index (k, μ) is the diagram representation for $\exp(\angle^{\circ}_{\mathbf{k},\mu}t)$. In view of (2.5), we can combine the three diagram terms in the curly bracket under one sum $\sum_{\mathbf{p}}$. Since the two sums over k are the same, (6.9) finally reduces to

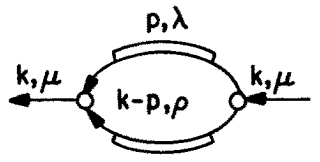
$$\mathcal{E}_0(t)=2\sum_{\mathbf{k}}^{+}\sum_{\mathbf{p}}\sum_{\mu,\lambda,\rho}$$



. (6.10)

Similarly, the simple transitions of Types II and III give the lowest-order diagonal fragments for $\{m\}=\{1_{\mathbf{k},\mu}\}$

$$\mathcal{E}_{1_{\mathbf{k},\mu}}(t)=\sum_{\mathbf{k}}^{+}\sum_{\mathbf{p}}\sum_{\mu,\lambda,\rho}$$



. (6.11)

Since the diagrams of (6.10) and (6.11) have the configuration of a closed loop connecting two vertices, they are referred to as the cycle approximation.

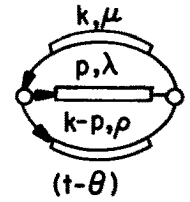
7. PERTURBATION ABOUT THE LAMINAR FLOW

The kinetic equation together with the Green's operator equation provide a sufficiently general basis for deriving the modal energy and averaged Green's equations. The theory of Balescu and Senatorski, however, involves only the kinetic equation with the viscous effect suppressed, and they introduce a propagator equation into the formulation during their "simple renormalization" process. As mentioned in Sec. 3, the viscous effect must be counteracted by suitable external forces. This is because the use of macroscopic viscosity amounts to injecting irreversibility into the statistical dynamics in an artificial manner. Nevertheless, the viscous term in perturbation suggests the structure of dynamic relaxation by the nonlinear interaction. Hence, the presence of viscous term, properly counteracted, will not detract from the conceptual consistency. In fact, the simple renormalization of Balescu-Senatorski is automatically accomplished by the viscous term.

A. Modal energy equation

Differentiate (3.9) with respect to t and insert (6.4) for $\partial f_{\{0\}}/\partial t$. With the use of (6.10), we then have

$$\frac{\partial I_{\mathbf{k},\mu}(t)}{\partial t}=\int d\eta\,\eta_{\mathbf{k},\mu}\Bigg\{\angle^{\circ}f_{\{0\}}(\eta,t)+2\int_0^t d\theta\times\sum_{\mathbf{k}}^{+}\sum_{\mathbf{p}}\sum_{\mu,\lambda,\rho}$$



$$f_{\{0\}}(\eta,\theta)+\cdots\Bigg\},\tag{7.1}$$

where the three dots are the higher-order terms of the diagonal fragments. Referring to the formulas in Table I, we can readily recover the corresponding mathematical expression for the diagram term (by dropping μ,λ,ρ and abbreviating $\bar{\phi}^{\mu|\lambda,\rho}_{\mathbf{k}|\mathbf{p},\mathbf{q}}$ by $\bar{\phi}_{\mathbf{k}}$, etc.):

$$\begin{aligned}(i)^2\left(\frac{2\pi}{L}\right)^3&\sum_{\mathbf{k}}^{+}\sum_{\mathbf{p}}\sum_{\mu,\lambda,\rho}\int_0^t d\theta\int d\eta\\&\times\eta_{\mathbf{k}}\left[\bar{\phi}_{\mathbf{k}}\frac{\partial}{\partial\eta_{\mathbf{k}}}-\bar{\phi}_{\mathbf{p}}\frac{\partial}{\partial\eta_{\mathbf{p}}}-\bar{\phi}_{\mathbf{k}-\mathbf{p}}\frac{\partial}{\partial\eta_{\mathbf{k}-\mathbf{p}}}\right]\eta_{\mathbf{k}}^{1/2}\eta_{\mathbf{p}}^{1/2}\eta_{\mathbf{k}-\mathbf{p}}^{1/2}\\&\times\exp[\angle^{\circ}_{\mathbf{k}}(t-\theta)]\exp[\angle^{\circ}_{\mathbf{p}}(t-\theta)]\exp[\angle^{\circ}_{\mathbf{k}-\mathbf{p}}(t-\theta)]\eta_{\mathbf{k}}^{1/2}\eta_{\mathbf{p}}^{1/2}\eta_{\mathbf{k}-\mathbf{p}}^{1/2}\\&\times\left[-\bar{\phi}_{\mathbf{k}}\frac{\partial}{\partial\eta_{\mathbf{k}}}+\bar{\phi}_{\mathbf{p}}\frac{\partial}{\partial\eta_{\mathbf{p}}}+\bar{\phi}_{\mathbf{k}-\mathbf{p}}\frac{\partial}{\partial\eta_{\mathbf{k}-\mathbf{p}}}\right]f_{\{0\}}(\eta,\theta).\end{aligned}\tag{7.2}$$

Carry out the η integration with the product hypothesis. Always assume $\varphi(\eta_{\mathbf{k},\mu},t)$ and $\partial\varphi/\partial\eta_{\mathbf{k},\mu}$ to vanish as $\eta_{\mathbf{k},\mu}\to\infty$. Multiplying out the square brackets, we find that the expression operated by $\bar{\phi}_{\mathbf{p}}\partial/\partial\eta_{\mathbf{p}}+\bar{\phi}_{\mathbf{k}-\mathbf{p}}\partial/\partial\eta_{\mathbf{k}-\mathbf{p}}$ of the first bracket integrates out to zero. After partial integration, the modal energy equation becomes

$$\begin{aligned}\left(\frac{\partial}{\partial t}+2\nu k^2\right)I_{\mathbf{k},\mu}(t)=\left(\frac{2\pi}{L}\right)^3&\sum_{\mathbf{p}}\sum_{\lambda,\rho}\int_0^t d\theta\\&\times\Big\{(\bar{\phi}^{\mu|\lambda,\rho}_{\mathbf{k}|\mathbf{p},\mathbf{k}-\mathbf{p}})^2\,G^{\circ}_{\mathbf{k},\mu}(t-\theta,\theta)\\&\times U^{\circ}_{\mathbf{p},\lambda}(t-\theta,\theta)\,U^{\circ}_{\mathbf{k}-\mathbf{p},\rho}(t-\theta,\theta)\\&-\bar{\phi}^{\mu|\lambda,\rho}_{\mathbf{k}|\mathbf{p},\mathbf{k}-\mathbf{p}}\,\bar{\phi}^{\lambda|\rho,\mu}_{\mathbf{p}|\mathbf{k}-\mathbf{p},\mathbf{k}}\,U^{\circ}_{\mathbf{k},\mu}(t-\theta,\theta)\\&\times G^{\circ}_{\mathbf{p},\lambda}(t-\theta,\theta)\,U^{\circ}_{\mathbf{k}-\mathbf{p},\rho}(t-\theta,\theta)\\&-\bar{\phi}^{\mu|\lambda,\rho}_{\mathbf{k}|\mathbf{p},\mathbf{k}-\mathbf{p}}\,\bar{\phi}^{\rho|\mu,\lambda}_{\mathbf{k}-\mathbf{p}|\mathbf{k},\mathbf{p}}\,U^{\circ}_{\mathbf{k},\mu}(t-\theta,\theta)\\&\times U^{\circ}_{\mathbf{p},\lambda}(t-\theta,\theta)\,G^{\circ}_{\mathbf{k}-\mathbf{p},\rho}(t-\theta,\theta)\Big\}+\cdots,\end{aligned}\tag{7.3}$$

where the three dots are the higher-order terms. The significance of the statistical functions

$$U^{\circ}_{\mathbf{k},\mu}(t-\theta,\theta)=\int d\eta_{\mathbf{k},\mu}\,\eta_{\mathbf{k},\mu}^{1/2}\exp[\angle^{\circ}_{\mathbf{k},\mu}(t-\theta)]\eta_{\mathbf{k},\mu}^{1/2}\varphi(\eta_{\mathbf{k},\mu},\theta),\tag{7.4}$$

$$G^{\circ}_{\mathbf{k},\mu}(t-\theta,\theta)=-\int d\eta_{\mathbf{k},\mu}\,\eta_{\mathbf{k},\mu}^{1/2}\exp[\angle^{\circ}_{\mathbf{k},\mu}(t-\theta)]\eta_{\mathbf{k},\mu}^{1/2}\frac{\partial\varphi(\eta_{\mathbf{k},\mu},\theta)}{\partial\eta_{\mathbf{k},\mu}},\tag{7.5}$$

will be explored presently.

First, consider the velocity components $\eta_{\mathbf{k},\mu}^{1/2}$ at time t and $t+\tau$. During the time interval τ, the flow system is assumed to evolve by the adjoint operator of $\mathcal{L}_{\mathbf{k},\mu}^{\circ}$. We then assert that $U_{\mathbf{k},\mu}^{\circ}(\tau,t)$ is the covariance of such two velocity components. For the simultaneous-time argument, $U_{\mathbf{k},\mu}^{\circ}(\tau,t)$ reduces to the modal energy $I_{\mathbf{k},\mu}(t)$. This is because $U_{\mathbf{k},\mu}^{\circ}(0,t)$ is the average of two $\eta_{\mathbf{k},\mu}^{1/2}$, both at time t, weighted according to $\varphi(\eta_{\mathbf{k},\mu},t)$;

$$U_{\mathbf{k},\mu}^{\circ}(0,t)=\int d\eta_{\mathbf{k},\mu}(\eta_{\mathbf{k},\mu}^{1/2})^2\varphi(\eta_{\mathbf{k},\mu},t). \tag{7.6}$$

Now suppose that the flow system under consideration has the Liouville operator $\mathcal{L}_{\mathbf{k},\mu}^{\circ}$. Then the average of $\eta_{\mathbf{k},\mu}^{1/2}$ at time t and $\eta_{\mathbf{k},\mu}^{1/2}$ at time $t+\tau$ is[13] given by

$$\int d\eta_{\mathbf{k},\mu}\varphi(\eta_{\mathbf{k},\mu},t)\eta_{\mathbf{k},\mu}^{1/2}\exp(\tilde{\mathcal{L}}_{\mathbf{k},\mu}^{\circ}\tau)\eta_{\mathbf{k},\mu}^{1/2}, \tag{7.7}$$

where $\tilde{\mathcal{L}}_{\mathbf{k},\mu}^{\circ}=-2\nu k^2\eta_{\mathbf{k},\mu}\partial/\partial\eta_{\mathbf{k},\mu}$ is the adjoint of $\mathcal{L}_{\mathbf{k},\mu}^{\circ}$. By partial integration, we can reduce (7.7) to the covariance

$$U_{\mathbf{k},\mu}^{\circ}(\tau,t)=\int d\eta_{\mathbf{k},\mu}\eta_{\mathbf{k},\mu}^{1/2}\exp(\mathcal{L}_{\mathbf{k},\mu}^{\circ}\tau)\eta_{\mathbf{k},\mu}^{1/2}\varphi(\eta_{\mathbf{k},\mu},t). \tag{7.8}$$

Secondly, we show that the averaged Green's function

$$G_{\mathbf{k},\mu}^{\circ}(\tau,t)=-\int d\eta_{\mathbf{k},\mu}\eta_{\mathbf{k},\mu}^{1/2}\exp(\mathcal{L}_{\mathbf{k},\mu}^{\circ}\tau)\eta_{\mathbf{k},\mu}^{1/2}\frac{\partial\varphi(\eta_{\mathbf{k},\mu},t)}{\partial\eta_{\mathbf{k},\mu}} \tag{7.9}$$

is identical to the phase-correlation function. Expanding out $\exp(\mathcal{L}_{\mathbf{k},\mu}^{\circ}\tau)$ in (7.8), we obtain by partial integration that

$$U_{\mathbf{k},\mu}^{\circ}(\tau,t)=I_{\mathbf{k},\mu}(t)[1-2\nu k^2\tau/2+(2\nu k^2\tau)^2/2!\,2^2+\cdots]. \tag{7.10}$$

The square bracket is the phase-correlation function. By the expansion of $\exp(\mathcal{L}_{\mathbf{k},\mu}^{\circ}\tau)$ in (7.9), we find that the square bracket of (7.10) is nothing but the averaged Green's function. Hence, this leads to the fluctuation–dissipation relation for a conservative system in thermal equilibrium[1]

$$U_{\mathbf{k},\mu}^{\circ}(\tau,t)=I_{\mathbf{k},\mu}(t)\,G_{\mathbf{k},\mu}^{\circ}(\tau,t). \tag{7.11}$$

With the use of (7.11) the right-hand side of (7.3) is expressed in terms of I and G°. To complete the statistical formulation, we must therefore derive an equation for $G_{\mathbf{k},\mu}^{\circ}$.

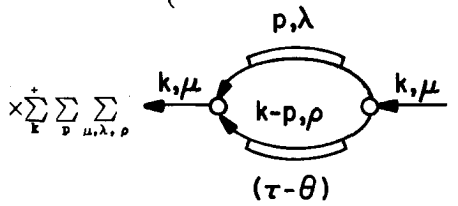
B. Averaged Green's equation

We begin with the turbulent averaged Green's function

$$G_{\mathbf{k},\mu}(\tau,t)=-\int d\eta_{\mathbf{k},\mu}\eta_{\mathbf{k},\mu}^{1/2}\mathcal{G}(\eta_{\mathbf{k},\mu},\tau)\eta_{\mathbf{k},\mu}^{1/2}\frac{\partial\varphi(\eta_{\mathbf{k},\mu},t)}{\partial\eta_{\mathbf{k},\mu}}, \tag{7.12}$$

which is defined by replacing $\exp(\mathcal{L}_{\mathbf{k},\mu}^{\circ}\tau)$ in (7.9) with the Green's operator $\mathcal{G}(\eta_{\mathbf{k},\mu},\tau)$. Now differentiate (7.12) with respect to τ and insert (6.7) for $\partial\mathcal{G}/\partial\tau$. With the use of (6.11) we then have

$$\begin{aligned}\frac{\partial G_{\mathbf{k},\mu}(\tau,t)}{\partial\tau}&=-\int d\eta_{\mathbf{k},\mu}\eta_{\mathbf{k},\mu}^{1/2}\Big\{\mathcal{L}_{\mathbf{k},\mu}^{\circ}\mathcal{G}(\eta_{\mathbf{k},\mu},\tau)+\int_0^{\tau}d\theta\int_{(\mathbf{k},\mu)}d\eta\\ &\times\sum_{\mathbf{k}}^{+}\sum_{\mathbf{p}}\sum_{\mu,\lambda,\rho}\Big(\prod_{(q,\nu)\neq(\mathbf{k},\mu)}\varphi(\eta_{q,\nu},\theta)\Big)\mathcal{G}(\eta_{\mathbf{k},\mu},\theta)\Big\}\eta_{\mathbf{k},\mu}^{1/2}\frac{\partial\varphi(\eta_{\mathbf{k},\mu},t)}{\partial\eta_{\mathbf{k},\mu}}.\end{aligned} \tag{7.13}$$



By consulting Table I, we can readily write down the mathematical expression for the diagram term

$$\begin{aligned}&-\frac{(i)^2}{2}\left(\frac{2\pi}{L}\right)^3\sum_{\mathbf{p}}\sum_{\lambda,\rho}\int_0^{\tau}d\theta\int d\eta_{\mathbf{k}}d\eta_{\mathbf{p}}d\eta_{\mathbf{k}-\mathbf{p}}\\ &\times\eta_{\mathbf{k}}\left[-\bar{\phi}_{\mathbf{k}}\frac{\partial}{\partial\eta_{\mathbf{k}}}+\bar{\phi}_{\mathbf{p}}\frac{\partial}{\partial\eta_{\mathbf{p}}}+\bar{\phi}_{\mathbf{k}-\mathbf{p}}\frac{\partial}{\partial\eta_{\mathbf{k}-\mathbf{p}}}\right]\\ &\times\eta_{\mathbf{p}}^{1/2}\eta_{\mathbf{k}-\mathbf{p}}^{1/2}\exp[\mathcal{L}_{\mathbf{p}}^{\circ}(\tau-\theta)]\exp[\mathcal{L}_{\mathbf{k}-\mathbf{p}}^{\circ}(\tau-\theta)]\eta_{\mathbf{p}}^{1/2}\eta_{\mathbf{k}-\mathbf{p}}^{1/2}\\ &\times\left[\bar{\phi}_{\mathbf{k}}\frac{\partial}{\partial\eta_{\mathbf{k}}}-\bar{\phi}_{\mathbf{p}}\frac{\partial}{\partial\eta_{\mathbf{p}}}-\bar{\phi}_{\mathbf{k}-\mathbf{p}}\frac{\partial}{\partial\eta_{\mathbf{k}-\mathbf{p}}}\right]\eta_{\mathbf{k}}^{1/2}\\ &\times\varphi(\eta_{\mathbf{p}},\theta)\varphi(\eta_{\mathbf{k}-\mathbf{p}},\theta)\mathcal{G}(\eta_{\mathbf{k}},\theta)\\ &\times\eta_{\mathbf{k}}^{1/2}\frac{\partial\varphi(\eta_{\mathbf{k}},t)}{\partial\eta_{\mathbf{k}}}.\end{aligned} \tag{7.14}$$

Again, the expression operated by $\bar{\phi}_{\mathbf{p}}\partial/\partial\eta_{\mathbf{p}}+\bar{\phi}_{\mathbf{k}-\mathbf{p}}\partial/\partial\eta_{\mathbf{k}-\mathbf{p}}$ of the first square bracket integrates out to zero. Further, $\bar{\phi}_{\mathbf{k}}\partial/\partial\eta_{\mathbf{k}}$ of the first square bracket operating on $\bar{\phi}_{\mathbf{k}}\partial/\partial\eta_{\mathbf{k}}$ of the second square bracket gives no contribution. Hence, the averaged Green's equation becomes

$$\begin{aligned}&\left(\frac{\partial}{\partial\tau}+\nu k^2\right)G_{\mathbf{k},\mu}(\tau,t)\\ &=-\frac{1}{2}\left(\frac{2\pi}{L}\right)^3\sum_{\mathbf{p}}\sum_{\lambda,\rho}\int_0^{\tau}d\theta\\ &\times[\bar{\phi}_{\mathbf{k}}^{\mu|\lambda,\rho}{}_{\mathbf{p},\mathbf{k}-\mathbf{p}}\,\bar{\phi}_{\mathbf{p}|\mathbf{k}-\mathbf{p},\mathbf{k}}^{\lambda|\rho,\mu}G_{\mathbf{k},\mu}(\theta,t)G_{\mathbf{p},\lambda}^{\circ}(\tau-\theta,\theta)\\ &\times U_{\mathbf{k}-\mathbf{p},\rho}^{\circ}(\tau-\theta,\theta)+\bar{\phi}_{\mathbf{k}}^{\mu|\lambda,\rho}{}_{\mathbf{p},\mathbf{k}-\mathbf{p}}\,\bar{\phi}_{\mathbf{k}-\mathbf{p}|\mathbf{k},\mathbf{p}}^{\rho|\mu,\lambda}G_{\mathbf{k},\mu}(\theta,t)\\ &\times U_{\mathbf{p},\lambda}^{\circ}(\tau-\theta,\theta)G_{\mathbf{k}-\mathbf{p},\rho}^{\circ}(\tau-\theta,\theta)]+\cdots,\end{aligned} \tag{7.15}$$

where the three dots represent the higher-order terms. The initial condition is $G_{\mathbf{k},\mu}(0,t)=1$.

C. Renormalization

The modal energy and averaged Green's equations are the laminar perturbation expansions; hence they are of no use unless the nonlinear interaction is weak. For strong turbulence, Wyld has shown that certain classes of expansion terms can be consolidated, thereby incorporating the dynamic effect of a certain kind of the nonlinear interactions of all orders. Although Wyld's

work is based on the moment formulation, his observation is valid for the distribution function formulation because of the diagram similarity. Of all the expansion terms in (7.3) and (7.15), there are those made up of compositing a number of the respective first-order terms. Then summing up such expansion terms amounts to replacing $U^\circ \to U$ and $G^\circ \to G$ in (7.3) and (7.15). In analogy to (7.12), we define the turbulent covariance

$$U_{\mathbf{k},\mu}(\tau,t)=\int d\eta_{\mathbf{k},\mu}\,\eta_{\mathbf{k},\mu}^{1/2}\mathcal{G}(\eta_{\mathbf{k},\mu},\tau)\eta_{\mathbf{k},\mu}^{1/2}\varphi(\eta_{\mathbf{k},\mu},t), \quad (7.16)$$

by replacing $\exp(\mathcal{L}^\circ_{\mathbf{k},\mu}\tau)$ in (7.8) with the Green's operator. Balescu and Senatorski justified this heuristic renormalization rule by invoking Resibois' factorization theorem[15] which asserts that if two subgroups of eddies are temporarily interacting independently one from the other, the time ordering between the interactions involving eddies of the first group and eddies of the second groups is completely irrelevant. In any event, the modal energy and averaged Green's equations have the renormalized lowest-order contribution:

$$\left(\frac{\partial}{\partial t}+2\nu k^2\right)I_{\mathbf{k},\mu}(t)$$
$$=\left(\frac{2\pi}{L}\right)^3\sum_{\mathbf{p}}\sum_{\lambda,\rho}\int_0^t d\theta$$
$$\times\Big\{(\bar\phi^{\mu|\lambda,\rho}_{\mathbf{k}|\mathbf{p},\mathbf{k}-\mathbf{p}})^2 G_{\mathbf{k},\mu}(t-\theta,\theta)U_{\mathbf{p},\lambda}(t-\theta,\theta)U_{\mathbf{k}-\mathbf{p},\rho}(t-\theta,\theta)$$
$$-\bar\phi^{\mu|\lambda,\rho}_{\mathbf{k}|\mathbf{p},\mathbf{k}-\mathbf{p}}\bar\phi^{\lambda|\ \rho,\mu}_{\mathbf{p}|\mathbf{k}-\mathbf{p},\mathbf{k}}U_{\mathbf{k},\mu}(t-\theta,\theta)G_{\mathbf{p},\lambda}(t-\theta,\theta)$$
$$\times U_{\mathbf{k}-\mathbf{p},\rho}(t-\theta,\theta)-\bar\phi^{\mu|\lambda,\rho}_{\mathbf{k}|\mathbf{p},\mathbf{k}-\mathbf{p}}\bar\phi^{\rho|\mu,\lambda}_{\mathbf{k}-\mathbf{p}|\mathbf{k},\mathbf{p}}U_{\mathbf{k},\mu}(t-\theta,\theta)$$
$$\times U_{\mathbf{p},\lambda}(t-\theta,\theta)G_{\mathbf{k}-\mathbf{p},\rho}(t-\theta,\theta)\Big\}, \quad (7.17)$$

$$\left(\frac{\partial}{\partial\tau}+\nu k^2\right)G_{\mathbf{k},\mu}(\tau,t)$$
$$=-\frac{1}{2}\left(\frac{2\pi}{L}\right)^3\sum_{\mathbf{p}}\sum_{\lambda,\rho}\int_0^\tau d\theta$$
$$\times\Big\{\bar\phi^{\mu|\lambda,\rho}_{\mathbf{k}|\mathbf{p},\mathbf{k}-\mathbf{p}}\bar\phi^{\lambda|\ \rho,\mu}_{\mathbf{p}|\mathbf{k}-\mathbf{p},\mathbf{k}}G_{\mathbf{k},\mu}(\theta,t)G_{\mathbf{p},\lambda}(\tau-\theta,\theta)$$
$$\times U_{\mathbf{k}-\mathbf{p},\rho}(\tau-\theta,\theta)+\bar\phi^{\mu|\lambda,\rho}_{\mathbf{k}|\mathbf{p},\mathbf{k}-\mathbf{p}}\bar\phi_{\mathbf{k}-\mathbf{p}|\mathbf{k},\mathbf{p}}^{\ \rho|\mu,\lambda}G_{\mathbf{k},\mu}(\theta,t)$$
$$\times U_{\mathbf{p},\lambda}(\tau-\theta,\theta)G_{\mathbf{k}-\mathbf{p},\rho}(\tau-\theta,\theta)\Big\}, \quad (7.18)$$

with $G_{\mathbf{k},\mu}(0,t)=1$. And, the fluctuation–dissipation relation in renormalized form,

$$U_{\mathbf{k},\mu}(\tau,t)=I_{\mathbf{k},\mu}(t)\,G_{\mathbf{k},\mu}(\tau,t), \quad (7.19)$$

provides the link.

We shall show in Secs. 8 and 9 that (7.17)–(7.19) are Herring's SCFA equations. They describe the evolution of covariance for the simultaneous-time arguments and prescribe the covariance for the nonsimultaneous-time arguments by (7.19). Since Kraichnan's DIA evolves the covariance in the entire t–t' plane, we show the correspondence between (7.17)–(7.19) and the DIA equations in the simultaneous-time limit. According to the definitions (7.12) and (7.16), the reference time t refers to the overall flow system, whereas the difference time τ reflects the dynamic relaxation. To accentuate the dynamic relaxation, we shall therefore suppress the reference time argument: $U_{\mathbf{k},\mu}(\tau,t)\to U_{\mathbf{k},\mu}(\tau)$, and $G_{\mathbf{k},\mu}(\tau,t)\to G_{\mathbf{k},\mu}(\tau)$. Invoking the isotropy requirements (3.10), we impose the rotational symmetry, $U_{\mathbf{k},\mu}(\tau)=U_{\mathbf{k}}(\tau)/2$, $I_{\mathbf{k},\mu}(t)=I_{\mathbf{k}}(t)/2$, $G_{\mathbf{k},\mu}(t)=G_{\mathbf{k}}(\tau)$; and the spherical symmetry, $U_{\mathbf{k}}(\tau)=U(k,\tau)$, $I_{\mathbf{k}}(t)=I(k,t)$, $G_{\mathbf{k}}(\tau)=G(k,\tau)$. Further, identify $I(k,t)$ with $U(k,t)$, implying that $I(k,t)$ is the simultaneous-time covariance $U(k,t)$. After summing over μ, we obtain the isotropic form of (7.17) and (7.18):

$$\left(\frac{\partial}{\partial t}+2\nu k^2\right)U(k,t)$$
$$=2\pi k\int_0^t d\theta\int_\Delta\int dp\,dq\,pq\{a(k,p,q)G(k,t-\theta)U(p,t-\theta)$$
$$\times U(k-p,t-\theta)-b(k,p,q)U(k,t-\theta)G(p,t-\theta)$$
$$\times U(k-p,t-\theta)\}, \quad (7.20)$$

$$\left(\frac{\partial}{\partial\tau}+\nu k^2\right)G(k,\tau)$$
$$=-\pi k\int_0^\tau d\theta\int_\Delta\int dp\,dq\,pqb(k,p,q)G(k,\tau-\theta)G(p,\theta)U(k-p,\theta). \quad (7.21)$$

Here we have used the usual notations:

$$\lim_{L\to\infty}(2\pi/L)^3\sum_{\mathbf{p}}\to\int d\mathbf{p}=2\pi\int_\Delta\int dp\,dq(pq/k),$$

$$a(k,p,q)=(1-xyz-2z^2y^2)/2,$$

and

$$b(k,p,q)=(p/k)(xy+z^3),$$

where x,y,z are the cosines of the interior angles opposite to the legs k,p,q, respectively. Note that a and b are derived from

$$\sum_{\mu,\lambda,\rho}(\bar\phi^{\mu|\lambda,\rho}_{\mathbf{k}|\mathbf{p},\mathbf{k}-\mathbf{p}})^2=2k^2(1-xyz-2z^2y^2)$$

and

$$\sum_{\mu,\lambda,\rho}\bar\phi^{\mu|\lambda,\rho}_{\mathbf{k}|\mathbf{p},\mathbf{k}-\mathbf{p}}\bar\phi^{\lambda|\ \rho,\mu}_{\mathbf{p}|\mathbf{k}-\mathbf{p},\mathbf{k}}=2kp(xy+z^3).$$

We see that (7.20) and (7.21) are identical to the isotropic DIA equation in the stationary turbulent field, evolving along the diagonal of the t–t' plane.[2]

8. PERTURBATION ABOUT A TURBULENT FLOW

In Sec. 7.3, renormalization was carried out by simply replacing $\exp(\mathcal{L}^\circ_{\mathbf{k},\mu}t)$ in the first-order laminar expansion terms by the Green's operator. Of course, this recipe is too heuristic to stand up to a rigorous argument. As in Wyld's work, it is necessary to demonstrate that the renormalized term actually includes a certain class of the expansion terms of all orders. Such a renormalization procedure is indeed very tedious. We shall therefore propose a way out of it, thereby avoiding the actual summation of laminar expansion terms. A lesson learned from the heuristic renormalization is this: If we would develop perturbation about a reference flow state which has the Green's operator $\mathcal{G}(\eta_{\mathbf{k},\mu},t)$, then the modal energy and averaged Green's equations will have the lowest-order terms given directly by (7.17) and (7.18). Unlike the laminar perturbation, we do not know *a priori* the reference flow state. In the present perturbation, therefore, we must treat the reference state as an unknown and determine it along with the turbulence dynamic equations. This sort of philosophy is common to recent turbulence

theories.[4,8,16]

Let us go back to the Liouville equation (5.2). Introduce an arbitrary operator $\mathcal{L}''$ with the requirement that it be decomposable $\mathcal{L}'' = \sum_{\mathbf{k}}^{+} \sum_{\mu} \mathcal{L}''_{\mathbf{k},\mu}$. We add $\mathcal{L}'' f_{\{m\}}(\eta, t)$ to both sides of (5.2) and put it in the form

$$\frac{\partial f_{\{m\}}(\eta, t)}{\partial t} = \mathcal{L}^{T} f_{\{m\}}(\eta, t) + \lambda \sum_{\{m'\}} \langle m | \mathcal{L}' | m' \rangle f_{\{m'\}}(\eta, t) - \lambda^2 \mathcal{L}'' f_{\{m\}}(\eta, t), \tag{8.1}$$

where $\mathcal{L}^{T} = \mathcal{L}^{\circ} + \mathcal{L}''$. The ordering parameter λ is introduced into the above. No claim is made of authority of the particular orders assigned to the right-hand side terms of (8.1): It is justified *a posteriori* from the structure of the nonlinear interaction terms. This arbitrariness cannot be eliminated from the dynamic consideration alone. Hence, herein lies the vulnerability of present turbulence theories which must cope with a strongly interacting many-body problem but with no parameter to guide us in developing a systematic approximation. For $\lambda = 1$, (8.1) reduces to the original problem (5.2). The zeroth-order problem ($\lambda = 0$) is the reference flow state. Since it is not known *a priori*, we do not seek a conventional perturbation expansion in powers of λ. Rather, the aim is to choose the zeroth-order problem in such a way that it can best approximate the actual turbulence dynamics in a statistical sense.

Following Sec. 5, we derive a kinetic equation with the diagonal fragments of $O(\lambda^2)$

$$\begin{aligned}\frac{\partial f_{\{m\}}(\eta, t)}{\partial t} &= \mathcal{L}^{T} f_{\{m\}}(\eta, t) + \lambda \sum_{\{m'\}} \langle m | \mathcal{L}' | m' \rangle \exp(\mathcal{L}^{T} t) f_{\{m'\}}(\eta, 0) \\ &+ \lambda^2 \Big(\sum_{\{m'\}} \int_0^t dt' \, \langle m | \mathcal{L}' | m' \rangle \exp[\mathcal{L}^{T}(t - t')] \langle m' | \mathcal{L}' | m \rangle \\ &\times f_{\{m\}}(\eta, t') - \mathcal{L}'' f_{\{m\}}(\eta, t) \Big).\end{aligned} \tag{8.2}$$

Suppose that we wish to describe the actual turbulence dynamics by the zeroth-order problem

$$\frac{\partial f_{\{m\}}(\eta, t)}{\partial t} = \mathcal{L}^{T} f_{\{m\}}(\eta, t). \tag{8.3}$$

Since $\mathcal{L}^{T}$ is decomposable by definition, we see that (8.3) is readily amenable to solution, whereas the original problem (5.2) may not. In order for (8.3) to represent the actual turbulence dynamics, we must require that the second and third terms of (8.2), both multiplied by λ, vanish identically:

$$\begin{aligned}0 &= \lambda \sum_{\{m'\}} \langle m | \mathcal{L}' | m' \rangle f_{\{m'\}}(\eta, t) \\ &+ \lambda^2 \Big(\sum_{\{m'\}} \int_0^t dt' \, \langle m | \mathcal{L}' | m' \rangle \exp[\mathcal{L}^{T}(t - t')] \langle m' | \mathcal{L}' | m \rangle \\ &\times f_{\{m\}}(\eta, t') - \mathcal{L}'' f_{\{m\}}(\eta, t) \Big).\end{aligned} \tag{8.4}$$

In the first term, we have replaced $\exp(\mathcal{L}^{T} t) f_{\{m'\}}(\eta, 0)$ by $f_{\{m'\}}(\eta, t)$ because they are related to each other by the solution of (8.3). Since the transition matrix does not allow $\{m'\} = \{m\}$, the first term of (8.4) makes no contribution. Therefore, the operator satisfying

$$\mathcal{L}'' f_{\{m\}}(\eta, t) = \sum_{\{m'\}} \int_0^t dt' \, \langle m | \mathcal{L}' | m' \rangle \exp[\mathcal{L}^{T}(t - t')] \times \langle m' | \mathcal{L}' | m \rangle f_{\{m\}}(\eta, t') \tag{8.5}$$

assures fulfillment of (8.4). The zeroth-order problem can thus approximate the acutal turbulence in that $\mathcal{L}''$ describes certain of the nonlinear dynamics.

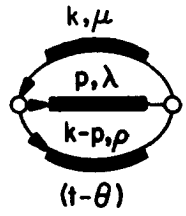
A. Turbulence dynamic equations

The introduction of (8.5) into (8.3) gives

$$\frac{\partial f_{\{m\}}(\eta, t)}{\partial t} = \mathcal{L}^{\circ} f_{\{m\}}(\eta, t) + \sum_{\{m'\}} \int_0^t dt' \, \langle m | \mathcal{L}' | m' \rangle \times \exp[\mathcal{L}^{T}(t - t')] \langle m' | \mathcal{L}' | m \rangle f_{\{m\}}(\eta, t'). \tag{8.6}$$

This is identical to the lowest-order of (6.4), if we replace $\exp(\mathcal{L}^{\circ} t)$ by $\exp(\mathcal{L}^{T} t)$. Consequently, the modal energy equation derived from (8.6) would be the same as (7.1) except the double lines are replaced by bold lines representing $\exp(\mathcal{L}^{T}_{\mathbf{k},\mu} t)$:

$$\frac{\partial I_{\mathbf{k},\mu}(t)}{\partial t} = \int d\eta \, \eta_{\mathbf{k},\mu} \Big(\mathcal{L}^{\circ} f_{\{0\}}(\eta, t) + 2 \int_0^t d\theta \times \sum_{\mathbf{k}}^{+} \sum_{\mathbf{p}} \sum_{\mu,\lambda,\rho} \quad f_{\{0\}}(\eta, \theta) \Big).$$



(8.7)

Hence, this should directly give the renormalized modal energy equation (7.17). The turbulent covariance and averaged Green's functions are defined respectively by (7.16) and (7.12) with the identification

$$\mathcal{G}(\eta_{\mathbf{k},\mu}, \tau) = \exp(\mathcal{L}^{T}_{\mathbf{k},\mu} \tau). \tag{8.8}$$

Next, the averaged Green's equation is given by the prescription

$$\frac{\partial G_{\mathbf{k},\mu}(\tau, t)}{\partial \tau} = -\int d\eta_{\mathbf{k},\mu} \, \eta^{1/2}_{\mathbf{k},\mu} \Big(\frac{\partial \mathcal{G}(\eta_{\mathbf{k},\mu}, \tau)}{\partial \tau} \Big) \eta^{1/2}_{\mathbf{k},\mu} \Big(\frac{\partial \varphi(\eta_{\mathbf{k},\mu}, t)}{\partial \eta_{\mathbf{k},\mu}} \Big). \tag{8.9}$$

We note from (8.8) that

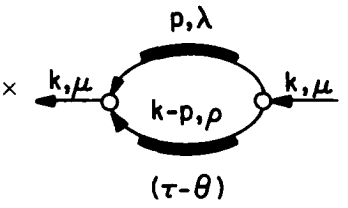
$$\frac{\partial \mathcal{G}(\eta_{\mathbf{k},\mu}, \tau)}{\partial \tau} = \mathcal{L}^{T}_{\mathbf{k},\mu} \mathcal{G}(\eta_{\mathbf{k},\mu}, \tau). \tag{8.10}$$

For the evaluation of the right-hand side, consider the identity

$$\mathcal{L}^{T} f_{\{1_{\mathbf{k},\mu}\}}(\eta, \tau) = (\mathcal{L}^{\circ} + \mathcal{L}'') f_{\{1_{\mathbf{k},\mu}\}}(\eta, \tau). \tag{8.11}$$

Under the product hypothesis (6.6), we integrate (8.11) over all action variables except $\eta_{\mathbf{k},\mu}$:

$$\begin{aligned}&\mathcal{L}^{T}_{\mathbf{k},\mu} \mathcal{G}(\eta_{\mathbf{k},\mu}, \tau) \\ &= \mathcal{L}^{\circ}_{\mathbf{k},\mu} \mathcal{G}(\eta_{\mathbf{k},\mu}, \tau) + \int_0^{\tau} d\theta \sum_{\mathbf{p}} \sum_{\lambda,\rho} \int d\eta_{\mathbf{p},\lambda} \, d\eta_{\mathbf{k}-\mathbf{p},\rho}\end{aligned}$$



$$\times \varphi(\eta_{\mathbf{p},\lambda}, \theta) \varphi(\eta_{\mathbf{k}-\mathbf{p},\rho}, \theta) \mathcal{G}(\eta_{\mathbf{k},\mu}, \theta). \tag{8.12}$$

After the final integration over $\eta_{\mathbf{p}}$ and $\eta_{\mathbf{k}-\mathbf{p}}$, we have

$$\begin{aligned}&\mathcal{L}^T_{\mathbf{k},\mu}\,\mathcal{G}(\eta_{\mathbf{k},\mu},\tau)\\&=\mathcal{L}^0_{\mathbf{k},\mu}\,\mathcal{G}(\eta_{\mathbf{k},\mu},\tau)+\tfrac{1}{2}(2\pi/L)^3\sum_{\mathbf{p}}\sum_{\lambda,\rho}\int_0^\tau d\theta\\&\times[\bar{\phi}^{\mu|\lambda,\rho}_{\mathbf{k}|\mathbf{p},\mathbf{k}-\mathbf{p}}\bar{\phi}^{\lambda|\ \rho,\mu}_{\mathbf{p}|\mathbf{k}-\mathbf{p},\mathbf{k}}G_{\mathbf{p},\lambda}(\tau-\theta,\theta)U_{\mathbf{k}-\mathbf{p},\rho}(\tau-\theta,\theta)\\&+\bar{\phi}^{\mu|\lambda,\rho}_{\mathbf{k}|\mathbf{p},\mathbf{k}-\mathbf{p}}\bar{\phi}^{\rho|\ \mu,\lambda}_{\mathbf{k}-\mathbf{p}|\mathbf{k},\mathbf{p}}U_{\mathbf{p},\lambda}(\tau-\theta,\theta)G_{\mathbf{k}-\mathbf{p},\rho}(\tau-\theta,\theta)]\\&\times\eta^{1/2}_{\mathbf{k},\mu}\,\partial/\partial\eta_{\mathbf{k},\mu}\,\eta^{1/2}_{\mathbf{k},\mu}\mathcal{G}(\eta_{\mathbf{k},\mu},\theta).\end{aligned}\tag{8.13}$$

Upon introducing (8.13) into (8.9), we recover at once the renormalized Green's equation (7.18).

Finally, (7.19) is recovered by a procedure similar to that we used to derive (7.11). Although $\mathcal{L}^T_{\mathbf{k},\mu}$ now involves $\eta^{1/2}_{\mathbf{k},\mu}\partial/\partial\eta_{\mathbf{k},\mu}\eta^{1/2}_{\mathbf{k},\mu}$ in addition to $\partial/\partial\eta_{\mathbf{k},\mu}\eta_{\mathbf{k},\mu}$, both the expansion of $\exp(\mathcal{L}^T_{\mathbf{k},\mu}\tau)$ and partial integration can be carried out just as in Sec. 7A. In this way, we have rederived the SCF equations as the first-order perturbation about a turbulent flow state but without summing the laminar expansion terms. Conceptually, we can extend the kinetic equation (8.2) to higher orders in λ, thereby including in the operator $\mathcal{L}^T$ the dynamic effect of arbitrarily complicated interactions. This, however, does not seem to be a workable way of refining the turbulence theory because such a series, even if we find one, is likely to diverge.[17]

B. Covariance equations

From the covariance definition, we can obtain two kinds of covariance equations. For the first kind, we differentiate (7.16) with respect to τ:

$$\begin{aligned}&\partial U_{\mathbf{k},\mu}(\tau,t)/\partial\tau\\&=\int d\eta_{\mathbf{k},\mu}\eta^{1/2}_{\mathbf{k},\mu}\mathcal{L}^T_{\mathbf{k},\mu}\,\mathcal{G}(\eta_{\mathbf{k},\mu},\tau)\eta^{1/2}_{\mathbf{k},\mu}\varphi(\eta_{\mathbf{k},\mu},t).\end{aligned}\tag{8.14}$$

Since it has the same form as (8.9), we can immediately write down the equation for $U_{\mathbf{k},\mu}(\tau,t)$ in τ by replacing $G_{\mathbf{k},\mu}(\theta,t)$ in (7.18) by $U_{\mathbf{k},\mu}(\theta,t)$. The use of (8.14) is in deriving the fluctuation–dissipation relation, which we have already deduced by the direct term-by-term comparison of the covariance and averaged Green's funtions. Therefore, (8.14) provides no new information in our theory. Without going into detail, we point out that the autocorrelation equation of Balescu–Senatorski is basically of this kind, although they have incorporated the evolution of $\varphi(\eta_{\mathbf{k},\mu},t)$ by another kinetic equation.

Next, to derive the second kind of covariance equation, we reintroduce into (7.16) the factor

$$\prod_{(\mathbf{p},\nu)\neq(\mathbf{k},\mu)}\int d\eta_{\mathbf{p},\nu}\varphi(\eta_{\mathbf{p},\nu},t),$$

which is unity. Now the differentiation with respect to t rather than τ gives

$$\begin{aligned}&\frac{\partial U_{\mathbf{k},\mu}(\tau,t)}{\partial t}\\&=\int d\eta\,\eta^{1/2}_{\mathbf{k},\mu}\,\mathcal{G}(\eta_{\mathbf{k},\mu},\tau)\eta^{1/2}_{\mathbf{k},\mu}\,[\partial f_{\{0\}}(\eta,t)/\partial t].\end{aligned}\tag{8.15}$$

The introduction of (8.6) into the above gives

$$\begin{aligned}&\left(\frac{\partial}{\partial t}+2\nu k^2\right)U_{\mathbf{k},\mu}(\tau,t)\\&=\left(\frac{2\pi}{L}\right)^3\sum_{\mathbf{p}}\sum_{\lambda,\rho}\int_0^t d\theta\\&\times\{(\bar{\phi}^{\mu|\lambda,\rho}_{\mathbf{k}|\mathbf{p},\mathbf{k}-\mathbf{p}})^2G_{\mathbf{k},\mu}(\tau+t-\theta,\theta)U_{\mathbf{p},\lambda}(t-\theta,\theta)U_{\mathbf{k}-\mathbf{p},\rho}(t-\theta,\theta)\\&-\bar{\phi}^{\mu|\lambda,\ \rho}_{\mathbf{k}|\mathbf{p},\mathbf{k}-\mathbf{p}}\bar{\phi}^{\lambda|\ \rho,\mu}_{\mathbf{p}|\mathbf{k}-\mathbf{p},\mathbf{k}}U_{\mathbf{k},\mu}(\tau+t-\theta,\theta)G_{\mathbf{p},\lambda}(t-\theta,\theta)\\&\times U_{\mathbf{k}-\mathbf{p},\rho}(t-\theta,\theta)-\bar{\phi}^{\mu|\lambda,\rho}_{\mathbf{k}|\mathbf{p},\mathbf{k}-\mathbf{p}}\bar{\phi}^{\rho|\ \mu,\lambda}_{\mathbf{k}-\mathbf{p}|\mathbf{k},\mathbf{p}}U_{\mathbf{k},\mu}(\tau+t-\theta,\theta)\\&\times U_{\mathbf{p},\lambda}(t-\theta,\theta)G_{\mathbf{k}-\mathbf{p},\rho}(t-\theta,\theta)\}.\end{aligned}\tag{8.16}$$

We have used the identity (dropping μ)

$$\begin{aligned}&\int d\eta_{\mathbf{k}}\eta^{1/2}_{\mathbf{k}}\exp(\mathcal{L}^T_{\mathbf{k}}\tau)\eta^{1/2}_{\mathbf{k}}\frac{\partial}{\partial\eta_{\mathbf{k}}}\eta^{1/2}_{\mathbf{k}}F\\&=-\int d\eta_{\mathbf{k}}\eta^{1/2}_{\mathbf{k}}\exp(\mathcal{L}^T_{\mathbf{k}}\tau)F,\end{aligned}$$

where F is a function of $f_{\{0\}}$. Note that (8.16) is a trivial generalization of (7.17). It evolves along the simultaneous-time argument just as (7.17). In the distribution function formalism, all that we can do is to describe the covariance evolution along the diagonal of the t–t' plane, whereas the relaxation in the off-diagonal is prescribed by the fluctuation–dissipation relation. This is the inherent limitation of the distribution function formulation. In contradistinction, the DIA can evolve the covariance in the entire t–t' plane by two statistical equations, one in t and the other in t'. Since the modal energy is the simultaneous-time limit of the covariance, the moment formulation can provide more complete statistical information than the distribution function formalism. In some applications, however, the one-time nature of the distribution function formulation is a blessing rather than a curse because it trades the ease of computation for the loss of statistical information.

9. HERRING'S SELF-CONSISTENT-FIELD APPROXIMATION

Herring's SCFA also represents perturbation about a turbulent flow field and hence is in spirit very similar to the present perturbation theory. To comment on the SCFA, we shall rederive (8.6) by transcribing Herring's procedure to the triad-interaction representation in action-angle variables. Again the Liouville equation (5.2) is the starting point. We propose to examine the following kinetic equation:

$$\begin{aligned}\partial f_{\{m\}}(\eta,t)/\partial t=&\int_0^t dt'\,\mathcal{L}^H(t-t')f_{\{m\}}(\eta,t')\\&+\sum_{\{m'\}}\int_0^t dt'\,\mathcal{V}_{mm'}(t-t')f_{\{m'\}}(\eta,t').\end{aligned}\tag{9.1}$$

Here, $\mathcal{L}^H(t-t')$ is an arbitrary, but decomposable Liouville operator $\mathcal{L}^H=\sum_{\mathbf{k}}^{+}\sum_{\mu}\mathcal{L}^H_{\mathbf{k},\mu}$, and the operator

$$\begin{aligned}\mathcal{V}_{mm'}(t-t')=&-\mathcal{L}^H(t-t')\,\delta(m-m')+\mathcal{L}^0\delta(t-t')\,\delta(m-m')\\&+\langle m|\mathcal{L}'|m'\rangle\,\delta(t-t')\end{aligned}\tag{9.2}$$

is defined so as to relate (9.1) with (5.2). In parallel to (8.3), the aim is to describe the actual turbulence by the kinetic equation of a simpler form

$$\partial f^H_{\{m\}}(\eta,t)/\partial t=\int_0^t dt'\,\mathcal{L}^H(t-t')f^H_{\{m\}}(\eta,t'). \tag{9.3}$$

By the superscript H, we are allowing for the possibility that $f^H_{\{m\}}$ may be different from the actual distribution function at any stage of approximation. Associated with (9.3) is the Green's equation

$$\partial g_{\{m\}}(\eta,t-t')/\partial t=\int_0^t d\theta\,\mathcal{L}^H(t-\theta)g_{\{m\}}(\eta,\theta-t')+\delta(t-t'), \tag{9.4}$$

whose solution is

$$g_{\{m\}}(\eta,t-t')=1+\int_{t'}^t d\theta\int_0^\theta d\theta'\,\mathcal{L}^H(\theta-\theta')g_{\{m\}}(\eta,\theta'-t'). \tag{9.5}$$

By treating the second term of (9.1) as the inhomogeneous term, the formal solution becomes

$$f_{\{m\}}(\eta,t)=f^H_{\{m\}}(\eta,t)+\sum_{\{m'\}}\int_0^t dt'\,g_{\{m\}}(\eta,t-t')\times\int_0^{t'}d\theta\,\mathcal{V}_{mm'}(t'-\theta)f_{\{m'\}}(\eta,\theta). \tag{9.6}$$

Now solve this by iteration

$$f_{\{m'\}}(\eta,t)=f^H_{\{m'\}}(\eta,t)+\sum_{\{m''\}}\int_0^t dt'\,g_{\{m'\}}(\eta,t-t')\times\int_0^{t'}d\theta\,\mathcal{V}_{m'm''}(t'-\theta)f^H_{\{m''\}}(\eta,\theta)+\cdots. \tag{9.7}$$

The iterative expansion (9.7) can be consolidated into a closed form by using the auxiliary operator satisfying a Dyson's equation; however, such is superfluous because the leading terms are all we need explicitly for the present purpose.

The formulation thus far has been formalistic. To initiate the SCFA, we apply $\sum_{\{m'\}}\langle m|\mathcal{L}'|m'\rangle$ to the right of (9.7):

$$\sum_{\{m'\}}\langle m|\mathcal{L}'|m'\rangle f_{\{m'\}}(\eta,t)=\sum_{\{m'\}}\langle m|\mathcal{L}'|m'\rangle f^H_{\{m'\}}(\eta,t)+\sum_{\{m'\}}\sum_{\{m''\}}\int_0^t dt'\int_0^{t'}d\theta\times\langle m|\mathcal{L}'|m'\rangle g_{\{m'\}}(\eta,t-t')\mathcal{V}_{m'm''}(t'-\theta)f^H_{\{m''\}}(\eta,\theta)+\cdots. \tag{9.8}$$

At this point in the perturbation theory, it is important to observe the following steps in the order to be stated. First, we identify the right-hand side of (9.8) with the last term of (5.2); hence we write

$$\partial f_{\{m\}}(\eta,t)/\partial t-\mathcal{L}^\circ f_{\{m\}}(\eta,t)=\text{rhs of (9.8)}. \tag{9.9}$$

Secondly, we replace $f_{\{m\}}$ with $f^H_{\{m\}}$ on the ground that the two distributions must eventually be indistinguishable, if the perturbation scheme is expected to work:

$$\partial f^H_{\{m\}}(\eta,t)/\partial t-\mathcal{L}^\circ f^H_{\{m\}}(\eta,t)=\sum_{\{m'\}}\langle m|\mathcal{L}'|m'\rangle f^H_{\{m'\}}(\eta,t)+\sum_{\{m'\}}\sum_{\{m''\}}\int_0^t dt'\int_0^{t'}d\theta\times\langle m|\mathcal{L}'|m'\rangle g_{\{m'\}}(\eta,t-t')\mathcal{V}_{m'm''}(t'-\theta)f^H_{\{m''\}}(\eta,\theta)+\cdots. \tag{9.10}$$

This second step should not preceed the first. That is, if we had applied $f_{\{m\}}\to f^H_{\{m\}}$ directly to (9.8), it would have resulted in cancellation of the two terms adjacent to the equality sign. Since the transition matrix $\langle m|\mathcal{L}'|m'\rangle$ does not allow $\{m\}=\{m'\}$, the first right-hand side term of (9.10) makes no contribution. Similarly, only the $\langle m'|\mathcal{L}'|m''\rangle\,\delta(t-t')$ of $\mathcal{V}_{m'm''}$ will contribute to the last term of (9.10):

$$\partial f^H_{\{m\}}(\eta,t)/\partial t=\mathcal{L}^\circ f^H_{\{m\}}(\eta,t)+\sum_{\{m'\}}\int_0^t dt'\,\langle m|\mathcal{L}'|m'\rangle\times g_{\{m'\}}(\eta,t-t')\langle m'|\mathcal{L}'|m\rangle f^H_{\{m\}}(\eta,t'). \tag{9.11}$$

Upon identifying $g_{\{m'\}}(\eta,t-t')$ with $\exp[\mathcal{L}^T(t-t')]$, (9.11) agrees with (8.6). We have thus demonstrated similarity between Herring's SCFA and the present perturbation procedure.

10. STATIONARY TURBULENCE DYNAMICS

Let us examine the stationary behavior of (7.17) and (7.18). Suppose that $\varphi(\eta_{\mathbf{k},\mu})$ has the stationary distribution

$$\varphi(\eta_{\mathbf{k},\mu})=q^{-1}_{\mathbf{k},\mu}\exp(-\eta_{\mathbf{k},\mu}/q_{\mathbf{k},\mu}). \tag{10.1}$$

Since $q_{\mathbf{k},\mu}=\int_0^\infty d\eta_{\mathbf{k},\mu}\eta_{\mathbf{k},\mu}\varphi(\eta_{\mathbf{k},\mu})$, it is the stationary value of the modal energy, $I_{\mathbf{k},\mu}\to q_{\mathbf{k},\mu}$ as $t\to\infty$. It is well known that (10.1) is the steady solution of a Fokker–Planck equation[7]

$$\frac{\partial\varphi}{\partial t}=\omega_{\mathbf{k},\mu}\frac{\partial}{\partial\eta_{\mathbf{k},\mu}}\left(q_{\mathbf{k},\mu}\eta_{\mathbf{k},\mu}\frac{\partial}{\partial\eta_{\mathbf{k},\mu}}+\eta_{\mathbf{k},\mu}\right)\varphi, \tag{10.2}$$

where $\omega_{\mathbf{k},\mu}$ is the dynamic friction. The random process cannot be Gaussian in action-angle variables; hence (10.2) is not the more familiar Gaussian Fokker–Planck equation. Since $\mathcal{L}^T$ is decomposable, the zeroth-order problem (8.3) can be written down for each (k, μ) (by dropping μ)

$$\partial\varphi(\eta_{\mathbf{k}},t)/\partial t=\mathcal{L}^T_{\mathbf{k}}\varphi(\eta_{\mathbf{k}},t). \tag{10.3}$$

We shall consider here a particular operator representation

$$\mathcal{L}^T_{\mathbf{k}}=\omega_{\mathbf{k}}\left[\frac{\partial}{\partial\eta_{\mathbf{k}}}\left(q_{\mathbf{k}}\eta_{\mathbf{k}}\frac{\partial}{\partial\eta_{\mathbf{k}}}+\eta_{\mathbf{k}}\right)-\frac{1}{2}\left(1+\frac{q_{\mathbf{k}}}{2\eta_{\mathbf{k}}}\right)\right]. \tag{10.4}$$

Note that the Fokker–Planck operator of (10.2) appears in the first parenthesis. On the other hand, we have introduced the second parenthesis so as to expand $\mathcal{L}^T_{\mathbf{k}}$ in terms of the associated Laguerre polynomials of order 1. Although this choice is not unique, the expansion of $\mathcal{L}^T_{\mathbf{k}}$ by the Laguerre polynomials of other orders will lead to the same stationary dynamics. Formally, the Green's operator of (10.3) is $\exp(\mathcal{L}^T_{\mathbf{k}}t)$. Since the Green's operator enters into the covariance and averaged Green's functions, our immediate goal is to derive a concrete representation for it.

To this end, we consider the eigenvalue problem[18,19]

$$\mathcal{L}^T_{\mathbf{k}}\psi_{r_{\mathbf{k}}}(\eta_{\mathbf{k}})=-\omega_{\mathbf{k}}r_{\mathbf{k}}\psi_{r_{\mathbf{k}}}(\eta_{\mathbf{k}}), \tag{10.5}$$

which for the eigenvalues $r_{\mathbf{k}}=1,2,3,\cdots$ has the eigenfunctions

$$\psi_{r_{\mathbf{k}}}(\eta_{\mathbf{k}})=\frac{(r_{\mathbf{k}}-1)!}{(r_{\mathbf{k}}!)^3}q^{-1}_{\mathbf{k}}\exp\left(-\frac{\eta_{\mathbf{k}}}{q_{\mathbf{k}}}\right)\left(\frac{\eta_{\mathbf{k}}}{q_{\mathbf{k}}}\right)^{1/2}L^1_{r_{\mathbf{k}}}\left(\frac{\eta_{\mathbf{k}}}{q_{\mathbf{k}}}\right), \tag{10.6}$$

where $L^1_{r_{\mathbf{k}}}(x)$ are the associated Laguerre polynomials of order 1. Here the associated Laguerre polynomials of

order s are defined by differentiating the Laguerre polynomials s times, $L_r^s(x) = d^s L_r(x)/dx^s$, and satisfy[20]

$$x\,d^2L_r^s(x)/dx^2 + (s+1-x)\,dL_r^s(x)/dx + (r-s)\,L_r^s(x) = 0.$$

Associated with (10.5) is the adjoint eigenvalue problem

$$\tilde{\mathcal{L}}^{\tau}_{\mathbf{k}}\,\tilde{\psi}_{r_{\mathbf{k}}}(\eta_{\mathbf{k}}) = -\omega_{\mathbf{k}} r_{\mathbf{k}}\,\tilde{\psi}_{r_{\mathbf{k}}}(\eta_{\mathbf{k}}), \tag{10.7}$$

where the adjoint operator is

$$\tilde{\mathcal{L}}^{\tau}_{\mathbf{k}} = \omega_{\mathbf{k}}\left[q_{\mathbf{k}}\eta_{\mathbf{k}}\frac{\partial^2}{\partial\eta_{\mathbf{k}}^2} + (q_{\mathbf{k}} - \eta_{\mathbf{k}})\frac{\partial}{\partial\eta_{\mathbf{k}}} - \frac{1}{2}\left(1 + \frac{q_{\mathbf{k}}}{2\eta_{\mathbf{k}}}\right)\right]. \tag{10.8}$$

The eigenfunctions of (10.7) are

$$\tilde{\psi}_{r_{\mathbf{k}}}(\eta_{\mathbf{k}}) = \left(\frac{\eta_{\mathbf{k}}}{q_{\mathbf{k}}}\right)^{1/2} L^1_{r_{\mathbf{k}}}\left(\frac{\eta_{\mathbf{k}}}{q_{\mathbf{k}}}\right) \tag{10.9}$$

for the eigenvalues $r_{\mathbf{k}} = 1, 2, 3, \cdots$. Note that $\psi_{r_{\mathbf{k}}}$ and $\tilde{\psi}_{r_{\mathbf{k}}}$ form a set of biorthogonal (normalized) eigenfunctions

$$\int_0^\infty d\eta_{\mathbf{k}}\,\psi_{r_{\mathbf{k}}}(\eta_{\mathbf{k}})\,\tilde{\psi}_{r'_{\mathbf{k}}}(\eta_{\mathbf{k}}) = \delta_{r_{\mathbf{k}} r'_{\mathbf{k}}}. \tag{10.10}$$

Now the Green's equation for (10.3) is

$$\partial g(\eta_{\mathbf{k}}, \eta'_{\mathbf{k}}, t, t')/\partial t = \mathcal{L}^{\tau}_{\mathbf{k}} g(\eta_{\mathbf{k}}, \eta'_{\mathbf{k}}, t, t') + \delta(\eta_{\mathbf{k}} - \eta'_{\mathbf{k}})\delta(t - t'). \tag{10.11}$$

In terms of the biorthogonal eigenfunctions, we can write down the solution

$$g(\eta_{\mathbf{k}}, \eta'_{\mathbf{k}}, t - t') = \sum_{r_{\mathbf{k}}=1}^{\infty} \exp[-\omega_{\mathbf{k}} r_{\mathbf{k}}(t - t')]\,\psi_{r_{\mathbf{k}}}(\eta_{\mathbf{k}})\,\tilde{\psi}_{r_{\mathbf{k}}}(\eta'_{\mathbf{k}}). \tag{10.12}$$

Hence, the Green's operator has the form

$$G(\eta_{\mathbf{k}}, t - t') = \sum_{r_{\mathbf{k}}=1}^{\infty} \exp[-\omega_{\mathbf{k}} r_{\mathbf{k}}(t - t')] \times \psi_{r_{\mathbf{k}}}(\eta_{\mathbf{k}}) \int_0^\infty d\eta'_{\mathbf{k}}\,\tilde{\psi}_{r_{\mathbf{k}}}(\eta'_{\mathbf{k}}). \tag{10.13}$$

By using this operator, the covariance under (10.1) becomes

$$U_{\mathbf{k}}(t - \theta, \theta) = q_{\mathbf{k}} \exp[-\omega_{\mathbf{k}}(t - \theta)], \tag{10.14}$$

and the averaged Green's function reduces to

$$G_{\mathbf{k}}(t - \theta, \theta) = \exp[-\omega_{\mathbf{k}}(t - \theta)]. \tag{10.15}$$

As anticipated, (10.14) and (10.15) satisfy the fluctuation–dissipation relation. The phase-correlation function has the form $\exp[-\omega_{\mathbf{k}}(t - \theta)]$, thereby representing the solution of a Fokker–Planck equation.

Let us now introduce (10.14) and (10.15) into (7.17) and (7.18). First, carry out the θ integration in (7.17) and let $t \to \infty$. By noting $\partial I_{\mathbf{k},\mu}/\partial t \to 0$ as $t \to \infty$, the modal energy relation in isotropic form becomes

$$\nu k^2 q(k) = \pi k \int\!\!\int_\Delta dp\,dq\,pq \times \left(\frac{a(k,p,q)q(p)q(k-p) - b(k,p,q)q(k)q(k-p)}{\omega(k) + \omega(p) + \omega(k-p)}\right). \tag{10.16}$$

Following Sec. 7C, we have extended the isotropic requirements to $q_{\mathbf{k},\mu} = q(k)/2$ and $\omega_{\mathbf{k},\mu} = \omega(k)$. Secondly, we perform the θ integration in (7.18), and then integrate the resulting equation in τ over $[0, \infty]$. The dynamic friction in isotropic form then becomes

$$\omega(k) = \nu k^2 + \pi k \int\!\!\int_\Delta dp\,dq\,pq\,\frac{b(k,p,q)q(k-p)}{\omega(p) + \omega(k-p)}. \tag{10.17}$$

The pair (10.16) and (10.17) describes the relationship among the triad modal energies in the stationary limit of (7.17) and (7.18). They agree in form with Herring's stationary SCFA[4] in that (10.16) has the relaxation factor $[\omega(k) + \omega(p) + \omega(k-p)]^{-1}$, whereas (10.17) has a slightly different one, $[\omega(p) + \omega(k-p)]^{-1}$. Although Edwards[8] also derives a pair relation similar to (10.16) and (10.17), his dynamic friction has the relaxation factor $[\omega(k) + \omega(p) + \omega(k-p)]^{-1}$.

11. CONCLUSIONS BY WAY OF A UNIFYING OBSERVATION

The turbulence theories of Balescu–Senatorski[6] and Herring[5] parallel the perturbation schemes of Wyld[3] and Kraichnan[1] developed in the moment formulation. In summary, the theories of Balescu–Senatorski and Wyld involve the laminar perturbation and renormalization. On the other hand, both Herring and Kraichnan develop perturbation about a reference flow state which is very close to the actual turbulent field, thereby avoiding the tedious renormalization. Two other theories of Edwards[8] and Phythian,[16] however, represent perturbation about the Gaussian random process. Edwards' theory belongs to the distribution function formalism, and Phythian's theory is based on the moment formulation. In Table II, the three turbulence theories in each formulation category are classified according to the underlying perturbation schemes.

The reference flow state of Phythian's theory is described by the Langevin equation (see, for instance, Ref. 21). Phythian was able to determine the dynamic friction and diffusion coefficient (the variance of external random forces) in terms of the triad interactions, thereby recovering the stationary DIA equations in frequency domain. In Edwards' theory, the reference flow state is governed by the Fokker–Planck equation (see, for instance, Ref. 22). Edwards derives one condition relating both the dynamic friction and diffusion coefficient with the triad interactions. He then splits this condition into two parts: one for the dynamic friction and the other for the diffusion coefficient. With this much latitude allowed, Edwards' theory cannot give the totally

TABLE II. Classification of turbulent theories.

Formulation	Moment	Distribution function
Perturbation		
About the laminar flow	Wyld (1961)	Balescu–Senatorski (1970)
About a turbulent flow	Kraichnan (1958)	Herring (1965–66)
About the Gaussian random process	Phythian (1969)	Edwards (1964)

correct stationary dynamics in that the relaxation factor of (10.17) becomes erroneously replaced by $[\omega(k)+\omega(p)+\omega(k-p)]^{-1}$. Here the trouble is the one-time nature of the distribution function formalism, which impedes the simultaneous determination of the dynamic friction and diffusion coefficient. Since the random process is not Gaussian in action-angle variables, the present statistical mechanical formulation cannot be used to investigate Edwards' theory. We must therefore go back to the triad-interaction representation (2.6) and start anew the theoretical formulation but without having access to the elegant transition matrix. We shall briefly sketch Edwards' theory in Appendix B, and show the predicament in determining the dynamic friction and diffusion coefficient simultaneously.

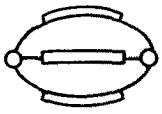
ACKNOWLEDGMENTS

Discussions with Dr. Jackson Herring were very helpful in clarifying certain aspects of his self-consistent-field approximation. Correspondence with Professor R. Balescu has been of great value. In particular, I appreciate very much the letter containing the opinions of his and Dr. A. Senatorski on the original manuscript. Also, correspondence with Professor S. F. Edwards is much appreciated. Finally, Dr. Uriel Frisch has made some suggestions on the presentation of the materials in the original manuscript.

APPENDIX A: THE PRODUCT HYPOTHESIS

Under the cycle approximation, the kinetic equation (6.4) becomes

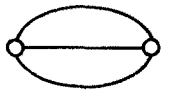
$$\frac{\partial f_{\{0\}}(\eta,t)}{\partial t} = \not\subset^{\circ} f_{\{0\}}(\eta,t) + 2\int_0^t d\theta \times \sum_{\mathbf{k}}^{+}\sum_{\mathbf{p}}\sum_{\mu,\lambda,\rho}$$



$$f_{\{0\}}(\eta,\theta). \tag{A1}$$

Consider the inviscid case. The first term drops out of (A1) and the laminar $\exp(\not\subset^{\circ}_{\mathbf{k},\mu} t)$ denoted by the double line degenerates to the unit operator. Hence, the inviscid kinetic equation becomes

$$\partial f_{\{0\}}(\eta,t)/\partial t = 2\int_0^t d\theta \sum_{\mathbf{k}}^{+}\sum_{\mathbf{p}}\sum_{\mu,\lambda,\rho}$$

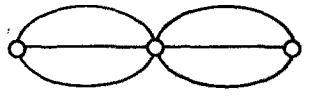


$$f_{\{0\}}(\eta,\theta). \tag{A2}$$

Let us solve this by iteration:

$$f_{\{0\}}(\eta,t) = f_{\{0\}}(\eta,0) + (t^2/2!) \times 2 \sum_{\mathbf{k}}^{+}\sum_{\mathbf{p}}\sum_{\mu,\lambda,\rho}$$

$$f_{\{0\}}(\eta,0) + (t^4/4!) \times 4 \sum_{\mathbf{k}}^{+} \cdots \sum$$



$$\times f_{\{0\}}(\eta,0) + \cdots. \tag{A3}$$

Suppose that the product hypothesis is initially imposed, $f_{\{0\}}(\eta,0) = \Pi_{\mathbf{k},\mu}\,\varphi(\eta_{\mathbf{k},\mu},0)$. Upon applying $\int_{(\mathbf{k},\mu)} d\eta$ to (A3), we find that $\varphi(\eta_{\mathbf{k},\mu},0)$ separates out from the factors $I_{\mathbf{p}}$ and $I_{\mathbf{k}-\mathbf{p}}$. Thus, in the cycle approximation the single mode distribution $\varphi(\eta_{\mathbf{k},\mu},t)$ will not induce statistical interaction with other φ's, if they are statistically independent at the initial time.

APPENDIX B: EDWARDS' GENERALIZED-RANDOM-PHASE APPROXIMATION

We shall briefly retrace the essential steps of Edwards' turbulence theory, thereby indicating the difficulty in simultaneously determining the dynamic friction and diffusion coefficient. Consider an abstraction of the triad-interaction representation

$$dx_k/dt + \nu k^2 x_k = \sum_{p,q} \overline{\phi}_{kpq} x_p x_q + r_k, \tag{B1}$$

where x_k denotes a suitable enumeration of $v^{\mu}(\mathbf{k})$ and $w^{\mu}(\mathbf{k})$, and r_k is the random force for mode k. The symmetrized coupling coefficient $\overline{\phi}_{kpq}$ is symmetric in p and q, and $\overline{\phi}_{kpq}+\overline{\phi}_{pqk}+\overline{\phi}_{qkp}=0$ assures the energy conservation. Now take arbitrary nonrandom function $R_k(x,t)$ and random function $s_k(t)$. After adding $R_k x_k + s_k$ to both sides of (B1), we put the resulting equation in the form[23]

$$dx_k/dt = -\omega_k x_k + h_k + \lambda \sum_{p,q} \phi_{kpq} x_p x_q + \lambda^2 (R_k x_k - s_k), \tag{B2}$$

where

$$\omega_k = \nu k^2 + R_k, \tag{B3}$$

$$h_k = r_k + s_k. \tag{B4}$$

We have introduced the ordering parameter λ into (B2) [see Eq. (8.1)].

The reference flow state ($\lambda=0$) is the Langevin equation

$$dx_k/dt = -\omega_k x_k + h_k. \tag{B5}$$

Since the stochastic solution of (B5) is completely known under the Gaussian h_k, the objective is to determine ω_k and the variance of h_k in such a way that (B5) can best approximate (B1) in a statistical sense. This is Phythian's *self-consistent* perturbation theory based on the moment formulation, which yields exactly the stationary DIA equations in frequency domain.

Under the Gaussian h_k the distribution function $F(x)$ satisfies the Fokker–Planck equation

$$\frac{\partial F}{\partial t} - \sum_k \frac{\partial}{\partial x_k}\left(\omega_k x_k + d_k \frac{\partial}{\partial x_k}\right) F = 0, \tag{B6}$$

where d_k is the diffusion coefficient ($2d_k$ is the variance of h_k). On the other hand, the Liouville equation for the phase space points which evolve according to (B5) is

$$\frac{\partial F}{\partial t} - \sum_k \frac{\partial}{\partial x_k}(\omega_k x_k - h_k) F = 0. \tag{B7}$$

Comparing (B6) and (B7), we can infer the correspondence between the forcing and equivalent diffusion terms:

$$\frac{\partial}{\partial x_k} h_k F \longrightarrow -\frac{\partial}{\partial x_k} d_k \frac{\partial}{\partial x_k} F. \tag{B8}$$

Edwards' theory begins with the Liouville equation for the phase space points evolving according to (B2):

$$\frac{\partial F}{\partial t} - \sum_k \frac{\partial}{\partial x_k}\left\{\left(\omega_k x_k + d_k \frac{\partial}{\partial x_k}\right) - \lambda \sum_{p,q} \overline{\phi}_{kpq} x_p x_q\right.$$

$$-\lambda^2\left(R_k x_k + S_k \frac{\partial}{\partial x_k}\right)\Big\} F = 0. \tag{B9}$$

In the above we have invoked another correspondence in analogy to (B8):

$$\frac{\partial}{\partial x_k} s_k F \longleftrightarrow -\frac{\partial}{\partial x_k} S_k \frac{\partial}{\partial x_k} F, \tag{B10}$$

where $2S_k$ is the variance of s_k. Since r_k and s_k are statistically independent, we find from (B4) that

$$d_k = e_k + S_k, \tag{B11}$$

where $2e_k$ is the variance of r_k.

For the steady-state $\partial F/\partial t = 0$, we look for a series solution[8,24]

$$F = F^0 + \lambda F^1 + \lambda^2 F^2 + \cdots. \tag{B12}$$

The zeroth-order F^0 has the solution

$$F^0(x) = \prod_k f(x_k) = \prod_k (2\pi q_k)^{-1/2} \exp(-x_k^2/2q_k). \tag{B13}$$

Here the modal energy $q_k = \int_{-\infty}^{\infty} dx_k x_k^2 f(x_k)$ is related to the ratio

$$q_k = d_k/\omega_k. \tag{B14}$$

The first-order F^1 becomes

$$F^1(x) = \sum_{k,p,q} \bar{\phi}_{kpq} q_k^{-1} x_k x_p x_q \Omega_{kpq}^{-1} F^0(x), \tag{B15}$$

where $\Omega_{kpq} = \omega_k + \omega_p + \omega_q$. The solution of F^2 is complicated because of the proliferation of product terms when $\partial/\partial x_k \bar{\phi}_{kpq} x_p x_q$ operates twice on F^0. Considering only the terms multiplied by $(1/4\omega_k)H_2(x_k/\sqrt{2q_k})$, where H_2 is the Hermite polynomial of order 2, we have

$$F^2(x) = \sum_k \Bigg(R_k - S_k q_k^{-1} + 2\sum_{p,q} (\bar{\phi}_{kpq})^2 \Omega_{kpq}^{-1} q_k^{-1} q_p q_q + 4\sum_{p,q} \bar{\phi}_{kpq}\bar{\phi}_{pqk}\Omega_{kpq}^{-1} q_q\Bigg)(1/4\omega_k) \times H_2(x_k/\sqrt{2q_k})\, F^0(x) + \cdots, \tag{B16}$$

where the three dots represent terms involving the coefficients different from $(1/4\omega_k)H_2(x_k/\sqrt{2q_k})$.

Since we shall require the zeroth-order problem to share the same modal energy with the actual turbulent flow, it is necessary that

$$\lambda \int dx\, x_k^2 F^1(x) + \lambda^2 \int dx\, x_k^2 F^2(x) = 0. \tag{B17}$$

The first integral is zero due to the antisymmetry. We must therefore demand that

$$R_k - S_k q_k^{-1} + 2\sum_{p,q} (\bar{\phi}_{kpq})^2 \Omega_{kpq}^{-1} q_k^{-1} q_p q_q + 4\sum_{p,q} \bar{\phi}_{kpq}\bar{\phi}_{pqk}\,\Omega_{kpq}^{-1} q_q = 0. \tag{B18}$$

Note that (B18) is not sufficient for (B17) because the additional terms denoted by the three dots in (B16) do not all drop out (in contrast to Edwards' claim). At any rate, let us accept (B18) as the unique condition. This is the first ambiguity of Edwards' theory. By physical arguments, Edwards breaks up (B18) into two relations:

$$R_k = -4\sum_{p,q} \bar{\phi}_{kpq}\bar{\phi}_{pqk}\Omega_{kpq}^{-1} q_q, \tag{B19}$$

$$S_k = 2\sum_{p,q} (\bar{\phi}_{kpq})^2 \Omega_{kpq}^{-1} q_p q_q. \tag{B20}$$

This is the second ambiguity. The dynamic friction (B3) then becomes

$$\omega_k = \nu k^2 - 4\sum_{p,q} \bar{\phi}_{kpq}\bar{\phi}_{pqk} q_q/(\omega_k + \omega_p + \omega_q). \tag{B21}$$

In view of (B11) and (B14), the modal energy relation becomes

$$\nu k^2 q_k = e_k + 2\sum_{p,q} (\bar{\phi}_{kpq})^2 q_p q_q/(\omega_k + \omega_p + \omega_q) + 4\sum_{p,q} \bar{\phi}_{kpq}\bar{\phi}_{pqk} q_k q_q/(\omega_k + \omega_p + \omega_q). \tag{B22}$$

Compare these with (10.16) and (10.17). The modal energy relation (B22) agrees in form with (10.16); however, the dynamic friction (B21) has a relaxation factor $(\omega_k + \omega_p + \omega_q)^{-1}$ apparently different from (10.17). The simultaneous determination of R_k and S_k is not possible because of the one-time nature of the distribution function formalism. We have indicated how the two-time formulation can rectify Edwards' theory, thereby yielding the correct relaxation factors for both the dynamic friction and diffusion coefficient. Since our argument is heuristic, we shall not present here the detail which may be found in the original manuscript.

[1] R. H. Kraichnan, Phys. Rev. **109**, 1407 (1958).
[2] R. H. Kraichnan, J. Fluid Mech. **5**, 497 (1959).
[3] H. W. Wyld, Jr., Ann. Phys. **14**, 143 (1961).
[4] J. R. Herring, Phys. Fluids **8**, 2219 (1965).
[5] J. R. Herring, Phys. Fluids **9**, 2106 (1966).
[6] R. Balescu and A. Senatorski, Ann. Phys. **58**, 587 (1970).
[7] I. Prigogine, *Non-equilibrium Statistical Mechanics* (Interscience, New York, 1962).
[8] S. F. Edwards, J. Fluid Mech. **18**, 239 (1964).
[9] J. Lee, "The Triad-Interaction Representation of Homogeneous Turbulence," (to be published).
[10] J. Lee, "The Triad-Interaction Representation and Direct-Interaction Approximation," (to be published).
[11] G. K. Batchelor, *The Theory of Homogeneous Turbulence* (Cambridge U. P., Cambridge, 1956).
[12] I. Prigogine and P. Resibois, Physica **27**, 629 (1961).
[13] R. W. Zwanzig, *Lectures in Theoretical Physics, Vol. III*, edited by W. Brittin, B. W. Downs, and J. Downs (University of Colorado Press, Boulder, 1961), p. 106.
[14] P. Resibois, *Physics of Many-Particle Systems*, edited by E. Meernon (Gordon and Breach, New York, 1967), Vol. 1, p. 495.
[15] P. Resibois, Phys. Fluids **6**, 817 (1963).
[16] R. Phythian, J. Phys. A **2**, 181 (1969).
[17] R. H. Kraichnan, *Dynamics of Fluids and Plasmas*, edited by S. I. Pai (Academic, New York, 1966), p. 239.
[18] B. Friedman, *Principles and Techniques of Applied Mathematics* (Wiley, New York, 1956).
[19] P. M. Morse and H. Feshbach, *Methods of Theoretical Physics* (McGraw-Hill, New York, 1953), Vol. 1, Chap. 7.
[20] I. N. Sneddon, *Special Functions of Mathematical Physics and Chemistry* (Oliver and Boyd, Edinburgh, 1956).
[21] S. Chandrasekhar, Rev. Mod. Phys. **15**, 1 (1943).
[22] M. C. Wang and G. E. Uhlenbeck, Rev. Mod. Phys. **17**, 323 (1945).
[23] J. Lee, J. Phys. A **4**, 73 (1971).
[24] J. R. Allen, Ph. D. thesis, University of Manchester (1963).

Quasiperiodic pointwise solutions of the periodic, time-dependent Schrödinger equation*

Joyce M. Okuniewicz

Department of Chemistry, University of Minnesota, Minneapolis, Minnesota 55455
(Received 21 December 1973)

For Hamiltonians periodic in time, we obtain under certain assumptions a condition which is necessary and sufficient for the existence of quasiperiodic pointwise solutions to the Schrödinger equation. Orthonormality and completeness of these functions in $L^2(R^n)$ are investigated, and the time-displacement operator is considered as a sum of quasiperiodic terms.

1. INTRODUCTION

Despite its antiquity, the periodic, time-dependent Schrödinger equation

$$i\frac{\partial}{\partial t}\psi = H(t)\psi, \quad H(t+\tau) = H(t) \tag{1.1}$$

remains a largely unsolved problem in quantum mechanics. It is also an important one, for it represents in particular the interaction of a system of quantum mechanical particles with classical monochromatic radiation. One question of interest is whether there are solutions to the equation which are quasiperiodic in form, that is, solutions which can be written as

$$\psi(t) = f(t)\exp(-i\alpha t), \quad f(t+\tau) = f(t), \tag{1.2}$$

where $f(t)$ is periodic in time and α is a real constant. Such functions would correspond to stable physical behavior and provide a counterpart to the stationary states of time-independent theory.

In this work we shall begin a function-analytic investigation of quasiperiodic solutions to the Schrödinger equation by an approach based on a spectral operator acting in $L^2(R^n\times[0,\tau])$. Under certain assumptions we obtain a criterion which is both necessary and sufficient for the existence of quasiperiodic solutions in $L^2(R^n)$. In addition, we investigate orthonormality and completeness of these solutions as well as the expression of the time-displacement operator as a sum of quasiperiodic terms.

Strong motivation for the existence of quasiperiodic solutions is provided by Floquet's theorem[1] which states that, for an n-dimensional homogeneous system of ordinary linear differential equations

$$i\frac{d}{dt}\mathbf{x} = \mathbf{Y}(t)\mathbf{x}, \quad \mathbf{Y}(t+\tau) = \mathbf{Y}(t), \tag{1.3}$$

the solution may be written as

$$\mathbf{x} = \mathbf{P}(t)\exp(i\mathbf{J}t)\mathbf{c}, \quad \mathbf{P}(t+\tau) = \mathbf{P}(t), \tag{1.4}$$

where $\mathbf{P}(t)$ is a periodic matrix, $\mathbf{c}$ is an arbitrary constant vector, and $\mathbf{J}$ is a constant matrix in Jordan form. When $\mathbf{Y}(t)$ is Hermitian, $\mathbf{J}$ is diagonal and real, and the components of (1.4) are given by

$$x_j = \sum_{k=1}^{n} p_{jk}(t)\exp(iJ_k t)c_k. \tag{1.5}$$

We may draw an analogy between the Schrödinger equation and the Floquet system by, for example, expanding the wavefunction in a coordinate basis set and noting that the time-dependent coefficients form a system of type (1.3) with Hermitian $\mathbf{Y}(t)$. The only difference is that the dimension is now infinite, and for this reason we cannot always expect quasiperiodic solutions in the Hilbert space of interest. For example, in the simple case of a free particle Hamiltonian there are no quasiperiodic solutions of (1.1) in $L^2(R^3)$. This points out the need for a condition which is necessary as well as sufficient.

Several other authors[2–6] have noted the relevance of quasiperiodic solutions to the Schrödinger equation and have developed perturbation schemes for obtaining them. Although the convergence of these series and, indeed, the existence of such solutions is in question, Young *et al.*[2] have pointed out that they may have asymptotic validity with respect to a perturbation parameter.

The spectral operator upon which we base our approach is introduced in the following section for Hamiltonians with regular coefficients and for those with many-body Coulombic singularities. The next three sections are largely preparatory to the existence theorems in Sec. 6, although the orthonormality and completeness theorems in Sec. 4 are of additional interest. In Sec. 7 the time-displacement operator is considered first as a sum of quasiperiodic terms and then in a more compact form.

2. THE SCHRÖDINGER OPERATOR

Assume for a moment that there exists a quasiperiodic function

$$\psi(t) = \exp(-i\lambda t)\zeta(t), \quad \zeta(t+\tau) = \zeta(t) \tag{2.1}$$

which satisfies the Schrödinger equation (1.1) in a pointwise sense. Then the periodic function $\zeta(t)$ satisfies the equation

$$A(t)\zeta(t) \equiv \left(H(t) - i\frac{\partial}{\partial t}\right)\zeta(t) = \lambda\zeta(t), \tag{2.2}$$

so $\zeta(t)$ is an eigenfunction of $A(t)$ with real eigenvalue. It is well known that if a linear operator in a separable Hilbert space is self-adjoint, then a complete set of eigenfunctions and eigenpackets of the operator exists in the space.[7] This suggests, then, that we attempt to make $A(t)$ into a self-adjoint operator, the existence of whose eigenfunctions will imply the existence of quasiperiodic solutions.

First we choose a suitable Hilbert space. Quantum mechanically we shall ultimately be interested in solutions in $L^2(R^n)$, so this must be a subspace. Periodicity in the t variable suggests we choose $L^2(R^n\times[0,\tau])$.[8] Norms and scalar products with subscripts x, t, and xt

refer respectively to $L^2(R^n)$, $L^2[0,\tau]$, and $L^2(R^n\times[0,\tau])$. Norms and scalar products without a subscript refer to $L^2(R^n)$.

Now we define the Schrödinger operators more precisely. Reasons for various qualifications are given subsequently. The operator A_1 in $D(A_1)$ below is characterized by regular coefficients.

$$A_1 \equiv \sum_{j,k=1}^{n} \frac{\partial}{\partial x_j} a_{jk}(x,t) \frac{\partial}{\partial x_k} + i\sum_{j=1}^{n}\left(b_j(x,t)\frac{\partial}{\partial x_j} + \frac{\partial}{\partial x_j} b_j(x,t)\right) + q(x,t) - i\frac{\partial}{\partial t}, \tag{2.3}$$

where

(α) a_{jk}, b_j, q real, $a_{jk} = a_{kj}$;

(β) $a_{jk}, b_j \in C^{1;0}(R^n\times[0,\tau])$, $q \in C^0[R^n\times[0,\tau])$;

(γ) a_{jk}, b_j, q satisfy $g(x,\tau) = g(x,0)$.

$C^{l;m}(R^n\times[0,\tau])$ is the set of all functions on $R^n\times[0,\tau]$ with continuous coordinate derivatives of lth order and a continuous time derivative of mth order. When l is equal to m, a single superscript is used. A subscript of zero denotes compact support in $R^n\times[0,\tau]$. If

$$I_1 = \{u \mid u \in C_0^\infty(R^n\times[0,\tau]),\ u(x,\tau) = u(x,0)\} \tag{2.4}$$

and

$$I_2 = \{u \mid u \in C_0^{2;1}(R^n\times[0,\tau]),\ u(x,\tau) = u(x,0)\}, \tag{2.5}$$

then D_α is defined to be any linear subspace of $L^2(R^n\times[0,\tau])$ such that

(α) $I_1 \subseteq D_\alpha \subseteq I_2$, (2.6)

(β) if $u \in D_\alpha$, then, for all integers n, $u\exp[2\pi i n t/\tau] \in D_\alpha$.

Clearly, I_1, I_2 themselves satisfy (2.6β). D_β is defined by the expression

$$D_\beta \equiv \{u \mid u \in C^{2;1}(R^n\times[0,\tau]),\ u(x,\tau) = u(x,0),\ u, A_1 u \in L^2(R^n\times[0,\tau])\}. \tag{2.7}$$

The domain $D(A_1)$ is then taken to be either D_α or D_β.

The operator A_2 in $D(A_2)$ below allows for the presence of many-body Coulombic potentials. We consider m particles and denote the coordinate vector of the kth particle by $\vec{r}_k$. Let $\{a_j^k\}$, where $j = 1,\ldots,m$ and $k = 1,\ldots,l$, be a set of real numbers such that for fixed k the numbers $\{a_j^k\}$ are not all zero. Let $\{\vec{b}^k\}$, where $k = 1,\ldots,l$, be a set of real three-dimensional vectors and $\{c_k\}$, where $k = 1,\ldots,l$, be a set of real numbers. Define

$$\vec{R}_k \equiv \Sigma_j a_j^k \vec{r}_j + \vec{b}^k \tag{2.8}$$

and

$$V_c \equiv \Sigma_k c_k / |\vec{R}_k|. \tag{2.9}$$

The operator A_2 is defined by

$$A_2 \equiv A_1 + V_c, \tag{2.10}$$

where A_1 is given by (2.3) with $n = 3m$ and V_c is given by (2.9). The domain $D(A_2)$ is taken to be D_α, which is given by (2.6) with $n = 3m$. For future reference we define R_s^{3m} to be the set of points in R^{3m} at which none of the vectors $\vec{R}_k$ is zero. The complement of this set is denoted by $(R_s^{3m})'$.

The precise nature of any singular coefficients in the Schrödinger operator will not be relevant to all points of investigation below. In such instances the subscript is omitted, and the operator A refers to A_1 in $D(A_1)$ with the regularity hypotheses (2.3β) weakened in some manner here unspecified.

The operators A_1, A_2 are formally self-adjoint. The domains $D(A_1)$, $D(A_2)$, as they include I_1, are dense in $L^2(R^n\times[0,\tau])$.[9] The boundary condition $u(x,\tau) = u(x,0)$ is a stepping stone to periodicity, and, further, a condition of this sort is needed to insure symmetry. Alternatively we could require

$$u(x,\tau) = \exp(i\alpha)u(x,0), \tag{2.11}$$

where α is any real number. However, any choice of $\alpha \neq 0$ will translate the eigenfunctions and eigenvalues in such a way that the final wavefunctions are independent of α. It is convenient to include I_1 in $D(A)$, as the members of I_1 are often used to establish regularity of eigenfunctions. The requirement (2.6β) is necessary to the development in Sec. 4. The requirement (2.6α) facilitates necessity proofs in Sec. 6.

Note that, by a change of scale

$$t' = t/\tau, \tag{2.12}$$

we may equivalently consider the operator

$$A'(t') \equiv A(t'\tau) = H(t'\tau) - \frac{i}{\tau}\frac{\partial}{\partial t'} \tag{2.13}$$

in the space $L^2(R^n\times[0,1])$. Therefore, the choice of space need not depend on the parameter τ, provided τ is neither zero nor infinity.

3. ESSENTIAL SELF-ADJOINTNESS OF THE SCHRÖDINGER OPERATOR

Although we shall not establish essential self-adjointness of the Schrödinger operator in any generality, we shall demonstrate this property for the special case in which the time-dependent potential associated with the operator A_2 is bounded. If the Coulombic term V_c is given by (2.9), we define

$$A_{20} \equiv \sum_{j,k=1}^{3m} a_{jk}\left(i\frac{\partial}{\partial x_j} - b_j(x)\right)\left(i\frac{\partial}{\partial x_k} - b_k(x)\right) + V_c + V_1(x) + V_2(x,t) - i\frac{\partial}{\partial t} \tag{3.1}$$

under the conditions

(α) the numbers a_{jk} form a real, positive definite, symmetric, constant matrix;

(β) $\{b_k(x)\} \in C^1(R^{3m})$ are real and independent of t;

(γ) $V_1 \in C^0(R^{3m})$ is real and independent of t; there exist M and $0 \leq \beta \leq 2$ such that $|V_1| \leq M|x|^\beta$;

(δ) $V_2 \in C^0(R^{3m}\times[0,\tau])$ is real and bounded; $V_2(\tau) = V_2(0)$,

and take $D(A_{20}) = I_1$.

Theorem 3.1: The Schrödinger operator A_{20} in I_1 is essentially self-adjoint.

Proof: The essential self-adjointness of $A_{20}-V_2+i\,\partial/\partial t$ in $C_0^\infty(R^{3m})$ is a special case of a theorem due to Ikebe and Kato.[10,11] It is well known that the operator $-i\,\partial/\partial t$ is essentially self-adjoint in the subspace

$$D_t=\{w(t)\,|\,w\in C^\infty[0,\tau];\ \ w(\tau)=w(0)\}$$

of $L^2[0,\tau]$. Essential self-adjointness of $A_{20}-V_2$ in

$$D_{xt}=\Big\{u(x,t)\,\Big|\,u=\sum_{j,k=1}^{M,N}c_{jk}v_jw_k;\ v_j\in C_0^\infty(R^{3m});\ w_k\in D_t;\ \text{all finite integers } M,N\Big\}$$

therefore follows from a theorem of Reed and Simon[12] and the natural isomorphism between $L^2(R^{3m}\times[0,\tau])$ and the tensor product space $L^2(R^{3m})\otimes L^2[0,\tau]$.[13]

It is evident that $D_{xt}\subset I_1$, so $A_{20}-V_2$ in I_1 is an extension of $A_{20}-V_2$ in D_{xt}. As the closure of an essentially self-adjoint operator contains every symmetric extension of the operator,[14] $A_{20}-V_2$ in I_1 is essentially self-adjoint if it is symmetric. It follows from the Gauss integral theorem[15] that $A_{20}-V_2$ in I_1 is symmetric. Finally, as an essentially self-adjoint operator retains this property when a real and bounded function is added,[16] the theorem is proved.

4. ORTHONORMALITY AND COMPLETENESS OF EIGENFUNCTIONS IN $L^2(R^n)$

The Schrödinger operator is here assumed to be essentially self-adjoint. Its closure $\bar{A}$ is then self-adjoint, and we shall work within the closure. Recall that $\bar{A}$ has an orthonormal system of eigenfunctions if it has discrete spectrum, and this system is complete in $L^2(R^n\times[0,\tau])$ if and only if its spectrum is purely discrete. We investigate similar eigenfunction properties in $L^2(R^n)$. Further characterization is needed, as stated in the following lemma. Define

$$\omega\equiv 2\pi/\tau. \tag{4.1}$$

Lemma 4.1: If ϕ is an eigenfunction of $\bar{A}$ with eigenvalue λ, then, for any integer n, $\exp(in\omega t)\phi$ is an eigenfunction of $\bar{A}$ with eigenvalue $\lambda+n\omega$. The orthonormal eigenfunctions of $\bar{A}$ may be partitioned into disjoint sets $S_1=\{\exp(in\omega t)\psi_1\}$, $S_2=\{\exp(in\omega t)\psi_2\},\cdots$.

Proof: Let $\{u_k\}$ be a Cauchy sequence in the domain of A which converges to ϕ. Then, for each n, $\{\exp(in\omega t)u_k\}$ is a Cauchy sequence in $D(A)$ which converges to $\exp(in\omega t)\phi$. We have

$$\bar{A}\exp(in\omega t)\phi=\lim_{k\to\infty}A\exp(in\omega t)u_k=\lim_{k\to\infty}\exp(in\omega t)(A+n\omega)u_k$$
$$=(\lambda+n\omega)\exp(in\omega t)\phi, \tag{4.2}$$

which establishes the first assertion.

Choose a normalized eigenfunction ψ_1 of $\bar{A}$. The eigenfunctions $\exp(in\omega t)\psi_1$ all belong to different eigenvalues of $\bar{A}$, so the members of S_1 are orthonormal. Choose another normalized eigenfunction ψ_2 which is orthogonal to all members of S_1. From the relation

$$(\exp(ik\omega t)\psi_1,\exp(ij\omega t)\psi_2)_{xt}=(\exp[i(k-j)\omega t]\psi_1,\psi_2)_{xt}=0 \tag{4.3}$$

we see the members of S_1, S_2 are all mutually orthogonal, and thus to sets S_1, S_2 are mutually disjoint. Continuing in this way we obtain the desired partition. Notice that degeneracy in the point spectrum of $\bar{A}$ causes no difficulty. However, in the degenerate case the eigenfunctions of $\bar{A}$ are not determined uniquely, so the partition is not unique.

Now choose one, and only one, member of each set S_j, multiply by $1/\sqrt{\tau}$, and denote this new collection by

$$J\equiv\{\zeta_j\}. \tag{4.4}$$

Due to different choices of n in the sets S_j, and, if applicable, degeneracy in the point spectrum of $\bar{A}$, the set J is not unique. We consider any particular choice to have been made.

Theorem 4.1: For almost every fixed t on $[0,\tau]$ the members of J form an orthonormal system in $L^2(R^n)$.

Proof: The orthogonality relation between elements of any particular set S_j is

$$(\exp(ik\omega t)\psi_j,\exp(il\omega t)\psi_j)_{xt}=\delta_{kl}, \tag{4.5}$$

which may be written as

$$1/\sqrt{\tau}\int_0^\tau\exp(-in\omega t)\,\|\zeta_j\|_x^2\,dt=\delta_{n0}\sqrt{\tau}. \tag{4.6}$$

$\|\zeta_j\|_x^2$ is a function of t, and the quantities on the left of (4.6) are its Fourier coefficients. Therefore $\|\zeta_j\|_x^2$ may be written as

$$\|\zeta_j\|_x^2=1+y(t) \tag{4.7}$$

almost everywhere in t. The function $y(t)$ is defined by (4.7) and the requirement

$$(u,y)_t=0 \tag{4.8}$$

for every $u\in L^2[0,\tau]$. We show $y(t)$ is zero almost everywhere. As $\zeta_j\in L^2(R^n\times[0,\tau])$, $\|\zeta_j\|_x^2$, and therefore $y(t)$, is measurable. We define a function $\alpha(t)$ by

$$\alpha(t)\equiv\begin{cases}1, & y(t)=0\\ y(t)/|y(t)|, & y(t)\neq 0\end{cases}. \tag{4.9}$$

As $y(t)$ is measurable, $\alpha(t)$ is also measurable.[17] Measurability of $\alpha^2(t)$, together with the fact that $\alpha(t)$ has unit magnitude, implies $\alpha\in L^2[0,\tau]$. Therefore, from (4.8) we find that

$$0=(\alpha,y)_t=\int_0^\tau|y(t)|\,dt, \tag{4.10}$$

so $y(t)$ is zero almost everywhere.

The orthogonality relation between elements of any two different sets S_j, $S_{j'}$ may be written as

$$\int_0^\tau\exp(-in\omega t)(\zeta_j,\zeta_k)_x\,dt=0. \tag{4.11}$$

In similar fashion, $(\zeta_j,\zeta_k)_x$ is shown to be zero almost everywhere in t, and the theorem is proved.

Theorem 4.2: The system $J=\{\zeta_j\}$ is complete in $L^2(R^n)$ for almost all t if and only if the spectrum of $\bar{A}$ is purely discrete.

Proof: Purely discrete spectrum is equivalent to completeness of the eigenfunction system $\{\exp(in\omega t)\zeta_j\}$ in $L^2(R^n\times[0,\tau])$. If $\{\exp(in\omega t)\zeta_j\}$ is complete in $L^2(R^n\times[0,\tau])$, completeness of $\{\zeta_j\}$ in $L^2(R^n)$ for almost all t follows from this and the fact that $L^2(R^n)$ is a subspace of $L^2(R^n\times[0,\tau])$. Next assume $\{\zeta_j\}$ is complete in $L^2(R^n)$ for almost all t. If $\{\chi_j(x)\}$ is a complete orthonormal system in $L^2(R^n)$, then it follows that

$$\chi_j(x) = \Sigma_k(\zeta_k, \chi_j)_x\zeta_k \tag{4.12}$$

almost everywhere in x for almost all t. Also, the functions

$$\exp(in\omega t)\chi_j = \exp(in\omega t)\Sigma_k(\zeta_k, \chi_j)_x\zeta_k \tag{4.13}$$

are complete in $L^2(R^n\times[0,\tau])$. Consider $\psi\in L^2(R^n\times[0,\tau])$ orthogonal to the set $\{\exp(in\omega t)\zeta_k\}$. We have

$$\int dt\exp(in\omega t)(\psi,\zeta_k)_x = 0. \tag{4.14}$$

Completeness of $\{\exp(in\omega t)\}$ in $L^2[0,\tau]$ implies $(\psi,\zeta_k)_x$ is zero almost everywhere in t. Consequently, the sum

$$S^{n,j}(t) = \Sigma_k\exp(in\omega t)(\psi,\zeta_k)_x(\zeta_k,\chi_j)_x \tag{4.15}$$

is zero almost everywhere in t and thus we find that

$$\int dtS^{n,j}(t) = \int dt\exp(in\omega t)\Sigma_k\int\psi^*(\zeta_k,\chi_j)_x\zeta_k\,dx = 0. \tag{4.16}$$

Now, that $\psi\in L^2(R^n\times[0,\tau])$ implies that $\psi\in L^2(R^n)$ for almost all t. Further, by assumption, the sequence of partial sums corresponding to (4.12) converges to χ_j in $L^2(R^n)$ for almost all t. Therefore, by continuity of the scalar product in $L^2(R^n)$, we may interchange the order of summation and integration over x in (4.16) for almost all t. Considering the time integral in (4.16), we may do so for all t. Therefore it follows that

$$(\psi,\exp(in\omega t)\chi_j)_{xt} = 0. \tag{4.17}$$

Completeness of $\{\exp(in\omega t)\chi_j\}$ implies ψ is the zero element, which in turn implies $\{\exp(in\omega t)\zeta_j\}$ is complete in $L^2(R^n\times[0,\tau])$.

It is evident that a further property is needed to establish orthonormality in $L^2(R^n)$ for all t. As we shall later be interested in eigenfunctions, if any, which are pointwise continuous, we see what can be said with this hypothesis. The following lemma shows it is sufficient to guarantee boundedness of the $L^2(R^n)$ norm.

Lemma 4.2: Assume that, for every $t\in[0,\tau]$, the members of $\{\zeta_j\}$ are continuous in t at every $x\in R^n$. Then, for every $t\in[0,\tau]$, ζ_j is an element of $L^2(R^n)$ and the inequality $|(\zeta_j,\zeta_k)_x|\le 1$ is satisfied.

Proof: Let M_j be the set of values of t for which $\|\zeta_j\|_x\neq 1$. As M_j is of measure zero and ζ_j is continuous in t, for each $t\in[0,\tau]$ we can find a sequence $u_n^j = |\zeta_j(x,t_n)|^2$ such that $u_n^j\to|\zeta_j(x,t)|^2$ for every $x\in R^n$, and $t_n\notin M_j$. Every member of this sequence is positive and measurable, so by Fatou's lemma,[18]

$$\int_{R^n}\liminf_{n\to\infty}u_n^j\,dx\le\liminf_{n\to\infty}\int_{R^n}u_n^j\,dx, \tag{4.18}$$

we have

$$\int_{R^n}|\zeta_j(x,t)|^2dx\le 1. \tag{4.19}$$

Therefore ζ_j and $|\zeta_j|$ are elements of $L^2(R^n)$ and we may use the Schwartz inequality in $L^2(R^n)$ to conclude that

$$|(\zeta_j,\zeta_k)_x|\le\int_{R^n}|\zeta_j||\zeta_k|\,dx\le\|\zeta_j\|_x\|\zeta_k\|_x\le 1. \tag{4.20}$$

Therefore the lemma is proved.

It is well known that, alone, pointwise continuity of $|\zeta_j|^2$ is insufficient to guarantee continuity of $\|\zeta_j\|_x^2$ in t. However, we have also the constancy of $\|\zeta_j\|_x^2$ for almost all t and the derivative results of Lemma 4.2. Unfortunately, this additional information remains insufficient to establish orthonormality, as shown by the following counterexample.[19] The functions

$$\psi_n(x,t) = \frac{t}{\pi^{1/4}2^{n/2}(n!)^{1/2}}H_n(xt^2)\exp(-t^4x^2/2), \tag{4.21}$$

where $H_n(xt^2)$ are the Hermite polynomials

$$H_n(z) = (-1)^n\exp(z^2)\frac{d^n}{dz^n}\exp(-z^2), \tag{4.22}$$

are infinitely differentiable. They form an orthonormal system in $L^2(R^1)$ when t is unequal to zero, but their norms are zero at the latter point.

The need for a further property to establish orthonormality in $L^2(R^n)$, which stems directly from use of the space $L^2(R^n\times[0,\tau])$, is a weak point in this formulation. It is unlikely that, without a complete reformulation of the problem, a criterion can be proposed which is better than the old standard of uniform convergence.

Theorem 4.3: Assume that, for every j, ζ_j is an element of $C^0(R^n\times[0,\tau])$ and the integral $\|\zeta_j\|_x^2$ converges uniformly in t on $[0,\tau]$. Then $\{\zeta_j\}$ is an orthonormal system in $L^2(R^n)$ for every t on $[0,\tau]$. Also, the members of $\{\zeta_j\}$ are strongly continuous functions of t with respect to the $L^2(R^n)$ norm. That is, given $\epsilon>0$, there exists δ_j such that

$$|t-t'|<\delta_j\Rightarrow\|\zeta_j(t)-\zeta_j(t')\|_x<\epsilon.$$

Proof: It is well known[20] that these conditions are sufficient to insure continuity of $\|\zeta_j(t)\|_x^2$ on $[0,\tau]$. As $[0,\tau]$ is compact, the resultant uniform continuity, together with the implication of Theorem 4.1 that $\|\zeta_j(t)\|_x$ is unity on a dense subset, implies $\|\zeta_j(t)\|_x$ is unity everywhere on $[0,\tau]$.

Proof is similar for the scalar product, except we must show the integral $(\zeta_j,\zeta_k)_x$ converges uniformly on $[0,\tau]$. Consider the one-dimensional case. By Lemma 4.2, $\zeta_j(t)$ is an element of $L^2(R^1)$ for every t. If N is a positive real number, this implies $\zeta_j(t)\in L^2(N,\infty)$ for every t. Therefore, we may use the Schwartz inequality in $L^2(N,\infty)$ to show that

$$\left|\int_N^\infty\zeta_j^*\zeta_k dx\right|\le\left(\int_N^\infty|\zeta_j|^2dx\right)^{1/2}\left(\int_N^\infty|\zeta_k|^2dx\right)^{1/2}. \tag{4.23}$$

A similar estimate holds for the interval $(-\infty,-N)$. These estimates, together with uniform convergence of $\|\zeta_j\|_x^2$ and $\|\zeta_k\|_x^2$, imply uniform convergence of $(\zeta_j,\zeta_k)_x$. Generalization to higher dimensions is obvious. The statement of strong continuity follows from similar considerations.

Just as pointwise continuity of the members of $\{\zeta_j\}$, combined with Theorem 4.1, is insufficient to guarantee orthonormality, it is evident that the strong continuity of Theorem 4.3, combined with Theorem 4.2, does not imply completeness. We do not expect that there is any additional internal criterion implying the latter property which is not essentially equivalent to the statement of completeness itself.

5. THE BOUNDARY CONDITION

As the boundary, that is the set of points $(x,t)\in[R^n\times 0]$, $(x,t)\in[R^n\times\tau]$, is a set of measure zero, it is clear

that additional properties are required to show eigenfunctions of $\bar{A}$ satisfy the boundary condition

$$u(x, \tau) = u(x, 0). \tag{5.1}$$

As before, we shall be interested in continuous eigenfunctions, and shall assume a degree of continuity consistent with the problem. Cases in which the Hamiltonian is discontinuous must be treated according to the nature and distribution of the discontinuities present.

Theorem 5.1: Assume $\bar{A}_1$ is symmetric and its eigenfunctions are members of $C^{2;1}(R^n \times [0, \tau])$. Then these eigenfunctions satisfy the boundary condition (5.1) at every $x \in R^n$.

Proof: Let $x^\alpha = (x_1^\alpha, \ldots, x_n^\alpha)$ be an arbitrary point in R^n, and $\gamma > 0$ be a fixed positive real number. Define

$$\begin{aligned} V &\equiv [x_1^\alpha - \gamma, x_1^\alpha + \gamma] \times \cdots \times [x_n^\alpha - \gamma, x_n^\alpha + \gamma] \times [0, \tau] \\ &\equiv V_n \times [0, \tau]. \end{aligned} \tag{5.2}$$

By hypothesis, $I_1 \subseteq D(A_1)$. Let u be an arbitrary element of I_1 with compact support in V and ζ_j be an eigenfunction of $\bar{A}_1$. It is easy to see from the Gauss integral theorem[15] that

$$(H_1 u, \zeta_j)_{xt} = (u, H_1 \zeta_j)_{xt}, \tag{5.3}$$

where $H_1 = A_1 + i\,\partial/\partial t$. Symmetry of $\bar{A}_1$ in $D(\bar{A}_1)$ and (5.3) imply that

$$0 = (A_1 u, \zeta_j)_{xt} - (u, A_1 \zeta_j)_{xt} = i \int_V \frac{\partial}{\partial t} (u^* \zeta_j)\, dx\, dt. \tag{5.4}$$

Again using the theorem of Gauss, we find that

$$\int_{V_n} u^*(x, 0)[\zeta_j(x, \tau) - \zeta_j(x, 0)]\, dx = 0. \tag{5.5}$$

Now, the members of I_1 with compact support in V, evaluated at $t = 0$, are dense in $L^2(V_n)$. The function $[\zeta_j(x, \tau) - \zeta_j(x, 0)]$, being continuous on the compact set V_n, is a member of $L^2(V_n)$. As, from (5.5), this function is orthogonal to a dense subset, it must be zero for almost all $x \in V_n$. As $[\zeta_j(x, \tau) - \zeta_j(x, 0)]$ is in fact uniformly continuous on V_n, it must be zero for every $x \in V_n$, and in particular for the point x^α. As x^α was arbitrary, the theorem is proved.

Theorem 5.2: Assume $\bar{A}_2$ is symmetric and its eigenfunctions are members of $C^{2;1}(R_s^{3m} \times [0, \tau])$ and $C^0(R^{3m} \times [0, \tau])$. Then these eigenfunctions satisfy the boundary condition (5.1) at every $x \in R^{3m}$.

Proof: Let x^α be an arbitrary fixed point in R_s^{3m}. Then we can find $\gamma > 0$ and sets V, V_n such that V_n is disjoint from $(R_s^{3m})'$. Therefore the proof of Theorem 5.1 applies, and the boundary condition is satisfied except perhaps on $(R_s^{3m})'$.

Clearly R_s^{3m} is dense in R^{3m}, so for arbitrary fixed $x^\beta \in (R_s^{3m})'$ we can find a sequence $\{x_n\} \subset R_s^{3m}$ such that $x_n \to x^\beta$. As the function $g(x) \equiv [\zeta_j(x, \tau) - \zeta_j(x, 0)]$, where $\zeta_j(x, t)$ is an eigenfunction of $\bar{A}_2$, is continuous on R^{3m}, for arbitrary ϵ we can fine $N(\epsilon)$ such that

$$n > N(\epsilon) \Rightarrow |g(x_n) - g(x^\beta)| < \epsilon. \tag{5.6}$$

As $g(x_n)$ is zero for every n, we conclude that the limit $g(x^\beta)$ is also zero, and the theorem is proved.

6. QUASIPERIODIC POINTWISE SOLUTIONS

In this section we find a criterion which is necessary and sufficient for the existence of quasiperiodic solutions to the Schrödinger equation in $L^2(R^n)$. This criterion is dependent on two requirements. The first of these, essential self-adjointness of the Schrödinger operator, was established in Sec. 3 for a limited class of problems. The second requirement, regularity of the eigenfunctions of that operator, will not be investigated here. However, it is well to point out that the corresponding time-independent problem has received much attention,[21,22] and results of that work lead us to expect that the regularity hypotheses made here can be shown to be satisfied. We also note that a typical regularity proof neither establishes nor requires the existence of eigenfunctions to have been previously established. The requirement of eigenfunction regularity is not in fact needed for the necessity part of our theorem.

By a pointwise solution to the Schrödinger equation $A_1\psi = 0$ is meant a function which is a member of $C^{2;1}(R^n \times R^1)$ and satisfies the equation in the usual sense. For the equation $A_2\psi = 0$ we mean instead that this solution is a member of $C^{2;1}(R_s^{3m} \times R^1)$ and $C^0(R^{3m} \times R^1)$.

One final comment is in order. It would be desirable in the following theorem to replace the statement that the $L^2(R^n)$ norm of the quasiperiodic function is bounded in time with the statement that it is constant in time. From Sec. 4 it is clear that this is impossible without additional information. However, the theorem can be stated in this form if, for example, the hypotheses of Theorem 4.3 are satisfied.

Theorem 6.1: Assume the Schrödinger operator A_1 is essentially self-adjoint and the eigenfunctions of its closure, if any exist, are members of $C^{2;1}(R^n \times [0, \tau])$. Then, in order that the corresponding Schrödinger equation have quasiperiodic pointwise solutions whose $L^2(R^n)$ norm is bounded in time, it is necessary and sufficient that the point spectrum of $\bar{A}_1$ not be empty.

Proof: Assume the point spectrum of $\bar{A}_1$ is not empty, and let ζ_j be an eigenfunction of $\bar{A}_1$ with eigenvalue λ_j. As $\zeta_j \in C^{2;1}(R^n \times [0, \tau])$, we conclude that the function $\psi_j \equiv \zeta_j \exp(-i\lambda_j t)$ is a pointwise solution of the equation $A_1\psi = 0$ when $t \in [0, \tau]$. Define an extension of ψ_j to the interval $[\tau, 2\tau]$ by

$$\psi_j \equiv \exp(-i\lambda_j t)\zeta_j(t - \tau), \quad \tau \leq t \leq 2\tau, \tag{6.1}$$

and let $t' = t - \tau$. Then ψ_j satisfies

$$\begin{aligned} A_1(t)\psi_j(t) &= A_1(t' + \tau) \exp[-i\lambda_j(t' + \tau)]\zeta_j(t') \\ &= \exp(-i\lambda_j \tau) A_1(t')\psi_j(t') \\ &= 0 \end{aligned} \tag{6.2}$$

on that interval. By Theorem 5.1 the function ζ_j satisfies $\zeta_j(\tau) = \zeta_j(0)$ on R^n, so ψ_j is continuous when $t = \tau$. Further, the relation

$$\begin{aligned} \frac{\partial}{\partial t}\zeta_j(t)\bigg|_{t=\tau} &= -i[H_1(\tau) - \lambda_j]\zeta_j(\tau) \\ &= -i[H_1(0) - \lambda_j]\zeta_j(0) \end{aligned} \tag{6.3}$$

$$= \frac{\partial}{\partial t}\zeta_j(t)\Big|_{t=0}$$

shows that ψ_j has a continuous first time derivative when $t=\tau$. Therefore, if we define a function ψ_j on $R^n\times R^1$ by

$$\psi_j(t) = \exp(-i\lambda_j t)\zeta_j(t-n\tau), \quad n\tau \leq t \leq (n+1)\tau, \tag{6.4}$$

it follows that this quasiperiodic function is a member of $C^{2;1}(R^n\times R^1)$ and a pointwise solution of the Schrödinger equation on its domain. By Lemma 4.2 ψ is a member of $L^2(R^n)$ and its norm there, which is constant for almost all t, is bounded in time by that constant. Therefore, sufficiency of the condition is proved.

Now assume ψ is a pointwise quasiperiodic solution of $A_1\psi=0$. Then we may write $\psi=\exp(-i\lambda t)\zeta$, where ζ is periodic in time and a member of $C^{2;1}(R^n\times R^1)$. If ζ_τ denotes the restriction of ζ to $R^n\times[0,\tau]$, then ζ_τ satisfies

$$A_1\zeta_\tau = \lambda\zeta_\tau \tag{6.5}$$

and

$$\zeta_\tau(\tau) = \zeta_\tau(0). \tag{6.6}$$

As $\|\psi\|_x$ is assumed bounded in time it follows that $\zeta_\tau \in L^2(R^n\times[0,\tau])$ and, from (6.5), $A_1\zeta_\tau\in L^2(R^n\times[0,\tau])$.

If $\zeta_\tau\in D(\bar{A}_1)$, it follows from this and (6.5) that ζ_τ is an eigenfunction of $\bar{A}_1$, and hence the point spectrum of $\bar{A}_1$ is not empty. If $D(A_1)=D_\beta$, it follows from above that $\zeta_\tau\in D(A_1)$, and thus $\zeta_\tau\in D(\bar{A}_1)$.

If $D(A_1)=D_\alpha$, let D'_α be the set of all finite linear combinations of members of D_α with ζ_τ. As ζ_τ, $A_1\zeta_\tau\in L^2(R^n\times[0,\tau])$, the operator A_1 in D'_α is an extension of A_1 in D_α. We show it is a symmetric extension. Evidently the equality $(A\zeta_\tau,\zeta_\tau)_{xt}=(\zeta_\tau,A\zeta_\tau)_{xt}$ follows from (6.5). Now consider the quantity

$$Q \equiv (A_1u,\zeta_\tau)_{xt} - (u,A_1\zeta_\tau)_{xt} \tag{6.7}$$

for arbitrary fixed $u\in D_\alpha$. As both u, ζ_τ satisfy the boundary condition (6.6), it is easy to show by the Gauss integral theorem[15] that Q is zero so A_1 in D'_α is symmetric. But, as A_1 in D_α is essentially self-adjoint, $\bar{A}_1$ has no proper symmetric extensions[14] and so $\zeta_\tau\in D(A_1)$. Therefore the theorem is proved.

In the many-body Coulomb problem a smallness condition on the first derivatives is needed to establish necessity, so we split the theorem into two parts. We say a function $g\in C^{1;0}(R^{3m}_s\times[0,\tau])$ satisfies condition (σ) if, for each compact subset Z of $R^{3m}\times[0,\tau]$, there exist numbers $F_k\geq 0$, where $k=1,\ldots,l+1$, and $0\leq\beta<2$ such that

$$\left|\frac{\partial g}{\partial x_j}\right| \leq \sum_{k=1}^{l}\frac{F_k}{|\vec{R}_k|^\beta} + F_{l+1} \tag{6.8}$$

for all $x,t\in Z\cap(R^{3m}_s\times[0,\tau])$. This is a reasonable condition in light of the form of the Coulombic singularity. Alternative smallness conditions would also allow our proof to succeed.

Theorem 6.2*a*: Assume the Schrödinger operator A_2 is essentially self-adjoint, the point spectrum of $\bar{A}_2$ is not empty, and the eigenfunctions of $\bar{A}_2$ are members of $C^{2;1}(R^{3m}_s\times[0,\tau])$ and $C^0(R^{3m}\times[0,\tau])$. Then the corresponding Schrödinger equation has quasiperiodic pointwise solutions whose $L^2(R^{3m})$ norm is bounded in time.

Proof: This is analogous to the sufficiency proof of Theorem 6.1, except Theorem 5.2 is used in place of Theorem 5.1.

Theorem 6.2*b*: Assume the Schrödinger operator A_2 is essentially self-adjoint and the corresponding Schrödinger equation has a quasiperiodic pointwise solution which satisfies condition (σ) and whose $L^2(R^{3m})$ norm is bounded in time. Then the point spectrum of $\bar{A}_2$ is not empty.

Proof: The proof is analogous to the demonstration of necessity in Theorem 6.1 except we must show, for arbitrary $u\in D_\alpha$, that the quantity

$$Q \equiv (A_2u,\zeta_\tau)_{xt} - (u,A_2\zeta_\tau)_{xt} \tag{6.9}$$

is zero when $\zeta_\tau\in C^{2;1}(R^{3m}_s\times[0,\tau])$, $C^0(R^{3m}_s\times[0,\tau])$ and satisfies (6.5), (6.6), and condition (σ). Let C^{3m}_M be a cube in R^{3m} with center at the origin and side of length M sufficiently large that the support of u is contained in the interior of the cube. Let $V_k(\epsilon)$ be the set of points in C^{3m}_M for which $|\vec{R}_k|<\epsilon$. Then Q may be written as

$$Q = \sum_{j=1}^{5} Q_j, \tag{6.10}$$

where

$$Q_1 = \int_{\cup_k V_k\times[0,\tau]} u^* V_c\zeta_\tau\, d\vec{r}_1\cdots d\vec{r}_m\, dt, \tag{6.11}$$

$$Q_2 = \int_{\cup_k V_k\times[0,\tau]} [(A_1u)^*\zeta_\tau - u^*A_2\zeta_\tau]\, d\vec{r}_1\cdots d\vec{r}_m\, dt, \tag{6.12}$$

$$Q_3 = i\int_{[C^{3m}_M-\cup_k V_k]\times[0,\tau]} \frac{\partial}{\partial t}(u^*\zeta_\tau)\, d\vec{r}_1\cdots d\vec{r}_m\, dt, \tag{6.13}$$

$$Q_4 = \int_{[C^{3m}_M-\cup_k V_k]\times[0,\tau]} \sum_{j=1}^{3m}\frac{\partial}{\partial x_j}\left[\sum_{k=1}^{3m} a_{jk}\frac{\partial u^*}{\partial x_k}\zeta_\tau - 2ib_j u^*\zeta_\tau\right] d\vec{r}_1\cdots d\vec{r}_m\, dt, \tag{6.14}$$

$$Q_5 = -\int_{[C^{3m}_M-\cup_k V_k]\times[0,\tau]} \sum_{j=1}^{3m}\frac{\partial}{\partial x_j} \times\left[\sum_{k=1}^{3m} a_{jk}u^*\frac{\partial\zeta_\tau}{\partial x_k}\right] d\vec{r}_1\cdots d\vec{r}_m\, dt. \tag{6.15}$$

That Q_2 is $O(\epsilon^3)$ follows from (6.5). That Q_3 is zero follows from (6.6) and the theorem of Gauss.

The term Q_1 is easily seen to satisfy the inequality

$$|Q_1| \leq B\sum_{i,k=1}^{l} Q_{ki}, \tag{6.16}$$

where B is a constant and

$$Q_{ki} = \int_{V_k\cap C^{3m}_M}\frac{1}{|\vec{R}_i|}\, d\vec{r}_1\cdots d\vec{r}_m. \tag{6.17}$$

By a coordinate transformation the term Q_{ki} can be shown to be $O(\epsilon^2)$. By means of the divergence theorem,[15] a coordinate transformation, and condition (σ) in the case of Q_5, the terms Q_4 and Q_5 can be shown to be $O(\epsilon^2)$ and $O(\epsilon^{2-\beta})$, respectively. Therefore Q is $O(\epsilon^{2-\beta})$. As Q is independent of ϵ, which can be made arbitrarily small, the theorem is proved.

Under the hypotheses of Theorems 6.1 and 6.2a the members of the set $J=\{\zeta_j\}$, multiplied respectively by $\exp(-i\lambda_j t)$ and extended periodically in time, are quasiperiodic pointwise solutions of the Schrödinger equation. From Theorem 4.1 these solutions form an orthonormal system in $L^2(R^n)$ for almost all t, and this is true for all t if the hypotheses of Theorem 4.3 are additionally satisfied. From Theorem 4.2 these quasiperiodic solutions are complete in $L^2(R^n)$ for almost all t if the spectrum of the Schrödinger operator is purely discrete.

7. THE TIME-DISPLACEMENT OPERATOR

A linear operator on $L^2(R^n)$ is a time-displacement operator for the Schrödinger equation (1.1) if it is unitary, strongly continuous in t, satisfies the composition law

$$T(t,t')T(t',t_0)=T(t,t_0) \tag{7.1}$$

with

$$T(t_0,t_0)=I, \tag{7.2}$$

and satisfies the Schrödinger equation in some sense when applied to a particular dense subset of $L^2(R^n)$. In existence proofs for time-displacement operators associated with evolution equations it is usually most suitable to require a strong solution property. Our development is oriented instead toward a pointwise solution property. In general, neither type of solution implies the other.

In this section we investigate expression of the time-displacement operator as a sum of quasiperiodic terms. Clearly, this is only possible when the spectrum of the Schrödinger operator is purely discrete. However, in the case of mixed spectrum we obtain an isometric operator which is a partial solution. It is to be noted that our results are somewhat weakened by the fact that Theorems 4.1 and especially 4.2 are true only almost everywhere in t. The final theorem in this section provides further insight into the form of the time-displacement operator.

We make the following assumptions, the first three of which are permanent (the subscript l may be taken to be either 1 or 2):

(α) A_l is essentially self-adjoint.

(β) The point spectrum of $\bar{A}_l$ is not empty.

(γ) The eigenfunctions of $\bar{A}_l$ are members of $C^{2;1}(R^n\times[0,\tau])$ when $l=1$ and are members of $C^{2;1}(R_s^{3m}\times[0,\tau])$ and $C^0(R^{3m}\times[0,\tau])$ when $l=2$. (7.3)

(δ) The integrals $\|\zeta_j\|^2$ converge uniformly in t.

The set $J=\{\zeta_j\}$ is given by (4.4) and extended periodically in t as specified in Sec. 6. The set of corresponding eigenvalues is denoted by $\{\lambda_j\}$. First we see what can be said without the hypothesis (7.3δ).

Theorem 7.1: Assume (7.3α–γ), and let $\{a_j\}$ be a set of complex numbers, independent of t, such that $\sum_j |a_j|^2$ is convergent. Then the function

$$\psi(t)\equiv\sum_{j=1}^{\infty} a_j\zeta_j(t)\exp(-i\lambda_j t) \tag{7.4}$$

is an element of $L^2(R^n)$ for every $t\in R^1$ and a pointwise solution of the Schrödinger equation when the sum is finite. The series defined by $\psi(t)$ converges in $L^2(R^n)$ uniformly in t and satisfies the inequality

$$\|\psi(t)\|^2\le\sum_j |a_j|^2. \tag{7.5}$$

Proof: Consider the sequence of partial sums

$$\psi_N=\sum_{j=1}^{N} a_j\zeta_j(t)\exp(-i\lambda_j t). \tag{7.6}$$

Taking $N>M$, we have

$$\|\psi_N-\psi_M\|^2=\int_{R^n}\left|\sum_{j=M+1}^{N} a_j\zeta_j(t)\exp(-i\lambda_j t)\right|^2 dx. \tag{7.7}$$

By Theorem 4.1 the members of J form an orthonormal system in $L^2(R^n)$ for almost all t. Let M be the set of values of t for which this is true. Then, as the function $\zeta_j(t)$ are continuous in t at every $x\in R^n$, for every t we can find a sequence

$$u_n\equiv\left|\sum_{j=M+1}^{N} a_j\zeta_j(t_n)\exp(-i\lambda_j t_n)\right|^2 \tag{7.8}$$

such that

$$u_n\to\left|\sum_{j=M+1}^{N} a_j\zeta_j(t)\exp(-i\lambda_j t)\right|^2 \tag{7.9}$$

pointwise, and $t_n\in M$. By the Fatou lemma[18] it follows that

$$\|\psi_N-\psi_M\|^2\le\liminf\int_{R^n}u_n\,dx. \tag{7.10}$$

Applying Theorem 4.1 to the right side of (7.10) we find that

$$\|\psi_N-\psi_M\|^2\le\sum_{j=M+1}^{N}|a_j|^2. \tag{7.11}$$

As $\sum_{j=1}^{\infty}|a_j|^2$ is convergent and $\{a_j\}$ is independent of t, we conclude from (7.11) that the sequence of partial sums converges in $L^2(R^n)$ uniformly in t. That is, for arbitrary $\epsilon>0$ there exists $N(\epsilon)$ independent of t such that

$$M>N(\epsilon)\Rightarrow\|\psi_M-\psi\|<\epsilon. \tag{7.12}$$

To obtain the estimate (7.5) we apply as above Theorem 4.1 and Fatou's lemma to the quantity $\|\psi_N\|^2$ to obtain the inequality

$$\|\psi_N\|_x^2\le\sum_{j=1}^{N}|a_j|^2, \tag{7.13}$$

and then take the limit on N. As $\{\psi_N\}$ converges to ψ in $L^2(R^n)$, we may interchange this limit process with the integration on the left of (7.13), and the desired result follows. That each partial sum (7.6) is a pointwise solution of the Schrödinger equation follows directly from Theorems 6.1 and 6.2a, and the theorem is proved.

Define a quantity for t, $t_0\in R^1$ on $L^2(R^n)$ by

$$T(t,t_0)\psi\equiv\sum_{j=1}^{\infty}\zeta_j(t)\exp[-i\lambda_j(t-t_0)](\zeta_j(t_0),\psi). \tag{7.14}$$

From Theorems 7.1 and 4.1 it is clear that $T(t,t_0)$ is a map of $L^2(R^n)$ into itself if and only if $\sum_{j=1}^{\infty}|(\zeta_j(t_0),\psi|^2$ converges for every ψ. The hypotheses (7.3α–γ) appear to be insufficient to guarantee this convergence for every t_0. Therefore we cannot claim that (7.14) is a well-defined operator on $L^2(R^n)$ for every t_0. Recall

from Theorem 4.3 that, under the additional hypothesis (7.3δ), $\{\zeta_j(t_0)\}$ is an orthonormal system in $L^2(R^n)$ for every t_0, which implies (7.14) is well defined on $L^2(R^n)$. In this circumstance we shall also consider the operator (7.14) on the subspaces

$$D_0(t_0) \equiv \left\{ u \,\middle|\, u = \sum_{j=1}^{N} b_j \zeta_j(t_0),\ N \text{ finite} \right\} \tag{7.15}$$

and

$$D(t_0) \equiv \bar{D}_0(t_0) \tag{7.16}$$

of $L^2(R^n)$. Here $\{b_j\}$ is a set of arbitrary complex constants. The closure $\bar{D}_0(t_0)$ of $D_0(t_0)$ is the span of the orthonormal system $\{\zeta_j(t_0)\}$.

Theorem 7.2a: Assume (7.3α—δ). Then the linear operator (7.14) on $L^2(R^n)$ into itself is strongly continuous in t and has the properties (7.1) and

$$(T(t, t_0)u, T(t, t_0)v) = (T(t_0, t_0)u, T(t_0, t_0)v) \tag{7.17}$$

for every t, t', $t_0 \in R^1$. $T(t, t_0)u$ is a pointwise solution of the associated Schrödinger equation for every $u \in D_0(t_0)$.

Proof: The properties (7.1), (7.17) are easy consequences of Theorems 4.3 and 7.1, and continuity of the scalar product in $L^2(R^n)$. The pointwise solution property was shown in Theorem 7.1. Now consider the strong continuity, and let ψ be an arbitrary element of $L^2(R^n)$. We have

$$\begin{aligned} &\| T(t, t_0)\psi - T(t', t_0)\psi \| \\ &\leq \left\| \sum_{j=1}^{N} (\zeta_j(t) \exp(-i\lambda_j t) - \zeta_j(t') \exp(-i\lambda_j t')) \exp(i\lambda_j t_0) \right. \\ &\quad \left. \times (\zeta_j(t_0), \psi) \right\| + \left\| \sum_{j=N+1}^{\infty} \zeta_j(t) \exp[-i\lambda_j(t - t_0)](\zeta_j(t_0), \psi) \right\| \\ &\quad + \left\| \sum_{j=N+1}^{\infty} \zeta_j(t') \exp[-i\lambda_j(t' - t_0)](\zeta_j(t_0), \psi) \right\|. \end{aligned} \tag{7.18}$$

By Theorem 7.1, $T(t, t_0)\psi$ converges in $L^2(R^n)$ uniformly in t, so, given $\epsilon > 0$, we can find $N(\epsilon)$ independent of t, t' such that the second and third terms on the right of (7.18) are each less than $\epsilon/3$ when $N > N(\epsilon)$. As $N(\epsilon)$ is finite and independent of t, t', we can, using Theorem 4.3, find $\delta(\epsilon) > 0$ such that $|t - t'| < \delta(\epsilon)$ implies the first term on the right of (7.18) is also less than $\epsilon/3$. Therefore, for arbitrary $\epsilon > 0$ we can find $\delta(\epsilon) > 0$ such that

$$|t - t'| < \delta(\epsilon) \Rightarrow \| T(t, t_0)\psi - T(t', t_0)\psi \| < \epsilon, \tag{7.19}$$

and the theorem is proved.

The following form of Theorem 7.2a applies when $T(t, t_0)$ is defined on $D(t_0)$.

Theorem 7.2b: Assume (7.3α—δ). Then, in addition to the properties in Theorem 7.2a, the operator (7.14) on $D(t_0)$ onto $D(t)$ is isometric and satisfies (7.2).

Proof: First we show the range of $T(t, t_0)$ on $D(t_0)$ is indeed $D(t)$. To see this, let $\chi = \sum_j (\zeta_j(t), \chi)\zeta_j(t)$ be an arbitrary element of $D(t)$. Then it is clear that the function

$$\phi = \sum_j (\zeta_j(t), \chi) \exp[i\lambda(t - t_0)]\zeta_j(t_0) \tag{7.20}$$

is an element of $D(t_0)$ and satisfies $T(t, t_0)\phi = \chi$. The relation (7.2) follows immediately from the definition of $D(t_0)$. For arbitrary $u, v \in D(t_0)$ (7.2) implies

$$(T(t_0, t_0)u, T(t_0, t_0)v) = (u, v). \tag{7.21}$$

Relations (7.21), (7.16) imply the isometry, and the theorem is proved.

The following theorem shows that, in the case of purely discrete spectrum, the operator (7.14) has the requisite properties of the time-displacement operator almost everywhere in t, t_0.

Theorem 7.3: Assume (7.3α—δ). Further, assume the spectrum of $\bar{A}_t$ is purely discrete. Then the operator (7.14) is strongly continuous in t and satisfies (7.1). Also, for almost all $t, t_0 \in R^1$, the operator (7.14) is a unitary map of $L^2(R^n)$ onto itself, has the property (7.2), and satisfies the Schrödinger equation pointwise when applied to the dense subset $D_0(t_0)$ of $L^2(R^n)$.

Proof: From Theorem 4.2 the spaces $D(t)$, $D(t_0)$ are equal to $L^2(R^n)$ for almost all t, t_0. Therefore, as an isometric map on a Hilbert space whose range is the same Hilbert space is unitary, and as $T(t, t_0)$ is isometric from Theorem 7.2b, we conclude that $T(t, t_0)$ on $L^2(R^n)$ is unitary for almost all t, t_0. As $D(t_0)$ is equal to $L^2(R^n)$ for almost all t_0, $D_0(t_0)$ is dense in $L^2(R^n)$ for Theorem 4.2 and the definition of $D(t_0)$. The other properties have been established in Theorem 7.2a.

Theorem 7.4: Assume (7.3α—δ). Then the operator (7.14) on $D(t_0)$ onto $D(t)$ may be written in the form

$$T(t, t_0) = P(t, t_0) \exp[-iB(t_0)(t - t_0)]. \tag{7.22}$$

The operator $B(t_0)$ in $D(t_0)$ is self-adjoint for every t_0, periodic in that parameter, and independent of t. Its spectrum is independent of t_0 and is related to the point spectrum of A_t by $\sigma(A_t) = \{\lambda_j + 2\pi n/\tau\}$, where n runs through all integers and $\{\lambda_j\}$ is the set of eigenvalues of $B(t_0)$. The operator $P(t, t_0)$ is isometric, strongly continuous in t, periodic in t, t_0, and satisfies

$$P(t, t')P(t', t_0) = P(t, t_0) \tag{7.23}$$

with

$$P(t_0, t_0) = I. \tag{7.24}$$

Proof: We define an operator $B(t_0)$ in the Hilbert space $D(t_0)$ by

$$B(t_0)u = \sum_{j=1}^{\infty} \lambda_j \zeta_j(t_0)(\zeta_j(t_0), u) \tag{7.25}$$

with domain

$$D(B(t_0)) = \{u \,|\, u, B(t_0)u \in D(t_0)\}.$$

As the members of $\{\lambda_j\}$ are real and, from Theorem 4.3, the system $\{\zeta_j(t_0)\}$ is orthonormal, it is clear that $B(t_0)$ in $D(t_0)$ is self-adjoint. By definition $B(t_0)$ is periodic in t_0. Relation between the spectra of A_t and $B(t_0)$ is clear from definitions and Lemma 4.1. By the operator $\exp[-iB(t_0)(t - t_0)]$ we mean

$$\exp[-iB(t_0)(t - t_0)]u = \sum_j \exp[-i\lambda_j(t - t_0)]\zeta_j(t_0)(\zeta_j(t_0), u) \tag{7.26}$$

or, equivalently,

$$\exp[-iB(t_0)(t - t_0)] = \int \exp[-i\lambda(t - t_0)]\, dE_\lambda(t_0), \tag{7.27}$$

where $E_\lambda(t_0)$ is the spectral family of $B(t_0)$. Clearly, the operator (7.26) maps $D(t_0)$ onto itself. We define $P(t, t_0)$ on $D(t_0)$ by

$$P(t, t_0)u = \sum_{j=1}^{\infty} \zeta_j(t)(\zeta_j(t_0), u). \tag{7.28}$$

As $\{\zeta_j(t)\}$, $\{\zeta_j(t_0)\}$ are orthonormal, when $u, v \in D(t_0)$ the sequences of partial sums

$$P_n(t, t_0)v = \sum_{j=1}^{n} \zeta_j(t)(\zeta_j(t_0), v), \tag{7.29}$$

$$(\exp[-iB(t_0)(t-t_0)])_m u = \sum_{k=1}^{m} \exp[-i\lambda_k(t-t_0)]\zeta_k(t_0)(\zeta_k(t_0), u) \tag{7.30}$$

converge in $D(t_0)$ respectively to $P(t, t_0)v$, $\exp[-iB(t_0) \times (t-t_0)]u$. Using these facts, along with continuity of the scalar product in $D(t_0)$, we see that (7.22) is indeed satisfied. Relation (7.24), as well as periodicity, immediately follows from the definition of $P(t, t_0)$. The proofs in Theorems 7.2a, b of strong continuity of $T(t, t_0)$, the isometry of $T(t, t_0)$, and relation (7.1) are readily seen to be applicable to $P(t, t_0)$, and the theorem is proved.

The corresponding statement of this theorem when the spectrum of A_I is purely discrete is evident from the development above and Theorem 7.3. In particular, $B(t_0)$ is self-adjoint in $L^2(R^n)$ for almost all t_0, and $P(t, t_0)$ is a unitary map of $L^2(R^n)$ onto itself for almost all t, t_0.

8. DISCUSSION

In this work we have found a criterion which is necessary and sufficient for the existence of quasiperiodic pointwise solutions in $L^2(R^n)$ to the periodic, time-dependent Schrödinger equation. This criterion rests on two hypotheses: essential self-adjointness of the Schrödinger operator defined in Sec. 2, and sufficient regularity of its eigenfunctions. The first hypothesis was shown to be satisfied for a limited class of problems. Although we consider it reasonable to expect that both hypotheses will be satisfied to a considerable degree of generality, this remains to be shown.

A strong point of our criterion is that it is in a form well suited to application. That is, the question of the existence of quasiperiodic solutions to a particular Schrödinger equation is reduced to the qualitative spectral analysis of an operator which can be written explicitly. For purposes of qualitative spectral analysis it is desirable to remove the period parameter τ from the underlying Hilbert space. This can be done as shown at the end of Sec. 2.

Although well suited to the investigation of quasiperiodic solutions, our approach is less successful with respect to the time-displacement operator. However, under an additional hypothesis of uniform convergence, we have shown that, when the spectrum of the Schrödinger operator is purely discrete, there exists an operator expressible as a sum of quasiperiodic terms which has all the requisite properties of the time-displacement operator almost everywhere in t. The form of the operator $T(t, t_0)$ expressed in Theorem 7.4 is highly suggestive. It leads us to suspect that, regardless of the nature of the spectrum of the Schrödinger operator, the time-displacement operator may be written as the product of a periodic unitary operator and the exponential of a self-adjoint operator dependent only on the initial time. We shall return to this question in a forthcoming publication.

ACKNOWLEDGMENTS

I wish to express my appreciation to Dr. Stephen Prager for a great deal of helpful advice during this work and to Dr. H. Ted Davis for a timely introduction to functional analysis. In addition to these, I thank Dr. Peter Rejto, Dr. Charles McCarthy, Mr. Murray Brockman, and Mr. Robert Numrich for much lively discussion and constructive criticism.

†This work was supported in part by National Science Foundation Grant GP 7552.

[1]E. A. Coddington and N. Levinson, *Theory of Ordinary Differential Equations* (McGraw-Hill, New York, 1955), pp. 78—81.
[2]R. H. Young, W. J. Deal, Jr., and N. R. Kestner, J. Mol. Phys. 17, 369 (1969).
[3]N. D. Sen Gupta, J. Phys. A 3, 618 (1970).
[4]N. D. Sen Gupta, J. Phys. A 5, 401 (1972).
[5]P. Langhoff, S. T. Epstein, and M. Karplus, Rev. Mod. Phys. 44, 602 (1972).
[6]H. Sambé, Phys. Rev. A 7, 2203 (1973).
[7]G. Hellwig, *Differential Operators of Mathematical Physics* (Addison-Wesley, Reading, Mass., 1967), p. 170.
[8]In a concurrent investigation Sambé, Ref. 6, notes the relevance of $L^2(R^n \times [0, \tau])$ and the Hermitian nature of $A(t)$, but he does not proceed with a rigorous treatment.
[9]Ref. 7, p. 24.
[10]T. Ikebe and T. Kato, Arch. Ratl. Mech. Anal. 9, 72 (1962).
[11]T. Kato, Supp. Prog. Theor. Phys. 40, 3 (1967).
[12]M. Reed and B. Simon, *Methods of Modern Mathematical Physics* (Academic Press, New York, 1972), Vol. I, p. 301.
[13]Ref. 12, pp. 51, 52.
[14]Ref. 7, p. 172.
[15]L. H. Loomis and S. Sternberg, *Advanced Calculus* (Addison-Wesley, Reading, Mass., 1968), pp. 424, 425.
[16]Ref. 7, p. 182.
[17]W. Rudin, *Real and Complex Analysis* (McGraw-Hill, New York, 1966), p. 12.
[18]Ref. 17, p. 22.
[19]Charles A. McCarthy, private communication.
[20]A. E. Taylor, *Advanced Calculus* (Blaisdell, Waltham, Mass., 1955), p. 667. Taylor does this in one dimension, but generalization to higher dimensions is obvious.
[21]K. O. Friedrichs, Commun. Pure Appl. Math. VI, 299 (1953).
[22]T. Kato, Commun. Pure Appl. Math. X, 151 (1957).

Elimination of a coordinate singularity in the three-body problem

Robert Nyden Hill

Department of Physics, University of Delaware, Newark, Delaware 19711
(Received 5 April 1974)

One of the coordinate systems commonly used in the three-body problem consists of three center-of-mass coordinates, three interparticle separations, and three Euler angles specifying the orientation of the triangle whose vertices are the three particles. The usual specification of the Euler angles for this system, which aligns the axes of the body-fixed coordinate system with the principal axes of the moment of inertia tensor, results in a coordinate singularity whenever two of the moments of inertia are equal. An alternative specification of the Euler angles for the equal mass case which treats the three particles symmetrically and eliminates the coordinate singularity at the equilateral triangle configuration is presented.

I. INTRODUCTION AND STATEMENT OF THE PROBLEM

One of the coordinate systems commonly used to describe the location of three (identical) particles lying at the points $\mathbf{r}_1$, $\mathbf{r}_2$, $\mathbf{r}_3$ consists of the position

$$\mathbf{R} \equiv \tfrac{1}{3}(\mathbf{r}_1 + \mathbf{r}_2 + \mathbf{r}_3) \tag{1}$$

of the center-of-mass, the three interparticle separations

$$\begin{aligned} s_1 &= |\mathbf{r}_3 - \mathbf{r}_2|, \\ s_2 &= |\mathbf{r}_1 - \mathbf{r}_3|, \\ s_3 &= |\mathbf{r}_2 - \mathbf{r}_1|, \end{aligned} \tag{2}$$

and a set of Euler angles α, β, γ describing the orientation of the triangle whose verticles are at $\mathbf{r}_1, \mathbf{r}_2, \mathbf{r}_3$. Edmonds' conventions[1] will be used for the Euler angles: a rotation through α about the z axis of the space fixed system is followed by a rotation through β about the new y axis and in turn by a rotation through γ about the new z axis (in the body system). If x_i, y_i, z_i are Cartesian coordinates of $\mathbf{r}_i$ in a right handed space fixed system, and ξ_i, η_i, ζ_i are the Cartesian coordinates of $\mathbf{r}_i$ in the body system, then

$$\begin{aligned} x_i &= (\cos\alpha \cos\beta \cos\gamma - \sin\alpha \sin\gamma)\,\xi_i \\ &\quad - (\cos\alpha \cos\beta \sin\gamma + \sin\alpha \cos\gamma)\,\eta_i + (\cos\alpha \sin\beta)\,\zeta_i, \\ y_i &= (\sin\alpha \cos\beta \cos\gamma + \cos\alpha \sin\gamma)\,\xi_i \\ &\quad - (\sin\alpha \cos\beta \sin\gamma - \cos\alpha \cos\gamma)\,\eta_i + (\sin\alpha \sin\beta)\,\zeta_i \\ z_i &= -(\sin\beta \cos\gamma)\,\xi_i + (\sin\beta \sin\gamma)\,\eta_i + (\cos\beta)\,\zeta_i. \end{aligned} \tag{3}$$

Nine conditions are needed to specify the coordinates ξ_i, η_i, ζ_i $(i = 1, 2, 3)$ in the body system. Three are provided by the (rotationally invariant) Eq. (2), three more by the conditions

$$\zeta_1 = \zeta_2 = \zeta_3 = 0, \tag{4}$$

which state that the $\mathbf{r}_1, \mathbf{r}_2, \mathbf{r}_3$ triangle lies in the $\xi - \eta$ plane, and two more by the conditions

$$\xi_1 + \xi_2 + \xi_3 = \eta_1 + \eta_2 + \eta_3 = 0 \tag{5}$$

which state that the center of mass lies at the origin of the $\xi - \eta - \zeta$ system. The positive direction on the ζ axis is specified by requiring that a circuit from $\mathbf{r}_1$ to $\mathbf{r}_2$ to $\mathbf{r}_3$ back to $\mathbf{r}_1$ encircle the origin of the $\xi - \eta$ plane in a counter clockwise direction.

This statement plus the conditions (2), (4), (5) fix the orientation of the $\mathbf{r}_1, \mathbf{r}_2, \mathbf{r}_3$ tirangle in the $\xi - \eta - \zeta$ system up to a rotation about the ζ axis. If polar coordinates are introduced in the $\xi - \eta$ plane via

$$\xi_i = \rho_i \cos\phi_i, \quad \eta_i = \rho_i \sin\phi_i, \tag{6}$$

it follows from (2) and (4)—(6) that ρ_1, ρ_2, ρ_3 are given by

$$\rho_1 = \tfrac{1}{3}(-s_1^2 + 2s_2^2 + 2s_3^2)^{1/2}, \tag{7a}$$

$$\rho_2 = \tfrac{1}{3}(2s_1^2 - s_2^2 + 2s_3^2)^{1/2}, \tag{7b}$$

$$\rho_3 = \tfrac{1}{3}(2s_1^2 + 2s_2^2 - s_3^2)^{1/2} \tag{7c}$$

and that the angles ϕ_{21}, ϕ_{32}, ϕ_{13} defined by

$$\phi_{21} = \phi_2 - \phi_1, \quad \phi_{32} = \phi_3 - \phi_2, \quad \phi_{13} = \phi_1 - \phi_3 \tag{8}$$

are fixed by

$$\sin\phi_{21} = (I_1 I_2)^{1/2}/(3^{1/2}\rho_1\rho_2), \tag{9a}$$

$$\cos\phi_{21} = (s_1^2 + s_2^2 - 5s_3^2)/(18\rho_1\rho_2), \tag{9b}$$

$$\sin\phi_{32} = (I_1 I_2)^{1/2}/(3^{1/2}\rho_2\rho_3), \tag{9c}$$

$$\cos\phi_{32} = (-5s_1^2 + s_2^2 + s_3^2)/(18\rho_2\rho_3), \tag{9d}$$

$$\sin\phi_{13} = (I_1 I_2)^{1/2}/(3^{1/2}\rho_3\rho_1), \tag{9e}$$

$$\cos\phi_{13} = (s_1^2 - 5s_2^2 + s_3^2)/(18\rho_3\rho_1), \tag{9f}$$

where I_1 and I_2 are the principal axis values of the moments of inertia in the $\xi - \eta$ plane divided by the particle mass. Explicitly,

$$I_1 = [s_1^2 + s_2^2 + s_3^2 + 2(s_1^4 + s_2^4 + s_3^4 - s_1^2 s_2^2 - s_2^2 s_3^2 - s_3^2 s_1^2)^{1/2}]/6 \tag{10}$$

and

$$I_2 = [s_1^2 + s_2^2 + s_3^2 - 2(s_1^4 + s_2^4 + s_3^4 - s_1^2 s_2^2 - s_2^2 s_3^2 - s_3^2 s_1^2)^{1/2}]/6 \tag{11}$$

One more condition is needed to pin down the rotation about the ζ axis and complete the specification of the orientation of the $\mathbf{r}_1$, $\mathbf{r}_2$, $\mathbf{r}_3$ triangle in the $\xi - \eta - \zeta$ plane. The usual condition

$$\xi_1\eta_1 + \xi_2\eta_2 + \xi_3\eta_3 = 0 \tag{12}$$

(vanishing of the products of inertia) aligns the $\xi - \eta - \zeta$ axes with the principal axes of the moment of inertia tensor and implies that

$$\phi_1 = \tfrac{1}{2}\tan^{-1}\left(\frac{6(s_3^2 - s_2^2)(3I_1I_2)^{1/2}}{2s_1^4 + 5s_2^4 + 5s_3^4 - 5s_1^2s_2^2 - 2s_2^2s_3^2 - 5s_3^2s_1^2}\right) \tag{13}$$

with similar expressions for ϕ_2 and ϕ_3. With the usual choice (12), the kinetic energy operator

$$H = -\frac{\hbar^2}{2m}\sum_{i=1}^{3}\left(\frac{\partial^2}{\partial x_i^2} + \frac{\partial^2}{\partial y_i^2} + \frac{\partial^2}{\partial z_i^2}\right) \tag{14}$$

can be expressed in the form[2,3]

$$H = H_{cm} + H_{int} \tag{15}$$

where

$$H_{cm} = -\frac{\hbar^2}{6m}\left(\frac{\partial^2}{\partial X^2} + \frac{\partial^2}{\partial Y^2} + \frac{\partial^2}{\partial Z^2}\right) \tag{16}$$

with X, Y, Z the Cartesian coordinates of the center-of-mass $\mathbf{R}$ and

$$H_{int} = -\frac{\hbar^2}{2m}(T_1 + T_2 + T_3), \tag{17}$$

where

$$T_1 = 2\left(\frac{\partial^2}{\partial s_1^2} + \frac{\partial^2}{\partial s_2^2} + \frac{\partial^2}{\partial s_3^2}\right) + \left(\frac{s_1^2 + s_2^2 - s_3^2}{s_1 s_2}\right)\frac{\partial^2}{\partial s_1 \partial s_2} + \left(\frac{-s_1^2 + s_2^2 + s_3^2}{s_2 s_3}\right)\frac{\partial^2}{\partial s_2 \partial s_3} + \left(\frac{s_1^2 - s_2^2 + s_3^2}{s_3 s_1}\right)\frac{\partial^2}{\partial s_3 \partial s_1} + 4\left(\frac{1}{s_1}\frac{\partial}{\partial s_1} + \frac{1}{s_2}\frac{\partial}{\partial s_2} + \frac{1}{s_3}\frac{\partial}{\partial s_3}\right), \tag{18a}$$

$$T_2 = \frac{4(I_1I_2)^{1/2}}{3^{1/2}(I_1 - I_2)^2}\left[\left(\frac{s_2^2 - s_3^2}{s_1}\right)\frac{\partial}{\partial s_1} + \left(\frac{s_3^2 - s_1^2}{s_2}\right)\frac{\partial}{\partial s_2} + \left(\frac{s_1^2 - s_2^2}{s_3}\right)\frac{\partial}{\partial s_3}\right]\left(\frac{i}{\hbar}L_\zeta\right), \tag{18b}$$

and

$$T_3 = -\tfrac{1}{2}(I_1^{-1} + I_2^{-1})\hbar^{-2}(L^2 - L_\zeta^2) - (I_1 + I_2)(I_1 - I_2)^{-2}(\hbar^{-2}L_\zeta^2) + \tfrac{1}{4}(I_2^{-1} - I_1^{-1})\hbar^{-2}(L_+^2 + L_-^2) \tag{18c}$$

with

$$L^2 = -\hbar^2\left[\frac{\partial^2}{\partial\beta^2} + \cot\beta\frac{\partial}{\partial\beta} + \csc^2\beta\left(\frac{\partial^2}{\partial\alpha^2} + \frac{\partial^2}{\partial\gamma^2} - 2\cos\beta\frac{\partial^2}{\partial\alpha\partial\gamma}\right)\right], \tag{19a}$$

$$L_z = -i\hbar\frac{\partial}{\partial\alpha}, \tag{19b}$$

$$L_\zeta = -i\hbar\frac{\partial}{\partial\gamma}, \tag{19c}$$

and

$$L_\pm = i\hbar\exp(\mp i\gamma)\left(\csc\beta\frac{\partial}{\partial\alpha} - \cot\beta\frac{\partial}{\partial\gamma} \mp i\frac{\partial}{\partial\beta}\right). \tag{19d}$$

Here L^2 is the total angular momentum operator, L_z is the z component of angular momentum in the space-fixed system, L_ξ, L_η, L_ζ are components of angular momentum in the body system, and $L_\pm = L_\xi \pm iL_\eta$ are raising and lowering operators in the body system.

The traditional condition (12) has the disadvantage of introducing a coordinate singularity at the equilateral triangle configuration $s_1 = s_2 = s_3$, where $I_1 = I_2$ and the right-hand side of (13) is undefined. This difficulty also shows up in the kinetic energy; the operators T_2 and T_3 defined in (18b) and (18c) contain coefficients which are infinite when $I_1 = I_2$. The line $s_1 = s_2 = s_3$ in the $s_1 - s_2 - s_3$ space is like a square root branch point in the complex plane: one circuit around this line changes the right-hand side of (13) by π. Infinitesimal changes in s_1, s_2, and/or s_3 in the neighborhood of $s_1 = s_2 = s_3$ can result in finite changes in the right-hand side of (41), making it awkward to expand about $s_1 = s_2 = s_3$.

This coordinate singularity has caused difficulties in the theory of the nonlinear triatomic molecule[4] and the triton.[5] The present work was motivated by a desire to avoid similar difficulties in the theory of the exchange third virial coefficient.[6] A canonical transformation in the form of an expansion about $s_1 = s_2 = s_3$ was used by Weiguny[3] to eliminate the singularity. However, such an expansion does not give a new coordinate system globally. In the next section, the usual condition (12) will be replaced by an alternative condition which eliminates the singularity.

A singularity is also present at those configurations in which the three particles are in a straight line so that I_1 or I_2 vanishes. This singularity, which is accompanied by a decrease in the number of rotational degrees of freedom from three to two, will not be discussed further in the present paper.

II. SOLUTION OF THE PROBLEM

The coordinate singularity at $s_1 = s_2 = s_3$ arises from the fact that when $I_1 = I_2$, all axes through the origin of the $\xi - \eta$ plane are principal axes of the moment of inertia tensor, so that aligning the principal axes of the moment of inertia tensor with the coordinate axes no longer fixes the coordinate system. An obvious way out of the difficulty would be to replace the usual condition (12) by a condition such as $\eta_1 = 0$, which places particle 1 on the ξ axis. This solution, however, has the defect of treating the particles asymmetrically, thus complicating the discussion of the symmetry of the wavefunction under particle permutation.

In the equilateral triangle configuration $s_1 = s_2 = s_3$ with ϕ_1 equal to an arbitrarily chosen constant c, the particles can be permuted by certain rotations which are equivalent to changes in Euler angles α, β, γ. In particular, the interchange of particles 2 and 3 results from a rotation of $\pi - 2c$ about the ξ axis followed by a rotation of π about the η axis; this is equivalent to the replacement of α, β, γ by α', β', γ' where

$$\alpha' = \pi + \alpha, \quad \beta' = \pi - \beta, \quad \gamma' = -2c - \gamma. \tag{20}$$

Similar formulas hold for other permutations of the particles: to interchange 1 and 3, replace c by $c + (2\pi/3)$ in the above (i.e., ϕ_1 by ϕ_2); to interchange 1 and 2, replace c by $c + (4\pi/3)$ in the above (ϕ_1 by ϕ_3).

The dependence of ϕ_1 on s_1, s_2, and s_3 for an arbitrary configuration will now be determined by requiring that a permutation P of the particles be accomplished by applying the permutation P to the interparticle separations s_1, s_2, s_3 and making the same changes in the Euler angles as are made for the equilateral triangle configuration. For example, 2 and 3 are to be inter-

changed by replacing s_1, s_2, s_3, α, β, γ by s_1', s_2', s_3', α', β', γ' where α', β', γ' are given by (20) and

$$s_1' = s_1, \quad s_2' = s_3, \quad s_3' = s_2. \tag{21}$$

A rotation of $\pi - 2c$ about the ξ axis followed by a rotation of π about the η axis carries $\phi_1(s_1, s_2, s_3)$ into $2c - \phi_1(s_1, s_2, s_3)$; interchange of s_2 and s_3 then carries it into $2c - \phi_1(s_1, s_3, s_2)$. The fact that particle 1 is left fixed by the interchange of 2 and 3 then implies the condition

$$\phi_1(s_1, s_2, s_3) = 2c - \phi_1(s_1, s_3, s_2). \tag{22}$$

Consideration of the interchange of 1 and 3 yields

$$\phi_2(s_1, s_2, s_3) = 2[c + (2\pi/3)] - \phi_2(s_3, s_2, s_1), \tag{23}$$

while considering the interchange of 1 and 2 leads to

$$\phi_3(s_1, s_2, s_3) = 2[c + (4\pi/3)] - \phi_3(s_2, s_1, s_3). \tag{24}$$

Because all other permutations can be built up from the interchanges, no further conditions need be imposed. The general solution of Eq. (22)—(24) is

$$\phi_1(s_1, s_2, s_3) = c + \tfrac{1}{3}(-\phi_{21} + \phi_{13}) + f(s_1, s_2, s_3), \tag{25a}$$

$$\phi_2(s_1, s_2, s_3) = c + 2\pi/3 + \tfrac{1}{3}(-\phi_{32} + \phi_{21}) + f(s_1, s_2, s_3), \tag{25b}$$

$$\phi_3(s_1, s_2, s_3) = c + 4\pi/3 + \tfrac{1}{3}(-\phi_{13} + \phi_{32}) + f(s_1, s_2, s_3), \tag{25c}$$

where f is an arbitrary antisymmetric function of s_1, s_2, and s_3 and ϕ_{21}, ϕ_{32}, ϕ_{13} are defined by Eqs. (9). It is easy to verify that (25) is consistent with (8). As a consequence of (9), ϕ_{21}, ϕ_{32}, ϕ_{13} all lie between 0 and π. Thus $\phi_i - f$ cannot stray from its value c at the equilateral triangle configuration by more than $\pi/3$, and is well behaved in the neighborhood of the equilateral triangle configuration.

Let ϕ_1^{old} stand for the right-hand side of (13), and ϕ_1^{new} for the right-hand side of (25a). It can then be shown that the difference $\theta = \phi_1^{\text{old}} - \phi_1^{\text{new}}$ between the two specifications of ϕ_1 is given by

$$\theta(s_1, s_2, s_3) = \tfrac{1}{6}\tan^{-1}[N(s_1, s_2, s_3)/D(s_1, s_2, s_3)] - c - f(s_1, s_2, s_3), \tag{26}$$

where

$$N(s_1, s_2, s_3) = 24(3I_1I_2)^{1/2}[3(I_1 + I_2)^2 + 4I_1I_2](s_1^2 - s_2^2)(s_2^2 - s_3^2)(s_3^2 - s_1^2) \tag{27}$$

and

$$D(s_1, s_2, s_3) = 27(I_1 - I_2)^6 + 4(I_1 + I_2)[(I_1 + I_2)^2 + 12I_1I_2](s_1^2 + s_2^2 - 2s_3^2)(s_2^2 + s_3^2 - 2s_1^2) \times (s_3^2 + s_1^2 - 2s_2^2). \tag{28}$$

The new coordinate system differs from the old only in the specification of γ: $\gamma^{\text{new}} = \gamma^{\text{old}} + \theta$. It is easy to show that θ reduces to the rotation generated by Weiguny's cannonical transformation[3] in the neighborhood of $s_1 = s_2 = s_3$. The effect of this rotation on the Hamiltonian given by Eqs. (15)—(19) is the replacement of T_2 and T_3 by T_2' and T_3' where

$$T_2' = \left(\frac{(s_1^2 - s_2^2)(s_2^2 - s_3^2)(s_3^2 - s_1^2)}{81\rho_1^2\rho_2^2\rho_3^2(3I_1I_2)^{1/2}} + \sum_{i=1}^{3}\Omega_i - T_1 f + \sum_{i=1}^{3}\Lambda_i f\right)\left(\frac{i}{\hbar}L_\xi\right) \tag{29}$$

and

$$T_3' = -\tfrac{1}{2}(I_1^{-1} + I_2^{-1})\hbar^{-2}(L^2 - L_\xi^2) + \left[-(54\rho_1^2\rho_2^2\rho_3^2)^{-1} \times [2(I_1 + I_2)^2 - (I_1 - I_2)^2] + \sum_{i=1}^{3}\left(\Omega_i f - \tfrac{1}{2}(\Lambda_i f)\frac{\partial f}{\partial s_i}\right)\right]\hbar^{-2}L_\xi^2 + \tfrac{1}{4}(I_2^{-1} - I_1^{-1})\hbar^{-2}[\exp(2i\theta)L_+^2 + \exp(-2i\theta)L_-^2], \tag{30}$$

where

$$\Omega_1 \equiv -2(27\rho_1^2\rho_2^2\rho_3^2)^{-1}(3I_1I_2)^{1/2}\rho_1^2(s_2^2 - s_3^2)s_1^{-1}\partial/\partial s_1, \tag{31a}$$

$$\Omega_2 \equiv -2(27\rho_1^2\rho_2^2\rho_3^2)^{-1}(3I_1I_2)^{1/2}\rho_2^2(s_3^2 - s_1^2)s_2^{-1}\partial/\nabla s_2, \tag{31b}$$

$$\Omega_3 \equiv -2(27\rho_1^2\rho_2^2\rho_3^2)^{-1}(3I_1I_2)^{1/2}\rho_3^2(s_1^2 - s_2^2)s_3^{-1}\partial/\partial s_3, \tag{31c}$$

$$\Lambda_1 \equiv 4\left(\frac{\partial}{\partial s_1}\right) + (s_1s_2)^{-1}(s_1^2 + s_2^2 - s_3^2)\left(\frac{\partial}{\partial s_2}\right) + (s_3s_1)^{-1} \times (s_1^2 - s_2^2 + s_3^2)\left(\frac{\partial}{\partial s_3}\right), \tag{32a}$$

$$\Lambda_2 \equiv (s_1s_2)^{-1}(s_1^2 + s_2^2 - s_3^2)\left(\frac{\partial}{\partial s_1}\right) + 4\left(\frac{\partial}{\partial s_2}\right) + (s_2s_3)^{-1} \times (-s_1^2 + s_2^2 + s_3^2)\left(\frac{\partial}{\partial s_3}\right), \tag{32b}$$

and

$$\Lambda_3 \equiv (s_3s_1)^{-1}(s_1^2 - s_2^2 + s_3^2)\left(\frac{\partial}{\partial s_1}\right) + (s_2s_3)^{-1}(-s_1^2 + s_2^2 + s_3^2) \times \left(\frac{\partial}{\partial s_2}\right) + 4\left(\frac{\partial}{\partial s_3}\right). \tag{32c}$$

If the arbitrary antisymmetric function f and its first two partial derivatives are nonsingular at $I_1 = I_2$, then the operators T_2' and T_3' no longer contain the singularity at $I_1 = I_2$ which was present in T_2 and T_3. Furthermore, if f and its first two derivatives vanish at $I_1 = I_2$, the term T_3' can be shown to reduce to the Hamiltonian for a plane rigid rotator when $I_1 = I_2$.

[1]A.R. Edmonds, *Angular Momentum in Quantum Mechanics* (Princeton U.P., Princeton, New Jersey, 1960).
[2]H. Diehl, J. Flügge, U. Schröder, A. Völkel, and A. Weiguny, Z. Physik 162, 1 (1961).
[3]A. Weiguny, J. Mol. Spectr. 15, 1 (1965).
[4]Carl Eckart, Phys. Rev. 46, 383 (1934) and 47, 552 (1935); J.H. Van Vleck, Phys. Rev. 47, 487 (1935).
[5]G.H. Derrick, Nucl. Phys. 16, 405 (1960) and 18, 303 (1960).
[6]R.N. Hill (accepted for publication in J. Stat. Phys.).

A study of the Hilbert space properties of the Veneziano model operator formalism

L. Brink, P. H. Frampton, and H. B. Nielsen

CERN, Geneva, Switzerland
(Received 3 January 1973)

The domains of definition of the operators used to factorize the generalized Veneziano model are studied within the Hilbert space defined by the harmonic oscillator creation and annihilation operators $a_\mu^{(r)\dagger}, a_\mu^{(r)}$. These individual operators may not be well behaved, although, of course, the matrix elements used in the conventional operational factorization are well defined. Concerning the individual operators, it is shown that the ground-state vertex written as $V(p)=\exp[-\Sigma_{r=1}^{\infty}(p\cdot a^{(r)+})/\sqrt{r}]\exp[\Sigma_{r=1}^{\infty}(p\cdot a^{(r)})/\sqrt{r}]$ is nowhere defined within the Hilbert space; the product with a twisting operator $\Omega(q)V(p)$ is, however, densely defined, as is the symmetrical three-reggeon vertex. The propagator $D(p)$ is bounded everywhere, away from its poles. The twisting operator $\Omega(p)$ is undefined on finite occupation states, but is densely defined on a subset of coherent states; its Hermitian conjugate $\Omega^+(p)$ is densely defined on both finite occupation and coherent states. It is found that a suitable rewritten form of the product $D(q)V(p)$ is densely defined for certain values of momenta; this relates to the fact that off-mass shell states satisfying $(L_0-L_{-r}-1)|\Phi>=0$, where L_n are the conventional gauge operators, are better defined than those satisfying $(L_0-L_{-r}+r-1|\Phi'>=0$.

1. INTRODUCTION

The study of the properties of the N-particle generalization of Veneziano's beta function dual model for two-body scattering has been greatly facilitated by the harmonic oscillator operator formalism.[1] This operator formalism makes manifest the factorization properties and the spectrum of states which are not obvious in the original integral representation. It can be written in a form where both the factorization and Möbius invariance properties are displayed simultaneously.[2] For a review of the formalism we refer the reader to the article of Alessandrini *et al.*,[3] and the references cited therein.

In the present paper we shall be concerned with the operator formalism developed in Ref. 1, together with the twisting operator and symmetric three-reggeon vertex of Ref. 4. With the three operators, the propagator $D(p)$, the symmetric vertex $V(p_1p_2p_3)$, and the twisting operator $\Omega(p)$, one can, aside from the gauge identities (which we shall consider toward the end of the article in discussing the physical states), build up the whole theory including loops.

We shall discuss the mathematical basis of the operator formalism, in particular the properties within the Hilbert space defined by the Fock space of harmonic oscillator states.

Concerning the three principal operators, regarded as operators acting on Hilbert space states, the results may be summarized: The propagator $D(p)$ is bounded over the whole space; the ground state vertex $V(p)$ is nowhere defined, while the symmetric vertex $\Omega(q)\,V(p)$ and its generalization to the symmetrical three-reggeon vertex $V(p_1p_2p_3)$ are densely defined; the twisting operator $\Omega(p)$ is not defined except on states with null (4-momentum, while its adjoint $\Omega^+(p)$ is densely defined.

Of course, the Hilbert space is rather a restricted concept and similar difficulties of staying within a Hilbert space occur already in nonrelativistic quantum mechanics.[5] The conventional usage of the operator formalism involves always matrix elements of strings of operators $(VD\,VD\cdots V)$ and these are well defined in terms of generalized beta functions and their analytic continuation; thus the results of the present paper do not, of course, cast any doubt on the validity of the normal use of the operator formalism (i.e., on the matrix elements). The mathematical properties of the specific operators are important to know, if one wants to extract as much as possible out of the operator formalism. It is also of importance in understanding dual models and might provide a means for further developments and for construction of other more realistic dual models.

The organization of the paper is as follows: In Sec. 2 we give some mathematical definitions of what we mean by certain classes of vectors within the Hilbert space and introduce some terminology useful for discussing the domains of definition of operators. We study the propagator and twisting operator in Sec. 3, while in Sec. 4 the vertex is investigated, firstly the ground-state vertex and then the fully symmetric three-reggeon vertex. Section 5 is concerned with the redefinition of the product $D(q)V(p)$ and with the alternative definitions of physical states. The final Sec. 6 is devoted to some discussion.

2. MATHEMATICAL DEFINITIONS

When we, in the following sections, are going to claim that certain operators are defined what we shall mean is really only that they are defined as operators mapping a Hilbert space into itself (or possibly into another Hilbert space).

The Hilbert space of interest for us is the Fock space in the operator formalism[1] of the Veneziano model. Let us first consider a set of occupation number states of the type

$$|\{l_1^1, l_1^2, \ldots, l_1^{\cdot}; l_2^1, \cdots\}\rangle = \prod_{n=1}^{\infty}\prod_{\mu=0}^{3}\frac{(a_\mu^{(n)\dagger})^{l_n^\mu}}{(l_n^\mu!)^{1/2}}\,|\,0\rangle, \tag{2.1}$$

where

$$[a_\mu^{(n)}, a_\nu^{(m)\dagger}] = -\delta_{nm}g_{\mu\nu} \tag{2.2}$$

with

$$g_{\mu\nu} = -\delta_{\mu\nu}(-1)^{\delta_{\mu 0}}.$$

Here the l_n^μ are zero for n sufficiently large. The state with all occupation numbers identically equal to zero is called the vacuum state $|\,0\rangle$.

We define the space $\mathfrak{F}$ as the vector space consisting of all (finite) linear combinations of the vectors (2.1). We

call $\mathfrak{F}$ the space of finite occupation states. A typical state $|f\rangle \in \mathfrak{F}$ may be written

$$|f\rangle = \sum_{\{l\}} C_{\{l\}} |\{l\}\rangle,$$

where only finitely many of the coefficients $C_{\{l\}}$ are nonzero.

The space is made a pre-Hilbert space by defining the inner product

$$\langle f' | \Gamma | f\rangle = \sum_{\{l\}} C'^{*}_{\{l\}} C_{\{l\}} \tag{2.3}$$

which follows from (2.1) when we put

$$\Gamma = (-1)^{\sum_{n=1}^{\infty} a_0^{(n)\dagger} a_0^{(n)}}. \tag{2.4}$$

A norm is defined by $\| |f\rangle \| = \sqrt{\langle f | \Gamma | f\rangle}$. Note that this definition of the metric is not Lorentz invariant.

Completing the space $\mathfrak{F}$, we obtain the full Hilbert Fock-space $\mathcal{H}$. The points $|h\rangle$ of the Hilbert space $\mathcal{H}$, which we will consider in this article, can be written as formally infinite linear combinations of states of type (2.1), i.e.,

$$|h\rangle = \sum_{\{l\}} C_{\{l\}} |\{l\}\rangle, \tag{2.5}$$

but now an infinite number of coefficients $C_{\{l\}}$ may be different from zero. However, the norms in $\mathcal{H}$ are bounded, and it is thus required that

$$\sum_{\{l\}} |C_{\{l\}}|^2 < \infty. \tag{2.6}$$

As a consequence of the noncovariance of the metric Γ, Eq. (2.4), the Hilbert space $\mathcal{H}$ is not covariant.

An example of a state in $\mathcal{H}$ is the coherent state defined by

$$|\alpha^{(n)}\rangle = \exp\left(\sum_n \alpha^{(n)} \cdot a^{(n)\dagger}\right) |0\rangle, \tag{2.7}$$

where we have required

$$\sum_{n=1}^{\infty} \sum_{\mu=0}^{3} |\alpha_\mu^{(n)}|^2 < \infty. \tag{2.8}$$

We shall denote the set of all finite linear combinations of finite-norm coherent states as $\mathcal{C}$.

A function $\mathcal{T}$ that maps every vector $|h\rangle \in \mathfrak{D}$ into a vector $\mathcal{T}|h\rangle \in \Delta \subseteq \mathcal{H}$ is called an operator in the space $\mathcal{H}$ defined on the domain $\mathfrak{D}$ and Δ is called the image when it is required that each element in Δ has the form $\mathcal{T}|h\rangle$.

In the following sections we shall be interested in whether the domain $\mathfrak{D}$ is a dense set in $\mathcal{H}$, i.e., whether

$$\overline{\mathfrak{D}} = \mathcal{H}, \tag{2.9}$$

where $\overline{\mathfrak{D}}$ is the closure of $\mathfrak{D}$, according to the topology defined by the norm.

We shall also be interested in whether the operators are bounded. A linear operator $\mathcal{T}$ is bounded when

$$\sup_{\substack{\| |h\rangle \| \leq 1 \\ |h\rangle \in \mathfrak{D}}} \| \mathcal{T} |h\rangle \| < \infty. \tag{2.10}$$

3. MÖBIUS GROUP OPERATORS

In the operator formalism of the dual resonance model certain representations of the group of Möbius transformations leaving a circle invariant $SL(2,R) \approx SO(2,1)$ homomorphic to $SU(1,1)$ play an important role.[3] In the conventional model the generators are

$$L_0(p) = -p^2 - \sum_{n=1}^{\infty} n a^{(n)\dagger} \cdot a^{(n)},$$

$$L_+(p) = -\sqrt{2}\, a^{(1)\dagger} \cdot p - \sum_{n=1}^{\infty} \sqrt{n(n+1)}\, a^{(n+1)\dagger} \cdot a^{(n)},$$

$$L_-(p) = (L_+(p))^{\dagger}. \tag{3.1}$$

Of particular interest are the following functions of these generators: the propagator

$$D = (L_0 - 1)^{-1} \tag{3.2}$$

and the twisting operators

$$\Omega = (-1)^{L_0} e^{-L_+} = e^{L_+} (-1)^{L_0},$$
$$\Omega^{+} = e^{-L_-} (-1)^{L_0} = (-1)^{L_0} e^{L_-}. \tag{3.3}$$

In the upper half of Table I the boundedness and domain properties of such operators are summarized.

We now indicate how these entries in the table were obtained.

(i) *$D(p)$ and $D^{-1}(p)$:*

On the space $\mathfrak{F}$, $D(p)$ is bounded, since the eigenvalue of $(L_0 - 1)^{-1}$ of an $\mathfrak{F}$ state is bounded. This is because the eigenvalue of $(L_0 - 1)^{-1}$ on a state $|f\rangle$ is given by

$$(L_0 - 1)^{-1} |f, p\rangle = \sum_{\{l\}} C_{\{l\}} (L_0 - 1)^{-1} |\{l\}, p\rangle$$
$$= \sum_{\{l\}} C_{\{l\}} (-p^2 - \sum_{n,\mu} n l_n^{\mu} - 1)^{-1} |\{l\}, p\rangle \tag{3.4}$$

and the norm of this state $\| (L_0 - 1)^{-1} |f\rangle \|$ is always bounded off-mass shell. Now we can apply a theorem about bounded operators to be found, for example, in Naimark's book.[6]

Theorem: In a Hilbert space $\mathcal{H}$, a bounded linear operator A is extendible by continuity from its domain $\mathfrak{D}_A$ to a bounded linear operator with $\overline{\mathfrak{D}}_A$, i.e., the closure of $\mathfrak{D}_A$, as its domain of definition.

In the present case, since $D(p)$ is bounded on $\mathfrak{F}$ it can be extended, therefore, by continuity to be bounded on $\overline{\mathfrak{F}} = \mathcal{H}$, the full Hilbert space.

$D^{-1}(p)$ is unbounded on $\mathfrak{F}$, but it is defined there, i.e., $\mathfrak{D}_{D^{-1}(p)} \supseteq \mathfrak{F}$, which is easily seen from Eq. (3.4) written

TABLE I. Boundedness and domain properties of operators.

Operator	Bounded on $\mathcal{H}$, the full Hilbert space	Defined on finite occupation states, i.e., $\mathfrak{D} \supseteq \mathfrak{F}$	Defined on coherent states, i.e., $\mathfrak{D} \supseteq \mathcal{C}$	Defined on a dense subset of $\mathcal{C}$	Defined (i.e., bounded) on the vacuum, i.e., $\mathfrak{D} \supseteq \|0\rangle$
$D(p)$	Yes[a]	Yes[a]	Yes[a]	Yes[a]	Yes[a]
$D^{-1}(p)$	No	Yes	No	Yes	Yes
$\Omega^{\dagger}(p)$	No	Yes	No	Yes	Yes
$\Omega(p)$	No	No	No	Yes	[b]
$V(p)$	No	No	No	No	No
$\Omega(q)V(p)$	No	No	No	Yes	[b]
$D(q)V(p)$ $\left(\begin{matrix}\|p\|^2 < +2 \\ q^2 < -1\end{matrix}\right)$[c]	No	Yes	No	Yes	Yes

[a] We are always working off-mass shell for $D(p)$.
[b] Yes, when $p_\mu = 0$; No, otherwise.
[c] $|p|^2 = p_0^2 + \underline{p}^2$.

for $(L_0 - 1)$. Now we consider $D^{-1}(p)$ acting on a coherent state

$$D^{-1}(p) \mid \alpha^{(n)}\rangle = (L_0 - 1) \mid \alpha^{(n)}\rangle = (-p^2 - \sum_{n=1}^{\infty} n a_n^\dagger \cdot \alpha_n - 1) \mid \alpha^{(n)}\rangle . \quad (3.5)$$

The norm of this state is given by

$$\| D^{-1}(p) \mid \alpha^{(n)}\rangle \|^2 = (p^2+1)^2 \langle \alpha^{(n)} \mid \alpha^{(n)}\rangle + 2(p^2+1) \times \sum_{n=1}^{\infty} n \mid \alpha^{(n)} \mid^2 \langle \alpha^{(n)} \mid \alpha^{(n)}\rangle + \left[\left(\sum_{n=1}^{\infty} n \mid \alpha^{(n)} \mid^2 \right)^2 - \sum_{n=1}^{\infty} n^2 \mid \alpha^{(n)} \mid^2 \right] \langle \alpha^{(n)} \mid \alpha^{(n)}\rangle , \quad (3.6)$$

and the summations can diverge, while still $\langle \alpha_n \mid \alpha_n \rangle$ is finite. Therefore, $D^{-1}(p)$ is not defined on $\mathcal{C}$. It is, however, defined on a dense subset, for example, the dense subset $\{ \mid \alpha_n = \hat{n} z^n \rangle \}$, where $\mid z \mid < 1$ and $\hat{n}$ is a constant 4-vector.

(ii) $\Omega^+(p)$:

On a state of the $\mathcal{F}$ space we have

$$\Omega^+(p) \mid f\rangle = (-1)^{L_0(p)} e^{L_-(p)} \mid f\rangle . \quad (3.7)$$

Now the exponential in $\mid L_-(p) \mid$ becomes a polynomial; also the mode summation in $L_-(p)$ is finite. Therefore, the state $e^{L_-(p)} \mid f\rangle$ has finite but unbounded norm. The operator $(-1)^{L_0(p)}$ is unitary and, therefore, norm preserving. It follows that $\Omega^+(p)$ is defined on $\mathcal{F}$, but unbounded on both $\mathcal{F}$ and $\bar{\mathcal{F}} = \mathcal{H}$. Since $\mid 0\rangle \subset \mathcal{F}$, $\Omega^+(p)$ is defined on the vacuum.

For a coherent state $\mid \alpha_n, p\rangle$

$$\Omega^+(p) \mid \alpha_n, p\rangle = (-1)^{L_0(p)} e^{L_-(0)} \exp\left(- \sum_{n=1}^{\infty} \frac{a_n \cdot p}{\sqrt{n}} \right) \mid \alpha_n, p\rangle . \quad (3.8)$$

Using the canonical formalism of Alessandrini *et al.*,[3] we now have

$$e^{L_-(0)} \mid \alpha_n p\rangle = \mid \sum_m C_{nm} \alpha_m, p\rangle , \quad (3.9)$$

where

$$\sum_{m=1}^{\infty} C_{nm} \frac{z^m}{\sqrt{m}} = \frac{1}{\sqrt{n}} \left(\frac{z}{1-z} \right)^n . \quad (3.10)$$

Thus,

$$\| e^{L_-(0)} \mid \alpha_n, p\rangle \|^2 = \exp\left(\sum_{n=1}^{\infty} \left| \sum_{m=n}^{\infty} \left(\frac{n}{m} \right)^{1/2} \binom{m}{n} \alpha_m \right|^2 \right) = \exp\left(\sum_{n=1}^{\infty} \mid \alpha_n \mid^2 + \sum_{n=1}^{\infty} \sqrt{n(n+1)} \mid \alpha_{n+1} \mid^2 + \cdots \right) \quad (3.11)$$

which for all $\alpha_n > 0$ can be divergent even for $\Sigma \mid \alpha_n \mid^2$ finite. To show that it can be defined on a dense set of $\mathcal{C}$, we use $\alpha_n = \hat{n} z^n / \sqrt{n}$ with $\mid z \mid < 1$; then we have

$$\| e^{L_-(p)} \mid \alpha_n p\rangle \|^2 = \exp\left(\mid \hat{n} \mid^2 \sum_{n=1}^{\infty} \frac{1}{\sqrt{n}} \left| \frac{z}{1-z} \right|^{2n} \right) \times \exp\left[-2 \operatorname{Re}\left(\sum_{n=1}^{\infty} \frac{z^n}{n} n \cdot p \right) \right] , \quad (3.12)$$

where $\mid \hat{n} \mid^2 = \hat{n}_0^2 + \hat{\mathbf{n}}^2$, which is convergent for $\operatorname{Re} z < \frac{1}{2}$. Therefore, $\Omega^+(p)$ is defined on a dense set in $\mathcal{C}$.

(iii) $\Omega(p)$:

For general p^2 we have

$$\| \Omega(p) \mid 0\rangle \|^2 = \exp\left(\mid p \mid^2 \sum_{n=1}^{\infty} \frac{1}{n} \right) . \quad (3.13)$$

Note that for $p^\mu = 0$ we have $\Omega(0) \mid 0\rangle = \mid 0\rangle$. Thus, on vacuum, $\Omega(p)$ is not defined within the Hilbert space for $p_\mu \neq 0$.

For an occupation state $\mid f\rangle$ we investigate first a singly occupied level, where

$$\| \Omega(0) a_q^\dagger \mid 0, 0\rangle \|^2 = \| e^{-L_+(0)} a_q^\dagger \mid 00\rangle \|^2 = \sum_{n=0}^{\infty} \left| \frac{(q+n)!}{q!\,n!} \left(\frac{q}{q+n} \right) \right|^{1/2} = \infty , \quad \text{for any } q \geq 1, \quad (3.14)$$

where we have used the explicit form of $L_+(0)$, Eq. (2.1), and the fact that $(-1)^{L_0(0)}$ is unitary. It is not difficult to convince oneself that this argument generalizes to any $\mid f\rangle$ state. Therefore, $\Omega(0)$ is unbounded on $\mathcal{F}$ and on its closure $\bar{\mathcal{F}} = \mathcal{H}$.

For a coherent state we can write

$$\| \Omega(p) \mid \alpha_n \rangle \|^2 = \| (-1)^{L_0(p)} e^{-L_+(0)} \exp\left(\sum_{n=1}^{\infty} \frac{a_n^\dagger \cdot p}{\sqrt{n}} \right) \mid \alpha_n \rangle \|^2 = \exp\left[\sum_{n=1}^{\infty} \frac{1}{n} \left| \sum_{m=1}^{n} \binom{n}{m} \times \sqrt{m} (-1)^m \alpha_m^\mu - p^\mu \right|^2 \right] . \quad (3.15)$$

If we choose $\alpha_m^\mu = - z^m p^\mu / \sqrt{m}$ such that $\mid z \mid < 1 \wedge \mid 1 - z \mid < 1$ then this leads to a finite norm. By adding to α_m^μ lower powers of z (obtained by differentiating α_m as is discussed in more detail for the vertex in Section 4 below) we can show that $\Omega(p)$ is defined on a dense subset of $\mathcal{C}$, although not on $\mathcal{C}$ itself since $\mathcal{C}$ includes states with α_m^μ such that, for example, $0 > z > -1$.

4. THE VERTEX

In this section we consider the conventional untwisted vertex $V(p)$ for emission of a scalar ground state meson; the result will be that $V(p)$ is nowhere defined within the Hilbert space. We shall, however, find that the twisted vertex $\Omega(p+q)V(p)$ is a densely defined unbounded operator. More generally the cyclically symmetric Caneschi–Schwimmer–Veneziano vertex will be found to be defined for a certain dense set of coherent states in the sense that, putting in one type of coherent state on one leg, together with another type on a second leg, one will obtain at the third leg a normalizable state.

We shall first give some rather simple and convincing arguments that the vertex $V(p)$ is not defined as long as the momentum p has no time component (which is possible for special spacelike momentum), and further is not defined, for general momentum, on any $\mathcal{C}$ or $\mathcal{F}$ state. Only then shall we introduce a more abstract approach to demonstrate that $V(p)$ is quite generally undefined.

The conventional ground state vertex in the operator formalism is written formally

$$V(p) = \exp\left(- \sum_{n=1}^{\infty} \frac{(a^{(n)+} \cdot p)}{\sqrt{n}} \right) \exp\left(\sum_{n=1}^{\infty} \frac{(a^{(n)} \cdot p)}{\sqrt{n}} \right) \quad (4.1)$$

so that the N point function can then be written formally (in a multiperipheral configuration)

$$A_N = \langle 0 | V(k_2) D(S_{12}) V(k_3) D(S_{13}) \ldots V(k_{N-2}) \times D(S_{N-1,N}) V(k_{N-1}) | 0 \rangle, \quad (4.2)$$

where $S_{ij} = (p_i + p_{i+1} + \ldots + p_j)^2$, and the bra and ket vacuum states have momenta k_1 and k_N, respectively.

We now will show the unhappy result that the operator $V(p)$ is not defined anywhere within the Hilbert space. Define an operator

$$V_x(p) = \exp\left(-\sum_{n=1}^{\infty} \frac{a^{(n)\dagger} \cdot p}{\sqrt{n}} x^n\right) \exp\left(\sum_{n=1}^{\infty} \frac{a^{(n)} \cdot p}{\sqrt{n}} x^n\right) \quad (4.3)$$

so that $\lim_{x \to 1} V_x(p) = V(p)$; also define, for momentum with no time component, the unitary operator

$$U(p, x) = \exp\left(p \cdot \sum_{n=1}^{\infty} \frac{1}{\sqrt{n}} x^n (a^{(n)} - a^{(n)\dagger})\right). \quad (4.4)$$

$U(p, x)$ is unitary (and bounded) in $\mathfrak{F}$. Now, we use the fact that

$$\forall |\psi\rangle \in \mathfrak{H} \exists \text{ a sequence } \{|\phi_n\rangle\}_{n=1}^{\infty}$$

such that

$$\lim_{n \to \infty} |\phi_n\rangle = |\psi\rangle \quad \text{where all } |\phi_i\rangle \in \mathfrak{F}. \quad (4.5)$$

Now we observe that

$$V_x(p) = C(p, x) U(p, x) \quad (4.6)$$

with the c number

$$c(p, x) = \exp\left(-\frac{p^2}{2} \sum_{n=1}^{\infty} \frac{x^n}{n}\right). \quad (4.7)$$

Now, since $U(p, x)$ is unitary for all x, we can consider

$$\lim_{n \to \infty} \| U(p, x) |\phi_n\rangle \|^2 = \lim_{n \to \infty} \| |\phi_n\rangle \|^2 = \| |\psi\rangle \|^2, \quad (4.8)$$

and then we see that

$$\| V_x(p) |\phi_n\rangle \|^2 = |c(p, x)|^2 \| U(p, x) |\phi_n\rangle \|^2 = |c(p, x)|^2 \| |\phi_n\rangle \|^2. \quad (4.9)$$

Therefore, $V_x(p)$ is bounded by $|c(p, x)|$ and, for $|x| < 1$, $V_x(p)$ is defined on $\mathfrak{F}$. For $x \to 1$, however, on a general state $|\psi\rangle$,

$$\lim_{x \to 1} \| V_x(p) |\psi\rangle \|^2 = \lim_{x \to 1} |c(p, x)|^2 \| |\psi\rangle \|^2 = \infty. \quad (4.10)$$

Thus we deduce, for momentum with no time component, that $V(p)$ is undefined everywhere in the Hilbert space.

For general $p\mu$, it is easy to show that $V(p)$ is not defined on any coherent state or any $\mathfrak{F}$ state, as a Hilbert space operator. In the case of a coherent state $|\{\alpha_n\}, q>$ there occurs, in the norm $\| V(p) | \{\alpha_n\}, q > \|$, an exponential of a term

$$\left[(\underline{p}^2 + p_0^2) \sum_{n=1}^{\infty} \frac{1}{n}\right]$$

which cannot be cancelled by any choice of the α_n such that

$$\sum_{n=1}^{\infty} |\alpha_n|^2 < \infty.$$

That $V(p)$ is not defined on $\mathfrak{F}$ states is seen by noticing that the harmonic oscillators above a certain mode number are excited by $V(p)$ in just the same way as mentioned already for the coherent state.

The more abstract general proof for the nonexistence of $V(p)$ proceeds in three steps: (i) the derivation of an operator identity for $V_x(p)^\dagger \Gamma V_x(p)$; (ii) the proof that the expectation value of an exponential of a Hermitian operator is strictly greater than zero; and (iii) the deduction from (i) and (ii) that any image of $V(p)$ has infinite norm.

(i) By using the commutation rules, Eq. (2.2), for the harmonic oscillator operators, it is straightforward to find that (formally), for $p_\mu = (p_{0\perp}, \underline{p})$,

$$V_x(p)^\dagger \Gamma V_x(p) = \exp\left(p_\mu \sum_{n=1}^{\infty} \frac{a_\mu^{n\dagger} x^n}{\sqrt{n}}\right) \exp\left(-p_\mu \sum_{n=1}^{\infty} \frac{a_{n\mu} x^n}{\sqrt{n}}\right) \Gamma \times \exp\left(-p_\mu \sum_{n=1}^{\infty} \frac{a_\mu^{n\dagger} x^n}{\sqrt{n}}\right) \exp\left(p_\mu \sum_{n=1}^{\infty} \frac{a_\mu^n x^n}{\sqrt{n}}\right) = (1 - x^2)^{-3 p_0^2 - \underline{p}} \tilde{U}_x \exp(-H_x) \Gamma \exp(H_x) \tilde{U}_x \quad (4.11)$$

as an operator identity, where

$$\tilde{U}_x = \exp\left(p_0 \sum_{n=1}^{\infty} \frac{a_{n0}^\dagger + a_{n0}}{\sqrt{n}} x^n\right) \quad (4.12)$$

is Hermitian with respect to the indefinite metric

$$\tilde{U}_x = \tilde{U}_x^\dagger \quad (4.13)$$

and unitary with respect to the Γ metric

$$\tilde{U}_x^{-1} = \Gamma \tilde{U}_x^\dagger \Gamma \quad (4.14)$$

and where

$$H_x = p_0 \sum_{n=1}^{\infty} \frac{a_{n0} - a_{n0}^\dagger}{\sqrt{n}} x^n \quad (4.15)$$

is Hermitian in the Γ metric

$$H_x = \Gamma H_x^\dagger \Gamma. \quad (4.16)$$

In deriving Eq. (4.11) we used

$$\exp\left(2 p_0 \sum_{n=1}^{\infty} \frac{a_{n0} x^n}{\sqrt{n}}\right) = (1 - x^2)^{-p_0^2} \exp(H_x) \tilde{U}_x.$$

The Γ norm squared of $V(p)|h\rangle$, where $|h> \in \mathfrak{H}$, is the expectation value of this operator (4.11), and from the ultimate step in Eq. (4.11) we see that this expectation value is equal to a diverging c number multiplied by the expectation value of an expontial $\exp(2H_x)$ for the state $\tilde{U}_x |h\rangle$. Since $\tilde{U}_x$ is unitary and so bounded, this state $\tilde{U}_x |h\rangle$ exists for all $|h\rangle \in \mathfrak{H}$ and is in fact different from zero, since $\| \tilde{U}_x |h\rangle \| = \| |h> \|$.

(ii) Now we wish to show that the expectation value $\langle h | \exp(H) | h \rangle$, for H a Hermitian operator, is greater than zero.

According to the spectral theorem of von Neumann[7] for Hermitian operators, we may write the Stieltjes integral form of H

$$H = \int_{-\infty}^{\infty} t dI_t, \quad (4.17)$$

where I_t is the family ofprojectors for the operator H. More generally, we may make such an integral repre-

sentation for any operator function of H, using the same spectral function I_t; in particular,

$$\exp(H) = \int_{-\infty}^{\infty} \exp(t) dI_t. \tag{4.18}$$

The resolution of the identity I_t is defined such that $\lim_{t\to-\infty} I_t = 0$ and $\lim_{t\to+\infty} I_t = 1$. Therefore, there exists a t' such that

$$\langle h|(1 - I_{t'})|h\rangle > 0,$$

whereupon, writing

$$\langle h|\exp(H)|h\rangle = \int_{-\infty}^{t'} \exp(t)\langle h|dI_t|h\rangle + \int_{t'}^{\infty} \exp(t)\langle h|dI_t|h\rangle, \tag{4.19}$$

we see that the first term on the right-hand side is greater than or equal to zero, while the second term satisfies

$$\int_{t'}^{\infty} \exp(t)\langle h|dI_t|h\rangle \geq \exp(t')\int_{t'}^{\infty}\langle h|dI_t|h\rangle = \exp(t')\langle h|(1 - I_{t'})|h\rangle. \tag{4.20}$$

It follows that

$$\langle h|\exp(H)|h\rangle > 0. \quad \text{QED} \tag{4.21}$$

(iii) Now we combine the results of Eqs. (4.11) and (4.21) to deduce that

$$\begin{aligned}\|V(p)|h\rangle\|^2 &= \lim_{x\to 1}[(1-x^2)^{-3p_0^2-p} \times \langle h|\tilde{U}_x^\dagger \Gamma \exp(2H_x)\tilde{U}_x|h\rangle] \\ &= \left(\lim_{x\to 1}(1-x^2)^{-3p_0^2-p}\right)\cdot\langle h|\tilde{U}_1^\dagger \Gamma\exp(2H_1(\tilde{U}_1|h\rangle \\ &= \infty.\end{aligned} \tag{4.22}$$

By this we have shown the the vertex $V(p)$ is nonexistent within the Hilbert space.

Despite the bad properties of $V(p)$ the operator ΩV corresponding to the cyclically symmetric vertex can in fact be defined somewhere even for spacelike momentum of the ground state particle. Note that we must regard the operator ΩV as a single entity rather than as a product of two operators, if we wish to remain in the Hilbert space. In fact let

$$|\alpha, q\rangle = \prod_{n=1}^{\infty} e^{\alpha_n \cdot a^{(n)\dagger}}|0, q\rangle \tag{4.23}$$

be a coherent state with 4-momentum q; the twisted ground state vertex is formally written $\Omega(q+p)V(p)$ when acting on $|\alpha, q\rangle$, and

$$\begin{aligned}\Omega(p+q)V(p)|\alpha_1 q\rangle &= \Omega(0)\exp\left((p+q)\cdot\sum\frac{a^{(n)\dagger}}{\sqrt{n}}\right)\left(-p\cdot\sum\frac{a^{(n)\dagger}}{\sqrt{n}}\right)\left(p\cdot\sum\frac{a^{(n)}}{\sqrt{n}}\right) \\ &\exp\left(-p\cdot\sum\frac{a^{(n)\dagger}}{\sqrt{n}}\right)\exp\left(p\cdot\sum\frac{a^{(n)}}{\sqrt{n}}\right)|\alpha, q\rangle = \\ &= \exp\left(-q\sum_{n=1}^{\infty}\frac{a^{(n)\dagger}}{\sqrt{n}}\right):\exp-\sum_{n,m}a^{(n)\dagger}(C_{nm}-\delta_{nm})a^{(m)}: \\ &\exp\left(p\cdot\sum\frac{a^{(n)}}{\sqrt{n}}|\alpha, q\rangle\right) \\ &= \exp\left(p\cdot\sum_{n=1}^{\infty}\frac{\alpha n}{\sqrt{n}}\right)\left|\left\{-\sum_j C_{ij}\alpha_j - q\frac{1}{\sqrt{i}}\right\}_{i=1,2,\ldots}, q\right\rangle.\end{aligned} \tag{4.24}$$

Here

$$C_{nm} = \left(\frac{m}{n}\right)^{1/2}(-1)^m\binom{n}{m}.$$

It is rather easy to see, by taking, for instance,

$$\alpha_n = Z^n/\sqrt{n}\cdot q \quad \text{with } |z| < 1 \wedge |1-z| < 1 \tag{4.25}$$

and using[3]

$$\sum_{m=1}^{\infty} nm\frac{Z^m}{\sqrt{m}} = \frac{(1-Z)^n - 1}{\sqrt{n}}, \tag{4.26}$$

that the expression for the formal symbol $\Omega(p+q)V(p)|\alpha, q\rangle$ becomes

$$\exp\left(p\cdot q\sum_{n=1}^{\infty}\frac{Z^n}{n}\right)\left|\left\{-\frac{(1-Z)^i}{\sqrt{i}}\right\}q\right._{i=1,2,\ldots}, q\Big\rangle, \tag{4.27}$$

which is a finite norm coherent state. To be specific, the norm is

$$\left\|\exp\left(p\cdot q\sum_{n=1}^{\infty}\frac{Z^n}{n}\right)\left|\left\{-\frac{(1-Z)^i}{\sqrt{i}}q\right\}_{i=1,2,\ldots}, q\right\rangle\right\| = \exp\left(\mathrm{Re}\, p\cdot q\sum_{n=1}^{\infty}\frac{Z^n}{n} + \frac{1}{2}\sum_{i=1}^{\infty}\frac{(1-Z)^{2i}}{i}|q|^2\right) \tag{4.28}$$

and this exponential is finite when $|z| < 1$ and $|1-z| < 1$.

The possibility (4.25) is only one out of infinitely many since we can add to the series α any series for which $C\cdot\beta$ has finite norm, i.e., for which

$$\sum_{n=1}^{\infty}\left|\sum_{m=1}^{\infty}C_{nm}\beta_m\right|^2 < \infty. \tag{4.29}$$

By using (4.26) it is easy by differentiation, for example, to find an infinite number of such series β, namely

$$\beta_n^{(r)} = n(n-1)\ldots(n-r+1)Z^{n-r}/\sqrt{n}, \tag{4.30}$$

where $z = 1, 2, \ldots$ and $|z|\langle 1 \wedge |1-z| < 1$. In fact one finds

$$\sum_{m=1}^{\infty}C_{nm}\beta_m^{(r)} = (-1)^r\frac{(1-Z)^{n-r}n(n-1)\ldots(n-r+1)}{\sqrt{n}}, \tag{4.31}$$

which is obviously of finite norm because $|1-z| < 1$. We remark for its own interest that for $z = \frac{1}{2}$ the series β_m is an eigenvector of thematrix C with eigenvalue $(-1)^r$.

We now want to show that by meansof the coherent states

$$\left|\left\{\frac{Z^i}{\sqrt{i}}q\right\}_{i=1,2,\ldots} + \sum_{r=1}^{\text{finite}}C_r\beta^{(r)}, q\right\rangle, \tag{4.32}$$

where c_r are 4-vectors, it is possible to argue that the operator ΩV is defined on a dense domain in the Hilbert space.

Since any coherent state $|\alpha\rangle$ (with $\sum|\alpha_n|^2 < \infty$) can be approximated by another one $|\alpha'\rangle$ provided we can approximate the series α by α' in the norm of the Hilbert space l_2 of series with convergent square sum, we can show that the states of type (4.32) can approximate any coherent state. By choosing z in (4.30) small

but different from zero, one easily shows that linear combinations of series β can approximate any series which has only zeros after a certain step and thus any series at all. By choosing z sufficiently small we can make the norm of

$$\{(Z^i/\sqrt{i}\,)q\}_{i=1,2,\ldots}$$

arbitrarily small too; thus states of the form (4.32) can approximate any coherent state and so by taking finite linear combinations we can approximate all states in the Hilbert space by states for which ΩV is defined. That is to say, ΩV is densely defined.

It is rather easy to see that also the Caneschi–Schwimmer–Veneziano three-reggeon vertex[4] is defined on a dense set in the following sense: There exists a dense set of vectors $|1\rangle$ for which the "vertex operator"

$$_2\langle 0|\,_1\langle 0|V(P_1,P_2,P_3)|1\rangle\,|2\rangle\,|0\rangle_3 \tag{4.33}$$

mapping the space 2 into space 3 is densely defined, i.e., has a dense domain in Hilbert space 2.

We shall see that the domain in space 2 can be chosen the same whatever the state $|1\rangle$ is as long as $|1\rangle$ belongs to the dense set mentioned in space 1.

In fact we can, for the cyclically symmetric vertex

$$V(P_1,P_2,P_3) = {}_a\langle 0|\,_b\langle 0|\exp\{-P_1\cdot c^\dagger + P_2\cdot a + P_3\cdot b + [ab]_- - [bc^\dagger]_- - [c^\dagger a]_-\}\,|a\rangle\,|b\rangle\,|0\rangle_c \tag{4.34}$$

in which

$$[\alpha\beta]_- = \sum_{m,n=1}^{\infty} \frac{\alpha_n\beta_m}{\sqrt{nm}B(-n,m)},$$

show that it is defined for the following coherent state,

$$|a\rangle = \left|\left\{\frac{Z^n}{\sqrt{n}}P_1 + \sum_{r=1}^{\text{finite}} C_r\beta_n^{(r)}\right\}_{n=1,2,\ldots},P_1\right\rangle \tag{4.35}$$

[a state of type (4.21)], where

$$|Z|<1\wedge|1-Z|<1, \tag{4.36}$$

and a state $|b\rangle$ that is either a finite occupation number state or a coherent state with only finitely many modes excited. We shall prove it explicitly for the latter case.

It is in fact rather easy to check that with such states $|a\rangle$ and $|b\rangle$ the vector (4.34) becomes of finite norm. First it is noticed that both $|a\rangle$ and $|b\rangle$ are eigenstates of, respectively, $e^{p_2\cdot a}$ and $e^{p_3\cdot b}$ with finite eigenvalues because of, respectively, the exponential convergence z^n and the cutoff. Secondly

$$\exp([ab]_-)\,|a\rangle\,|b\rangle = \exp(\text{convergent } c \text{ number})\,|a\rangle\,|b\rangle, \tag{4.37}$$

because of the cutoff in the excitations in $|b\rangle$ and the exponential decrease from $|a\rangle$. Thirdly, also in the factors involving $c^\dagger$ can the a's and b's be replaced by c numbers and the overlap with the vacuum states $_a\langle 0|\,_b\langle 0|$ just results in a finite c number too. So the whole expression becomes a formal coherent state in space 3, i.e., for the form

$$\exp\left(\sum_{n=1}^{\infty}\gamma_n c^{(n)\dagger}\right)|0\rangle_c, \tag{4.38}$$

and the only thing to be checked is that the norm is finite, and that will be the case provided

$$\sum_{n=1}^{\infty}|\gamma_n|^2<\infty. \tag{4.39}$$

The contribution to $\sum_{n=1}^{\infty}\gamma_n c^{(n)\dagger}$ from $[bc^\dagger]_-$ is only different from zero for a finite number of γ_n, i.e., $n\le N$. It is thus not able to spoil (4.39). The main trick is that we have arranged it so that although both $-[c^\dagger a]$ and $-p_1\cdot c^\dagger$ give contributions that violate (4.39) the sum

$$\{-p_1\cdot c^\dagger - [c^\dagger a]_-\}$$

gives a contribution that obeys restriction (4.39). In fact the term $(z^n/\sqrt{n})p_1$ is accurately constructed to provide this cancellation; the terms $\sum_{r=1}^{\text{finite}} C_r\beta^{(r)}$ give rise through $-[c^\dagger a]_-$ only to a contribution obeying (4.39). This completes the proof that the cyclically symmetric vertex is densely defined.

5. THE PRODUCT $D(q)V(p)$; PHYSICAL STATES

We have shown that the ground state vertex $V(p)$ is in general nonexistent in the Hilbert space. However, consider the combination

$$D(q)V(p) = \int_0^1 dx\, x^{L_0-2}V(p) = \int_0^1 dx\, V(P_1x)x^{L_0-2}, \tag{5.1}$$

where

$$V(p_1x) = \exp\left(-\sum_{r=1}^{\infty}\frac{a_{r\mu}p^\mu}{\sqrt{r}}x^r\right)\exp\left(\sum_{r=1}^{\infty}\frac{a_{r\mu}p^\mu}{\sqrt{r}}x^{-r}\right). \tag{5.2}$$

We can use the combination of operators in Eq. (5.1), defined by the integral expression containing $V(p,x)$. The rule should be that the integration is done *after* the integrand has operated. Consider this operator acting now on a coherent state

$$\left\|\int_0^1 dx\exp\left(-\sum_r\frac{p\cdot a_r^\dagger x^r}{\sqrt{r}}\right)\exp\left(\sum_r\frac{p\cdot a_r\cdot x^{-r}}{\sqrt{r}}\right)x^{L_0(q)-2}\,|\alpha_n\rangle\right\|^2 \times\exp\left(\sum_{n=1}^{\infty}(x_1x_2)^n\left(|\alpha_n|^2-\frac{2\,\mathrm{Re}\,|\alpha_n\cdot p|}{\sqrt{n}}\right)\right) \times\exp\left(2p_\mu\sum_{n=1}^{\infty}\frac{\alpha_n^\mu}{\sqrt{n}}\right), \tag{5.3}$$

where $|p|^2 = p_0^2 + \underline{p}^2$ and $|\alpha_n\cdot p| = (\alpha_{n0}p_0 + \underline{\alpha}_n\cdot\underline{p})$. The last two exponentials are finite on a dense subset of $\mathcal{C}$ (namely provided $\sum\alpha_n/\sqrt{n}<\infty$).

Hence, on this dense set the norm is finite if $q^2<-1$ and $|p|^2<+2$ to avoid singularities at the lower and upper end points, respectively. This redefinition of the product $D(q)V(p)$ is thus defined on a dense subset of $\mathcal{C}$ for these momentum values.

The fact that the product of propagator times ground state vertex is better defined than the vertex alone has some interesting consequences on the definition of an off-mass shell physical state, if we require that the off-shell state remains normalizable within the Hilbert space. In general a physical state defined by its coupling to N ground state particles,

$$|\Phi\rangle = DV(k_1)DV(k_2)D\cdots DV(k_{N-1})\,|0\rangle, \tag{5.4}$$

satisfies the gauge condition[8]

$$w_{-r}|\Phi\rangle = (L_0 - L - r - 1)\,|\Phi\rangle = 0. \tag{5.5}$$

If we redefine a physical state without the final propagator, then it satisfies instead the conditions

$$|\Phi'\rangle = V(k_1)DV(k_2)D\cdots DV(k_{N-1})|0\rangle,$$
$$W_{-r}|\Phi'\rangle = (L_0 - L - r + r - 1)|\Phi'\rangle = 0. \tag{5.6}$$

In view of the nonexistence of $V(p)$ as a Hilbert space operator, we expect that the $|\phi'\rangle$ states defined by Eq. (5.6) be not normalizable, and it is amusing to confirm this by constructing such states within the irreducible representations of the gauge algebra.

In Ref. 9, it is described how to analyze the spectrum of states in terms of irreducible representations of the Virasoro algebra, with generators L_r $r = 0, \pm 1, \pm 2, \pm 3, \cdots$. For the present purpose we note that an exactly similar analysis for spacelike momentum can be made using irreducible representation of the Gliozzi algebra[10] with generators $L_0, L_{\pm 1}$. Each irreducible representation of the Gliozzi algebra then contains one and only one state (that having lowest L_0 eigenvalue) which satisfies $L_{-1}|\Phi''\rangle = 0$. All other states, obtained by raising with L_{+1} are the σ states.

Within each representation of the Gliozzi algebra it is straightforward to determine the unique state which is a physical state according to the definitions (5.5), (5.6), respectively (for $r = 1$). We may write for the former case

$$|\Phi\rangle = \sum_{n=0}^{\infty} \alpha_n (L_1)^n |\Phi'', c\rangle, \tag{5.7}$$

where

$$L_0|\Phi'', c\rangle = c|\Phi'', c\rangle, \quad L_{-1}|\Phi'', c\rangle = 0.$$

Using the commutation relation

$$[L_{-1}, L_1^n] = 2\sum_{q=0}^{n-1} L_1^q L_0 L_1^{n-q-1}, \tag{5.8}$$

one finds that the condition (5.5) for $r = 1$ gives, putting $\alpha_0 = 1$,

$$\alpha_n = (1/n!)[\Gamma(c + h - 1)\Gamma(2c)/\Gamma(c - 1)\Gamma(2c + n)]. \tag{5.9}$$

If

$$\| |\Phi'', c\rangle \| = 1,$$

One then finds, always for spacelike momentum with zero energy, that

$$\langle \Phi | \Phi \rangle = \sum_{n=0}^{\infty} \left(\prod_{q=1}^{n} \frac{(c + q - 2)^2}{q(2c + q - 1)} \right). \tag{5.10}$$

For large n the square bracket behaves as

$$\prod_{q=1}^{n} \frac{(c + q - 2)^2}{q(2c + q - 1)} \overset{n\to\infty}{\sim} \frac{1}{n^3} \tag{5.11}$$

so that the state $|\Phi\rangle$ has finite norm.

If we write, however,

$$|\Phi'\rangle = \sum_{n=0}^{\infty} \alpha_n' (L_1)^n |\Phi'', c\rangle, \tag{5.12}$$

satisfying (5.6), then one finds

$$|\Phi'\rangle = \sum_{n=0}^{\infty} \frac{1}{n!} \frac{\Gamma(c + n)\Gamma(2c)}{\Gamma(2c + n)\Gamma(c)} (L_1)^n |\Phi'', c\rangle, \tag{5.13}$$

and hence

$$\langle \Phi' | \Phi' \rangle = \sum_{n=0}^{\infty} \left(\prod_{q=1}^{n} \frac{(c + q - 1)^2}{q(2c + q - 1)} \right) \overset{n \text{ large}}{\sim} \sum^{\infty} \frac{1}{n} \tag{5.14}$$

giving a logarithmically divergent norm.

We deduce that unless c is a negative integer, whereupon the summations in Eqs. (5.7) and (5.12) may cut off, all states satisfying Eq. (5.6) are nor normalizable. The states satisfying Eq. (5.5), on the other hand, can be normalizable. Thus, the w_{-r} condition for an offshell physical state is more satisfactory in this respect than the W_{-r} definition. This is as expected from the better definition of $D(q)V(p)$ in Eq. (5.1) than that of $V(p)$ alone of Eq. (4.1).

6. SUMMARY AND DISCUSSION

We have studied the properties of the three fundamental operators in the operator formalism: $D(p)$, $V(p_1p_2p_3)$, and $\Omega(p)$. We have found that the vertex for ground state emission $V(p)$, is strictly speaking nonexistent within the Hilbert space. The twisting operator $\Omega(p)$ is also nonexistent when acting on any state with nonnull 4-momentum; its Hermitian conjugate is, however, densely defined. The propagator $D(p)$ is everywhere defined off-mass shell.

We have noted that the product $\Omega(q)V(p)$ and, more generally, the symmetric three-reggeon vertex $V(123)$ are densely defined. Also, the product $D(q)V(p)$ can be well defined and this was related to the recognition that states annihilated by $w_{-r} = (L_0 - L_{-r} - 1)$ were more suitable candidates for off-shell physical states than states annihilated by $W_{-r} = (L_0 - L_{-r} + r - 1)$.

The fact that not all operators and their image states can be represented within the Hilbert space is not very surprising because similar difficulties already occur in nonrelativistic quantum mechanics, where for example, the position operator acting on a square-integrable wavefunction can give a new function outside of the Hilbert space spanned by the set of all square-integrable functions.[5] In that case, extension to a larger space has proved useful.[11]

To conclude, we re-emphasize that the usual operator factorization (with matrix elements taken) of the generalized Veneziano model is well defined; it is only when we study the operators D, V, and Ω, in isolation, as Hilbert space operators, that the question of good definition arises. The matrix elements usually considered are scattering amplitudes, and for these we know the analytic structure and can continue analytically to any kinematical region. If we isolate operators or operator products, then there are no similar analyticity assumptions for these and, therefore, we have to understand their mathematical properties in order to use them correctly.

[1] Y. Nambu, *Symmetries and Quark Models*, edited by R. Chand (Gordon and Breach, New York, 1969), p. 269; S. Fubini, D. Gordon, and G. Veneziano, Phys. Lett. B **29**, 679 (1969).
[2] S. Fubini and G. Veneziano, Nuovo Cimento A **67**, 29 (1970).
[3] V. Alessandrini, D. Amati, M. Le Bellac, and D. Olive, Phys. Rep. **1**, 269 (1971).
[4] L. Caneschi, A. Schwimmer, and G. Veneziano, Phys. Lett. B **30**, 351 (1969).
[5] J. von Neumann, *Mathematical Foundations of Quantum Mechanics* (Princeton U. P., Princeton, N.J., 1955), Sec. II.5.
[6] M. A. Naimark, *Normed Rings*, translated by L. F. Boron (Noordhoff, New York, 1964), p. 74, Theorem II.
[7] J. von Neumann, Math. Ann. **102**, 49 (1929); L. V. Kantorovich and G. P. Akilov, *Functional Analysis in Normed Spaces* (Pergamon, New York, 1964), Chap. IX, Sec. 5.
[8] M. A. Virasoro, Phys. Rev. A **1**, 2933 (1970).
[9] P. H. Frampton and H. B. Nielsen, Nucl. PhysB **45**, 318 (1972).
[10] F. Gliozzi, Nuovo Cimento Lett. **22**, 846 (1969).
[11] I. M. Gel'fand and N. Ya. Vilenkin, *Generalized Functions* (Academic, New York, 1964), Vol. 4, Chap. I.

Electromagnetic scattering from two circular, coaxial disks

Lennart Marin

The Dikewood Corporation, Westwood Research Branch, Los Angeles, California 90024
(Received 6 March 1974)

The electromagnetic field scattered from two perfectly conducting, circular, coaxial disks is calculated when the incident field is a plane wave. Two quantities are of special interest: namely, the electric field at the center of the axis of the two disks and the integral of the electric field along the axis between the two disks. Analytical expressions for the low-frequency behavior of these two quantities are derived in the case where the distance between the two disks is large compared to the radii of the disks. For other frequencies and for separations of the two disks not too large these two quantities of interest are calculated numerically. Expressions for the scattered far field are also derived.

I. INTRODUCTION

One type of sensor for measuring the electric field is the parallel plate dipole. This sensor consists of two thin, perfectly conducting, parallel plates with some suitable electrical device for picking up the electric field or voltage between the two plates. In this paper we will analyze the characteristics of a symmetrical, parallel plate dipole consisting of two perfectly conducting circular disks and obtain quantitative information on the behavior of this dipole for a wide range of frequencies.

Electromagnetic scattering from one circular disk, two or more coaxial disks, or an annular ring has been treated by many investigators using a special integral equation technique.[1–6] A somewhat different integral-equation approach is discussed in Ref. 7. In all these references the electromagnetic scattering problem is reduced to the solution of a Fredholm integral equation of the second kind. The problem of acoustic scattering from any number of equal circular holes arbitrarily distributed in an infinitely large, rigid plate can be solved by a method of expansion in hypergeometrical polynomials.[8] An account of all methods available for solving the classical problem of acoustic and electromagnetic scattering from one disk can be found in Ref. 9.

However, no quantitative results seem to exist, except in some limiting cases, for the scattering from two or more coaxial disks. In this paper we will use the theory developed in Refs. 1—6 to obtain quantitative information about the scattered field for a wide range of frequencies. The approach we use is general and can be applied to the problem of scattering from any number of coaxial disks having different radii.

To begin with, the two disks are assumed to have the same radius, to be of zero thickness, and to be exposed to an incident plane wave. Specifically, we will calculate the electric field at the center of the dipole and the integral of the electric field between the two plates along the axis of the dipole. Expressions for the scattered far field are also derived.

In Sec. II we first scalarize the problem of two parallel, coaxial disks of different radii by expressing the scattered electromagnetic field in terms of the components of the Hertz potentials. From the solutions of these differential equations together with Green's theorem, some suitable transformation, and the edge conditions, we formulate, in Secs. III and IV, pairs of simultaneous Fredholm integral equations of the second kind. From a knowledge of the solution of these integral equations the scattered electromagnetic field can be calculated everywhere by performing simple integrations.

In Sec. V we express both the axial component of the scattered electric field on the axis of the two disks and the integral of this field along the axis between the two plates in terms of single integrals which involve the solution of the integral equations formulated in Secs. III and IV. These equations are solved iteratively for low frequencies and for large separations between the disks. These iterations are used to derive asymptotic expressions for both the electric field at the center of the dipole and the integral of the electric field along the axis between the two plates. For other frequencies and separations not too large compared with the disks' radii these two quantities are calculated numerically and graphed as a function of frequency for different sizes of the sensor and angles of the incident field.

II. FORMULATION OF THE BOUNDARY CONDITIONS IN TERMS OF THE HERTZ POTENTIALS

The electromagnetic boundary conditions on two perfectly conducting, circular, coaxial disks will be invoked to derive and subsequently solve a set of ordinary differential equations for certain components of the Hertz potentials for the scattered field. The two disks taken to be infinitely thin are separated by a distance $2d$ and their radii are denoted by a_+ and a_-, respectively (see Fig. 1). In cylindrical coordinates, the locations $\Sigma_\pm$ of the two disks can be expressed mathematically in the following

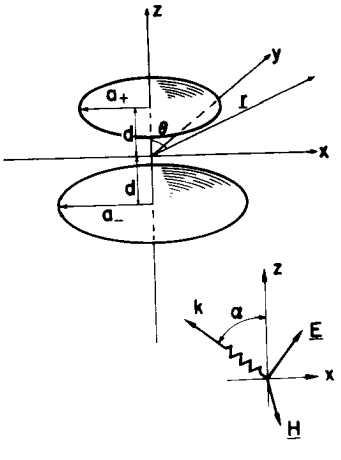


FIG. 1. The geometry of the problem.

way: $\Sigma_\pm=\{(\rho,\phi,z):0\leq\rho\leq a_\pm,\ 0\leq\phi\leq 2\pi,\ z=\pm d\}$. We will calculate the scattered field when the two disks are illuminated by a plane, monochromatic wave. The harmonic-time dependence $\exp(-i\omega t)$ will be understood and suppressed throughout the paper.

It is convenient to assume that the incident field can be split into two separate fields: one with the magnetic field parallel to the disks and the other with the electric field parallel to the disks. The boundary-value problems for these two different incident fields can be treated separately and the solution for an arbitrary incident field may then be obtained by superposition. We will first treat the case where the incident magnetic field is parallel to the disks.

A. Incident magnetic field parallel to the disks

Let the incident field be a plane wave such that

$$\begin{aligned}\mathbf{E}^{\mathrm{inc}}&=E_0'(\hat{x}\cos\alpha+\hat{z}\sin\alpha)\exp(ikz\cos\alpha-ikx\sin\alpha)\\ \mathbf{H}^{\mathrm{inc}}&=E_0'Z_0^{-1}\hat{y}\exp(ikz\cos\alpha-ikx\sin\alpha)\end{aligned}\tag{1}$$

where $k=\omega/c$, c being the vacuum speed of light and Z_0 the free-space wave impedance. The angle of incidence α is defined in Fig. 1. The scattered field can be determined from the magnetic and electric Hertz potentials, $\pi^{(m)}$ and $\pi^{(e)}$, as follows:

$$\begin{aligned}\mathbf{E}^{\mathrm{sc}}&=i\omega\nabla\times\pi^{(m)}+\nabla\times\nabla\times\pi^{(e)},\\ \mathbf{B}^{\mathrm{sc}}&=\nabla\times\nabla\times\pi^{(m)}-ikc^{-1}\nabla\times\pi^{(e)},\end{aligned}$$

where $\pi^{(m)}$ and $\pi^{(e)}$ both satisfy the Helmholtz equation $\nabla^2\pi+k^2\pi=0$.

Let us now choose some suitable series representations for the components of the Hertz potentials. If these representations satisfy all the required conditions, i.e., the Helmholtz equation, the boundary conditions on $\Sigma_\pm$, the edge conditions at the boundary of $\Sigma_\pm$, and the radiation condition at infinity, then, in view of the uniqueness of the solution of the electromagnetic scattering problem, the assumed representations are justified. Assume that the Hertz vectors for the scattered field can be represented in the following way:

$$\begin{aligned}\pi_\rho^{(m)}&=E_0'(i\omega)^{-1}\sum_{m=1}^{\infty}\xi_{m-1}(\rho,z)\sin m\phi,\\ \pi_\phi^{(m)}&=E_0'(i\omega)^{-1}\sum_{m=1}^{\infty}\xi_{m-1}(\rho,z)\cos m\phi,\\ \pi_z^{(m)}&=E_0'(i\omega)^{-1}\sum_{m=1}^{\infty}\eta_m(\rho,z)\sin m\phi,\quad \pi^{(e)}=E_0'\psi(\rho,z)\hat{z},\end{aligned}\tag{3}$$

where ξ_m and η_m satisfy the differential equation

$$\left\{\frac{\partial^2}{\partial\rho^2}+\frac{1}{\rho}\frac{\partial}{\partial\rho}-\frac{m^2}{\rho^2}+\frac{\partial^2}{\partial z^2}+k^2\right\}\begin{matrix}\xi_m\\ \eta_m\end{matrix}=0\tag{4}$$

and ψ satisfies (4) with $m=0$.

To determine the Hertz potentials for the scattered field on $\Sigma_\pm$, we will use the boundary conditions on the two disks,

$$E_\rho=E_\rho^{\mathrm{inc}}+E_\rho^{\mathrm{sc}}=0,\quad E_\phi=E_\phi^{\mathrm{inc}}+E_\phi^{\mathrm{sc}}=0,\quad \mathbf{r}\in\Sigma_\pm.\tag{5}$$

Expressing the incident field (1) in cylindrical coordinates, making use of the expressions (2)—(4) for the scattered field, and invoking the boundary conditions (5), we derive three ordinary uncoupled differential equations for $\eta_m(\rho,z)$, $\xi_m(\rho,z)$ which are valid on $\Sigma_\pm$. A solution of these differential equations is given by

$$\begin{aligned}\eta_m^\pm(\rho)&=\eta_m(\rho,\pm d)=B_m^\pm\rho^m+A_m^\pm J_m(k\rho\sin\alpha),\\ \alpha_m^\pm(\rho)&=\frac{\partial\xi_m}{\partial z}(\rho,\pm d)\\ &=(m+1)B_{m+1}^\pm\rho^m+k\sin\alpha A_{m+1}^\pm J_m(k\rho\sin\alpha),\\ \frac{\partial\psi}{\partial z}(\rho,\pm d)&=B_0^\pm-A_0^\pm J_0(k\rho\sin\alpha),\end{aligned}\tag{6}$$

where $0\leq\rho\leq a_\pm$, $A_m^\pm=\epsilon_m i^{-m+1}k^{-1}\cot\alpha\exp(\pm ikd\cos\alpha)$, $\epsilon_m=2$, $m\geq 1$, and $\epsilon_0=1$. Moreover, $B_m^\pm$ are unknown constants of integration to be determined later from the edge conditions at $\rho=a_\pm$, and $J_m(x)$ are Bessel functions of the first kind.

Having derived the expression (6) for the Hertz potentials of the scattered field when the incident magnetic field is parallel to the disks, we now go on to derive a corresponding expression when the incident electric field is parallel to the disks.

B. Incident electric field parallel to the disks

Let the incident field be a plane wave such that

$$\mathbf{E}^{\mathrm{inc}}=E_0'\hat{y}\exp(ikz\cos\alpha-ikx\sin\alpha).\tag{7}$$

In this case, the scattered field can be obtained from the magnetic Hertz potential alone,

$$\mathbf{E}^{\mathrm{sc}}=i\omega\nabla\times\pi^{(m)},\quad \mathbf{B}^{\mathrm{sc}}=\nabla\times\nabla\times\pi^{(m)}\tag{8}$$

and $\pi^{(m)}$ can be expanded in the following Fourier series:

$$\begin{aligned}\pi_\rho^{(m)}&=E_0''(i\omega)^{-1}\sum_{m=1}^{\infty}\chi_{m-1}(\rho,z)\cos m\phi,\\ \pi_\phi^{(m)}&=-E_0''(i\omega)^{-1}\sum_{m=1}^{\infty}\chi_{m-1}(\rho,z)\sin m\phi,\\ \pi_z^{(m)}&=E_0''(i\omega)^{-1}\sum_{m=0}^{\infty}\tau_m(\rho,z)\cos m\phi.\end{aligned}\tag{9}$$

The boundary conditions (5) on $\Sigma_\pm$, an expansion of the incident field (7) in cylindrical coordinates, and the expansions (8)—(9) of the scattered field enable us to derive the following expressions:

$$\begin{aligned}\tau_m^\pm(\rho)&=\tau_m(\rho,\pm d)=D_m^\pm\rho^m+C_m^\pm J_m(k\rho\sin\alpha),\\ \beta_m^\pm(\rho)&=\frac{\partial\chi_m}{\partial z}(\rho,\pm d)=(m+1)D_{m+1}^\pm\rho^m,\end{aligned}\tag{10}$$

where $0\leq\rho\leq a_\pm$, $C_m^\pm=\epsilon_m i^{-m+1}(k\sin\alpha)^{-1}\exp(\pm ikd\cos\alpha)$, and $D_m^\pm$ are unknown constants to be determined later from the edge conditions at $\rho=a_\pm$.

To sum up this section, the boundary conditions on the disks combined with a suitable choice of the form of the Hertz potential for the scattered field [(3) and (9)] have enabled us to derive the expressions (6) and (10) for the Hertz potentials on the disks. From these expressions and the scattered field being continuous off the disks and satisfying the Sommerfeld radiation condition, it is clear that η_m, τ_m, $\partial\xi_m/\partial z$, and $\partial\chi_m/\partial z$ are continuous functions of ρ and z everywhere (including $\Sigma_\pm$).

III. REDUCTION OF THE SCATTERING PROBLEM TO INTEGRAL EQUATIONS

In this section, we will reduce the electromagnetic boundary-value problem of scattering from two circular disks to the solution of Fredholm integral equations of the second kind. In deriving these integral equations we will make use of the boundary conditions derived in the previous section for the Hertz potentials. Once the solution of these integral equations are found, the scattered field can be obtained from a simple integration. We will start with the derivation of integral equations for $\eta_m(\rho, z)$ and $\tau_m(\rho, z)$, and then for $\xi_m(\rho, z)$, $\chi_m(\rho, z)$, and $\psi(\rho, z)$.

A. Integral equations for $\eta_m(\rho,z)$ and $\tau_m(\rho,z)$

By applying the Green's theorem to the function $\eta_m(\rho, z)\exp(im\phi)$ in the region outside the disks and observing that this function is a continuous function of z on $\Sigma_\pm$ and that it satisfies the radiation condition at infinity we arrive, after some algebraic manipulations, at the following integral expression:

$$\eta_m(\rho, z) = \tfrac{1}{2}\int_0^{a_+}\int_0^\infty pw^{-1}J_m(p\rho)J_m(p\rho')\exp(-w|z-d|)\rho' y^+(\rho')d\rho' dp + \tfrac{1}{2}\int_0^{a_-}\int_0^\infty pw^{-1}J_m(p\rho)J_m(p\rho')\exp(-w|z+d|)\times\rho' y^-(\rho')d\rho' dp, \tag{11}$$

where $w=(p^2-k^2)^{1/2}$, $p>k$, and $w=-i(k^2-p^2)^{1/2}$, $p<k$. The path of integration in (11) is along the real axis in the complex p plane and with a downward indentation at $p=k$. Moreover, $y_m^+(y_m^-)$ is the discontinuity at $z=d$ $(z=-d)$ of $\partial\eta_m/\partial z$, i.e.,

$$y_m^+(\rho)=\lim_{\epsilon\to 0+}\frac{\partial\eta_m}{\partial z}(\rho, d-\epsilon)-\frac{\partial\eta_m}{\partial z}(\rho, d+\epsilon)\ . \tag{12}$$

Substituting into (11) the expression (6) for $\eta_m(\rho, z)$ on $\Sigma_\pm$, we arrive at the following set of coupled integral equations of the first kind for $y_m^\pm(\rho)$:

$$\begin{aligned}&\int_0^{a_+}K_m(\rho,\rho',0)\rho' y_m^+(\rho')d\rho'+\int_0^{a_-}K_m(\rho,\rho',2d)\rho' y_m^-(\rho')d\rho' = \eta_m^+(\rho),\quad 0\le\rho\le a_+,\\ &\int_0^{a_+}K_m(\rho,\rho',2d)\rho' y_m^+(\rho')d\rho'+\int_0^{a_-}K_m(\rho,\rho',0)\rho' y_m^-(\rho')d\rho' = \eta_m^-(\rho),\quad 0\le\rho\le a_-,\end{aligned} \tag{13}$$

where

$$K_m(\rho,\rho',z)=\tfrac{1}{2}\int_0^\infty pw^{-1}J_m(p\rho)J_m(p\rho')\exp(-w|z|)dp$$

and the functions $\eta_m^\pm(\rho)$ are defined by (6).

The kernel, $K_m(\rho,\rho',0)$, in (13) has a logarithmic singularity at $\rho'=\rho$. A general procedure for transforming a certain class of one-dimensional integral equations into a set of Fredholm integral equations of the second kind has been developed in Ref. 4. An extension of this method shows that we can obtain the solution of (13) from the integral expression

$$y_m^\pm(\rho)=-\frac{2}{\pi}\rho^{m-1}\frac{d}{d\rho}\int_\rho^{a_\pm}(u^2-\rho^2)^{-1/2}\times\cosh[k(u^2-\rho^2)^{1/2}]u^{1-m}Y_m^\pm(u)\,du \tag{14}$$

provided that the functions $Y_m^\pm(u)$ satisfy the following set of Fredholm integral equations of the second kind:

$$\begin{aligned}&Y_m^+(u)+\int_0^{a_+}L_m(u,v)Y_m^+(v)dv+\int_0^{a_-}M_m(u,v)Y_m^-(v)dv=F_m^+(u),\quad 0\le u\le a_+,\\ &Y_m^-(u)+\int_0^{a_+}M_m(u,v)Y_m^+(v)dv+\int_0^{a_-}L_m(u,v)Y_m^-(v)dv=F_m^-(u),\quad 0\le u\le a_-.\end{aligned} \tag{15}$$

The kernels $L_m(u,v)$, $M_m(u,v)$ and the right-hand side $F_m^\pm(u)$ in (15) are given by the integral expressions $L_m(u,v)=L(u,v;m,m-\frac{1}{2})$, $M_m(u,v)=M(u,v;m,m-\frac{1}{2})$, where

$$L(u,v;m,\nu)=(uv)^{1/2}\int_0^\infty w^{1-2m}[(k^2+w^2)^m-w^{2m}]J_\nu(wu)J_\nu(wv)dw + i(uv)^{1/2}\int_0^k w^{1-2m}(k^2-w^2)^m I_\nu(wu)I_\nu(wv)dw, \tag{16}$$

$$M(u,v;m,\nu)=(uv)^{1/2}\int_0^\infty p^{2m+1}w^{-2m}J_\nu(wu)J_\nu(wv)\exp(-2wd)dp,$$

and

$$F_m^\pm(u)=2u^{-m}\frac{d}{du}\int_0^u(u^2-\rho^2)^{-1/2}\cosh[k(u^2-\rho^2)^{1/2}]\times\rho^{m+1}\eta_m^\pm(\rho)d\rho. \tag{17}$$

Thus, the functions $\eta_m(\rho, z)$ can readily be calculated from the integral expressions (11) and (14) once we have obtained the solution of (15). In this sense we have reduced the problem of finding $\eta_m(\rho, z)$ to solving a Fredholm integral equation of the second kind. An alternative form for the kernels in (15) is derived in Ref. 4.

We will now go on to show how the methods described above can be used to determine $\tau_m(\rho, z)$. Let $t_m^\pm(\rho)$ denote the discontinuity of $\partial\tau_m/\partial z$ on $\Sigma_\pm$ [cf. (12)]. A procedure analagous to the one used to determine $\eta_m(\rho, z)$ shows that one can obtain $t_m^\pm(\rho)$ from (14)—(17) by making the following substitutions in those equations: $y_m^\pm(\rho)\to t_m^\pm(\rho)$, $Y_m^\pm(u)\to T_m^\pm(u)$, and $\eta_m^\pm(\rho)\to\tau_m^\pm(\rho)$, where $\tau_m^\pm(\rho)$ is given by (10). The functions $\tau_m(\rho, z)$ can then be determined from (11) by making the substitutions $\eta_m(\rho, z)\to\tau_m(\rho, z)$ and $y_m^\pm(\rho)\to t_m^\pm(\rho)$.

Having derived two sets of Fredholm integral equations of the second kind for $\eta_m(\rho, z)$ and $\tau_m(\rho, z)$ we will now derive similar integral equations, the solutions of which give $\xi_m(\rho, z)$, $\chi_m(\rho, z)$, and $\psi(\rho, z)$.

B. Integral equations for $\xi_m(\rho,z)$, $\chi_m(\rho,z)$, and $\psi(\rho,z)$

In deriving integral equations for $\xi_m(\rho, z)$, $\chi_m(\rho, z)$, and $\psi(\rho, z)$ we first note that $\partial\xi_m/\partial z$, $\partial\chi_m/\partial z$, and $\partial\psi/\partial z$ are continuous functions for all values of z. Next, we let $x_m^\pm(\rho)$ denote the discontinuity of $\xi_m(\rho, z)$ on $\Sigma_\pm$, i.e.,

$$x_m^\pm(\rho)=\lim_{\epsilon\to 0}[\xi_m(\rho, d-\epsilon)-\xi_m(\rho, d+\epsilon)]. \tag{18}$$

From the Green's theorem we can then derive the following set of differential—integral equations for $x_m^\pm(\rho)$:

$$\begin{aligned}&\mathcal{L}\int_0^{a_+}K_m(\rho,\rho',0)\rho' x_m^+(\rho')d\rho'+\mathcal{L}\int_0^{a_-}K_m(\rho,\rho',2d)\rho' x_m^-(\rho')d\rho' = -\alpha_m^+(\rho),\quad 0\le\rho\le a_+,\\ &\mathcal{L}\int_0^{a_+}K_m(\rho,\rho',2d)\rho' x_m^+(\rho')d\rho'+\mathcal{L}\int_0^{a_-}K_m(\rho,\rho',0)\rho' x_m^-(\rho')d\rho' = -\alpha_m^-(\rho),\quad 0\le\rho\le a_-,\end{aligned} \tag{19}$$

where the differential operator $\mathcal{L}$ is given by $\mathcal{L}=d^2/d\rho^2$

$+\rho^{-1}d/d\rho - m^2\rho^{-2} + k^2$ and the functions $K_m(\rho, \rho', z)$ and $\alpha_m^{\pm}(\rho)$ are defined by (13) and (7), respectively. A transformation similar to the one used in transforming (13) into the set of equations (14)—(15) enables us to transform (19) into the following set of integral equations[4]:

$$x_m^{\pm}(\rho) = -\frac{2}{\pi}\rho^m \int_{\rho}^{a_{\pm}} (u^2-\rho^2)^{-1/2}\cosh[k(u^2-\rho^2)^{1/2}] \times u^{-m}X_m^{\pm}(u)du, \tag{20}$$

where the functions $X_m^{\pm}(u)$ satisfy the set of Fredholm integral equations of the second kind:

$$X_m^{+}(u) + \int_0^{a_+} U_m(u,v)X_m^{+}(v)dv + \int_0^{a_-} P_m(u,v)X_m^{-}(v)dv = G_m^{+}(u), \quad 0 \le u \le a_+,$$

$$X_m^{-}(u) + \int_0^{a_+} P_m(u,v)X_m^{+}(v)dv + \int_0^{a_-} U_m(u,v)X_m^{-}(v)dv = G_m^{-}(u), \quad 0 \le u \le a_-. \tag{21}$$

The kernels in (21) are given by the expressions, cf. (16), $U_m(u,v) = L(u,v;m,m+\frac{1}{2})$ and $P_m(u,v) = M(u,v;m,m+\frac{1}{2})$, and the right-hand side has the following integral representation:

$$G_m^{\pm}(u) = u^{-m-1}\frac{d}{du}\int_0^u (u^2-\rho^2)^{-1/2}\cosh[k(u^2-\rho^2)^{1/2}]g_m^{\pm}(\rho)d\rho,$$

where

$$g_m^{\pm}(\rho) = B_{m+1}^{\pm}\rho^{2m+3} + 2A_{m+1}^{\pm}\rho^{m+2}J_{m+1}(k\rho\sin\alpha). \tag{22}$$

Similarly, to determine $\chi_m(\rho, z)$ we denote the discontinuity of this function on $\Sigma_{\pm}$ by $w_m^{\pm}(\rho)$ [cf. (18)] and $w_m^{\pm}(\rho)$ can be obtained from (20)—(21) by making the following substitutions into those equations: $x_m^{\pm}(\rho) \to w_m^{\pm}(\rho)$, $X_m^{\pm}(u) \to W_m^{\pm}(u)$, and $g_m^{\pm}(\rho) \to h_m^{\pm}(\rho)$, where $h_m^{\pm}(\rho) = D_{m+1}^{\pm}\rho^{2m+3}$. After evaluating the expression (20) with the solution of (21) the functions $\xi_m(\rho, z)$ and $\chi_m(\rho, z)$ $(m \ge 0)$ can be calculated from (11) if we make the following substitutions in (11): $(\partial/\partial z)\eta_m(\rho, z) \to \xi_m(\rho, z)$, $y_m^{\pm}(\rho) \to x_m^{\pm}(\rho)$ and $(\partial/\partial z)\eta_m(\rho, z) \to \chi_m(\rho, z)$, $y_m^{\pm}(\rho) \to w_m^{\pm}(\rho)$, respectively.

It now remains to determine $\psi(\rho, z)$. To this end we note that $\partial\psi/\partial z$ is continuous on $\Sigma_{\pm}$ and we denote the discontinuity of $\psi(\rho, z)$ on $\Sigma_{\pm}$ by $z_{\pm}(\rho)$ [cf. (18)]. The functions $z_{\pm}(\rho)$ can be obtained from (20) by putting $m = 0$ in (20)—(22) and also substitute into those equations the following expressions: $x_0^{\pm}(\rho) \to z_{\pm}(\rho)$, $X_0^{\pm}(u) \to Z_{\pm}(u)$, and $g_0^{\pm}(\rho) \to B_0^{\pm}\rho^3 - 2(k\sin\alpha)^{-1}A_0^{\pm}\rho^2 J_1(k\rho\sin\alpha)$. Thus, we have reduced the problem of finding $\psi(\rho, z)$ to the solution of the integral equation (21). It should also be pointed out here that for $m = 0$ we have the following explicit representations of the kernels and the right-hand side in (21):

$$U_0(u,v) = N_1(u-v) - N_1(u+v),$$
$$N_1(u) = (i\pi u)^{-1}\sinh(ku),$$
$$P_0(u,v) = N_2(u-v) - N_2(u+v), \tag{23}$$
$$N_2(u) = \pi^{-1}\exp(2ikd)[2d\cosh(ku) + iu\sinh(ku)]/(u^2+4d^2),$$
$$G_0^{\pm}(u) = 2k^{-1}B_0^{\pm}\sinh(ku) - 2(k\cos\alpha)^{-1}A_0^{\pm}\sinh(ku\cos\alpha).$$

We have now completed our derivation of Fredholm integral equations of the second kind the solutions of which give the Hertz potentials of the scattered field. As we have seen in Sec. II, it still remains to determine some unknown constants of integration. These constants can be determined from the edge conditions, and in the next section we will formulate these conditions mathematically in terms of constraints on the solution of the integral equations (15) and (21).

IV. DETERMINATION OF THE UNKNOWN CONSTANTS FROM THE EDGE CONDITIONS

The surface current densities $i^{\pm}(\rho, \phi)$ on $\Sigma_{\pm}$ can be determined from the discontinuity of the tangential component of the scattered magnetic field since the incident electromagnetic field is continuous everywhere. In the case where the incident magnetic field is parallel to the disks, the expansions of $i_{\rho}^{\pm}(\rho, \phi)$ in a Fourier series,

$$i_{\rho}^{\pm}(\rho, \phi) = \sum_{m=0}^{\infty} i_{\rho m}^{\pm}(\rho)\cos m\phi, \tag{24}$$

combined with Eqs. (2), (3), (12), (14), and (20) give

$$i_{\rho 0}^{\pm} = (kE_0/i\pi Z_0)/[a_{\pm}(a_{\pm}-\rho)/2]^{1/2}Z_{\pm}(a_{\pm}) + O[(a_{\pm}-\rho)^{1/2}],$$

$$i_{\rho m}^{\pm} = (mE_0/i\omega\mu_0\pi a_{\pm})/[a_{\pm}(a_{\pm}-\rho)/2]^{1/2}[Y_m^{\pm}(a_{\pm}) - X_{m-1}^{\pm}(a_{\pm})] + O[(a_{\pm}-\rho)^{1/2}], \quad m \ge 1, \tag{25}$$

as $\rho \to a_{\pm}$. Thus, the edge conditions demand that

$$Z_{\pm}(a_{\pm}) = 0, \quad Y_m^{\pm}(a_{\pm}) = X_{m-1}^{\pm}(a_{\pm}), \quad m \ge 1, \tag{26}$$

and from (26) we can determine the constants of integration $B_m^{\pm}$, $m \ge 0$, in (6).

Similarly, in the case where the incident field is parallel to the disks, the edge conditions imply that

$$T_m^{\pm}(a_{\pm}) = W_{m-1}^{\pm}(a_{\pm}), \quad m \ge 1. \tag{27}$$

Again, (27) gives us the relationship that is needed to determine the constants $D_m^{\pm}$, $m \ge 0$, in (10). We wish to point out in passing that the edge condition for $i_{\phi}^{\pm}(\rho, \phi)$ is automatically satisfied by the conditions (26) and (27).

We have now concluded the reduction of the boundary-value problem of electromagnetic scattering from two circular, coaxial disks to the mathematical problem of solving two sets of Fredholm integral equations of the second kind. In the next section we will use the solution of these integral equations to calculate the fields along the axis of the two disks.

V. THE ELECTRIC FIELD ON THE AXIS OF THE DISKS

The electric field on the axis of the two disks is given by, cf. (1)—(3),

$$E_z(z) = E_z^{\mathrm{inc}}(z) + E_0' e(z), \quad e(z) = \frac{\partial^2}{\partial z^2}\psi(0, z) + k^2\psi(0, z), \tag{28}$$

where

$$\psi(\rho, z) = \psi^{+}(\rho, z) + \psi^{-}(\rho, z)$$

and

$$\psi^{\pm}(\rho, z) = \frac{\partial}{\partial z}\int_{\Sigma_{\pm}} G(\rho, \rho', \phi, z \mp d)\rho' z_{\pm}(\rho')d\rho' d\phi. \tag{29}$$

Some manipulations on (29) together with the Sonine's second integral[10] give

$$\psi^{\pm}(\rho, z) = (2\pi i)^{-1}\,\mathrm{sgn}(\pm d - z) \times \int_0^{a_{\pm}} [L^{\pm}(\rho, z, d, u) - L^{\pm}(\rho, z, d, -u)]Z_{\pm}(u)du, \tag{30}$$

where

$$L^{\pm}(\rho, z, d, u) = \exp\{ik[\rho^2 + (|z \mp d| - iu)^2]^{1/2}\} \times [\rho^2 + (|z \mp d| - iu)^2]^{-1/2},$$

$\mathrm{sgn}(x) = 1(-1)$ for $x > 0$ (< 0), and the real part of the square root is positive. From the integral equation (21) it is easy to see that $Z_{\pm}(u) = O(u)$ as $u \to 0$, so that the integral in (30) exists for all values of ρ and z. We also note that the expression (30) satisfies the ϕ-independent wave equation off the disks. The electric field on the axis of the disks can be obtained by differentiating (30). In order to obtain expressions more suitable for numerical calculations, we proceed as follows: suppose $Z'_{\pm}(u)$ and $Z''_{\pm}(u)$ are differentiable twice. Integrating (30) by parts and keeping in mind that $Z_{\pm}(a_{\pm}) = Z_{\pm}(0) = 0$, we get

$$\frac{d^2\psi}{dz^2}(\rho, z) = (2\pi i)^{-1}\,\mathrm{sgn}(\pm d - z) \times [L^{\pm}(\rho, z, d, a_{\pm}) - L^{\pm}(\rho, z, d, -a_{\pm})]Z'_{\pm}(a_{\pm}) - (2\pi i)^{-1}\,\mathrm{sgn}(\pm d - z) \times \int_0^{a_{\pm}} [L^{\pm}(\rho, z, d, u) - L^{\pm}(\rho, z, d, -u)]Z''_{\pm}(u)\,du, \tag{31}$$

where the prime denotes differentiation with respect to u. The functions $Z'_{\pm}(u)$ and $Z''_{\pm}(u)$ satisfy the integral equations

$$Z'_{+}(u) + \int_0^{a_+} [N_1(u-v) + N_1(u+v)]Z'_{+}(v)\,dv + \int_0^{a_-} [N_2(u-v) + N_2(u+v)]Z'_{-}(v)\,dv = G_0^{+\prime}(u), \quad 0 \le u \le a_+, \tag{32}$$

$$Z'_{-}(u) + \int_0^{a_+} [N_2(u-v) + N_2(u+v)]Z'_{+}(v)\,dv + \int_0^{a_-} [N_1(u-v) + N_1(u+v)]Z'_{-}(v)\,dv = G_0^{-\prime}(u), \quad 0 \le u \le a_-,$$

and

$$Z''_{+}(u) + \int_0^{a_+} [N_1(u-v) - N_1(u+v)]Z''_{+}(v)\,dv + \int_0^{a_-} [N_2(u-v) - N_2(u+v)]Z''_{-}(v)\,dv = Q^{+}(u), \quad 0 \le u \le a_+, \tag{33}$$

$$Z''_{-}(u) + \int_0^{a_+} [N_2(u-v) - N_2(u+v)]Z''_{+}(v)\,dv + \int_0^{a_-} [N_1(u-v) - N_1(u+v)]Z''_{-}(v)\,dv = Q^{-}(u), \quad 0 \le u \le a_-,$$

where

$$Q^{\pm}(u) = G_0^{\pm\prime\prime}(u) + [N_1(u + a_{\pm}) - N_1(u - a_{\pm})]Z'_{\pm}(a_{\pm}) + [N_2(u + a_{\mp}) - N_2(u - a_{\mp})]Z'_{\mp}(a_{\mp})$$

and $N_1(u)$, $N_2(u)$, and $G_0^{\pm}(u)$ are defined by (23). We note that the only difference between the set of integral equations for $Z_{\pm}(u)$ and that for $Z''_{\pm}(u)$ is in the right-hand side.

The scattered electric field on the axis of the disks can be cast into the following form for $-d < z < d$,

$$e(z) = \pi^{-1}\exp[ik(d-z)]\int_0^{a_+} N^{+}(z,u)[k^2 Z_{+}(u) - Z''_{+}(u)]\,du + \pi^{-1}\exp[ik(d-z)]N^{+}(z,a_{+})Z'_{+}(a_{+}) - \pi^{-1}\exp[ik(d+z)]\int_0^{a_-} N^{-}(z,u)[k^2 Z_{-}(u) - Z''_{-}(u)]\,du - \pi^{-1}\exp[ik(d+z)]N^{-}(z,a_{-})Z'_{-}(a_{-}), \tag{34}$$

where

$$N^{\pm}(z,u) = [u\cosh(ku) + i(z \pm d)\sinh(ku)]/[u^2 + (z \mp d)^2].$$

Another quantity of interest is $v(d)$, given by

$$v(d) \equiv E_0^{-1}\int_{-d}^{d} E_z(z)\,dz = 2k^{-1}\tan\alpha\,\sin(kd\sin\alpha) + [Z'_{+}(0) - Z'_{-}(0)]/2 - \int_0^{a_+}[N_1(u) + N_2(u)]Z'_{+}(u)\,du + \int_0^{a_-}[N_1(u) + N_2(u)]Z'_{-}(u)\,du + (2\pi i)^{-1}k^2\int_0^{a_+} K(u)Z_{+}(u)\,du - (2\pi i)^{-1}k^2\int_0^{a_-} K(u)Z_{-}(u)\,du, \tag{35}$$

where

$$K(u) = E_1(ku - 2ikd) - E_1(ku) + E_1(-ku - i0) - E_1(-ku - 2ikd)$$

and $E_1(\zeta)$ is the exponential integral.[11] Equations (34) and (35) are suitable for numerical treatment.

We will now discuss the solutions of the integral equations for $Z_{\pm}(u)$, $Z'_{\pm}(u)$, and $Z''_{\pm}(u)$. The kernels of these integral equations are small when the normalized wavenumber of the incident wave is small and the distance between the two plates is sufficiently large, i.e., when $\beta = ka \ll 1$ and $\eta (= kd)$ is of order unity. In this case an iterative solution of the integral equations can be obtained. For the special but important case where $a_+ = a_- = a$ this iterative solution gives

$$E_z(0) = E_0[\sin\alpha + 2\sin 2\alpha\,\sin(\eta\cos\alpha) \times e^{i\eta}\beta^5(\eta^4 + i\eta^3 - \eta^2 - 2)/(45\pi\eta^6)] + O(\beta^7) \tag{36}$$

$$v(d) = d\eta^{-1}\sin(2\alpha)\sin(\eta\cos\alpha) \times [\cos^{-2}\alpha - \beta^2/6 + \beta^4(11 + 6\cos^2\alpha)/360] + O(\beta^6).$$

For other frequencies and when the separation between the two plates is not large the integral equations can not be solved analytically. In this case we solved the integral equations for $Z_{\pm}(u)$, $Z'_{\pm}(u)$, and $Z''_{\pm}(u)$ numerically for $a_+ = a_- = a$, $0 < ka < 10$, $d/a = 0.1$, 0.05, 0.02, and $\alpha = 18°$, $36°$, $54°$, $72°$. From these numerical solutions we then, by simple integrations, calculated $e(0)$ and $v(d)$ from (34) and (35). The results of these calculations are shown in Figs. 2 and 3. In these figures we have used the normalized quantities $e' = 1 + e(0)/\sin\theta$ and $v' = v(d)/(2d\sin\alpha)$. Figure 2 shows e' as a function of ka with d/a as a parameter for different values of α. It was found that the curves for v' are very similar to those for e'. Therefore, the difference quantity $e' - v'$ is plotted in Fig. 3 as a function of ka with $d/a = 0.1$, 0.05. For $d/a = 0.02$ the difference between e' and v' is negligible, and hence the corresponding curves are omitted in the

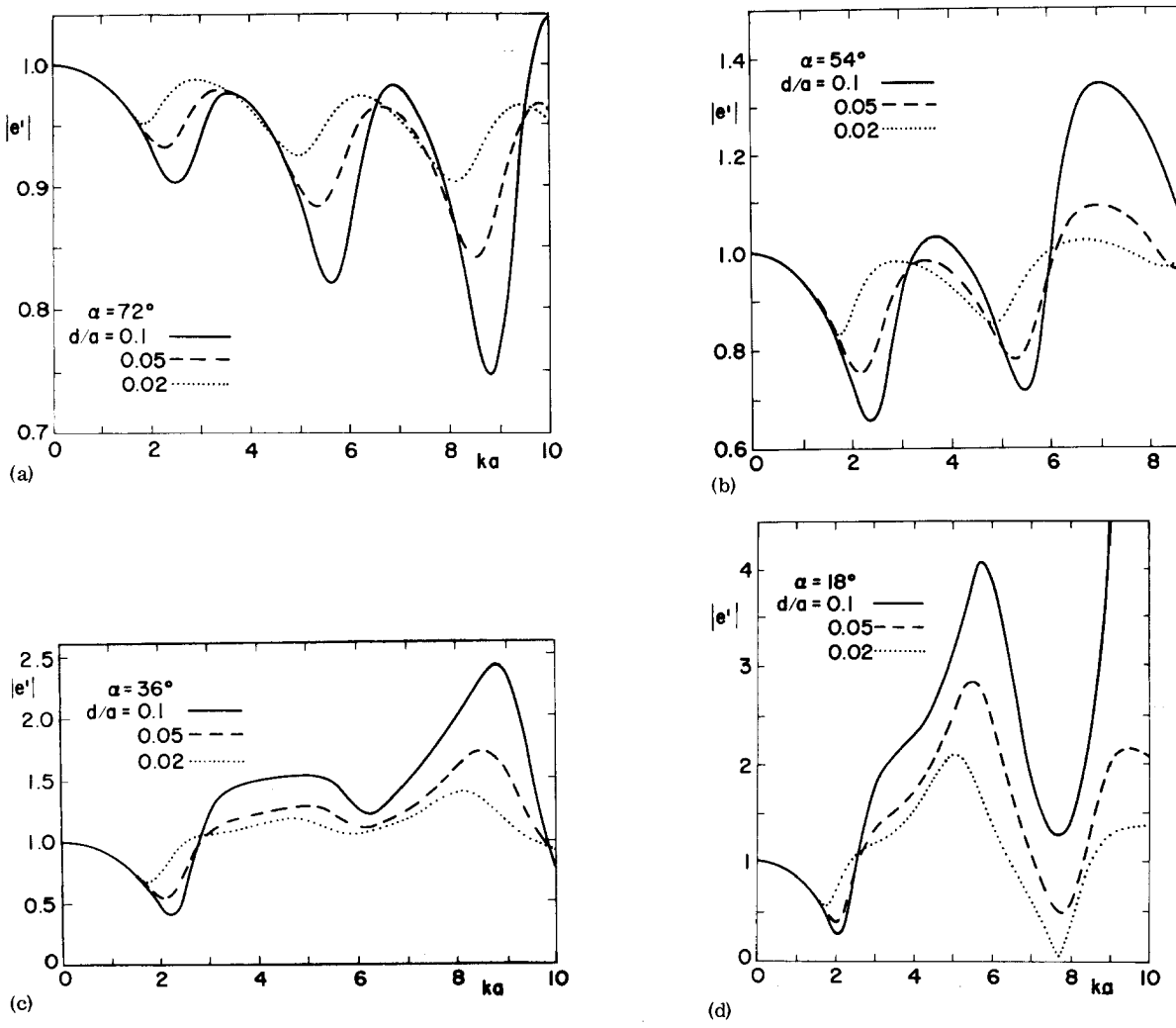


FIG. 2. The normalized electric field at the center of the two disks.

figures. It was found numerically that for $ka < 1$ we have $|e'| \approx 1 - k^2a^2\cos^2\alpha/6$. To conclude this section, we wish to point out that $e' = v' = 1$ for $\alpha = 90°$, and $E_z(0, z) = 0$ for $\alpha = 0°$, as expected.

VI. THE SCATTERED FAR FIELD

In this section we will express the scattered far field and the total scattered power in terms of the solutions of the integral equations (15) and (21). To this end we will first find a far-field expression for the scattered Hertz potentials and then use these expressions to find the scattered electric and magnetic far fields. We will first treat the case where the incident magnetic field is parallel to the disks. The case where the incident electric field is parallel to the disks is then treated in an analogous manner.

From the Green's theorem we can derive the following far-field representation of $\eta_m^{\pm}(\rho, z)$,

$$\eta_m^{\pm}(r\sin\theta, r\cos\theta) \sim (4\pi r)^{-1}\exp(ikr)\exp(\mp ikd\cos\theta) \times \int_0^{a_\pm} \rho' J_m(k\rho'\sin\theta) y_m^{\pm}(\rho')\,d\rho', \tag{37}$$

where (r, θ, ϕ) are the spherical coordinates (see Fig. 1). Substituting the expression (14) for $y_m^{\pm}(\rho)$ into (37) and using the Sonine formula, we get, after some algebraic manipulations,

$$\eta_m^{\pm} \sim \frac{1}{\sqrt{\pi}} \frac{\exp(ikr)}{kr} Q_m^{\pm}(\theta, k), \tag{38}$$

where

$$Q_m^{\pm}(\theta, k) = (4\pi)^{-1} k\sqrt{2\cos\theta}\exp(\mp ikd\cos\theta)\tan^m\theta \times \int_0^{a_\pm} I_{m-1/2}(ku\cos\theta) Y_m^{\pm}(u)\,du. \tag{39}$$

Similarly, one can derive the following far-field expressions:

$$\xi_m^{\pm} \sim \frac{-i}{\sqrt{\pi}} \frac{\exp(ikr)}{kr} R_m^{\pm}(\theta, k), \quad \psi^{\pm} \sim \frac{-i}{\sqrt{\pi}} \frac{\exp(ikr)}{kr} S^{\pm}(\theta, k), \tag{40}$$

where

$$R_m^{\pm}(\theta, k) = (4\pi)^{-1} k\sqrt{2\cos\theta}\exp(\mp ikd\cos\theta)\tan^m\theta \times \int_0^{a_\pm} I_{m+1/2}(ku\cos\theta) X_m^{\pm}(u)\,du \tag{41}$$

and $S^{\pm}(\theta, k)$ is obtained from (41) by making the substitutions $m = 0$, $Z_{\pm} \to X_m^{\pm}$, and $S^{\pm} \to R_m^{\pm}$ in this equation.

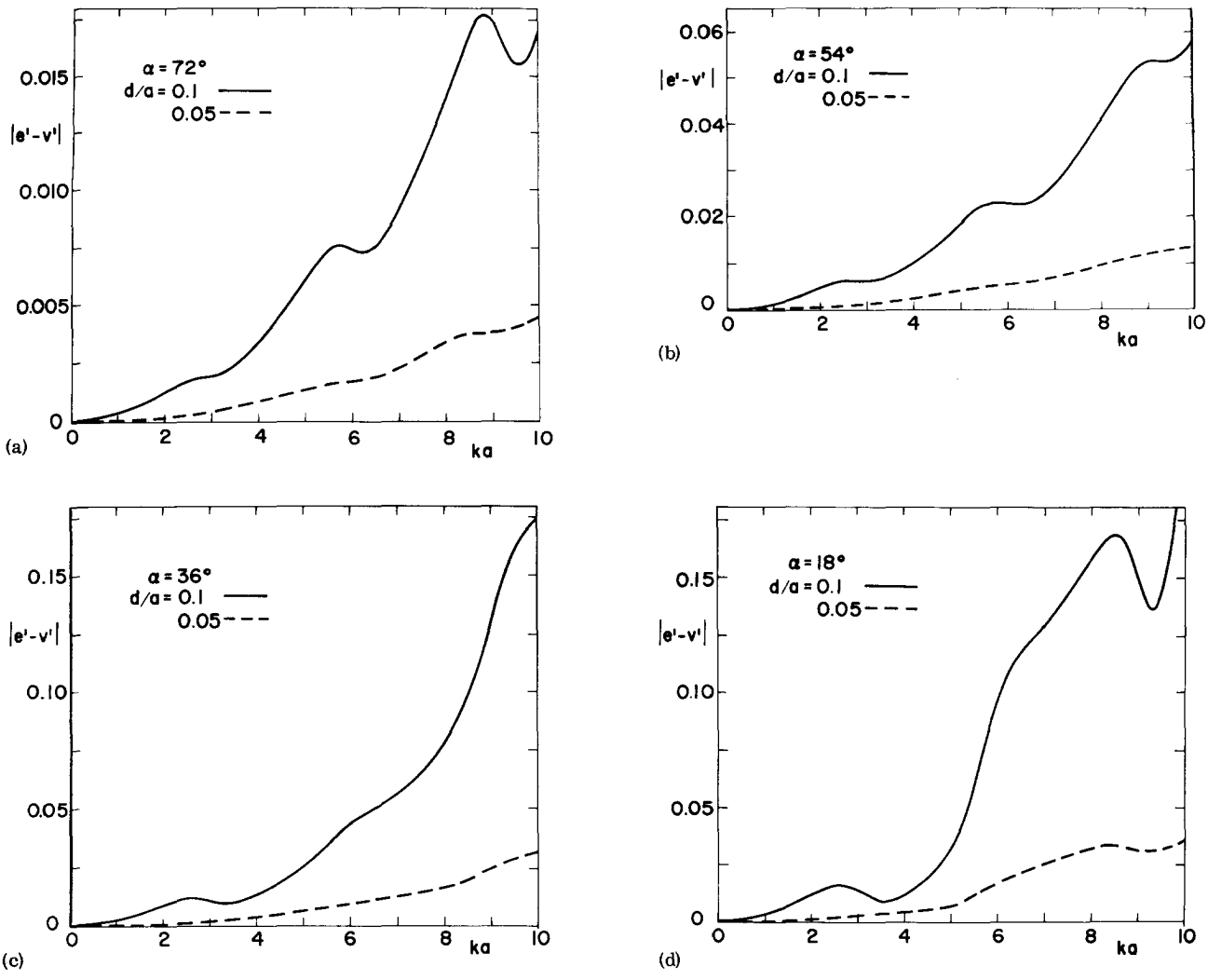


FIG. 3. The difference quantity $|e'-v'|$.

The scattered electromagnetic field can now be obtained from (2), (3), (38), and (39):

$$E_\theta^{sc} \sim E_0' \frac{\exp(ikr)}{r\sqrt{\pi}} \sum_{m=1}^{\infty} [R'_{m-1}(\theta,k)\cos m\phi + ik\sin\theta S'(\theta,k)],$$

$$E_\phi^{sc} \sim E_0' \frac{\exp(ikr)}{r\sqrt{\pi}} \sum_{m=1}^{\infty} [\cos\theta R'_{m-1}(\theta,k) - i\sin\theta Q'_m(\theta,k)]\sin m\phi, \tag{42}$$

$$H_\theta = -Z_0^{-1}E_\phi, \quad H_\phi = Z_0^{-1}E_\theta,$$

where $R'_m = R_m^+ + R_m^-$ and, similarly, for Q'_m and S'.

The total scattered power P' is given by

$$P' = Z_0^{-1}\int (|E_\theta|^2 + |E_\phi|^2) r^2 \sin\theta\, d\theta d\phi$$
$$= E_0'^2 Z_0^{-1} \int_{-\pi}^{\pi} [2k^2\sin^2\theta |S'|^2 + \sum_{m=1}^{\infty} (|R'_{m-1}|^2 + |\cos\theta R'_{m-1} + i\sin\theta Q'_m|^2)]\sin\theta\, d\theta. \tag{43}$$

Similarly, when the incident electric field is parallel to the disks we obtain the following expressions:

$$E_\theta^{sc} \sim -E_0'' \frac{\exp(ikr)}{r\sqrt{\pi}} \sum_{m=1}^{\infty} R''_{m-1}(\theta,k)\sin m\phi,$$

$$E_\phi^{sc} \sim -E_0'' \frac{\exp(ikr)}{r\sqrt{\pi}} \sum_{m=1}^{\infty} [\cos\theta R''_{m-1}(\theta,k) - i\sin\theta Q''_m(\theta,k)]\cos m\phi, \tag{44}$$

$$P'' = E_0''^2 Z_0^{-1} \int_{-\pi}^{\pi} \sum_{m=1}^{\infty} [|R''_{m-1}|^2 + |\cos\theta R''_{m-1} - i\sin\theta Q''_m|^2]\sin\theta\, d\theta,$$

where Q''_m and R''_m have been obtained from (39) and (41) by making the following substitutions in those formulas: $Y_m^{\pm}(u) \to T_m^{\pm}(u)$ and $X_m^{\pm}(u) \to W_m^{\pm}(u)$.

Before concluding this section it is worth pointing out that the total scattered power can also be obtained by integrating the real part of the Poynting vector over the surface of the disks. This method leads to a different representation of the total scattered power.[7]

VII. CONCLUDING REMARKS

The problem of electromagnetic scattering from two perfectly conducting, coaxial, circular disks can be reduced to the solution of a set of Fredholm integral equations of the second kind. Once the solutions of these equations have been determined, any field quantity can

be obtained by performing simple integrations on these solutions. The frequency variation of the field along the axis of the two plates has been obtained by solving the integral equations numerically for different values of radius-to-wavelength ratio as well as separation-to-radius ratio of two equal, coaxial disks.

Using the principle of analytic continuation we observe that the kernels of the integral equations (15) and (21) are entire functions of the complex frequency variable s. Moreover, the kernels are finite for $d \neq 0$ and, hence, the solutions of these integral equations are meromorphic functions of s. This result is, of course, in agreement with the results reported previously[12] concerning the analytical properties of the field scattered from a perfectly conducting, finite body.

It should also be pointed out that the integral equations determining electrostatic scattering from two perfectly conducting, circular, coaxial disks can be obtained by taking the limit as the radius-to-wavelength ratio tends to zero of the dynamic integral equations (15) and (21). The integral equations thus obtained resemble those obtained by Love[13] when calculating the electrostatic potential of two equal, circular, coaxial, conducting disks equally or oppositely charged.

ACKNOWLEDGMENTS

Thanks are due to Dr. K.S.H. Lee and Dr. C.E. Baum for their suggestions and Mr. R.W. Sassman for his assistance with the computer programming.

[1]D.S. Jones, Comm. Pure Appl. Math. 9, 713 (1956).
[2]W.E. Williams, Proc. Roy. Soc. Ser. A 267, 77 (1962).
[3]W.E. Williams, Proc. Camb. Phil. Soc. 58, 625 (1962).
[4]W.E. Williams, Z. Angew. Math. Phys. 13, 133 (1962).
[5]D.P. Thomas, Proc. Camb. Phil. Soc. 60, 621 (1964).
[6]D.L. Jain and R.P. Kanwal, J. Math. Phys. 12, 723 (1971).
[7]N.N. Lebedev and I.P. Skal'skaya, Zh. Tekh. Fiz. 29, 700 (1959) [Sov. Phys. Tech. Phys. 4, 627 (1959)].
[8]Y. Nomura and T. Osanai, J. Phys. Soc. Jpn. 16, 819 (1961).
[9]F.B. Sleator, in *Electromagnetic and Acoustic Scattering by Simple Shapes*, edited by J.J. Bowman, T.B.A. Senior, and P.L.E. Uslenghi (North-Holland, Amsterdam, 1969), Chap. 14.
[10]G.N. Watson, *Theory of Bessel Functions* (Camb. U.P., Cambridge, 1944).
[11]M. Abramowitz and I.N. Stegun, Eds., *Handbook of Mathematical Functions* (National Bureau of Standards, Washington, D.C., 1964).
[12]L. Marin, IEEE Trans. Antennas Prop. AP-21, 809 (1973).
[13]E.R. Love, Quart. J. Mech. Appl. Math. 2, 428 (1949).

CUMULATIVE AUTHOR INDEX

All authors published so far in the current volume are listed alphabetically with the issue and page numbers following the dash. A cumulative author and subject index covering issues 1 through 12 is published yearly. An (E) *after the page number indicates an Erratum.*