JOURNAL OF
MATHEMATICAL
PHYSICS

VOLUME 13, NUMBER 4 APRIL 1972

The Angular Spectrum Representation of Electromagnetic Fields in Crystals.* I. Uniaxial Crystals

Éamon Lalor†
Department of Physics and Astronomy, University of Rochester, Rochester, New York 14627.
(Received 20 October 1970)

The electromagnetic field in a linear, nonmagnetic, nonabsorbing uniaxial crystal which fills the entire half-space $z \geq 0$ and whose optic axis is perpendicular to the plane $z = 0$ is represented as an angular spectrum of plane waves. The angular spectrum representation consists of a superposition of plane electromagnetic waves expressed as the sum of two integrals. In general both homogeneous and evanescent plane waves are required in each integral. Each plane wave of the spectrum satisfies the identical equations obeyed by the entire field. The homogeneous plane waves of the first integral are all ordinary waves and those of the second integral all extraordinary waves. The spectral amplitudes of the field are explicitly expressed in terms of the Fourier transform of the field in the place $z = 0$. The method of stationary phase is applied to the integral representation and it is thereby shown that in the far zone the field may be expressed as the sum of an outgoing (nonuniform) spherical wave and an outgoing (nonuniform) ellipsoidal wave. The amplitude of these waves, at each point on the wave surface, is expressed in terms of the Fourier transform of the field in the plane $z = 0$.

1. INTRODUCTION

It has been known for some time that it is possible, under rather general conditions, to represent a monochromatic electromagnetic field in an empty half-space, $z \geq 0$ say, as an angular spectrum of plane waves.[1-3] The angular spectrum representation of such fields consists in general of a superposition of homogeneous plane waves which propagate in all directions into the half-space and a superposition of evanescent waves which propagate in directions parallel to the xy plane and decay exponentially in the positive z direction. Each plane wave component of the spectrum, whether homogeneous or evanescent, satisfies the *same* set of equations which are obeyed by the entire field; in this sense the angular spectrum representation is a true mode expansion of the field.

Since the early 1950's this representation has been successfully employed, with increasing frequency (sometimes in a restricted form), in the solution of a great variety of physical problems such as diffraction theory,[4] the theory of optical instruments,[5,6] radio propagation,[7] antenna theory,[8,9] holography,[10-12] and many other topics.[13]

The usefulness of the angular spectrum representation of electromagnetic fields in vacuo suggests that it may be desirable to investigate the existence of a similar representation of fields inside material media. Such a representation would be helpful in the treatment of various problems involving the interaction of electromagnetic radiation and bulk matter. If the half-space $z \geq 0$ contains for example a homogeneous isotropic, linear, nonabsorbing, nonmagnetic dielectric, it is easily seen that the angular spectrum representation is identical with that for the vacuum case except for the change of wave number from ω/c for the vacuum case $n\omega/c$ for the dielectric case where ω is the frequency, c is the vacuum velocity of the radiation, and n is the index of refraction at frequency ω. If, however, the dielectric has the properties mentioned above except that it is nonisotropic, the situation is quite different, and this problem has up to now received little attention in the literature.[14]

In this paper and the next we shall extend the angular spectrum representation to cover fields in anisotropic media. Paper I will be devoted to uniaxial crystals and Paper II mostly to biaxial crystals.

2. DERIVATION OF THE ANGULAR SPECTRUM REPRESENTATION

Monochromatic electric and magnetic fields

$$\mathbf{E}(\mathbf{r}, t) = \mathbf{E}_0(\mathbf{r}) \exp(-i\omega t) \tag{1a}$$

and

$$\mathbf{H}(\mathbf{r}, t) = \mathbf{H}_0(\mathbf{r}) \exp(-i\omega t) \tag{1b}$$

in a (macroscopically) homogeneous, linear, nonmagnetic, nonabsorbing anisotropic source-free dielectric filling a volume V obey Maxwell's equations which in the Gaussian system of units have the form[15]:

$$\operatorname{curl} \mathbf{H}_0(\mathbf{r}) + (i\omega/c)\, \boldsymbol{\epsilon} \cdot \mathbf{E}_0(\mathbf{r}) = 0, \tag{2a}$$

$$\operatorname{curl} \mathbf{E}_0(\mathbf{r}) - (i\omega/c)\mathbf{H}_0(\mathbf{r}) = 0, \tag{2b}$$

$$\operatorname{div}[\epsilon \cdot \mathbf{E}_0(\mathbf{r})] = 0 \tag{2c}$$

$$\operatorname{div} \mathbf{H}_0(\mathbf{r}) = 0, \tag{2d}$$

inside the volume V. We have made use of the relationship

$$\mathbf{D}_0 = \epsilon \cdot \mathbf{E}_0 \tag{3}$$

between the electrical displacement $\mathbf{D}_0$ and the electric vector $\mathbf{E}_0$, where ϵ is the (real) dielectric tensor of the medium. Upon substituting Eq. (2b) into Eq. (2a), we see that $\mathbf{E}_0(\mathbf{r})$ satisfies the equation

$$\operatorname{curl}\operatorname{curl} \mathbf{E}_0(\mathbf{r}) - k^2 \epsilon \cdot \mathbf{E}_0(\mathbf{r}) = 0, \tag{4}$$

where

$$k = \omega/c. \tag{5}$$

If we choose as our coordinate axes the principal dielectric axes of the medium, the vector $\epsilon \cdot \mathbf{E}_0$ takes the form $(\epsilon_1 E_{ox}, \epsilon_2 E_{oy}, \epsilon_3 E_{oz})$. We now make the assumptions that

$$\epsilon_1 = \epsilon_2 \neq \epsilon_3 \tag{6}$$

and that the volume V is the entire half-space $z \geq 0$. Thus we are restricting our attention to a semiinfinite uniaxial crystal whose optic axis is perpendicular to the face of the crystal.

We shall confine our attention in what follows mostly to the electric field; the magnetic field is obtained in a straightforward manner by the use of Eq. (2b). Let us now write the electric field $\mathbf{E}_0(x, y, z)$ as a two-dimensional Fourier integral with respect to the first two variables, i.e.,

$$\mathbf{E}_0(x, y, z) = \iint_{-\infty}^{\infty} \mathcal{E}(\alpha, \beta; z) \exp\{ik(\alpha x + \beta y)\} d\alpha d\beta, \tag{7}$$

and, since $\mathbf{E}_0$ satisfies Eq. (4) we may show, using Eq. (7) and the Fourier inversion formula, that the three components of the vector function $\mathcal{E}(\alpha, \beta; z)$ satisfy the following differential equations:

$$k^2(\alpha^2 + \beta^2)\mathcal{E}_x - \frac{\partial^2 \mathcal{E}_x}{\partial z^2} + ik\alpha\left(ik\alpha\, \mathcal{E}_x + ik\beta \mathcal{E}_y + \frac{\partial \mathcal{E}_z}{\partial z}\right) - k^2 \epsilon_1 \mathcal{E}_x = 0, \tag{8a}$$

$$k^2(\alpha^2 + \beta^2)\, \mathcal{E}_y - \frac{\partial^2 \mathcal{E}_y}{\partial z^2} + ik\beta\left(ik\alpha \mathcal{E}_x + ik\beta \mathcal{E}_y + \frac{\partial \mathcal{E}_z}{\partial z}\right) - k^2 \epsilon_1 \mathcal{E}_y = 0, \tag{8b}$$

$$k^2(\alpha^2 + \beta^2)\mathcal{E}_z - \frac{\partial^2 \mathcal{E}_z}{\partial z^2} + \frac{\partial}{\partial z}\left(ik\alpha \mathcal{E}_x + ik\beta \mathcal{E}_y + \frac{\partial \mathcal{E}_z}{\partial z}\right) - k^2 \epsilon_3 \mathcal{E}_z = 0. \tag{8c}$$

In a similar manner Eq. (2c) implies that

$$ik\alpha\epsilon_1 \mathcal{E}_x + ik\beta\epsilon_1 \mathcal{E}_y + \epsilon_3 \frac{\partial \mathcal{E}_z}{\partial z} = 0. \tag{9}$$

Using Eq. (9) and rearranging terms, we may rewrite Eqs. (8a)–(8c) as

$$\frac{\partial^2 \mathcal{E}_x}{\partial z^2} - k^2(\alpha^2 + \beta^2 - \epsilon_1)\mathcal{E}_x - ik\alpha\left(1 - \frac{\epsilon_3}{\epsilon_1}\right)\frac{\partial \mathcal{E}_z}{\partial z} = 0, \tag{10a}$$

$$\frac{\partial^2 \mathcal{E}_y}{\partial z^2} - k^2(\alpha^2 + \beta^2 - \epsilon_1)\mathcal{E}_y - ik\beta\left(1 - \frac{\epsilon_3}{\epsilon_1}\right)\frac{\partial \mathcal{E}_z}{\partial z} = 0, \tag{10b}$$

and

$$\frac{\partial^2 \mathcal{E}_z}{\partial z^2} = -\frac{\epsilon_1}{\epsilon_3} k^2(\epsilon_3 - \alpha^2 - \beta^2)\mathcal{E}_z. \tag{10c}$$

We shall proceed to solve the set of coupled differential equations (10a)–(10c) as follows: We may solve Eq. (10c) immediately since it involves only one unknown $\mathcal{E}_z$, and, having substituted the solution for $\mathcal{E}_z$ into Eqs. (10a) and (10b), we will then solve these equations to obtain a solution for $\mathcal{E}_x$ and $\mathcal{E}_y$, respectively.

The general solution of Eq. (10c) is

$$\mathcal{E}_z(\alpha, \beta; z) = e_z^+(\alpha, \beta) \exp(ik\gamma_3 z) + e_z^-(\alpha, \beta) \exp(-ik\gamma_3 z), \tag{11}$$

where

$$\gamma_3 = [\epsilon_1 - (\epsilon_1/\epsilon_3)(\alpha^2 + \beta^2)]^{1/2} \quad \text{if} \quad \epsilon_3 \geq \alpha^2 + \beta^2, \tag{12a}$$

$$= +i[(\epsilon_1/\epsilon_3)(\alpha^2 + \beta^2) - \epsilon_1]^{1/2} \quad \text{if} \quad \epsilon_3 \leq \alpha^2 + \beta^2, \tag{12b}$$

and e_z^+ and e_z^- are arbitrary functions of α and β.[16] The exponentials on the right-hand side of Eq. (11) are oscillatory or growing and decaying, depending on whether $\alpha^2 + \beta^2$ is less than or greater than ϵ_3.

Upon inserting Eq. (11) into Eqs. (10a) and (10b) we obtain the following equations for $\mathcal{E}_x$ and $\mathcal{E}_y$:

$$\frac{\partial^2 \mathcal{E}_x}{\partial z^2} - k^2(\alpha^2 + \beta^2 - \epsilon_1)\mathcal{E}_x - ik\alpha\left[1 - \left(\frac{\epsilon_3}{\epsilon_1}\right)\right][ik\gamma_3 e_z^+ \times \exp(ik\gamma_3 z) - ik\gamma_3 \epsilon_z^- \exp(-ik\gamma_3 z)] = 0 \tag{13a}$$

and

$$\frac{\partial^2 \mathcal{E}_y}{\partial z^2} - k^2(\alpha^2 + \beta^2 - \epsilon_1)\mathcal{E}_y - ik\beta\left(1 - \frac{\epsilon_3}{\epsilon_1}\right)[ik\gamma_3 e_z^+ \exp(ik\gamma_3 z) - ik\gamma_3 e_z^- \exp(-ik\gamma_3 z)] = 0. \tag{13b}$$

These equations are solved in Appendix A; the solutions are shown to be

$$\mathcal{E}_x(\alpha, \beta; z) = e_x^+(\alpha, \beta)\exp(ik\gamma_1 z) + e_x^-(\alpha, \beta)\exp(-ik\gamma_1 z) - \frac{\epsilon_3}{\epsilon_1}\left(\frac{\alpha\gamma_3}{\alpha^2 + \beta^2}\right) e_z^+(\alpha, \beta) \exp(ik\gamma_3 z) + \frac{\epsilon_3}{\epsilon_1}\left(\frac{\alpha\gamma_3}{\alpha^2 + \beta^2}\right) e_z^-(\alpha, \beta) \exp(-ik\gamma_3 z) \tag{14a}$$

and

$$\mathcal{E}_y(\alpha, \beta; z) = e_y^+(\alpha, \beta) \exp(ik\gamma_1 z) + e_y^-(\alpha, \beta) \exp(-ik\gamma_1 z) - \frac{\epsilon_3}{\epsilon_1}\left(\frac{\beta\gamma_3}{\alpha^2 + \beta^2}\right) e_z^+(\alpha, \beta) \exp(ik\gamma_3 z) + \frac{\epsilon_3}{\epsilon_1}\left(\frac{\beta\gamma_3}{\alpha^2 + \beta^2}\right) e_z^-(\alpha, \beta) \exp(-ik\gamma_3 z), \tag{14b}$$

where

$$\gamma_1 = (\epsilon_1 - \alpha^2 - \beta^2)^{1/2} \quad \text{if} \quad \epsilon_1 \geq \alpha^2 + \beta^2, \tag{15a}$$

$$= + i(\alpha^2 + \beta^2 - \epsilon_1)^{1/2} \quad \text{if} \quad \epsilon_1 < \alpha^2 + \beta^2. \tag{15b}$$

γ_3 is defined by Eqs. (12a) and (12b) and e_x^+, e_x^-, e_y^+, and e_y^- are arbitrary functions of α and β.

If we substitute Eqs. (11), (14a), and (14b) into Eq. (7), we see that we have succeeded in representing the electric field $\mathbf{E}_0$ as a superposition of the following four types of plane waves:

(i) $(e_x^+(\alpha,\beta), e_y^+(\alpha,\beta), 0) \exp[ik(\alpha x + \beta y + \gamma_1 z)] \equiv \mathbf{E}_1$,

(ii) $(e_x^-(\alpha,\beta), e_y^-(\alpha,\beta), 0) \exp[ik(\alpha x + \beta y - \gamma_1 z)] \equiv \mathbf{E}_2$,

(iii) $$\left[-\frac{\epsilon_3}{\epsilon_1}\left(\frac{\alpha\gamma_3}{\alpha^2+\beta^2}\right), -\frac{\epsilon_3}{\epsilon_1}\left(\frac{\beta\gamma_3}{\alpha^2+\beta^2}\right), 1\right] e_z^+(\alpha,\beta) \times \exp[ik(\alpha x + \beta y + \gamma_3 z)] \equiv \mathbf{E}_3,$$

and

(iv) $$\left[\frac{\epsilon_3}{\epsilon_1}\left(\frac{\alpha\gamma_3}{\alpha^2+\beta^2}\right), \frac{\epsilon_3}{\epsilon_1}\left(\frac{\beta\gamma_3}{\alpha^2+\beta^2}\right), 1\right] e_z^-(\alpha,\beta) \times \exp[ik(\alpha x + \beta y - \gamma_3 z)] \equiv \mathbf{E}_4.$$

It may be verified that each of the waves $\mathbf{E}_1, \mathbf{E}_2, \mathbf{E}_3, \mathbf{E}_4$ is a solution to the original equation (4) and it may thus be regarded as a mode of the electric field in the half-space $z \geq 0$.

We have thus far not imposed any boundary conditions on the field and we shall address ourselves to the problem now. Physical considerations demand

(i) The field should remain bounded at infinity in the right half-space;

(ii) since there are no free sources anywhere in the right half-space, we would expect that in the right half-space all energy would be radiated *away* from the plane $z = 0$.

The first condition implies that

$$e_x^-(\alpha,\beta) = e_y^-(\alpha,\beta) = 0 \quad \text{for} \quad \alpha^2 + \beta^2 > \epsilon_1 \tag{16a}$$

and

$$e_z^-(\alpha,\beta) = 0 \quad \text{for} \quad \alpha^2 + \beta^2 > \epsilon_3. \tag{16b}$$

Considerable care must be exercised in the application of the "radiation condition" [condition (ii) above] to propagation in crystalline media since, as is well known, the directions of the ray vector (Poynting vector) will not in general coincide with the direction of the wave normal in such media. It will be necessary for us to examine the Poynting vector associated with each type of homogeneous plane wave[17] $\mathbf{E}_j$, $j = 1, 2, 3, 4$, and to accept as physical solutions only those waves whose ray vectors point away from the plane $z = 0$ into the half-space $z \geq 0$.

The time averaged Poynting vector may be defined as

$$\langle \mathbf{S} \rangle = (c/8\pi)\, \mathrm{Re}(\mathbf{E}_0 \times \mathbf{H}_0^*). \tag{17}$$

For each of the waves $\mathbf{E}_j$, $j = 1, 2, 3, 4$, we may calculate the magnetic field using Eq. (2b) and the ray vector using Eq. (17). It is found that in the case of the waves $\mathbf{E}_1$ and $\mathbf{E}_2$, the ray vector is in fact in the direction of the wave normal and thus we reject solutions of the type $\mathbf{E}_2$ since these would imply energy propagating in the medium towards the plane $z = 0$. Waves of the type $\mathbf{E}_1$ and $\mathbf{E}_2$ are known as *ordinary waves* whereas those of type $\mathbf{E}_3$ and $\mathbf{E}_4$ are called *extraordinary waves*; these concepts will be discussed in more detail shortly. For the extraordinary waves $\mathbf{E}_3$ and $\mathbf{E}_4$, the directions of the ray vector and the wave normal do not in general coincide. However, if we calculate the z component at the Poynting vector for $\mathbf{E}_3$, we find that

$$\langle S_z \rangle = \gamma_3 (\epsilon_3^2/\epsilon_1)\, |e_z^+(\alpha,\beta)|^2/(\alpha^2 + \beta^2), \tag{18a}$$

while for $\mathbf{E}_4$ we find that

$$\langle S_z \rangle = -\gamma_3 (\epsilon_3^2/\epsilon_1)\, |e_z^-(\alpha,\beta)|^2/(\alpha^2 + \beta^2). \tag{18b}$$

Thus since the right-hand side of (18a) is positive and the right-hand side of (18b) is negative, we see that all plane waves with wave normals pointing away from the plane $z = 0$ have ray vectors which also point away from the plane $z = 0$, while all waves with wave normals pointing towards the $z = 0$ have ray vectors also pointing towards that plane. It should be emphasized that this result holds only because the plane $z = 0$ is the plane perpendicular to the optic axis of the crystal; one would not expect it to hold for an arbitrary plane.

On the basis of the preceding arguments we reject solutions of the type $\mathbf{E}_4$. We may then using Eqs. (7), (11), (14a), and (14b) and conditions (i) and (ii) write the electric field $\mathbf{E}_0$ as a superposition of plane wave of the type $\mathbf{E}_1$ and $\mathbf{E}_3$ as follows:

$$\mathbf{E}_0(x,y,z) = \iint_{-\infty}^{\infty} \mathbf{e}^{(\mathrm{or})}(\alpha,\beta) \times \exp[ik(\alpha x + \beta y + \gamma_1 z)]\, d\alpha d\beta + \iint_{-\infty}^{\infty} \mathbf{e}^{(\mathrm{ex})}(\alpha,\beta) \times \exp[ik(\alpha x + \beta y + \gamma_3 z)]\, d\alpha d\beta, \tag{19}$$

where

$$\mathbf{e}^{(\mathrm{or})}(\alpha,\beta) = (e_x(\alpha,\beta), e_y(\alpha,\beta), 0), \tag{20a}$$

$$\mathbf{e}^{(\mathrm{ex})}(\alpha,\beta) = e_z(\alpha,\beta)\left[-\left(\frac{\alpha\gamma_3}{\alpha^2+\beta^2}\right)\frac{\epsilon_3}{\epsilon_1}, -\left(\frac{\beta\gamma_3}{\alpha^2+\beta^2}\right)\frac{\epsilon_3}{\epsilon_1}, 1\right] \tag{20b}$$

(the + superscripts on e_x^+, e_y^+, e_z^+ have been dropped), and γ_1 and γ_3 are defined in Eqs. (15) and (12), respectively. In Eqs. (19), (20a), and (20b) e_x, e_y, and e_z are arbitrary functions of the parameters α and β.

In the mode expansion (19), real values of γ_1 and γ_3 are associated with homogeneous plane waves with the direction cosines of the wave normals being proportional to $(\alpha, \beta, \gamma_1)$ and $(\alpha, \beta, \gamma_3)$, respectively. Imaginary values of γ_1 and γ_3 are both associated with evanescent waves which run up and down the xy plane and decay exponentially with increasing z. We may say, by analogy with the accepted terminology for electromagnetic fields in vacuo, that Eq. (19) rep-

resents the electric field as an angular spectrum of plane waves.

In order that the representation (19) have more than formal significance, the spectral amplitudes $\mathbf{e}^{(or)}$ and $\mathbf{e}^{(ex)}$, or equivalently the functions e_x, e_y, and e_z should be expressible in a simple manner. In fact these functions may be expressed in terms of the boundary value of the field $\mathbf{E}_0$ in the plane $z = 0$ by setting $z = 0$ in Eq. (19) and taking the Fourier inverse. The resulting expressions are

$$e_z(\alpha,\beta) = \left(\frac{k}{2\pi}\right)^2 \iint_{-\infty}^{\infty} E_{0z}(x,y,0) \times \exp[-ik(\alpha x + \beta y)]\,dxdy, \tag{21a}$$

$$e_x(\alpha,\beta) = \left(\frac{k}{2\pi}\right)^2 \iint_{-\infty}^{\infty} E_{0_x}(x,y,0) \times \exp[-ik(\alpha x + \beta y)]\,dxdy + \left(\frac{\alpha\gamma_3}{\alpha^2+\beta^2}\right)\frac{\epsilon_3}{\epsilon_1}\left(\frac{k}{2\pi}\right)^2 \iint_{-\infty}^{\infty} E_{0z}(x,y,0) \times \exp[-ik(\alpha x + \beta y)]\,dxdy, \tag{21b}$$

$$e_y(\alpha,\beta) = \left(\frac{k}{2\pi}\right)^2 \iint_{-\infty}^{\infty} E_{0y}(x,y,0) \times \exp\{-ik(\alpha x + \beta y)\}\,dxdy + \left(\frac{\beta\gamma_3}{\alpha^2+\beta^2}\right)\frac{\epsilon_3}{\epsilon_1}\left(\frac{k}{2\pi}\right)^2 \iint_{-\infty}^{\infty} E_{0z}(x,y,0) \times \exp[-ik(\alpha x + \beta y)]\,dxdy. \tag{21c}$$

Thus Eq. (19) together with Eqs. (20) and (21) represents the electric field everywhere in the half-space $z \geq 0$ in terms of its boundary value on the plane $z = 0$.

An expression for the magnetic field $\mathbf{H}_0$ may be derived from Eq. (19) using Eq. (2b), the result is

$$\mathbf{H}_0(x,y,z) = \iint_{-\infty}^{\infty} [\mathbf{s}_1 \times \mathbf{e}^{(or)}(\alpha,\beta)] \times \exp[ik(\alpha x + \beta y + \gamma_1 z)]\,d\alpha d\beta + \iint_{-\infty}^{\infty} [\mathbf{s}_3 \times \mathbf{e}^{(ex)}(\alpha,\beta)] \times \exp[ik(\alpha x + \beta y + \gamma_3 z)]\,d\alpha d\beta, \tag{22}$$

where

$$\mathbf{s}_1 \equiv (\alpha,\beta,\gamma_1) \tag{22'a}$$

and

$$\mathbf{s}_3 \equiv (\alpha,\beta,\gamma_3). \tag{22'b}$$

3. DISCUSSION OF THE ORDINARY AND EXTRAORDINARY WAVES OF THE ANGULAR SPECTRUM

It was shown in Sec. 2 of this paper that the electric and magnetic fields in a uniaxial crystal may be represented as a superposition of both homogeneous and evanescent plane waves of two different types. In this section we shall briefly examine the properties of these two types of plane waves.

Let us consider first a typical plane wave component in the first integral on the right-hand side of Eq. (19); it is of the form

$$\mathbf{E}^{(or)}(x,y,z) = (e_x(\alpha,\beta), e_y(\alpha,\beta), 0) \times \exp[ik(\alpha x + \beta y + \gamma_1 z)]. \tag{23}$$

Equation (2c) implies that

$$\operatorname{div} \mathbf{E}^{(or)}(x,y,z) = 0; \tag{24}$$

thus homogeneous plane waves of this type are transverse. It may be verified that the field $\mathbf{E}^{(or)}$ satisfies the reduced wave equation

$$(\nabla^2 + k^2\epsilon_1)\mathbf{E}^{(or)}(x,y,z) = 0, \tag{25}$$

since by Eq. (15)

$$\alpha^2 + \beta^2 + \gamma_1^2 = \epsilon_1 k^2. \tag{26}$$

Thus we see that the phase velocity of the wave $\mathbf{E}^{(or)}$ is

$$v_0 = c/\epsilon_1^{1/2}, \tag{27}$$

irrespective of its direction of propagation. This result implies that the first integral on the right-hand side of Eq. (19) is a superposition of ordinary waves.[18]

A typical component of the second integral in Eq. (19) has the form

$$\mathbf{E}^{(ex)}(x,y,z) = e_z(\alpha,\beta) \times \left[-\frac{\epsilon_3}{\epsilon_1}\left(\frac{\alpha\gamma_3}{\alpha^2+\beta^2}\right), -\frac{\epsilon_3}{\epsilon_1}\left(\frac{\beta\gamma_3}{\alpha^2+\beta^2}\right), 1\right] \times \exp[ik(\alpha x + \beta y + \gamma_3 z)]. \tag{28}$$

A wave of this type is *not* transverse since Eq. (2c) implies that

$$\operatorname{div} \mathbf{E}^{(ex)} = ike_z(\alpha,\beta)[1 - (\epsilon_3/\epsilon_1)]\gamma_3 \neq 0. \tag{29}$$

$\mathbf{E}^{(ex)}$ satisfies the reduced wave equation

$$[\nabla^2 + k^2\epsilon(\alpha,\beta)]\mathbf{E}^{(ex)}(x,y,z) = 0, \tag{30}$$

where

$$\epsilon(\alpha,\beta) = \epsilon_1 + (1 - \epsilon_1/\epsilon_3)(\alpha^2+\beta^2). \tag{31}$$

The phase velocity of $\mathbf{E}^{(ex)}$,

$$v_e = c/\epsilon(\alpha,\beta)^{1/2} \tag{32}$$

depends, through α and β, on its direction of propagation. The last integral in Eq. (19) is then a superposition of extraordinary waves.[18]

It may easily be verified that $\mathbf{E}^{(or)}$ and $\mathbf{E}^{(ex)}$ satisfy all the other well-known properties of ordinary and extraordinary waves, we shall not, however, carry this discussion any further here.

Since the magnetic field is related to the electric field by the Maxwell equation (2b), it is easily shown that the partial (**E**, **H**) fields, consisting of each of the two types of **E** waves discussed in this section and their corresponding **H** waves, have the true structure of plane electromagnetic waves.

4. ASYMPTOTIC BEHAVIOR OF THE ANGULAR SPECTRUM

In this section we shall investigate the asymptotic behavior of the field as $kR \to \infty$ along any line specified by fixed direction cosines $x/R, y/R, z/R, z > 0$. For the homogeneous ordinary wave contribution to the electric field, i.e.,

$$\mathbf{E}_{(h)}^{(or)}(x,y,z) = \iint_{\alpha^2+\beta^2 \leq \epsilon_1} e^{or}(\alpha,\beta) \times \exp\{ik(\alpha x + \beta y + \gamma_1 z)\}\, d\alpha d\beta, \quad (33)$$

the integral is essentially of the form whose asymptotic behavior was considered by Miyamoto and Wolf.[19] Using their results we may immediately write the asymptotic behavior of the electric field as

$$\mathbf{E}_{(h)}^{(or)}(x,y,z) \sim -\frac{2\pi i \epsilon_1^{1/2}}{k}\left(\frac{z}{R}\right) e^{or}\left(\epsilon_1^{1/2}\frac{x}{R}, \epsilon_1^{1/2}\frac{y}{R}\right) \times \frac{\exp(ik\epsilon_1^{1/2}R)}{R}. \quad (34)$$

Similarly for the magnetic field

$$\mathbf{H}_{(h)}^{(or)}(x,y,z) \sim -\frac{2\pi i \epsilon_1^{1/2}}{k}\left(\frac{z}{R}\right) \mathbf{s}_0 \times \mathbf{e}^{or}\left(\epsilon_1^{1/2}\frac{x}{R}, \epsilon_1^{1/2}\frac{y}{R}\right) \times \frac{\exp\{ik\epsilon_1^{1/2}R\}}{R}, \quad (35)$$

where

$$\mathbf{s}_0 = \epsilon_1^{1/2}(x/R, y/R, z/R). \quad (36)$$

The homogeneous extraordinary wave contribution to the electric field

$$\mathbf{E}_{(h)}^{(ex)}(x,y,z) = \iint_{\alpha^2+\beta^2 \leq \epsilon_3} \mathbf{e}^{(ex)}(\alpha,\beta) \times \exp[ik(\alpha x + \beta y + \gamma_3 z)]\, d\alpha d\beta \quad (37)$$

cannot however be cast in the form of Miyamoto and Wolf's integral. The asymptotic approximation to an integral of the type appearing in Eq. (37) is derived in Appendix B where it is shown that

$$\mathbf{E}_{(h)}^{(ex)}(x,y,z) \sim -\frac{2\pi i \epsilon_1^{1/2}}{k\xi^2}\left(\frac{z}{R}\right) \mathbf{e}^{(ex)}(\alpha',\beta') \times \frac{\exp[ik\epsilon_3^{1/2}\xi R]}{R}, \quad (38)$$

where

$$\alpha' = (\epsilon_3^{1/2}/\xi)(x/R), \quad (39)$$

$$\beta' = (\epsilon_3^{1/2}/\xi)(y/R), \quad (40)$$

and

$$\xi = \left[1 - \left(1 - \frac{\epsilon_1}{\epsilon_3}\right)\left(\frac{z}{R}\right)^2\right]^{1/2}. \quad (41)$$

Similarly for the magnetic field

$$\mathbf{H}_{(h)}^{(ex)} \sim -\frac{2\pi i \epsilon_1^{1/2}}{k\xi^2}\left(\frac{z}{R}\right) \mathbf{s}_e \times \mathbf{e}^{(ex)}(\alpha',\beta') \times \frac{\exp[ik\epsilon_3^{1/2}\xi R]}{R}, \quad (42)$$

where

$$\mathbf{s}_e = [\alpha', \beta', [\epsilon_1 - (\epsilon_1/\epsilon_3)(\alpha'^2 + \beta'^2)]^{1/2}]. \quad (43)$$

The surfaces of constant phase of the fields $\mathbf{E}_{(h)}^{(ex)}$ and $\mathbf{H}_{(h)}^{ex}$ at a great distance from the origin are seen to be

$$k\epsilon_3 R\xi = \text{const}, \quad (44)$$

which may be rewritten as

$$\frac{x^2}{\epsilon_1} + \frac{y^2}{\epsilon_1} + \frac{z^2}{\epsilon_3} = \text{const}. \quad (45)$$

Thus the surface of constant phase is an ellipsoid which has the same orientation and the same ratio between the length of its principal axes as the *ellipsoid of wave normals* (Ref. 15, p. 673) of the crystal.

Since the evanescent portion of the field is an integral with an exponentially decaying integrand, it may be shown that under reasonable assumptions, we may neglect the contribution of the evanescent waves to the field in the far zone. The asymptotic approximations, as $kR \to \infty$ along any line specified by the field direction cosines $[x/R, y/R, z/R, (z > 0)]$, to the electric and magnetic fields are therefore

$$\mathbf{E}_0(x,y,z) \sim -\frac{2\pi i \epsilon_1^{1/2}}{k}\left(\frac{z}{R}\right) \mathbf{e}^{(or)}\left(\epsilon_1^{1/2}\frac{x}{R}, \epsilon_1^{1/2}\frac{y}{R}\right) \times \frac{\exp(ik\epsilon_1^{1/2}R)}{R} - \frac{2\pi i \epsilon_1^{1/2}}{k\xi^2}\left(\frac{z}{R}\right) \mathbf{e}^{ex}(\alpha',\beta') \frac{\exp[ik\epsilon_3^{1/2}\xi R]}{R} \quad (46)$$

and

$$\mathbf{H}_0(x,y,z) \sim -\frac{2\pi i \epsilon_1^{1/2}}{k}\left(\frac{z}{R}\right) \mathbf{s}_0 \times \mathbf{e}^{or}\left(\epsilon_1^{1/2}\frac{x}{R}, \epsilon_1^{1/2}\frac{y}{R}\right) \times \frac{\exp[ik\epsilon_1^{1/2}R]}{R} - \frac{2\pi i \epsilon_1^{1/2}}{k\xi^2}\left(\frac{z}{R}\right) \mathbf{s}_e \times \mathbf{e}^{(ex)}(\alpha',\beta') \times \frac{\exp[ik\epsilon_3^{1/2}\xi R]}{R}. \quad (47)$$

We see that the over-all far zone structure of the field is that of the sum of an outgoing (nonuniform) spherical wave, associated with the ordinary waves of the spectrum and an outgoing (nonuniform) ellipsoidal wave associated with the extraordinary waves of the spectrum. In the neighborhood of any point in the far zone, the structure of the field is that of the sum of an ordinary and an extraordinary plane electromagnetic wave.

5. CONCLUSION

We have shown that the electromagnetic field in a uniaxial crystal, which is linear, nonmagnetic, and nonabsorbing and which occupies the entire half-space $z \geq 0$, and which furthermore has its optic axis perpendicular to the plane $z = 0$, may be represented as an angular spectrum of plane waves provided only that the field have an invertible Fourier integral representation in the plane $z = 0$ [Eq. (17)] and that it obeys the two physically reasonable boundary conditions (i) and (ii). The angular spectrum representation of the electric field [Eq. (19)] or the magnetic

field [Eq. (22)] consists of a superposition of plane waves expressed as the sum of two integrals, each integral containing in general, both homogeneous and evanescent plane waves. The homogeneous waves in the first integrand are ordinary waves and the homogeneous waves in the second integral are extraordinary waves. Each plane wave (whether homogeneous or evanescent) in each of the integrals obeys the identical equations satisfied by the entire field; the angular spectrum representation we have derived is then a mode expansion of the field. We have furthermore explicitly expressed the spectral amplitudes of the field in terms of the Fourier transform of the field in the plane $z = 0$ [Eqs. (21)].

In Sec. 3 we have applied the method of stationary phase to the integrals in the representation and have found that in the far zone, the field may be expressed as the sum of an outgoing (nonuniform) spherical wave and an outgoing (nonuniform) ellipsoidal wave. The complex vector amplitude of these waves, at each point on the wave surface, is expressed in terms of the Fourier transform of the field in the plane $z = 0$.

The next paper in this series will be devoted to the angular spectrum representation of the electromagnetic field in a biaxial crystal, one of whose dielectric axes is perpendicular to the face of the crystal. As a special case of that analysis we shall derive results, similar to those of this paper, for the case of a uniaxial crystal whose optic axis is *parallel* to the face of the crystal.

ACKNOWLEDGMENTS

The author would like to thank Professor Emil Wolf and Dr. Robin Asby for helpful discussions relating to this work.

APPENDIX A: SOLUTIONS OF THE DIFFERENTIAL EQS. (13a) and (13b)

Equations (13a) and (13b) are both of the form

$$y'' = -A^2 y + Be^{iCz} + De^{-iCz}, \tag{A1}$$

where $y'' = d^2y/dz^2$ and A, B, C, and D are constants. If we make the substitution $y = ue^{iAz}$, Eq. (A1) may be rewritten as

$$u'' + 2iAu' = Be^{i(C-A)z} + De^{-i(C+A)z}. \tag{A2}$$

Upon setting $u' = we^{-2iAz}$, Eq. (A2) becomes

$$w' = Be^{i(C+A)z} + De^{-i(C-A)z}, \tag{A3}$$

which may be solved to yield

$$w = \frac{B}{i(C+A)} e^{i(C+A)z} - \frac{D}{i(C-A)} e^{-i(C-A)z} + E, \tag{A4}$$

where E is a constant.

Thus

$$u' = \frac{B}{i(C+A)} e^{i(C-A)z} - \frac{D}{i(C-A)} e^{-i(C+A)} + Ee^{-2iAz}, \tag{A5}$$

which gives on integration

$$u = \frac{-B}{C^2 - A^2} e^{i(C-A)z} - \frac{D}{C^2 - A^2} e^{-i(C+A)z} + \frac{iE}{2A} e^{-2iAz} + F, \tag{A6}$$

where F is a constant.

Finally using the definition of u we obtain the solution of Eq. (A1) as

$$y = \frac{B}{A^2 - C^2} e^{iCz} + \frac{D}{A^2 - C^2} e^{-iCz} + \frac{iE}{2A} e^{-iAz} + Fe^{iAz}. \tag{A7}$$

If we substitute into Eq. (A7) the specific values of the constants A, B, C, and D that appear in Eqs. (13a) and (13b), we obtain the solutions for $\mathcal{E}_x$ and $\mathcal{E}_y$ given in Eqs. (14a) and (14b).

APPENDIX B: THE ASYMPTOTIC BEHAVIOR OF $\mathbf{E}_{(h)}^{(ex)}(x, y, z)$

We derive here the asymptotic approximation for large values of kR to the integral

$$\mathbf{E}_{(h)}^{(ex)}(x, y, z) = \iint_{\alpha^2+\beta^2<\epsilon_3} e^{(ex)}(\alpha, \beta) \times \exp\{ikR\mathbf{s}\cdot\mathbf{u}\} d\alpha d\beta, \tag{B1}$$

where

$$\mathbf{s} = \{\alpha, \beta, [\epsilon_1 - (\epsilon_1/\epsilon_3)(\alpha^2 + \beta^2)]^{1/2}\} \tag{B2}$$

and

$$\mathbf{u} = \mathbf{R}/R. \tag{B3}$$

The approximation may readily be found by the application of the principle of stationary phase.[19,20] On applying this principle we find the asymptotic approximation to (B1). We obtain

$$\mathbf{E}_{(h)}^{(ex)}(x, y, z) \sim \frac{2\pi i\epsilon}{k|\Delta^{1/2}|} \mathbf{e}^{(ex)}(\alpha', \beta') \frac{\exp\{ikR\mathbf{s}'\cdot\mathbf{u}\}}{R}. \tag{B4}$$

In Eq. (B4),

$$\mathbf{s}' = \{\alpha', \beta', [\epsilon_1 - (\epsilon_1/\epsilon_3)(\alpha'^2 + \beta'^2]^{1/2}\}, \tag{B5}$$

where α' and β' are the values of α and β which make the phase factor

$$\Phi = \alpha u_x + \beta u_y + [\epsilon_1 - (\epsilon_1/\epsilon_3)(\alpha^2 + \beta^2)]^{1/2} u_z \tag{B6}$$

stationary within the domain of integration, i.e., α' and β' are the roots of the equation

$$\frac{\partial\Phi}{\partial\alpha} = \frac{\partial\Phi}{\partial\beta} = 0, \tag{B7}$$

$$\Delta = \left[\frac{\partial^2\Phi}{\partial\alpha^2}\frac{\partial^2\Phi}{\partial\beta^2} - \left(\frac{\partial^2\Phi}{\partial\alpha\partial\beta}\right)^2\right]', \tag{B8}$$

and

$$\epsilon = \begin{cases} +1 \text{ when } \Delta > 0 & \left(\dfrac{\partial^2\Phi}{\partial\alpha^2}\right)' > 0, \\ -1 \text{ when } \Delta > 0 & \left(\dfrac{\partial^2\Phi}{\partial\beta^2}\right)' < 0, \\ i \text{ when } \Delta < 0. \end{cases} \tag{B9}$$

In (B8) and (B9) the prime denotes values at the stationary points.

The roots of Eq. (B7) may be shown to be

$$\alpha' = \frac{\epsilon_3^{1/2} u_x}{\{1 - [1 - (\epsilon_1/\epsilon_3)] u_z^2\}^{1/2}} \tag{B10}$$

and

$$\beta' = \frac{\epsilon_3^{1/2} u_y}{\{1 - [1 - (\epsilon_1/\epsilon_3)] u_z^2\}^{1/2}} . \tag{B11}$$

Hence

$$\mathbf{s}' \cdot \mathbf{u} = \epsilon_3^{1/2} \{1 - [1 - (\epsilon_1/\epsilon_3)] u_z^2\}^{1/2} . \tag{B12}$$

Further calculation leads to the results

$$\Delta = \frac{\{1 - [1 - (\epsilon_1/\epsilon_3)] u_z^2\}^2}{\epsilon_1 u_z^2} \tag{B13}$$

and

$$\epsilon = -1. \tag{B14}$$

The asymptotic approximation to $E_{(h)}^{(ex)}(x,y,z)$ as $kR \to \infty$ therefore is

$$\mathbf{E}_{(h)}^{(ex)}(x,y,z) \sim \frac{-2\pi i \epsilon_1^{1/2}}{k\xi^2} \left(\frac{z}{R}\right) \mathbf{e}^{(ex)}(\alpha', \beta') \times \frac{\exp[ik\epsilon_3^{1/2} \xi R]}{R}, \tag{B15}$$

where α' and β' are given by (B10), (B14), and

$$\xi = \left[1 - \left(1 - \frac{\epsilon_1}{\epsilon_3}\right)\left(\frac{z}{R}\right)^2\right]^{1/2} . \tag{B16}$$

---

* Research supported by the Air Force Office of Scientific Research.

† Present address: Department of Physics, University College, Belfield, Dublin 4, Ireland.

1 E. Wolf, Proc. Phys. Soc. (London) **74**, 269 (1959), especially the Appendix.

2 E. Lalor, J. Opt. Soc. Am. **58**, 1235 (1968).

3 A history of the angular spectrum and its application in the treatment of physical problems is given in E. Lalor "The Angular Spectrum Representation of Electromagnetic Fields and Its Use in the Treatment of Interaction Problems," (Ph.D. thesis, University of Rochester, 1970), Chap. I.

4 C. J. Bouwkamp, Rept. Progr. Phys. **17**, 39 (1954), and other works quoted in this report.

5 R. K. Luneburg, *Mathematical Theory of Optics* (U. of California Press, Berkeley, 1966), Chap. VI.

6 M. Kline and I. W. Kay, *Electromagnetic Theory and Geometrical Optics* (Interscience, New York, 1965), Chap. XI.

7 J. A. Ratcliffe, Rept. Progr. Phys. **19**, 188 (1956), and other works quoted in this article.

8 G. Borgiotti, IEEE Trans. Antennas Propagation (Communications) **11**, 94 (1967); Alta Frequenza **32**, 196 (1963); IEEE Trans. Antennas Propagation **15**, 565 (1967).

9 D. R. Rhodes, Proc. IEEE **52**, 1013 (1964); IEEE Trans. Antennas Propagation **14**, 676 (1966).

10 G. C. Sherman, J. Opt. Soc. Am. **57**, 1160 (1967).

11 R. Mittra and P. L. Ransom, in *Proceedings of the Symposium on Modern Optics*, edited by J. Fox (Polytechnic, Brooklyn, N.Y., 1967), p. 619.

12 E. Wolf and J. R. Shewell, J. Math. Phys. **11**, 2254 (1970).

13 A more complete list of topics covered and of contributions within each separate topics is given in Ref. 3.

14 See in this context P. C. Clemmow, *The Plane Wave Spectrum Representation of Electromagnetic Fields* (Pergamon, New York, 1966).

15 An excellent treatment of crystal optics is given in M. Born and E. Wolf, *Principles of Optics* (Pergamon, New York, 1970), 4th ed., Chap. 14.

16 If $\alpha^2 + \beta^2 = \epsilon_3$, Eq. (10c) has the solution $\mathcal{E}_z = Az + B$, where A and B are arbitrary. In the present context we neglect a solution of this form since the set of points $\alpha^2 + \beta^2 = \epsilon_3$ contribute a set of measure zero as far as the integrals in Eq. (7) are concerned.

17 Homogeneous plane waves in this context are those plane waves for which γ_1 or γ_3 (whichever is appropriate) is real.

18 Ref. 15, p. 680.

19 K. Miyamoto and E. Wolf, J. Opt. Soc. Am. **52**, 615 (1962). See particularly the Appendix.

20 M. Born and E. Wolf, Ref. 15, p. 753.

The Angular Spectrum Representation of Electromagnetic Fields in Crystals. II. Biaxial Crystals*

Éamon Lalor†

Department of Physics and Astronomy, University of Rochester, Rochester, N.Y. 14627

(Received 20 October 1970)

The electromagnetic field in a linear, nonmagnetic nonabsorbing biaxial crystal which fills the entire half-space $z \geq 0$ and which has one of its principal dielectric axes perpendicular to the plane $z = 0$ is represented as an angular spectrum of plane waves. The angular spectrum representation consists of a superposition of plane waves expressed as the sum of two integrals. Each integral contains in general both homogeneous and evanescent plane waves. Each plane wave of the angular spectrum (whether homogeneous or evanescent) satisfies the identical equations which are obeyed by the entire field. The spectral amplitudes of the field are explicitly expressed in terms of the Fourier transform of the field in the plane $z = 0$. The special case of a uniaxial crystal whose optic axis is parallel to the plane $z = 0$ is treated in some detail. The far zone structure of the field in such a crystal is determined using the method of stationary phase. The field in the far zone is expressed explicitly in terms of the Fourier transform of the field in the plane $z = 0$.

1. INTRODUCTION

In an earlier paper[1] (which we shall refer to as I), we derived and discussed the angular spectrum representation of electromagnetic fields in the half-space $z \geq 0$, which is assumed to be entirely filled by a linear, nonmagnetic, nonabsorbing uniaxial crystal whose optic axis is perpendicular to the plane $z = 0$.

In the present paper we provide a generalization of this work to cover the mathematically much more difficult case of a biaxial crystal with any one of its principle dielectric axes perpendicular to the plane $z = 0$. The special case of a uniaxial crystal having its optic axis *parallel* to the plane $z = 0$ is discussed in some detail in Sec. 3.

In (B8) and (B9) the prime denotes values at the stationary points.

The roots of Eq. (B7) may be shown to be

$$\alpha' = \frac{\epsilon_3^{1/2} u_x}{\{1 - [1 - (\epsilon_1/\epsilon_3)] u_z^2\}^{1/2}} \tag{B10}$$

and

$$\beta' = \frac{\epsilon_3^{1/2} u_y}{\{1 - [1 - (\epsilon_1/\epsilon_3)] u_z^2\}^{1/2}} . \tag{B11}$$

Hence

$$\mathbf{s}' \cdot \mathbf{u} = \epsilon_3^{1/2} \{1 - [1 - (\epsilon_1/\epsilon_3)] u_z^2\}^{1/2} . \tag{B12}$$

Further calculation leads to the results

$$\Delta = \frac{\{1 - [1 - (\epsilon_1/\epsilon_3)] u_z^2\}^2}{\epsilon_1 u_z^2} \tag{B13}$$

and

$$\epsilon = -1. \tag{B14}$$

The asymptotic approximation to $E_{(h)}^{(ex)}(x,y,z)$ as $kR \to \infty$ therefore is

$$\mathbf{E}_{(h)}^{(ex)}(x,y,z) \sim \frac{-2\pi i \epsilon_1^{1/2}}{k\xi^2} \left(\frac{z}{R}\right) \mathbf{e}^{(ex)}(\alpha', \beta') \times \frac{\exp[ik\epsilon_3^{1/2} \xi R]}{R}, \tag{B15}$$

where α' and β' are given by (B10), (B14), and

$$\xi = \left[1 - \left(1 - \frac{\epsilon_1}{\epsilon_3}\right)\left(\frac{z}{R}\right)^2\right]^{1/2} . \tag{B16}$$

---

* Research supported by the Air Force Office of Scientific Research.

† Present address: Department of Physics, University College, Belfield, Dublin 4, Ireland.

1 E. Wolf, Proc. Phys. Soc. (London) **74**, 269 (1959), especially the Appendix.

2 E. Lalor, J. Opt. Soc. Am. **58**, 1235 (1968).

3 A history of the angular spectrum and its application in the treatment of physical problems is given in E. Lalor "The Angular Spectrum Representation of Electromagnetic Fields and Its Use in the Treatment of Interaction Problems," (Ph.D. thesis, University of Rochester, 1970), Chap. I.

4 C. J. Bouwkamp, Rept. Progr. Phys. **17**, 39 (1954), and other works quoted in this report.

5 R. K. Luneburg, *Mathematical Theory of Optics* (U. of California Press, Berkeley, 1966), Chap. VI.

6 M. Kline and I. W. Kay, *Electromagnetic Theory and Geometrical Optics* (Interscience, New York, 1965), Chap. XI.

7 J. A. Ratcliffe, Rept. Progr. Phys. **19**, 188 (1956), and other works quoted in this article.

8 G. Borgiotti, IEEE Trans. Antennas Propagation (Communications) **11**, 94 (1967); Alta Frequenza **32**, 196 (1963); IEEE Trans. Antennas Propagation **15**, 565 (1967).

9 D. R. Rhodes, Proc. IEEE **52**, 1013 (1964); IEEE Trans. Antennas Propagation **14**, 676 (1966).

10 G. C. Sherman, J. Opt. Soc. Am. **57**, 1160 (1967).

11 R. Mittra and P. L. Ransom, in *Proceedings of the Symposium on Modern Optics*, edited by J. Fox (Polytechnic, Brooklyn, N.Y., 1967), p. 619.

12 E. Wolf and J. R. Shewell, J. Math. Phys. **11**, 2254 (1970).

13 A more complete list of topics covered and of contributions within each separate topics is given in Ref. 3.

14 See in this context P. C. Clemmow, *The Plane Wave Spectrum Representation of Electromagnetic Fields* (Pergamon, New York, 1966).

15 An excellent treatment of crystal optics is given in M. Born and E. Wolf, *Principles of Optics* (Pergamon, New York, 1970), 4th ed., Chap. 14.

16 If $\alpha^2 + \beta^2 = \epsilon_3$, Eq. (10c) has the solution $\mathcal{E}_z = Az + B$, where A and B are arbitrary. In the present context we neglect a solution of this form since the set of points $\alpha^2 + \beta^2 = \epsilon_3$ contribute a set of measure zero as far as the integrals in Eq. (7) are concerned.

17 Homogeneous plane waves in this context are those plane waves for which γ_1 or γ_3 (whichever is appropriate) is real.

18 Ref. 15, p. 680.

19 K. Miyamoto and E. Wolf, J. Opt. Soc. Am. **52**, 615 (1962). See particularly the Appendix.

20 M. Born and E. Wolf, Ref. 15, p. 753.

The Angular Spectrum Representation of Electromagnetic Fields in Crystals. II. Biaxial Crystals*

Éamon Lalor†

Department of Physics and Astronomy, University of Rochester, Rochester, N.Y. 14627

(Received 20 October 1970)

The electromagnetic field in a linear, nonmagnetic nonabsorbing biaxial crystal which fills the entire half-space $z \geq 0$ and which has one of its principal dielectric axes perpendicular to the plane $z = 0$ is represented as an angular spectrum of plane waves. The angular spectrum representation consists of a superposition of plane waves expressed as the sum of two integrals. Each integral contains in general both homogeneous and evanescent plane waves. Each plane wave of the angular spectrum (whether homogeneous or evanescent) satisfies the identical equations which are obeyed by the entire field. The spectral amplitudes of the field are explicitly expressed in terms of the Fourier transform of the field in the plane $z = 0$. The special case of a uniaxial crystal whose optic axis is parallel to the plane $z = 0$ is treated in some detail. The far zone structure of the field in such a crystal is determined using the method of stationary phase. The field in the far zone is expressed explicitly in terms of the Fourier transform of the field in the plane $z = 0$.

1. INTRODUCTION

In an earlier paper[1] (which we shall refer to as I), we derived and discussed the angular spectrum representation of electromagnetic fields in the half-space $z \geq 0$, which is assumed to be entirely filled by a linear, nonmagnetic, nonabsorbing uniaxial crystal whose optic axis is perpendicular to the plane $z = 0$.

In the present paper we provide a generalization of this work to cover the mathematically much more difficult case of a biaxial crystal with any one of its principle dielectric axes perpendicular to the plane $z = 0$. The special case of a uniaxial crystal having its optic axis *parallel* to the plane $z = 0$ is discussed in some detail in Sec. 3.

2. DERIVATION OF THE ANGULAR SPECTRUM REPRESENTATION

Monochromatic electric and magnetic fields

$$\mathbf{E}(\mathbf{r}, t) = \mathbf{E}_0(\mathbf{r}) \exp(-i\omega t) \tag{1a}$$

and

$$\mathbf{H}(\mathbf{r}, t) = \mathbf{H}_0(\mathbf{r}) \exp(-i\omega t), \tag{1b}$$

in a (macroscopically) homogeneous, linear, nonmagnetic, nonabsorbing anisotropic dielectric filling a volume V obey Maxwell's equations in the form[2]

$$\text{curl } \mathbf{H}_0(\mathbf{r}) + (i\omega/c)\epsilon \cdot \mathbf{E}_0(\mathbf{r}) = 0, \tag{2a}$$

$$\text{curl } \mathbf{E}_0(\mathbf{r}) - (i\omega/c)\mathbf{H}_0(\mathbf{r}) = 0, \tag{2b}$$

$$\text{div } [\epsilon \cdot \mathbf{E}_0(\mathbf{r})] = 0, \tag{2c}$$

$$\text{div } \mathbf{H}_0(\mathbf{r}) = 0, \tag{2d}$$

inside the volume V. We have made use of the relationship

$$\mathbf{D}_0 = \epsilon \cdot \mathbf{E}_0 \tag{3}$$

between the electrical displacement $\mathbf{D}_0$ and the electric vector $\mathbf{E}_0$ where ϵ is the (real) dielectric tensor of the medium. Upon substituting Eq. (2b) into Eq. (2a), we see that $\mathbf{E}_0(r)$ satisfies the equation

$$\text{curl curl } \mathbf{E}_0(\mathbf{r}) - k^2 \epsilon \cdot \mathbf{E}_0(\mathbf{r}) = 0, \tag{4}$$

where

$$k = \omega/c. \tag{5}$$

If we choose as our coordinate axes the principal dielectric axes of the medium, the vector $\epsilon \cdot \mathbf{E}_0$ takes the form $(\epsilon_1 E_{ox}, \epsilon_2 E_{oy}, \epsilon_3 E_{oz})$. If we now assume that no two of the components ϵ_1, ϵ_2, and ϵ_3 are equal and that the volume V is the entire half-space $z \geq 0$, we are discussing a semi-infinite biaxial crystal one of whose principal dielectric axes is perpendicular to the face of the crystal.

Let us now represent the electric field $\mathbf{E}_0(x, y, z)$ as a two-dimensional Fourier integral with respect to the first two variables, i.e.,

$$\mathbf{E}_0(x, y, z) = \iint_{-\infty}^{\infty} \epsilon(\alpha, \beta; z) \exp[ik(\alpha x + \beta y)] d\alpha d\beta. \tag{6}$$

If we make use of Eq. (6) and the Fourier inversion formula, we may rewrite Eq. (2c) as follows

$$ik\alpha\epsilon_1 \mathcal{E}_x + ik\beta\epsilon_2 \mathcal{E}_y + \epsilon_3 \frac{\partial \mathcal{E}_z}{\partial z} = 0. \tag{7}$$

Using Eq. (4), Eq. (6), the Fourier inversion formula, and Eq. (7), we may derive the following set of coupled differential equations for the three components of the vector function $\mathcal{E}(\alpha, \beta; z)$;

$$\frac{\partial^2 \mathcal{E}_x}{\partial z^2} + A_1 \mathcal{E}_x + B_2 \mathcal{E}_y = 0, \tag{8}$$

$$\frac{\partial^2 \mathcal{E}_y}{\partial z^2} + A_2 \mathcal{E}_y + B_1 \mathcal{E}_x = 0, \tag{9}$$

and

$$\frac{\partial^2 \mathcal{E}_z}{\partial z^2} + A_3 \mathcal{E}_z - \frac{iB_1}{k\beta} \frac{\partial \mathcal{E}_x}{\partial z} - \frac{iB_2}{k\alpha} \frac{\partial \mathcal{E}_y}{\partial z} = 0, \tag{10}$$

where

$$A_1 = k^2 \{\epsilon_1 - [\alpha^2(\epsilon_1/\epsilon_3) + \beta^2]\}, \tag{11a}$$

$$A_2 = k^2 \{\epsilon_2 - [\alpha^2 + (\epsilon_2/\epsilon_3)\beta^2]\}, \tag{11b}$$

$$A_3 = k^2 [\epsilon_3 - (\alpha^2 + \beta^2)], \tag{11c}$$

$$B_1 = k^2 \alpha\beta [1 - (\epsilon_1/\epsilon_3)], \tag{11d}$$

and

$$B_2 = k^2 \alpha\beta [1 - (\epsilon_2/\epsilon_3)]. \tag{11e}$$

Equations (8)–(10) are coupled in a more complicated way than the corresponding equations for uniaxial crystals [see I, Eqs. (10)] and their solution is not quite so straightforward. The method of solution adopted here is as follows: Eq. (8) is first solved for $\mathcal{E}_x$ in terms of $\mathcal{E}_y$, the solution is substituted into Eq. (9) and the resulting equations may be solved immediately since it depends only on $\mathcal{E}_y$. Once we have solved for $\mathcal{E}_y$ we may find $\mathcal{E}_x$ by straightforward substitution. Upon substituting both these solutions into Eq. (10) we may then solve the resulting equation for $\mathcal{E}_z$.

The first two steps in the above procedure are outlined in Appendix A where the following equation for $\mathcal{E}_y$ is derived:

$$\frac{\partial^4 \mathcal{E}_y}{\partial z^4} + (A_1 + A_2) \frac{\partial^2 \mathcal{E}_y}{\partial z^2} + (A_1 A_2 - B_1 B_2) \mathcal{E}_y = 0. \tag{12}$$

It is obvious from symmetry considerations that $\mathcal{E}_x$ also satisfies Eq. (12). Equation (12) is a fourth order, linear differential equation with constant coefficients and its solution is straightforward. Following standard practice[3] we factorize Eq. (12) as follows:

$$\left(\frac{\partial^2}{\partial z^2} + k^2\gamma_+^2\right)\left(\frac{\partial^2}{\partial z^2} + k^2\gamma_-^2\right) \mathcal{E}_y = 0, \tag{13}$$

where

$$k^2\gamma_\pm^2 = \tfrac{1}{2}\{A_1 + A_2 \pm [(A_1 - A_2)^2 + 4B_1B_2]^{1/2}\}. \tag{14}$$

The + subscript on γ_+ implies that on the right-hand side of Eq. (14), one takes the positive sign and the − subscript that the negative sign is taken. Equation (13) may be factorized further as follows to yield

$$\left(\frac{\partial}{\partial z} + ik\gamma_+\right)\left(\frac{\partial}{\partial z} - ik\gamma_+\right)\left(\frac{\partial}{\partial z} + ik\gamma_-\right)\left(\frac{\partial}{\partial z} - ik\gamma_-\right) \mathcal{E}_y = 0. \tag{15}$$

The solution of Eq. (15) for $\gamma_+ \neq \gamma_-$ is[4]

$$\mathcal{E}_y = L_1^{(+)} \exp(ik\gamma_+ z) + L_2^{(+)} \exp(ik\gamma_- z) + L_1^{(-)} \exp(-ik\gamma_+ z) + L_2^{(-)} \exp(-ik\gamma_- z) \tag{16}$$

and

$$\mathcal{E}_x = K_1^{(+)} \exp(ik\gamma_+ z) + K_2^{(+)} \exp(ik\gamma_- z) + K_1^{(-)} \exp(-ik\gamma_+ z) + K_2^{(-)} \exp(-ik\gamma_- z), \tag{17}$$

where $L_{1,2}^{(\pm)}$ and $K_{1,2}^{(\pm)}$ are constants for each value of α and β. The form of $\mathcal{E}_z$ may now be found in two different ways. In the first place we may substitute Eqs. (16) and (17) into Eq. (10) and solve for $\mathcal{E}_z$ following, essentially, the argument used in Appendix A of I to solve a different though similar equation. The result is

$$\mathcal{E}_z = M_1^{(+)} \exp(ik\gamma_+ z) + M_2^{(+)} \exp(ik\gamma_- z) + M_1^{(-)} \exp(-ik\gamma_+ z) + M_2^{(-)} \exp(-ik\gamma_- z) + P \exp(iA_3^{1/2} z) + Q \exp(-iA_3^{1/2} z), \quad (18)$$

where $M_1^{(\pm)}, P$, and Q are constants for each value of α and β.

On the other hand we may substitute Eqs. (16) and (17) into Eq. (7) and solve for $\mathcal{E}_z$. We then obtain

$$\mathcal{E}_z = M_1^{(+)} \exp(ik\gamma_+ z) + M_2^{(+)} \exp(ik\gamma_- z) + M_1^{(-)} \exp(-ik\gamma_+ z) + M_2^{(-)} \exp(-ik\gamma_- z) + R, \quad (19)$$

where R is a constant depending possibly on α and β.

For these two equations to be consistent for $k\gamma_+ \neq k\gamma_- \neq A_3^{1/2}$, we require that

$$P = Q = R = 0. \quad (20)$$

The functions $M_1^{(+)}, M_2^{(+)}, M_1^{(-)}, M_2^{(-)}$ are given by

$$M_1^{(+)} = -\left(\frac{\alpha\epsilon_1 K_1^{(+)} + \beta\epsilon_2 L_1^{(+)}}{\gamma_+ \epsilon_3}\right), \quad (21a)$$

$$M_2^{(+)} = -\left(\frac{\alpha\epsilon_1 K_2^{(+)} + \beta\epsilon_2 L_2^{(+)}}{\gamma_- \epsilon_3}\right), \quad (21b)$$

$$M_1^{(-)} = \frac{\alpha\epsilon_1 K_1^{(-)} + \beta\epsilon_2 L_1^{(-)}}{\gamma_+ \epsilon_3}, \quad (21c)$$

$$M_2^{(-)} = \frac{\alpha\epsilon_1 K_2^{(-)} + \beta\epsilon_2 L_2^{(-)}}{\gamma_- \epsilon_3}. \quad (21d)$$

We thus see, on using Eq. (6) and Eqs. (16), (17), (19), and (20), that the field is represented as a superposition of plane waves of the following four types:

$$\mathbf{E}_1^{(+)} = (K_1^{(+)}, L_1^{(+)}, M_1^{(+)}) \exp[ik(\alpha x + \beta y + \gamma_+ z)], \quad (22a)$$

$$\mathbf{E}_2^{(+)} = (K_2^{(+)}, L_2^{(+)}, M_2^{(+)}) \exp[ik(\alpha x + \beta y + \gamma_- z)], \quad (22b)$$

$$\mathbf{E}_1^{(-)} = (K_1^{(-)}, L_1^{(-)}, M_1^{(-)}) \exp[ik(\alpha x + \beta y - \gamma_+ z)], \quad (22c)$$

$$\mathbf{E}_2^{(-)} = (K_2^{(-)}, L_2^{(-)}, M_2^{(-)}) \exp[ik(\alpha x + \beta y - \gamma_- z)]. \quad (22d)$$

The fact that the K and L functions for each wave are not independent of each other may be seen by substituting Eqs. (16) and (17) into Eq. (8) [or Eq. (9)] and comparing the coefficients of each exponential factor. The following relationships may be shown to hold:

$$K_1^{(\pm)} = [B_2/(k^2\gamma_+^2 - A_1)] L_1^{(\pm)} \quad (23a)$$

and

$$K_2^{(\pm)} = [B_2/(k^2\gamma_-^2 - A_1)] L_2^{(\pm)}. \quad (23b)$$

The $(\pm)$ notation stands for two equations, one with the $(+)$ superscript on both sides, the other with the $(-)$ superscript.

We shall retain here the same conditions on the behavior of the field we demanded in I, namely:

(i) The field is bounded at infinity in the right half-space.

(ii) In the right half-space no energy is carried towards the plane $z = 0$, i.e., the Poynting vectors for the allowed waves of types (22a)–(22d) are directed away from the plane $z = 0$.

The algebra involved in the application of these two conditions to the field is quite lengthy. This is [in the case of condition (ii)] a reflection of the extremely complicated relationship between the directions of the wave normals and the ray vectors for biaxial crystals.[5] The detailed analysis is carried out in Appendix B. Here we will only give the results.

The discussion in Appendix B, taken together with Eqs. (6), (16), (17), (19)–(21), implies that the electric field $\mathbf{E}_0$ may be represented as

$$\mathbf{E}_0(x,y,z) = \iint_{-\infty}^{\infty} (K_1, L_1, -\eta_+ \left(\frac{\alpha\epsilon_1 K_1 + \beta\epsilon_2 L_1}{\gamma_+ \epsilon_3}\right) \times \exp[ik(\alpha x + \beta y + \eta_+\gamma_+ z)] d\alpha d\beta + \iint_{-\infty}^{\infty} \left[K_2, L_2, -\eta_- \left(\frac{\alpha\epsilon_1 K_2 + \beta\epsilon_2 L_2}{\gamma_- \epsilon_3}\right)\right] \times \exp[ik(\alpha x + \beta y + \eta_-\gamma_- z)] d\alpha d\beta, \quad (24)$$

where the K and L are related by

$$K_1 = [B_2/(k^2\gamma_+^2 - A_1)] L_1 \quad (25a)$$

and

$$K_2 = [B_2/(k^2\gamma_-^2 - A_1)] L_2. \quad (25b)$$

Further,

$$\eta_+ = \begin{cases} +1 \text{ if the point } (\alpha,\beta) \text{ is in } D_{h1}^{(+)}(\alpha,\beta) \\ \text{or } D_e^{(+)}(\alpha,\beta) \\ -1 \text{ if the point } (\alpha,\beta) \text{ is in } D_{h1}^{(+)}(\alpha,\beta) \end{cases} \quad (26a)$$

and

$$\eta_- = \begin{cases} +1 \text{ if the point } (\alpha,\beta) \text{ is in } D_{h1}^{(-)}(\alpha,\beta) \\ \text{or } D_e^{(-)}(\alpha,\beta) \\ -1 \text{ if the point } (\alpha,\beta) \text{ is in } D_{h2}^{(-)}(\alpha,\beta) \end{cases} \quad (26b)$$

The three domains $D_{h1}^{(+)}, D_{h2}^{(-)}$, and $D_e^{(+)}$, which are defined in Appendix B are mutually exclusive and cover the entire domain of α and β. $D_{h1}^{(+)}$ and $D_{h2}^{(+)}$ are associated with homogeneous waves whose wave normals are in the direction whose direction cosines are proportional to $(\alpha, \beta, \eta_+\gamma_+)$ and whose ray vectors point into the right half-space. $D_e^{(+)}$ is associated with evanescent waves which run up and down the plane $z = 0$ and decay exponentially with increasing z. Similar remarks may be made about $D_{h1}^{(-)}, D_{h2}^{(-)}$, and $D_e^{(-)}$ with $\eta_-\gamma_-$ taking the place of $\eta_+\gamma_+$.

Equation (24) is the required *angular spectrum representation for the electric field in a biaxial crystal*, which has one of its principal dielectric axes perpendicular to the face of the crystal. The angular spectrum representation for the magnetic field may be obtained from Eqs. (24) and (2b).

For Eq. (24) to be a useful representation of the field, it remains only to relate the functions K and L to the boundary value of the field in the plane $z = 0$. This may be accomplished by setting $z = 0$ in Eq. (24) and taking the Fourier inverse. We thereby obtain

$$K_1(\alpha,\beta)+K_2(\alpha,\beta)=\left(\frac{k}{2\pi}\right)^2\iint_{-\infty}^{\infty}E_{0x}(x,y,0)\times\exp[-ik(\alpha x+\beta y)]dxdy \quad (27a)$$

and

$$L_1+(\alpha,\beta)+L_2(\alpha,\beta)=\left(\frac{k}{2\pi}\right)^2\iint_{-\infty}^{\infty}E_{0y}(x,y,0)\times\exp[-ik(\alpha x+\beta y)]dxdy. \quad (27b)$$

We may solve explicitly for K_1 and K_2 by using Eqs. (25a), (25b), (27a), and (27b). The results are

$$K_1=\left[\mathcal{E}_y(\alpha,\beta;0)-\left(\frac{k^2\gamma_-^2-A_1}{B_2}\right)\mathcal{E}_x(\alpha,\beta;0)\right]\frac{B_2}{k^2(\gamma_+^2-\gamma_-^2)}, \quad (28a)$$

and

$$K_2=\left[\mathcal{E}_y(\alpha,\beta;0)-\left(\frac{k^2\gamma_+^2-A_1}{B_2}\right)\mathcal{E}_x(\alpha,\beta;0)\right]\frac{B_2}{k^2[\gamma_-^2-\gamma_+^2]}. \quad (28b)$$

Similarly

$$L_1=\left[\mathcal{E}_x(\alpha,\beta;0)-\left(\frac{B_2}{k^2\gamma_-^2-A_1}\right)\mathcal{E}_y(\alpha,\beta;0)\right]\frac{B_1}{k^2[\gamma_-^2-\gamma_+^2]} \quad (28c)$$

and

$$L_2=\left[\mathcal{E}_x(\alpha,\beta;0)-\left(\frac{B_2}{k^2\gamma_+^2-A_1}\right)\mathcal{E}_y(\alpha,\beta;0)\right]\frac{B_1}{k^2[\gamma_+^2-\gamma_-^2]} \quad (28d)$$

where

$$\mathcal{E}(\alpha,\beta;0)=\left(\frac{k}{2\pi}\right)^2\iint_{-\infty}^{\infty}\mathbf{E}_0(x,y,0)\times\exp[-ik(\alpha x+\beta y)]dxdy. \quad (29)$$

Thus we have succeeded in representing the electric and magnetic fields everywhere in the right half-space in terms of their boundary values in the plane $z=0$. Analysis similar to that carried out in Sec. 3 of I, results in the conclusion that for biaxial crystals all the plane waves of the spectrum are extraordinary waves, whose phase velocity depends on their direction of propagation. There are in general two phase velocities associated with each direction of propagation, one velocity is connected with a plane wave component of the first integral in Eq. (24), the other with a component of the second integral. There are, however, two specific directions, the optic axes of the crystal for which these two velocities are equal. The directions of the optic axes are given by the vectors $\mathbf{s}$ which satisfy the following equation;

$$\mathbf{s}\equiv(\alpha,\beta,\gamma_+)=(\alpha',\beta',\gamma_-), \quad (30)$$

i.e., we wish to find the values of α and β for which

$$\gamma_+=\gamma_- \quad (31)$$

or using Eq. (14),

$$(A_1-A_2)^2+4B_1B_2=0. \quad (32)$$

The solution of this equation for the specific case when $\epsilon_3>\epsilon_2>\epsilon_1$ gives the following two values for $\mathbf{s}$,

$$\mathbf{s}_\pm=\left\{\pm\left[\left(\frac{\epsilon_2-\epsilon_1}{\epsilon_3-\epsilon_1}\right)\epsilon_3\right]^{1/2},0,\left[\left(\frac{\epsilon_3-\epsilon_2}{\epsilon_3-\epsilon_1}\right)\epsilon_1\right]^{1/2}\right\}, \quad (33)$$

i.e., the optic axes lie in the xz-plane and are symmetrically situated about the z axis. The angle β which the axes make with the z axis is given by

$$\tan\beta=\pm\left[\left(\frac{\epsilon_2-\epsilon_1}{\epsilon_3-\epsilon_2}\right)\frac{\epsilon_3}{\epsilon_1}\right]^{1/2}. \quad (34)$$

This is essentially the result quoted in Ref. 2, p. 682.

3. SOME FURTHER DISCUSSIONS OF UNIAXIAL CRYSTALS

In I we derived an angular spectrum representation for the field in a uniaxial crystal whose optic axis is perpendicular to the xy-plane. If we set $\epsilon_1=\epsilon_2$ in the formulae derived in this paper for biaxial crystals, we recover the results of I as one would expect. However, we may do something further—if we set $\epsilon_1=\epsilon_3$ in the formulae of the preceding section, we would expect to obtain the angular spectrum representation for the field in a uniaxial crystal occupying the half-space $z\geq 0$, where now the optic axis lies in the xy-plane.

For $\epsilon_1=\epsilon_3$, Eq. (14) reduces to

$$\gamma_+^2=\epsilon_1-(\alpha^2+\beta^2)\quad[\equiv\gamma_0^2] \quad (35)$$

and

$$\gamma_-^2=\epsilon_2-[\alpha^2+(\epsilon_2/\epsilon_1)\beta^2]\quad[\equiv\gamma_e^2], \quad (36)$$

An examination of the functions $f(\alpha,\beta)$ and $g(\alpha,\beta)$ (see Appendix B), shows that for all points in $D_h^{(+)}(\alpha,\beta)$,

$$f(\alpha,\beta)>0\quad\text{and}\quad g(\alpha,\beta)>0; \quad (37)$$

thus $D_{h2}^{(+)}(\alpha,\beta)$ and $D_{h2}^{(-)}(\alpha,\beta)$ are null sets; this result implies that the Poynting vector of a plane wave whose wave vector points into the right half-space also points into the right half-space, when the plane $z=0$ contains the optic axis. We have therefore for this case,

$$\eta_\pm\equiv 1,\quad\text{all }\alpha\text{ and }\beta. \quad (38)$$

Equation (24) may now be shown to reduce to

$$\mathbf{E}_0(x,y,z)=\iint_{-\infty}^{\infty}K_1(\alpha,\beta)\left(1,0,-\frac{\alpha}{\gamma_0}\right)\times\exp[ik(\alpha x+\beta y+\gamma_0 z)]d\alpha d\beta+\iint_{-\infty}^{\infty}K_2(\alpha,\beta)\left(1,-\frac{\epsilon_1-\beta^2}{\alpha\beta},\frac{\gamma_e}{\alpha}\right)\times\exp[ik(\alpha x+\beta y+\gamma_e z)]d\alpha d\beta, \quad (39)$$

where

$$\gamma_0=[\epsilon_1-(\alpha^2+\beta^2)]^{1/2}\quad\text{for }\epsilon_1\geq\alpha^2+\beta^2$$
$$=+i[\alpha^2+\beta^2-\epsilon_1]^{1/2}\quad\text{for }\epsilon_1<\alpha^2+\beta^2 \quad (40)$$

and

$$\gamma_e=\left[\epsilon_2-\left(\alpha^2+\frac{\epsilon_2}{\epsilon_1}\beta^2\right)\right]^{1/2}\quad\text{for }1\geq\frac{\alpha^2}{\epsilon_2}+\frac{\beta^2}{\epsilon_1}$$
$$=+i\left(\alpha^2+\frac{\epsilon_2}{\epsilon_1}\beta^2-\epsilon_2\right)^{1/2}\quad\text{for }1<\frac{\alpha^2}{\epsilon_2}+\frac{\beta^2}{\epsilon_1}. \quad (41)$$

Furthermore,

$$K_2(\alpha,\beta) = -\left(\frac{\alpha\beta}{\epsilon_1-\beta^2}\right)\left(\frac{k}{2\pi}\right)^2 \iint_{-\infty}^{\infty} E_{0y}(x,y,0) \times \exp[-ik(\alpha x+\beta y)]\,dxdy \quad (42)$$

and

$$K_1(\alpha,\beta) = \left(\frac{k}{2\pi}\right)^2 \iint_{-\infty}^{\infty} E_{0x}(x,y,0)\exp[-ik(\alpha x+\beta y)]\,dxdy + \left(\frac{\alpha\beta}{\epsilon_1-\beta^2}\right)\left(\frac{k}{2\pi}\right)^2 \iint_{-\infty}^{\infty} E_{0y}(x,y,0) \times \exp[-ik(\alpha x+\beta y)]\,dxdy. \quad (43)$$

The representation for the magnetic field may be found in a straightforward manner using Eq. (2b). It is easily shown that the first integral in Eq. (39) is a superposition of ordinary waves and the second integral a superposition of extraordinary waves.

The far zone behavior of the field may be found by a stationary phase calculation similar to that performed in I. The result for the electric field is

$$\mathbf{E}_0(x,y,z) \sim -\frac{2\pi i\epsilon_1^{1/2}}{k}\left(\frac{z}{R}\right)\mathbf{e}^{\mathrm{or}}\left(\epsilon_1^{1/2}\frac{x}{R},\epsilon_1^{1/2}\frac{y}{R}\right) \times \frac{\exp(ik\epsilon_1^{1/2}R)}{R} - \frac{2\pi i\epsilon_1^{1/2}}{k\zeta^2}\left(\frac{z}{R}\right)\mathbf{e}^{\mathrm{ex}}(\alpha',\beta') \times \frac{\exp[ik\epsilon_3^{1/2}\zeta R]}{R}, \quad (44)$$

where

$$\zeta = \left[1-\left(1-\frac{\epsilon_1}{\epsilon_2}\right)\left(\frac{y}{R}\right)^2\right]^{1/2}, \quad (45)$$

$$\alpha' = \frac{\epsilon_2^{1/2}(x/R)}{\zeta}, \quad (46a)$$

and

$$\beta' = \frac{\epsilon_1(y/R)}{\epsilon_2^{1/2}\zeta}. \quad (46b)$$

Furthermore, we have defined

$$\mathbf{e}^{(\mathrm{or})}(\alpha,\beta) \equiv K_1(\alpha,\beta)(1,0,-\alpha/\gamma_0) \quad (47a)$$

and

$$\mathbf{e}^{(\mathrm{ex})}(\alpha,\beta) \equiv K_2(\alpha,\beta)\,(1,-(\epsilon_1-\beta^2)/\alpha\beta,\gamma_e/\alpha). \quad (47b)$$

Thus in the far zone, the field behaves as the superposition of an outgoing (nonuniform) spherical wave and an outgoing wave whose surfaces of constant phase are given by the equation

$$x^2/\epsilon_1 + y^2/\epsilon_2 + z^2/\epsilon_1 = \text{const.} \quad (48)$$

Equation (48) is the equation of an ellipsoid related to the ellipsoid of wave normals.[2]

4. CONCLUSION

We have shown that the electromagnetic field in a biaxial crystal, which is linear, nonmagnetic and nonabsorbing, and which occupies the entire half-space $z \geq 0$, and which furthermore has any one of its three principal dielectrics axes perpendicular to the plane $z = 0$, may be represented as an angular spectrum of plane waves, provided only that the field have an invertible Fourier integral representation in the plane $z = 0$ [Eq. (6)] and that it obeys the two physically reasonable boundary conditions (i) and (ii). The angular spectrum representation of the electric field [Eq. (24)] consists of a superposition of plane waves expressed as the sum of two integrals, each integral containing in general both homogeneous and evanescent plane waves. Each plane wave (whether homogeneous or evanescent) in each of the integrals obeys the identical equations satisfied by the entire field; the angular spectrum representation we have derived is thus a mode expansion of the field. We have furthermore explicitly expressed the spectral amplitudes of the field in terms of the Fourier transform of the field in the boundary plane $z = 0$ [Eqs. (27) and (28)].

In Sec. 3 we specialized the results already obtained to consider the case of a uniaxial crystal whose optic axis is parallel to the plane $z = 0$. Applying the method of stationary plane to the integrals of the representation we have shown that in the far zone the structure of the field is that of the superposition of an outgoing (nonuniform) spherical wave and an outgoing (nonuniform) ellipsoidal wave. The amplitudes of these waves for each point on the wave surface is given in terms of the Fourier transform of the field in the plane $z = 0$.

In another paper[6] we shall make use of the angular spectrum representation of the field in a biaxial crystal derived in this paper to treat the problem of internal conical refraction.

ACKNOWLEDGMENTS

The author would like to thank Professor Emil Wolf and Dr. Robin Asby for helpful discussions relating to this work.

APPENDIX A: DERIVATION OF EQ. (12)

We wish to solve the equations

$$\frac{\partial^2\mathcal{E}_x}{\partial z^2} + A_1\mathcal{E}_x + B_2\mathcal{E}_y = 0, \quad (A1)$$

$$\frac{\partial^2\mathcal{E}_y}{\partial z^2} + A_2\mathcal{E}_y + B_1\mathcal{E}_x = 0, \quad (A2)$$

for $\mathcal{E}_x$ and $\mathcal{E}_y$. We first substitute the trial solution

$$\mathcal{E}_x = ue^{iA_1^{1/2}z} \quad (A3)$$

into Eq. (A1) and obtain

$$u'' + 2iA_1^{1/2}u' + B_2\mathcal{E}_y e^{-iA_1^{1/2}z} = 0. \quad (A4)$$

Now substitution of

$$u' = we^{-2iA_1^{1/2}z} \quad (A5)$$

into (A4) results in

$$w' = -B_2\mathcal{E}_y e^{iA_1^{1/2}z}, \quad (A6)$$

which may be integrated to give

$$w = -B\int^z \mathcal{E}_y(z')e^{iA_1^{1/2}z'}dz' + C, \quad (A7)$$

where C is a constant. Substituting from (A5) into (A7), and integrating, we obtain

$$u = -B_2 \int^z e^{-2iA_1^{1/2}z'} \int^{z'} \mathcal{E}_y(z'') e^{iA_1^{1/2}z''} dz'' dz' + \frac{C}{-2iA_1^{1/2}} e^{-2iA_1^{1/2}z} + D, \quad \text{(A8)}$$

where D is a constant. We finally obtain, using (A3),

$$\mathcal{E}_x = -B_2 e^{iA_1^{1/2}z} \int^z e^{-2iA_1^{1/2}z'} \int^{z'} \mathcal{E}_y(z'') \times e^{iA_1^{1/2}z''} dz'' dz' + \frac{iC}{2A_1^{1/2}} e^{-iA_1^{1/2}z} + De^{iA_1^{1/2}z}. \quad \text{(A9)}$$

If we now substitute Eq. (A9) into Eq. (A2), divide through by $e^{iA_1^{1/2}z}$, differentiate each term with respect to z, divide through by $e^{-2iA_1^{1/2}z}$ and again differentiate each term with respect to z, we obtain a linear, fourth order differential equation with constant coefficient for $\mathcal{E}_y$:

$$\frac{\partial^4 \mathcal{E}_y}{\partial z^2} + (A_1 + A_2)\frac{\partial^2 \mathcal{E}_y}{\partial z^2} + (A_1A_2 - B_1B_2)\mathcal{E}_y = 0. \quad \text{(A10)}$$

It is obvious from the symmetric way in which the constants appear in (A10) that if we had started with Eq. (A2) instead of (A1) and performed similar operations, we would have arrived at an equation for $\mathcal{E}_x$ that is identical to (A10).

APPENDIX B: IMPLICATIONS OF TWO PHYSICAL RESTRICTIONS ON THE FIELD IN A BIAXIAL CRYSTAL

This appendix is concerned with the implications for the field in a biaxial crystal of the two conditions (i) and (ii) described in the text

Let us consider first the expression for $\gamma_\pm^2$. Equation (14) may be rewritten with the help of Eqs. (11a), (11b), (11d), and (11e) as

$$\gamma_\pm^2 = \tfrac{1}{2}\left((\epsilon_1 + \epsilon_2) - \alpha^2\left(\frac{\epsilon_1}{\epsilon_3} + 1\right) - \beta^2\left(\frac{\epsilon_2}{\epsilon_3} + 1\right) \pm \left\{\left[(\epsilon_1 - \epsilon_2) - \alpha^2\left(\frac{\epsilon_1}{\epsilon_3} - 1\right) - \beta^2\left(1 - \frac{\epsilon_2}{\epsilon_3}\right)\right]^2 + 4\alpha^2\beta^2\left(1 - \frac{\epsilon_1}{\epsilon_3}\right)\left(1 - \frac{\epsilon_2}{\epsilon_3}\right)\right\}^{1/2}\right). \quad \text{(B1)}$$

We shall now divide the domain of α and β into two mutually exclusive domains $D_h^{(\pm)}$ and $D_e^{(+)}$ (the $^{(\pm)}$ superscript indicates that we are dealing simultaneously with the same considerations for the two different functions γ_+ and γ_-). $D_h^{(\pm)}(\alpha, \beta)$ is the set of points in the $\alpha\beta$-plane for which the right-hand side of Eq. (B1) is real and positive. $D_e^{(\pm)}(\alpha, \beta)$ is that subset of the $\alpha\beta$-plane for which $\gamma_\pm$ is negative or complex. [Complex values of $\gamma_\pm$ may occur when $\epsilon_1 < \epsilon_3 < \epsilon_2$ or when $\epsilon_2 < \epsilon_3 < \epsilon_1$, since in these cases the right-hand side of Eq. (B1) may contain the square root of a negative quantity for some subset of $D_e^{(\pm)}(\alpha, \beta)$]. Points in $D_h^{(\pm)}$ are associated with homogeneous waves since $\gamma_\pm$ is real. Points in $D_e^{(\pm)}$ are associated with waves which either grow or decay exponentially with increasing z. For points in $D_e^{(\pm)}$ we define $\gamma_\pm$ as that square root of $\gamma_\pm^2$ for which the imaginary part of $\gamma_\pm$ is positive. Thus condition (i), which requires that the field remains finite at infinity in the right half-space, requires that we reject solutions of the form $\mathbf{E}_1^{(-)}$ and $\mathbf{E}_2^{(-)}$ [see Eqs. (22)] for points in $D_e^{(\pm)}$ since these waves grow exponentially with increasing z.

We now consider condition (ii) which demands that for points in $D_h^{(\pm)}$ (i.e., for homogeneous waves) we accept, of the four types of waves described in Eqs. (22), only those waves whose Poynting vector is directed away from the plane $z = 0$. The relationship between the directions of the wave normals and the ray vectors is extremely complicated for biaxial crystals.

We need, however, consider only the z components of the Poynting vector and we shall accept those solutions for which this component is positive and reject those for which it is negative. For a plane wave of the form

$$\mathbf{E}(x, y, z) = \mathbf{E}_0 \exp[ik(lx + my + nz)], \quad \text{(B2)}$$

the z component of the time-averaged Poynting vector

$$\langle \mathbf{S} \rangle = (c/8\pi)\mathrm{Re}(\mathbf{E} \times \mathbf{H}^*) \quad \text{(B3)}$$

may be written using Eq. (2b) as

$$\langle S_z \rangle = (c/8\pi)\mathrm{Re}[l(\mathbf{E}\cdot\mathbf{E}^*) - E_z^*(\mathbf{E}\cdot\mathbf{l})], \quad \text{(B4)}$$

where

$$\mathbf{l} = (l, m, n). \quad \text{(B5)}$$

For waves of the form $\mathbf{E}_1^{(+)}$ [see Eq. (22a)], $\langle \mathbf{S}_z \rangle$ becomes after some calculation involving Eqs. (21a), (22a), and (23a),

$$\langle S_z \rangle = \gamma_+(c/8\pi)\,|K_1^{(+)}|^2 f(\alpha, \beta), \quad \text{(B6)}$$

where

$$f(\alpha, \beta) = 1 + \left(\frac{k^2\gamma_+^2 - A_1}{B_2}\right)^2 + \frac{\{\alpha + [(k^2\gamma_+^2 - A_1)/B_2]\beta\}\{\alpha\epsilon_1 + [(k^2\gamma_+^2 - A_1)/B_2]\epsilon_2\beta\}}{\gamma_+^2\epsilon_3}. \quad \text{(B7)}$$

Similarly we may show that for waves of the form $\mathbf{E}_1^{(-)}$ [see Eq. (22c)],

$$\langle S_z \rangle = -\gamma_+(c/8\pi)\,|K_1^{(-)}|^2 f(\alpha, \beta), \quad \text{(B8)}$$

where $f(\alpha, \beta)$ is defined in Eq. (B7). We now divide $D_{h2}^{(+)}$, into the two mutually exclusive domains $D_{h1}^{(+)}$ and $D_{h2}^{(+)}$, where $D_{h1}^{(+)}$ is that subset of $D_h^{(+)}$ for which $f(\alpha, \beta) \geq 0$ and $D_{h2}^{(+)}$ is that subset of $D_h^{(+)}$ for which $f(\alpha, \beta) < 0$. Thus for points in $D_{h1}^{(+)}$ we accept solutions of the form $\mathbf{E}_1^{(+)}$ and reject those of the form $\mathbf{E}_1^{(-)}$, whereas for points in $D_{h2}^{(+)}$ we accept solutions of the form $\mathbf{E}_1^{(-)}$ and reject those of the form $\mathbf{E}_1^{(+)}$. In a similar manner we may calculate $\langle S_z \rangle$ for the waves $\mathbf{E}_2^{(+)}$ and $\mathbf{E}_2^{(-)}$. For the wave $\mathbf{E}_2^{(+)}$, we obtain

$$\langle S_z \rangle = \gamma_-(c/8\pi)\,|K_2^{(+)}|^2 g(\alpha, \beta) \quad \text{(B9)}$$

and for $\mathbf{E}_2^{(-)}$

$$\langle S_z \rangle = -\gamma_-(c/8\pi)\,|K_2^{(-)}|^2 g(\alpha, \beta), \quad \text{(B10)}$$

where

$$g(\alpha,\beta) = 1 + [(k^2\gamma_{\pm}^2 - A_1)/B_2]^2 + \frac{\{\alpha + [(k^2\gamma_{\pm}^2 - A_1)/B_2]\beta\}\{\alpha\epsilon_1 + [(k^2\gamma_{\pm}^2 - A_1)/B_2]\epsilon_2\beta\}}{\gamma_{\pm}^2\epsilon_3} \quad \text{(B11)}$$

We defined $D_{h1}^{(-)}$ as that subset of $D_h^{(-)}$ for which $g(\alpha,\beta) \geq 0$ and $D_{h2}^{(-)}$ as that subset of $D_h^{(-)}$ for which $g(\alpha,\beta) < 0$. Obviously

$$D_h^{(-)} = D_{h1}^{(-)} + D_{h2}^{(-)} \quad \text{(B12)}$$

For points in $D_{h1}^{(-)}$, condition (ii) implies that we accept solutions $\mathbf{E}_2^{(+)}$ and reject solutions $\mathbf{E}_2^{(-)}$, while for points in $D_{h2}^{(-)}$ we accept solutions $E_2^{(-)}$ and reject those of form $E_2^{(+)}$.

* Research supported by the Air Force Office of Scientific Research.

† Present address: Department of Physics, University College, Belfield, Dublin 4, Ireland.

1 E. Lalor, J. Math Phys. **13**, 437(1972) this paper is referred to in the text as I.

2 An excellent treatment of crystal optics is given in M. Born and E. Wolf, *Principles of Optics* (Pergamon, New York, 1970), 4th ed. Chap. 14.

3 E. L. Ince, *Integration of Ordinary Differential Equations* (Oliver and Boyd, Edinburgh, 1956), 7th ed., Chap. 5.

4 The solution for $\gamma_+ = \gamma_-$ has a different form. This solution is not considered here since the set of points in the $\alpha\beta$ plane for which it is valid constitutes a set of measure zero for the integral.

5 M. Born, *Optik* (Springer, Berlin, 1933), pp. 235-37, reprinted 1965.

6 E. Lalor, J. Math. Phys. **13**, 449 (1972).

An Analytical Approach to the Theory of Internal Conical Refraction*

Éamon Lalor†

Department of Physics and Astronomy, University of Rochester, Rochester, New York 14627

(Received 20 October 1970)

An analytical, quantitative treatment of internal conical refraction is presented. The recently derived angular spectrum representation of electromagnetic fields in biaxial crystals is employed. Explicit expressions for the electric field in the far zone of a conically refracted beam of light are found by applying the principle of stationary phase to the integrals of the representation. A close examination of the expressions for the field in the far zone yields all the known results of internal conical refraction including the Poggendorf dark band. The present treatment, however, goes beyond previous studies since, with our expressions, it is possible to investigate the detailed structure of the conically refracted field.

1. INTRODUCTION

In his investigations into the propagation of light in crystalline media on the basis of the theory of the elastic ether, Fresnel derived an equation for the wave surface (the locus at any instant t_1 of a disturbance originated at some previous time t_0 at some particular point r_0) in a biaxial crystal.[1] This Fresnel wave surface is a two-sheeted surface of the fourth degree, which reduces for uniaxial crystals to a sphere and an ellipsoid of revolution. Furthermore, a geometrical construction was proposed, which consisted of drawing appropriately positioned tangent planes to the wave surface, and which led to the determination of the velocities and planes of polarization of the refracted rays within the crystal. This construction was Fresnel's successful generalization of a construction given more than a century earlier by Huygens[2] for determining the directions of the refracted rays in isotropic media and uniaxial crystals.

In 1832 Sir William Rowan Hamilton discovered some remarkable properties of the Fresnel wave surface for biaxial crystals.[3] In the first place he noticed that the wave surface had four conoidal cusps at the points where the optic axes intersect the wave surface. Since to each of these points one may draw an infinite number of tangent planes, Fresnel's construction implies that an appropriately positioned cone of rays external to the crystal would be refracted into a single ray along the optic axis of the crystal. To this phenomenon Hamilton gave the name *external conical refraction*. Further, Hamilton found that there are four planes which are tangent to the wave surface at an infinite number of points constituting a circle of contact. The implications of the Fresnel construction in this case are that an appropriately positioned ray incident on the crystal would be refracted into a complete cone of rays within the crystal. This phenomenon was named *internal conical refraction*. At Hamilton's suggestion, Lloyd[4] performed an experiment in 1833 and with the configuration required for internal conical refraction observed a bright circle of light on a screen placed behind the crystal, thus confirming Hamilton's prediction. The success of this experiment contributed greatly to the general acceptance of the Fresnel wave theory of light.

As the experiments demonstrating internal conical refraction became more refined,[5,6] however, two bright circles of light separated by a dark circle were observed. These observations received no adequate explanation until Voigt[7] in 1905, realizing that any beam of light, no matter how well collimated, consists of a superposition of plane waves whose wave normals have a finite angular spread, calculated the distribution of the rays associated with the wave normals which are slightly inclined to the optic axis. As a result of these calculations he offered a more or less satisfactory explanation of the observed light pattern.

However, as has recently been pointed out,[8] Voigt's treatment is still largely qualitative in nature and may be unsatisfactory for some purposes. In view of the increased recent interest in the phenomenon,[8-10] we consider it desirable to have a complete quantitative analytical description of internal conical refraction using modern techniques. In this paper we pre-

$$g(\alpha,\beta) = 1 + [(k^2\gamma_{\pm}^2 - A_1)/B_2]^2 + \frac{\{\alpha + [(k^2\gamma_{\pm}^2 - A_1)/B_2]\beta\}\{\alpha\epsilon_1 + [(k^2\gamma_{\pm}^2 - A_1)/B_2]\epsilon_2\beta\}}{\gamma_{\pm}^2\epsilon_3} \quad \text{(B11)}$$

We defined $D_{h1}^{(-)}$ as that subset of $D_h^{(-)}$ for which $g(\alpha,\beta) \geq 0$ and $D_{h2}^{(-)}$ as that subset of $D_h^{(-)}$ for which $g(\alpha,\beta) < 0$. Obviously

$$D_h^{(-)} = D_{h1}^{(-)} + D_{h2}^{(-)} \quad \text{(B12)}$$

For points in $D_{h1}^{(-)}$, condition (ii) implies that we accept solutions $\mathbf{E}_2^{(+)}$ and reject solutions $\mathbf{E}_2^{(-)}$, while for points in $D_{h2}^{(-)}$ we accept solutions $E_2^{(-)}$ and reject those of form $E_2^{(+)}$.

* Research supported by the Air Force Office of Scientific Research.

† Present address: Department of Physics, University College, Belfield, Dublin 4, Ireland.

1 E. Lalor, J. Math Phys. **13**, 437(1972) this paper is referred to in the text as I.

2 An excellent treatment of crystal optics is given in M. Born and E. Wolf, *Principles of Optics* (Pergamon, New York, 1970), 4th ed. Chap. 14.

3 E. L. Ince, *Integration of Ordinary Differential Equations* (Oliver and Boyd, Edinburgh, 1956), 7th ed., Chap. 5.

4 The solution for $\gamma_+ = \gamma_-$ has a different form. This solution is not considered here since the set of points in the $\alpha\beta$ plane for which it is valid constitutes a set of measure zero for the integral.

5 M. Born, *Optik* (Springer, Berlin, 1933), pp. 235-37, reprinted 1965.

6 E. Lalor, J. Math. Phys. **13**, 449 (1972).

An Analytical Approach to the Theory of Internal Conical Refraction*

Éamon Lalor†

Department of Physics and Astronomy, University of Rochester, Rochester, New York 14627

(Received 20 October 1970)

An analytical, quantitative treatment of internal conical refraction is presented. The recently derived angular spectrum representation of electromagnetic fields in biaxial crystals is employed. Explicit expressions for the electric field in the far zone of a conically refracted beam of light are found by applying the principle of stationary phase to the integrals of the representation. A close examination of the expressions for the field in the far zone yields all the known results of internal conical refraction including the Poggendorf dark band. The present treatment, however, goes beyond previous studies since, with our expressions, it is possible to investigate the detailed structure of the conically refracted field.

1. INTRODUCTION

In his investigations into the propagation of light in crystalline media on the basis of the theory of the elastic ether, Fresnel derived an equation for the wave surface (the locus at any instant t_1 of a disturbance originated at some previous time t_0 at some particular point r_0) in a biaxial crystal.[1] This Fresnel wave surface is a two-sheeted surface of the fourth degree, which reduces for uniaxial crystals to a sphere and an ellipsoid of revolution. Furthermore, a geometrical construction was proposed, which consisted of drawing appropriately positioned tangent planes to the wave surface, and which led to the determination of the velocities and planes of polarization of the refracted rays within the crystal. This construction was Fresnel's successful generalization of a construction given more than a century earlier by Huygens[2] for determining the directions of the refracted rays in isotropic media and uniaxial crystals.

In 1832 Sir William Rowan Hamilton discovered some remarkable properties of the Fresnel wave surface for biaxial crystals.[3] In the first place he noticed that the wave surface had four conoidal cusps at the points where the optic axes intersect the wave surface. Since to each of these points one may draw an infinite number of tangent planes, Fresnel's construction implies that an appropriately positioned cone of rays external to the crystal would be refracted into a single ray along the optic axis of the crystal. To this phenomenon Hamilton gave the name *external conical refraction*. Further, Hamilton found that there are four planes which are tangent to the wave surface at an infinite number of points constituting a circle of contact. The implications of the Fresnel construction in this case are that an appropriately positioned ray incident on the crystal would be refracted into a complete cone of rays within the crystal. This phenomenon was named *internal conical refraction*. At Hamilton's suggestion, Lloyd[4] performed an experiment in 1833 and with the configuration required for internal conical refraction observed a bright circle of light on a screen placed behind the crystal, thus confirming Hamilton's prediction. The success of this experiment contributed greatly to the general acceptance of the Fresnel wave theory of light.

As the experiments demonstrating internal conical refraction became more refined,[5,6] however, two bright circles of light separated by a dark circle were observed. These observations received no adequate explanation until Voigt[7] in 1905, realizing that any beam of light, no matter how well collimated, consists of a superposition of plane waves whose wave normals have a finite angular spread, calculated the distribution of the rays associated with the wave normals which are slightly inclined to the optic axis. As a result of these calculations he offered a more or less satisfactory explanation of the observed light pattern.

However, as has recently been pointed out,[8] Voigt's treatment is still largely qualitative in nature and may be unsatisfactory for some purposes. In view of the increased recent interest in the phenomenon,[8-10] we consider it desirable to have a complete quantitative analytical description of internal conical refraction using modern techniques. In this paper we pre-

sent such a description. The method used employs the angular spectrum representation of electromagnetic fields in biaxial crystals derived recently by the author.[11]

2. DERIVATION OF INTERNAL CONICAL REFRACTION FROM THE ANGULAR SPECTRUM REPRESENTATION OF THE FIELD

Let us assume that the half-space $z \geq 0$ is entirely filled with a linear, nonmagnetic nonabsorbing biaxial crystal whose principal dielectric axes are in the directions of the Cartesian coordinate axes. In this coordinate system the dielectric tensor ϵ is diagonal[12] and has the three components ϵ_{xx}, ϵ_{yy}, and ϵ_{zz}. Let us write

$$\epsilon_{xx} \equiv \epsilon_1, \quad \epsilon_{yy} \equiv \epsilon_2, \quad \epsilon_{zz} \equiv \epsilon_3 \tag{1}$$

and let us demand, for the sake of definiteness, that

$$\epsilon_3 > \epsilon_2 > \epsilon_1. \tag{2}$$

Under these circumstances it may be shown[11] that the directions of the optic axes are given by the vectors

$$\mathbf{s}_\pm = (\pm\, \alpha_0, 0, \gamma_0), \tag{3}$$

where

$$\alpha_0 = \left[\left(\frac{\epsilon_2 - \epsilon_1}{\epsilon_3 - \epsilon_1}\right)\epsilon_3\right]^{1/2}, \tag{3'a}$$

$$\gamma_0 = \left[\left(\frac{\epsilon_3 - \epsilon_2}{\epsilon_3 - \epsilon_1}\right)\epsilon_1\right]^{1/2}. \tag{3'b}$$

We shall restrict our attention to the optic axis given by the + sign in Eq. (3).

In view of the physical impossibility of generating an electromagnetic field in the form of a plane wave of infinite extent, we shall define a *well collimated beam* in vacuo as a field which may be expressed as an angular spectrum of plane waves containing only homogeneous plane wave components whose wave normals make at most a very small angle with some fixed direction in space.

It may be verified by standard procedures that a plane electromagnetic wave incident on the crystal propagating in the direction given by the unit vector $\mathbf{s}_0$, where

$$\mathbf{s}_0 = (\alpha_0, 0, (1 - \alpha_0^2)^{1/2}), \tag{4}$$

is refracted into a plane wave inside the crystal whose wave normal lies along the optic axis $\mathbf{s}_+$. Furthermore, it may be shown that an incident plane wave inclined at a small angle to $\mathbf{s}_0$ will be refracted into two plane waves inside the crystal each of whose wave normals make a small angle with the optic axis $\mathbf{s}_+$. Thus a well collimated beam centered around the direction $\mathbf{s}_0$ incident on the crystal will be refracted into an angular spectrum of plane waves within the crystal, the wave normal of each component plane wave making a small angle with the optic axis. It is then reasonable to assume that the electric field inside the crystal will be given by an expression of the form of Eq. (24) of Ref. 11, where the boundary values $\mathcal{E}_x(\alpha, \beta; 0)$ and $\mathcal{E}_y(\alpha, \beta; 0)$ vanish for points α and β outside of some small two-dimensional domain D surrounding the point $(\alpha_0, 0)$. Thus, changing slightly the notation of Ref. 11, we may represent the (monochromatic) electric field inside the crystal in the form

$$\begin{aligned}\mathbf{E}_0(x, y, z) &= \iint_{D_+} \mathbf{P}_{(+)}(\alpha, \beta) \exp\left[ikR\left(\alpha\frac{x}{R} + \beta\frac{y}{R} + \gamma_+\frac{z}{R}\right)\right] d\alpha d\beta \\ &\quad + \int_{D_-}\int \mathbf{P}_{(-)}(\alpha, \beta) \exp\left[ikR\left(\alpha\frac{x}{R} + \beta\frac{y}{R} + \gamma_-\frac{z}{R}\right)\right] d\alpha d\beta,\end{aligned} \tag{5}$$

where

$$\mathbf{P}_{(\pm)}(\alpha, \beta) = K_{(\frac{1}{2})}(\alpha, \beta)\left[1, \frac{k^2\gamma_\pm^2 - A_1}{B_2}, -\left(\frac{\alpha\epsilon_1 + \beta\epsilon_2(k^2\gamma_\pm^2 - A_1)/B_2}{\gamma_\pm\epsilon_3}\right)\right]. \tag{6}$$

Equation (6) stands for two equations, one with the upper subscript on both sides of the equality sign and the other with the lower subscript. In Eqs. (6) the symbols are defined as follows (see Ref. 11):

$$K_{(\frac{1}{2})}(\alpha, \beta) = \left[\mathcal{E}_y(\alpha, \beta; 0) - \left(\frac{k^2\gamma_\mp^2 - A_1}{B_2}\right)\mathcal{E}_x(\alpha, \beta; 0)\right]\frac{B_2}{k^2[\gamma_\pm^2 - \gamma_\mp^2]}, \tag{7}$$

where

$$\mathcal{E}(\alpha, \beta; 0) = \left(\frac{k}{2\pi}\right)^2 \int_{-\infty}^{\infty}\iint \mathbf{E}_0(x, y, 0) \times \exp[-ik(\alpha x + \beta y)]\, d\alpha d\beta, \tag{8}$$

i.e., $\mathcal{E}(\alpha, \beta; 0)$ is the Fourier transform of the boundary value of the electric field in the plane $z = 0$; i.e.,

$$k^2\gamma_\pm^2 = \tfrac{1}{2}\{A_1 + A_2 \pm [(A_1 - A_2)^2 + 4B_1B_2]^{1/2}\}, \tag{9}$$

where

$$A_1 = k^2\{\epsilon_1 - [\alpha^2(\epsilon_1/\epsilon_3) + \beta^2]\}, \tag{10}$$

$$A_2 = k^2\{\epsilon_2 - [\alpha^2 + (\epsilon_2/\epsilon_3)\beta^2]\}, \tag{11}$$

$$B_1 = k^2\alpha\beta[1 - (\epsilon_1/\epsilon_3)], \tag{12}$$

and

$$B_2 = k^2\alpha\beta[1 - (\epsilon_2/\epsilon_3)]. \tag{13}$$

In Eqs. (5)–(13), $k = \omega/c$ where ω is the frequency and c the vacumn velocity of the radiation.

The domains of integration $D_\pm$ are small domains in the $\alpha\beta$ plane centered on the point $(\alpha_0, 0)$.[13] The boundary functions $\mathcal{E}_x(\alpha, \beta; 0)$ and $\mathcal{E}_y(\alpha, \beta; 0)$ may be found from a knowledge of the exact structure of the external light beam and the application of the boundary conditions for Maxwell's equations. We shall not make any further specific assumptions about the detailed structure of external beam incident on the crystal; we demand only that $\mathcal{E}_x$ and $\mathcal{E}_y$ are continuous and have a sufficient number of continuous derivatives to ensure the validity of the following analysis.

We shall find it convenient to make the change of variables

$$\begin{aligned}\alpha - \alpha_0 &= \rho\cos\theta, \\ \beta &= (\rho/G^{1/2})\sin\theta,\end{aligned} \tag{14}$$

where

$$G = (\epsilon_3 - \epsilon_2)/(\epsilon_3 - \epsilon_1). \tag{15}$$

Equation (5) then becomes

$$\begin{aligned}\mathbf{E}_0(x,y,z) &= \int_0^{2\pi} d\theta \int_0^{\bar\rho} \mathbf{P}'_{(+)}(\rho,\theta) \\ &\times \exp\left[ikR\left((\alpha_0 + \rho\cos\theta)\frac{x}{R} \right.\right. \\ &\left.\left. + \frac{\rho}{G^{1/2}}\sin\theta\,\frac{y}{R} + \gamma'_+\,\frac{z}{R}\right)\right]\rho d\rho \\ &+ \int_0^{2\pi} d\theta \int_0^{\bar\rho} P'_{(-)}(\rho,\theta)\exp\left[ikR\left((\alpha_0 + \rho\cos\theta)\frac{x}{R}\right.\right. \\ &\left.\left. + (\rho/G^{1/2})\sin\theta\,\frac{y}{R} + \gamma'_-\,\frac{z}{R}\right)\right]\rho d\rho,\end{aligned} \tag{16}$$

where $\bar\rho$ is sufficiently large to cover the area $D_\pm$ but at the same time

$$\bar\rho \ll 1. \tag{17}$$

The primed functions are the unprimed functions written in terms of the new variables defined by Eqs. (14).

We shall now proceed to calculate the approximate value of the field in the far zone by the application of the principle of stationary phase to the integrals in Eq. (16). We shall then examine the structure of the far field in sufficient detail to elucidate the essential aspects of internal conical refraction.

The method of stationary phase makes it possible to obtain the asymptotic approximation to integrals of the type appearing in Eq. (16) as $kR \to \infty$ along a direction specified by the *fixed* unit vector[14]

$$\mathbf{u} = (x/R, y/R, z/R), \qquad z > 0. \tag{18}$$

The result is

$$\begin{aligned}\mathbf{E}_0(x,y,z) \sim{}& \frac{2\pi i\epsilon_+ k}{|\Delta_+^{1/2}|}\,\rho'\mathbf{P}'_{(+)}(\rho',\theta')\,\frac{\exp(ikR\mathbf{t}'_+\cdot\mathbf{u})}{R} \\ &+ \frac{2\pi i\epsilon_- k}{|\Delta_-^{1/2}|}\,\rho'\mathbf{P}'_{(-)}(\rho,\theta')\,\frac{\exp(ikR\mathbf{t}'_-\cdot\mathbf{u})}{R},\end{aligned} \tag{19}$$

where

$$t'_\pm = (\alpha_0 + \rho'\cos\theta', (\rho'/G^{1/2})\sin\theta', \gamma'_\pm(\rho',\theta')). \tag{20}$$

Here ρ' and θ' are those values of ρ and θ which make the phase factors

$$\Phi_\pm = \mathbf{t}_\pm \cdot \mathbf{u} \tag{21}$$

stationary within the domain of integration, i.e., ρ' and θ' are the roots of the equations

$$\frac{\partial\Phi_\pm}{\partial\rho} = \frac{\partial\Phi_\pm}{\partial\theta} = 0. \tag{22}$$

Furthermore

$$\Delta_\pm = \left[\frac{\partial^2\Phi_\pm}{\partial\rho^2}\,\frac{\partial^2\Phi_\pm}{\partial\theta^2} - \frac{\partial^2\Phi_\pm}{\partial\rho\partial\theta}\right] \tag{23}$$

and

$$\epsilon_\pm = \begin{cases} +1 & \text{when } \Delta_\pm > 0,\ \left(\dfrac{\partial^2\Phi_\pm}{\partial\rho^2}\right)' > 0 \\ -1 & \text{when } \Delta_\pm > 0,\ \left(\dfrac{\partial^2\Phi_\pm}{\partial\rho^2}\right)' < 0 \\ -i & \text{when } \Delta_\pm < 0, \end{cases} \tag{24}$$

where the prime in Eqs. (23) and (24) denotes values at the stationary points.

The algebra involved in the solution of Eqs. (22) for the stationary points is lengthy and cumbersome due to the complicated form of $\gamma'_\pm(\rho,\theta)$. We give the detailed derivation in the Appendix and limit our present discussion to the general behavior of the solution.

After lengthy calculation, Eq. (22) yields the relations

$$u_x \simeq (1/\gamma_0)[B\alpha_0 \mp E\alpha_0\cos\theta + \rho J_\pm(\theta)]u_z \tag{25}$$

and

$$u_y \simeq (G^{1/2}/\gamma_0)[\mp E\alpha_0\sin\theta + \rho K_\pm(\theta)]u_z \tag{26}$$

where the qualities $J_\pm(\theta)$ and $K \pm (\theta)$ and the constants B and E are defined in the Appendix.

In deriving the approximate equations (25) and (26) we have neglected powers of ρ higher than the first because of the constraint (17).

Equations (25) and (26) define the stationary points ρ' and θ' for each direction of observation (u_x, u_y, u_z); however if these stationary points lie outside the domain of integration the leading term in the asymptotic expansion of our integral will be of higher order than $1/R$.

Thus Eqs. (25) and (26) may also be viewed as defining those directions in space along which the electric field in the far zone falls off as $1/R$, by demanding that the values of ρ and θ in Eqs. (25) and (26) are in the domain of integration. In all other directions of observation, the field in the far zone falls off faster than $1/R$. Our object is therefore to solve Eqs. (22) for u_x, u_y, and u_z for each point (ρ,θ) in the domain of integration.

Recalling the relation (17), we shall first consider Eqs. (25) and (26) in the zeroth-order approximation where we neglect ρ entirely. These equations may then be written as

$$u_x \simeq 1/\gamma_0(B \mp E\alpha_0\cos\theta)u_z \tag{25'}$$

and

$$u_y \simeq (1/\gamma_0)[\mp (EF)^{1/2}\alpha_0\sin\theta]u_z. \tag{26'}$$

The constant F is defined in the Appendix. Upon eliminating θ between Eqs. (25a) and (26a) we find that

$$\left(\frac{\gamma_0}{\alpha_0 E}\,u_x - \frac{B}{E}\,u_z\right)^2 + \left(\frac{\gamma_0}{\alpha_0(EF)^{1/2}}\,u_y\right)^2 = u_z^2, \tag{27}$$

which is the equation of a cone through the origin. We also have the relation

$$u_x^2 + u_y^2 + u_z^2 = 1. \tag{28}$$

The cone defined by Eq. (27) and the sphere by Eq. (28) intersect in a simple closed curve which we shall call the *critical curve* and is designated by C_0.

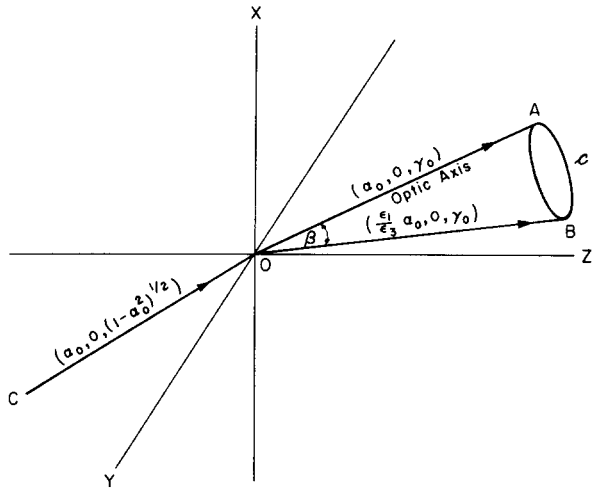


FIG. 1. Conical refraction in the zeroth-order approximation. The plane $z = 0$ is the face of the crystal. CO is the direction of the incident beam. OA and OB are the lines in which the cone [Eq. (27)] intersects the plane $y = 0$. OA is also the optic axis. The directions of these lines are defined by the vectors lying along them. The curve $\mathcal{C}$ is the intersection of the cone and a screen (not shown) placed in the path of the radiation.

We shall make a few general remarks about the cone defined by Eq. (27).

(i) It has its apex in the origin.

(ii) Each plane $z = \text{const}$ intersects it in an ellipse.

(iii) It is symmetrical about the plane $y = 0$.

(iv) It intersects the plane $y = 0$ in two lines whose directions are

$$\mathbf{l}_1 = (\alpha_0, 0, \gamma_0) \tag{29}$$

and

$$\mathbf{l}_2 = [(\epsilon_1/\epsilon_3)\alpha_0, 0, \gamma_0]. \tag{30}$$

The direction $\mathbf{l}_1$ is seen to be the optic axis.

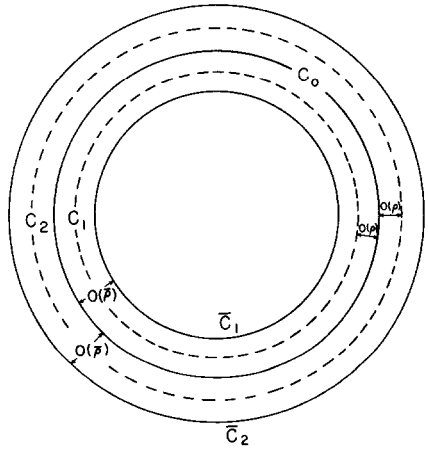


FIG. 2. Illustration of the area occupied by the endpoints of the vector $\mathbf{u}$ associated with stationary points (ρ, θ) inside the domain of integration. C_0 is the critical curve. C_1 and C_2 are the outer limits of the accessible area and are associated with the largest values of ρ which remains within the domain of integration. The dashed curves C_1 and C_2 are the two curves associated with one fixed value of ρ.

(v) The aperature β of the cone in the plane $y = 0$ may be calculated from Eqs. (29) and (30). We obtain

$$\tan\beta = [(\epsilon_3 - \epsilon_2)(\epsilon_2 - \epsilon_1)/(\epsilon_3\epsilon_1)]^{1/2}. \tag{31}$$

Equation (31) is in agreement with the result quoted in the literature.[12] The conclusion is, therefore, that a screen placed in the far zone of the field will be illuminated only on a curve where the cone defined by Eq. (27) intersects the screen. Thus all the general features observed in Lloyd's experiment have been verified. The results obtained so far are illustrated in Fig. 1.

In order to understand the fine structure of the radiation pattern, it is necessary to examine the solution to Eqs. (22) in the first order approximation, i.e., we must retain terms independent of ρ and those proportional to the first power of ρ only. These solutions are given as Eqs. (25) and (26). We shall now consider the solutions for u_x, u_y, and u_z for a fixed ρ within the range of integration. In this case it may be shown that the critical curve is replaced by two curves C_1 and C_2. C_1 is inside C_0 and C_2 is outside C_0. Furthermore, the distance from a point on C_0 to a corresponding point on the curves C_1 or C_2 is proportional to ρ. Thus as we vary ρ throughout the domain of integration, we fill in an area whose thickness is proportional to ρ and which is bisected by the curve C_0 (see Fig. 2). Hence a screen placed in the far zone of the field will be illuminated in an annular area and dark elsewhere.

We shall now show that the intensity of the illumination throughout the annular area is by no means constant, but varies in a manner compatible with the experimental observations. The relevant functions to consider in this regard are the amplitude factors $\rho/|\Delta_{\pm}^{1/2}|$ appearing in the expression (19) for the electric field. It is shown in the Appendix that

$$\rho/|\Delta_{\pm}^{1/2}| = \rho^{1/2} M_{\pm}(\theta), \tag{32}$$

to first order, where $M_{\pm}(\theta)$ in a bounded function of θ. Thus we see the origin of the two bright rings separated by a dark ring observed in the experiments; for, close to the critical curve, i.e., for very small ρ the amplitude of the electric field goes to zero, whereas as we move away from the critical curve towards C_1 and C_2, the amplitude grows proportionally to $\rho^{1/2}$ (i.e., the intensity grows proportionally to ρ). We should therefore expect to observe a dark band around the critical curve, which is precisely what is observed.

We shall conclude our analysis of conical refraction with a brief discussion of the polarization of the electric field in the far zone. If we consider the functions $\mathbf{P}'_{(\pm)}(\rho, \theta)$ defined by Eqs. (6) and (7) and we find, keeping only the zeroth order term, i.e., neglecting ρ entirely that

$$\begin{aligned}
&\mathbf{P}'_{+}(\rho, \theta) \\
&\simeq \tfrac{1}{2}\Big(G^{1/2}\mathcal{E}_y(\alpha, \beta; 0)\sin\theta + \mathcal{E}_x(\alpha, \beta; 0)(\cos\theta + 1), \\
&\quad (1 - \cos\theta)\,\mathcal{E}_y(\alpha, \beta; 0) + \frac{1}{G^{1/2}}\mathcal{E}_x(\alpha, \beta; 0)\sin\theta, \\
&\quad -\frac{\alpha_0}{\gamma_0} G^{1/2}\frac{\epsilon_1}{\epsilon_3}\mathcal{E}_y(\alpha, \beta; 0)\sin\theta \\
&\quad -\frac{\alpha_0}{\gamma_0}\frac{\epsilon_1}{\epsilon_3}\mathcal{E}_x(\alpha, \beta; 0)(\cos\theta + 1)\Big)
\end{aligned} \tag{33}$$

and

$$\mathbf{P}'_{-}(\rho,\theta) \simeq \tfrac{1}{2}\Big(-G^{1/2}\mathcal{E}_y(\alpha,\beta;0)\sin\theta + \mathcal{E}_x(\alpha,\beta;0)(1-\cos\theta), \times(\cos\theta+1)\mathcal{E}_y(\alpha,\beta;0) - \frac{1}{G^{1/2}}\mathcal{E}_x(\alpha,\beta;0)\sin\theta, \times\frac{\alpha_0}{\gamma_0}G^{1/2}\frac{\epsilon_1}{\epsilon_3}\mathcal{E}_y(\alpha,\beta;0)\sin\theta + \frac{\alpha_0}{\gamma_0}\frac{\epsilon_1}{\epsilon_3}\mathcal{E}_x(\alpha,\beta;0)(\cos\theta-1)\Big), \quad (34)$$

where α and β are related to ρ and θ by Eq. (14). The point in the far field for which Eq. (33) holds lies in the directions given by the unit vector $\mathbf{u}$ which is a solution of Eqs. (25′) and (26′) with the minus sign; the point at which Eq. (34) holds is determined by the solution of Eqs. (25′) and (26′) with the plus sign. Thus the orientation of the electric vector is determined at all points in the far zone by Eqs. (25), (26), and the boundary value of the field in the plane $z=0$. Finally it is shown in the Appendix that

$$\epsilon_+ = -i \quad (35)$$

and

$$\epsilon_- = -1. \quad (36)$$

3. DISCUSSION

We have considered a well collimated beam incident in such a manner upon the face of a biaxial crystal that the wave normals of the refracted waves associated with each plane wave component of the incident beam make at most a small angle with one of the optic axes of the crystal. The field inside the crystal was represented as an angular spectrum of plane waves. By applying the principle of stationary phase to the integrals of the representation, we have found explicit expressions for the electric field in the far zone. A close examination of the radiation pattern in the far zone [see Eqs. (19), (32)–(36)] yields all the known results connected with internal conical refraction. We believe that this is the first totally analytical and quantitative analysis of internal conical refraction and that the explicit expressions derived in this paper go beyond the largely qualitative treatments given previously.

APPENDIX: CALCULATIONS RELATING TO THE DERIVATION OF THE ASYMPTOTIC APPROXIMATION TO THE INTEGRALS IN EQ. (16)

The phase factors $\Phi_\pm$ may be written explicitly in the form

$$\Phi_\pm = \alpha u_x + \beta u_y + [A - B\alpha^2 - C\beta^2 \pm ([D + E\alpha^2 - F\beta^2]^2 + 4EF\alpha^2\beta^2)^{1/2}]^{1/2}u_z, \quad (A1)$$

where

$$A = (\epsilon_1+\epsilon_2)/2, \quad (A2a)$$
$$B = (\epsilon_1+\epsilon_2)/2\epsilon_3, \quad (A2b)$$
$$C = (\epsilon_1+\epsilon_3)/2, \quad (A2c)$$
$$D = (\epsilon_1-\epsilon_2)/2, \quad (A2d)$$
$$E = (\epsilon_3-\epsilon_1)/2\epsilon_3, \quad (A2e)$$
$$F = (\epsilon_3-\epsilon_2)/2\epsilon_3. \quad (A2f)$$

With the change of variables

$$\alpha - \alpha_0 = \rho\cos\theta, \quad \beta = (\rho/G)^{1/2}\sin\theta, \quad (A3a)$$

where

$$G = (\epsilon_3-\epsilon_2)/(\epsilon_3-\epsilon_1), \quad (A3b)$$

the phase factors become

$$\Phi_\pm = \alpha_0 u_x + \rho\cos\theta u_x + (\rho/G^{1/2})\sin\theta u_y + \gamma_0[1 - (\rho/\gamma_0^2)(\beta\alpha_0\cos\theta \mp E\alpha_0) - (\rho^2/2\gamma_0^4)(B\alpha_0\cos\theta \mp E\alpha_0)^2 - (B/2\gamma_0^2)\cos^2\theta\rho^2 - (C/2G\gamma_0^2)\sin^2\theta\rho^2 \pm (E/2\gamma_0^2)\rho^2\cos\theta + O(\rho^3)]u_z, \quad (A4)$$

where $O(\rho^3)$ means terms involving the third and higher powers of ρ.

We need the first and second order derivatives of $\Phi_\pm$ with respect to ρ and θ. We have

$$\frac{\partial\Phi_\pm}{\partial\rho} = \cos\theta u_x + [\sin(\theta)/G^{1/2}]u_y + \Big[-\frac{1}{\gamma_0}(B\alpha_0\cos\theta \mp E\alpha_0) - \rho/\gamma_0^3(B\alpha_0\cos\theta \mp E\alpha_0)^2 - (B/\gamma_0)\rho\cos^2\theta - (C/G)\gamma_0\rho\sin^2\theta \pm (E/\gamma_0)\rho\cos\theta + O(\rho^2)\Big]u_z, \quad (A5)$$

and

$$\frac{\partial^2\phi_\pm}{\partial\rho^2} = F_\pm(\theta)u_z, \quad (A6)$$

where

$$F_\pm(\theta) = -1/\gamma_0[(1/\gamma_0^2)(B\alpha_0\cos\theta \mp E\alpha)^2 + B\cos^2\theta + (C/G)\sin^2\theta \pm E\cos\theta + O(\rho)]. \quad (A7)$$

It should be noted that $F_\pm(\theta)$ does not vanish for any value of θ and in fact

$$F_\pm(\theta) < 0. \quad (A8)$$

We also find that

$$\frac{\partial\Phi_\pm}{\partial\theta} = -\rho\sin\theta u_x + (\rho/G^{1/2})\cos\theta u_y + [(\rho/\gamma_0)B\alpha_0\sin\theta + \frac{\rho^2}{\gamma_0^2}B\alpha_0\sin\theta(B\alpha_0\cos\theta \mp E\alpha_0) + (B/\gamma_0)\cos\theta\sin\theta\rho^2 - (C/G\gamma_0)\sin\theta\cos\theta\rho^2 \mp (E/2\gamma_0)\rho^2\sin\theta + O(\rho^3)]u_z, \quad (A9)$$

$$\frac{\partial^2\Phi_\pm}{\partial\theta^2} = -\rho\cos^2\theta u_x - (\rho/G^{1/2})\sin^2\theta + [\rho/\gamma_0 B\alpha_0\cos\theta + O(\rho^2)]u_z \quad (A10)$$

and

$$\frac{\partial^2\Phi}{\partial\rho\partial\theta} = -\sin\theta u_x + \frac{\cos\theta}{G^{1/2}}u_y + \Big(\frac{B\alpha_0}{\gamma_0}\sin\theta + O(\rho)\Big)u_z. \quad (A11)$$

The stationary points of the integrals are defined by those values of ρ and θ which satisfy the equations

$$\frac{\partial\Phi_\pm}{\partial\rho} = \frac{\partial\Phi_\pm}{\partial\theta} = 0. \tag{A12}$$

Using Eqs. (A5) and (A9), Eq. (A12) becomes, on retaining only first-order terms in ρ,

$$\cos\theta u_x + [\sin(\theta)/G^{1/2}]u_y \simeq (1/\gamma_0)[(B\alpha_0\cos\theta \mp E\alpha_0) + (\rho/\gamma_0^2)(B\alpha_0\cos\theta \mp E\alpha_0)^2 + B\rho\cos^2\theta + (C/G)\rho\sin^2\theta \mp E\rho\cos\theta]u_z \tag{A13}$$

and

$$-\sin\theta u_x + (\cos\theta/G^{1/2})u_y \simeq -(1/\gamma_0)[B\alpha_0\sin\theta + [\rho B\alpha_0\sin(\theta)/\gamma_0^2](B\alpha_0\cos\theta \mp E\alpha_0) + B\cos\theta \times \sin\theta\rho - (C/G)\sin\theta\cos\theta\rho \mp \tfrac{1}{2}E^2\rho\sin\theta]u_z. \tag{A14}$$

Eliminating u_y from (A13) and u_x from (A14), these equations become

$$u_y \simeq (G^{1/2}/\gamma_0)[\mp E\alpha_0\sin\theta + \rho K_\pm(\theta)]u_z \tag{A15}$$

and

$$u_x \simeq (1/\gamma_0)\Big([B\alpha_0 \mp E\alpha_0\cos\theta] + \rho J_\pm(\theta)\Big)u_z, \tag{A16}$$

where

$$J_\pm(\theta) = \left(\frac{B^2\alpha_0^2}{\gamma_0^2} + \frac{E^2\alpha_0^2}{\gamma_0^2} + B\right)\cos\theta \mp \left(E/2 + \frac{BE\alpha_0^2}{\gamma_0^2}\right)(1 + \cos^2\theta) \tag{A15'}$$

and

$$K_\pm(\theta) = \sin\theta\left[\left(\frac{C}{G} + \frac{E^2\alpha_0^2}{\gamma_0^2}\right) \mp \left(\frac{E}{2} + \frac{BE\alpha_0^2}{\gamma_0^2}\right)\cos\theta\right]. \tag{A16'}$$

Using Eqs. (A10) and (A13) we find

$$\left[\frac{\partial^2\Phi_\pm}{\partial\theta^2}\right]' = \pm\frac{E\alpha_0}{\gamma_0}\rho u_z + O(\rho^2), \tag{A17}$$

where the prime indicates that the derivative is evaluated at the stationary point. Using Eqs. (A11) and (A14) we obtain

$$\left[\frac{\partial^2\Phi_\pm}{\partial\rho\partial\theta}\right]' = O(\rho). \tag{A18}$$

It follows that

$$\Delta_\pm \equiv \left\{\frac{\partial^2\Phi_\pm}{\partial\rho^2}\,\frac{\partial^2\Phi_\pm}{\partial\theta^2} - \left[\frac{\partial^2\Phi_\pm}{\partial\rho\partial\theta}\right]^2\right\}' = \frac{E\alpha_0}{\gamma_0}F_\pm(\theta)\rho u_z^2 + O(\rho^2), \tag{A19}$$

and hence

$$1/|\Delta_\pm^{1/2}| \simeq [M_\pm(\theta)/u_z]\rho^{-1/2} \tag{A20}$$

where

$$M_\pm(\theta) = |\gamma_0/E\alpha_0F_\pm(\theta)|^{1/2}. \tag{A21}$$

Since $F_\pm(\theta) < 0$, we have

$$\Delta_+ < 0 \tag{A22a}$$

and

$$\Delta_- > 0. \tag{A22b}$$

Thus according to Eq. (24),

$$\epsilon_+ = -i \tag{A23a}$$

and

$$\epsilon_- = -1. \tag{A23b}$$

---

* Research supported by the Air Force Office of Scientific Research.

† Present address: Department of Physics, University College, Belfield, Dublin 4, Ireland.

1 An excellent account of the work of Fresnel and much background material is presented in E. T. Whittaker, *History of the Theories of the Aether and Electricity, Vol. I: The Classical Theories* (Nelson, London, 1951), Chap. IV.
2 See Ref. 1, Chap. 1.
3 W. R. Hamilton, Trans. Roy. Irish Acad. **17**, 1 (1833); also *Hamilton's Mathematical Papers*, edited by J. L. Synge and W. Conway (Cambridge U. P., Cambridge, 1931), Vol. I, p. 285.
4 H. Lloyd, Trans. Roy. Irish Acad. **17**, 145 (1833).
5 J. C. Poggendorf, Pogg. Ann. **48**, 461 (1839).
6 W. V. Haidinger, Wien. Ber. (2) **16**, 129 (1854); Pogg. Ann. (4) **96**, 486 (1855).
7 W. Voigt, Phys. Z. **6**, 672, 818 (1905); Ann. Physik **18**, 645 (1905).
8 D. L. Portigal and E. Burstein, J. Opt. Soc. Am. **59**, 1567 (1969).
9 H. Shih and N. Bloembergen, Phys. Rev. **184**, 895 (1969).
10 R. Goyal and H. Prakash, Phys. Letters **31A**, 506 (1970).
11 E. Lalor, J. Math. Phys. **13**, 443 (1972).
12 An excellent treatment of crystal optics is given in M. Born and E. Wolf, *Principles of Optics* (Pergamon, New York, 1970), 4th ed., Chap. 14.
13 The domains $D\pm$ are assumed to be contained within the domains $D_{hl}^{(\pm)}$ defined in Appendix B of Ref. 11.
14 See Ref. 12, Appendix III, Sec. 3.

Internal Labeling and the Group G_2

B. G. Wybourne

Physics Department, University of Canterbury, Christchurch, New Zealand

(Received 4 January 1971)

The states associated with the group–subgroup $R_7 \supset G_2$ are unambiguously representable as the stretched product states of just seven elementary multiplets. This result is used to derive a number of branching rules for the decomposition $R_7 \rightarrow G_2$ in a simpler manner than hitherto available. An example of the decomposition of the spin representations of R_7 is given.

I. INTRODUCTION

Faced with the problem of labeling the orbital states of the atomic f-shell, Racah[1] found it useful to consider the chain of groups

$$R_7 \supset G_2 \supset R_3,$$

where Cartan's exceptional group G_2 occurs as a subgroup of the seven-dimensional rotation group. By using the irreducible representations of R_7 and G_2 as auxiliary labels, Racah was able to uniquely label all the atomic states of the f-shell apart from a few duplications associated with the reduction to R_3 of the (31) and (40) representations of G_2.

Later, Flowers[2] applied the same chain of groups to the labeling of the orbital states of the nuclear f-shell. In this case representations of R_7 arise which, upon restriction to G_2, give rise to duplicated G_2 representations. Furthermore, when the representations of G_2 are restricted to R_3 representations of R_3 with multiplicities of ten or more are common.[3]

Recently Sharp and Lam[4,5] have discussed the internal-labeling problem using the stretched products of elementary multiplets. For each group–subgroup combination they considered, they found that a finite number of elementary multiplets sufficed and the solution to the internal labeling was complete. To date no proof that a finite set will always suffice has been given, though it is clear that a finite number will certainly solve the problem up to irreducible representations of a preassigned degree.

In this paper we establish a finite set of elementary multiplets, for the case of $R_7 \supset G_2$. This set has been used to give a number of branching rules in a simpler form than previously available.[6,7] While we have established that the finite set suffices to generate all known tables of $R_7 \supset G_2$ branching and to generally cover a wide range of special cases, a general proof of the completeness of our set has eluded us. Brief consideration is given to the internal labeling problem for the cases of $R_{14} \supset G_2$ and $G_2 \supset R_3$. For general definitions we follow the paper of Sharp.[5]

II. ELEMENTARY MULTIPLETS FOR $R_7 \supset G_2$

The elementary multiplets required to describe the G_2 content of R_7 representations were initially determined by forming stretched states from the basic representations of R_7 and G_2 and then inspecting tables of branching rules for unaccounted G_2 representations. The set of elementary multiplets was then enlarged until all G_2 representations were accounted for. The final set of elementary multiplets was established as

$$[\tfrac{1}{2}\tfrac{1}{2}\tfrac{1}{2}](00),\ [\tfrac{1}{2}\tfrac{1}{2}\tfrac{1}{2}](10),\ [100](10),\ [110](10),\ [110](11),\ [\tfrac{3}{2}\tfrac{1}{2}\tfrac{1}{2}](11),\ [210](11). \quad (1)$$

Stretched states associated with a given R_7 representation $[\lambda_1\lambda_2\lambda_3]$ and a G_2 representation (u_1u_2) may be formed from the product of the elementary multiplets given in Eq. (1):

$$\{[\tfrac{1}{2}\tfrac{1}{2}\tfrac{1}{2}](00)\}^a\{[\tfrac{1}{2}\tfrac{1}{2}\tfrac{1}{2}](10)\}^b\{[100](10)\}^c\{[110](10)\}^d \times\{[110](11)\}^e\{[\tfrac{3}{2}\tfrac{1}{2}\tfrac{1}{2}](11)\}^f\{[210](11)\}^g, \quad (2)$$

where necessarily

$$\lambda_1 = \tfrac{1}{2}a + \tfrac{1}{2}b + c + d + e + \tfrac{3}{2}f + 2g,$$
$$\lambda_2 = \tfrac{1}{2}a + \tfrac{1}{2}b + d + e + \tfrac{1}{2}f + g, \quad (3)$$
$$\lambda_3 = \tfrac{1}{2}a + \tfrac{1}{2}b + \tfrac{1}{2}f,$$

and

$$u_1 = b + c + d + e + f + g \quad \text{and} \quad u_2 = e + f + g. \quad (4)$$

In applying Eq. (2) certain states become over represented. These redundant states are excluded by the introduction of the subsidiary condition

$$fd = 0. \quad (5)$$

We note that had we confined our attention to only the true representations of R_7, we would have required the the stretched product states

$$\{[100](10)\}^a\{[110](10)\}^b\{[110](11)\}^c\{[111](00)\}^d \times\{[111](10)\}^e\{[111](20)\}^f\{[210](11)\}^g\{[211](11)\}^h \times\{[211](21)\}^i\{[311](22)\}^j \quad (6)$$

with the restrictions

$$e, h, i = 0 \text{ or } 1$$

and

$$hb = hc = he = hi = ei = ej = bi = bj = 0. \quad (7)$$

Equation (2) taken together with Eqs. (3)–(5) generate the known G_2 content of all the representations of R_7 with $6 \geq \lambda_1 \geq \lambda_2 \geq \lambda_3 \geq 0$. The results of the next section would suggest that the above results cover all representations of R_7 and G_2, but the complete proof has eluded us and would obviously involve a rather complex combinatorial treatment.

III. BRANCHING RULES FOR $R_7 \rightarrow G_2$

The general branching rule for making the decomposition $R_7 \rightarrow G_2$ has been given by Judd[6] and Stone.[7] However, a casual inspection of the tables of branchings[3] suggests that for most representations the decompositions can be expressed more simply. Indeed, via elementary properties of the partition of integers,[8,9] it is very easy to establish from Eq. (2) the following $R_7 \rightarrow G_2$ branching rules:

$$[\lambda 00] \rightarrow (\lambda, 0), \quad (8)$$

$$[\lambda\lambda 0] \rightarrow \sum_{\alpha=0}^{\lambda} (\lambda, \alpha), \quad (9)$$

TABLE I. Decomposition of the Spin Representations under $R_7 \supset G_2$.

$D_{[\lambda]}$	$[\lambda_1\lambda_2\lambda_3]$	(u_1u_2)
8	$[\frac{1}{2}\frac{1}{2}\frac{1}{2}]$	(00) + (10)
48	$[\frac{3}{2}\frac{1}{2}\frac{1}{2}]$	(10) + (11) + (20)
112	$[\frac{3}{2}\frac{3}{2}\frac{1}{2}]$	(10) + (11) + (20) + (21)
112	$[\frac{3}{2}\frac{3}{2}\frac{3}{2}]$	(00) + (10) + (20) + (30)
168	$[\frac{5}{2}\frac{1}{2}\frac{1}{2}]$	(20) + (21) + (30)
512	$[\frac{5}{2}\frac{3}{2}\frac{1}{2}]$	(11) + (20) + 2(21) + (22) + (30) + (31)
560	$[\frac{5}{2}\frac{3}{2}\frac{3}{2}]$	(10) + (11) + (21) + (30) + (31) + (40)
720	$[\frac{5}{2}\frac{5}{2}\frac{1}{2}]$	(20) + (21) + (22) + (30) + (31) + (32)
1008	$[\frac{5}{2}\frac{5}{2}\frac{3}{2}]$	(10) + (11) + (20) + (21) + (30) + (31) + (40) + (41)
672	$[\frac{5}{2}\frac{5}{2}\frac{5}{2}]$	(00) + (10) + (20) + (30) + (40) + (50)
448	$[\frac{7}{2}\frac{1}{2}\frac{1}{2}]$	(30) + (31) + (40)
1512	$[\frac{7}{2}\frac{3}{2}\frac{1}{2}]$	(21) + (22) + (30) + 2(31) + (32) + (40) + (41)
1728	$[\frac{7}{2}\frac{3}{2}\frac{3}{2}]$	(20) + (21) + (22) + (30) + (31) + (32) + (40) + (41) + (50)
2800	$[\frac{7}{2}\frac{5}{2}\frac{1}{2}]$	(21) + (22) + (30) + 2(31) + 2(32) + (33) + (40) + (41) + (42)
4096	$[\frac{7}{2}\frac{5}{2}\frac{3}{2}]$	(11) + (20) + 2(21) + (22) + (30) + 2(31) + (32) + (40) + 2(41) + (42) + (50) + (51)
3024	$[\frac{7}{2}\frac{5}{2}\frac{5}{2}]$	(10) + (11) + (20) + (21) + (30) + (31) + (40) + (41) + (50) + (51) + (60)
3080	$[\frac{7}{2}\frac{7}{2}\frac{1}{2}]$	(30) + (31) + (32) + (33) + (40) + (41) + (42) + (43)
4928	$[\frac{7}{2}\frac{7}{2}\frac{3}{2}]$	(20) + (21) + (22) + (30) + (31) + (32) + (40) + (41) + (42) + (50) + (51) + (52)
4752	$[\frac{7}{2}\frac{7}{2}\frac{5}{2}]$	(10) + (11) + (20) + (21) + (30) + (31) + (40) + (41) + (50) + (51) + (60) + (61)
1320	$[\frac{7}{2}\frac{7}{2}\frac{7}{2}]$	(00) + (10) + (20) + (30) + (40) + (50) + (60) + (70)

$$[\lambda\lambda\lambda] \to \sum_{\rho=0}^{2\lambda} (\rho, 0), \tag{10}$$

$$[\lambda\lambda'0] \to \sum_{\alpha=0} \sum_{\beta=\alpha} (\lambda - \alpha, \beta), \tag{11}$$

(here both summations are to the lesser of λ' and $\lambda - \lambda'$)

$$[\lambda\lambda\lambda'] \to \sum_{\alpha=0}^{2\lambda'} \sum_{\beta=0}^{\lambda-\lambda'} (\lambda - \lambda' + \alpha, \beta), \tag{12}$$

$$[\lambda\lambda'\lambda'] \to \sum_{\alpha=0}^{2\lambda'} \sum_{\beta=0} (\lambda + \lambda' - \alpha, \alpha - \beta) \tag{13}$$

(here the summation over β is to the lesser of $\lambda - \lambda'$ and α).

The above results may be checked by the computation of the dimension of the R_7 representations in terms of those of the sum of G_2 representations given in the decomposition. The above rules exhaust the possibilities for all R_7 representations other than those involving partitions into three distinct nonzero parts. The above rules were used to construct the branching rules for the spin representations of R_7 given in Table I. The branching rules given for partitions into three distinct parts were constructed using Eq. (2), as indeed was the entire table prior to the discovery of the above rules.

Duplications of G_2 representations can only arise in those representations of R_7 involving partitions into three distinct nonzero parts. Table I and tables published elsewhere[3] show a number of striking features. For example, in the R_7 representations $[\lambda + 1, \lambda, \lambda - 1]$ no G_2 representation ever occurs more than twice and then only if it is of the form $(\rho, 1)$ with $2\lambda - 1 \geq \rho \geq 2$. Using Eq. (2), we can show that the appropriate branching rule is

$$[\lambda + 1, \lambda, \lambda - 1] \to (11) + \sum_{b=0}^{2\lambda-3} [(b + 2, 0) + 2(b + 2, 1) + (b + 2, 2)] + (2\lambda, 0) + (2\lambda, 1), \tag{14}$$

where necessarily $b \geq 0$. This rule is a special case of the more general rule

$$[\lambda + 1, \lambda, \lambda - x] \to \sum_{\alpha=0}^{2(\lambda-x)} \sum_{\beta=0}^{x-1} (\alpha + x, \beta + 1) + \sum_{\alpha=0}^{2(\lambda-x)} \sum_{\beta=0}^{x} (\alpha + 1 + x, \beta) + \sum_{\alpha=0}^{2(\lambda-x)-1} (\alpha + x + 1, x + 1). \tag{15}$$

This latter rule shows that these representations of R_7 can never yield G_2 representations with multiplicities in excess of two. More complex branching rules can be derived but with greater combinatorial complexity. For these cases the general results of Judd[6] and Stone[7] are relevant.

IV. OTHER GROUP–SUBGROUP COMBINATIONS

Our success in representing the states associated with the $R_7 \supset G_2$ group–subgroup combination might encourage the belief that other combinations involving G_2 could equally well be considered. Judd[10] has recently given some branching rules for the $R_{14} \supset G_2$ combination. In particular, he finds that the (30) and (31) representations of G_2 occur twice in the decomposition of the (1111100) representation of R_{14}. However, there is no way of representing this elementary representation of R_{14} in terms of the stretched product of $R_{14} \supset G_2$ states, and hence the internal labeling problem is insoluble in terms of elementary multiplets.

The group–subgroup combination $G_2 \supset R_3$ is well known with the representations (31) and (40) containing twice repeated R_3 representations. We find that 40 $G_2 \supset R_3$ states are required to write the states arising in the (31) and (41) representations of G_2 as stretched product states. It is simple to see that the multiplet (81)S can not be expressed in terms of simpler multiplets, and hence the number of elementary multiplets required to solve the internal labeling problem for $G_2 \supset R_3$ must be very large if indeed it is finite.

V. CONCLUSION

The states associated with the group–subgroup combination $R_7 \supset G_2$ have been expressed as the stretched product states of just seven elementary multiplets with one restriction. The other group–subgroup combinations involving G_2, apart from that of $G_2 \supset SU_3$ treated by Sharp and Lam,[4] do not seem to be readily amenable to our approach. The branching rules given

yield an added insight into the multiplicity and internal labeling problems. The use of stretched product states in atomic and nuclear calculation could well lead to substantial simplifications since it is generally easiest to compute isoscalar factors for fully stretched states.[11]

[1] G. Racah, Phys. Rev. **76**, 1342 (1949).
[2] B. H. Flowers, Proc. Roy. Soc. (London) **A210**, 497 (1952).
[3] B. G. Wybourne, *Symmetry Principles in Atomic Spectroscopy* (Wiley, New York, 1970).
[4] R. T. Sharp and C. S. Lam, J. Math. Phys. **10**, 2033 (1969).
[5] R. T. Sharp, Proc. Camb. Phil. Soc. **68**, 571 (1970).
[6] B. R. Judd, *Operator Techniques in Atomic Spectroscopy* (McGraw-Hill, New York, 1963).
[7] A. P. Stone, J. Math. Phys. **11**, 29 (1970).
[8] B. G. Wybourne, J. Math. Phys. **10**, 467 (1969).
[9] J. G. Cleary and B. G. Wybourne, J. Math. Phys. **12**, 45 (1971).
[10] B. R. Judd, J. Phys. (Paris) C **31**, 4 (1970).
[11] S. J. Alisauskas and A. P. Jucys, J. Math. Phys. **10**, 2227 (1969).

On the Existence of Relations among the Eigenvalues of Difference Equations

P. Tsilimigras

Nuclear Research Center "Democritos," Aghia Paraskevi Attikis, Athens, Greece

(Received 24 September 1971; Revised Manuscript Received 30 November 1971)

By translating the eigenvalue problem of a difference equation to one of abstract operators, we give a method of finding relations among the eigenvalues of the difference equations.

Let us consider an operator B on any normed space having discrete eigenvalues $\lambda_1, \lambda_2, \lambda_3, \cdots$ with corresponding eigenfunctions $\varphi_1, \varphi_2, \varphi_3, \cdots$. Suppose that there is an operator U such that

$$BU = UF(B) \tag{1}$$

and with the additional property that the $\varphi_i, i = 1, 2, \ldots,$ do not belong to the null space of U. The function F suffices for our application to be a polynomial.

From the above equation (1) we obtain

$$B(U\varphi_i) = F(\lambda_i)(U\varphi_i) \tag{2}$$

Equation (2) means that if λ_i is an eigenvalue of the operator B, then $F(\lambda_i)$ is also an eigenvalue. By continuing in this way we conclude that if we know one eigenvalue λ_i, we can find a sequence of eigenvalues $F(\lambda_i), F(F(\lambda_i)) \cdots$. This sequence can be finite or infinite.

In the case

$$U^n\varphi_i \neq 0, \quad n \geq 1, \tag{3}$$

the eigenvalue λ_i would generate an infinite sequence of eigenvalues.

In the case

$$U^n\varphi_i = 0 \tag{4}$$

for some n, the eigenvalue λ_i would generate a finite sequence of eigenvalues.

A sufficient condition, which we shall use later, in order that (3) be valid is that the point spectrum of the operator U be the empty set.

In many problems of physics we encounter the eigenvalue problem of a difference equation of the form

$$\sum_{i=1}^{N} f(n + k_i) + a(n)f(n) = Ef(n) \tag{5}$$

in the space of square-summable sequences. For simplicity we consider the $k_i, i = 1, \ldots, N$, to be positive integers.

The above equation can be considered the realization of the eigenvalue problem of the operator[1,2]

$$T = \sum_{i=1}^{N} V^{k_i} + A, \tag{6}$$

where V is the unilateral shift operator on a separable Hilbert space H, over the complex field C, with an orthonormal basis $\{e_n\}_1^\infty$, e.g., $Ve_n = e_{n+1}$ and A is a diagonal operator defined as $Ae_n = a(n)e_n$. The form (6) is very convenient for studying relations among the eigenvalues. As an elementary application we shall prove the following theorem.

Theorem: If A is such that $Ae_n = bn + D$ where b and D are arbitrary constants, then for each eigenvalue λ_i of T there is an infinite sequence of eigenvalues $\{\lambda_i + mb\}_{m=1}^\infty$. In particular, if $b \geq 2\sqrt{N}$, the eigenvalues of T form an infinite sequence and $\lambda_{i+1} = \lambda_i + b$.

Proof: Let V be the above-mentioned unilateral shift operator. Then it follows directly that

$$TV = V(T + b) \tag{7}$$

The spectrum of[3] V is the closed unit disk in the complex plane and is purely continuous except for the points $z: |z| < 1$ which belong to the residual spectrum. The point spectrum of V is the empty set. This means that for each eigenfunction φ_i of T

$$V^n\varphi_i \neq 0, \quad n \geq 1. \tag{8}$$

Relations (7) and (8) imply that to each eigenvalue λ_i of T there belongs an infinite sequence of eigenvalues $\{\lambda_i + mb\}_{m=1}^\infty$ and the first part of the theorem is proved.

For the second part we write the operator T in the form $T = T_0 + A$ where $T_0 = \sum_{i=1}^{N} V^{k_i}$. We may assume $b > |D|$ without loss of generality, since addition of a constant does not change the nature of the spectrum. The obvious relations $bn + D > 0$ $\sum_{n=1}^{\infty} (bn + D)^{-2} < \infty$ imply that the operator A^{-1} exists and is a positive definite Hilbert–Schmidt operator.

yield an added insight into the multiplicity and internal labeling problems. The use of stretched product states in atomic and nuclear calculation could well lead to substantial simplifications since it is generally easiest to compute isoscalar factors for fully stretched states.[11]

[1] G. Racah, Phys. Rev. **76**, 1342 (1949).
[2] B. H. Flowers, Proc. Roy. Soc. (London) **A210**, 497 (1952).
[3] B. G. Wybourne, *Symmetry Principles in Atomic Spectroscopy* (Wiley, New York, 1970).
[4] R. T. Sharp and C. S. Lam, J. Math. Phys. **10**, 2033 (1969).
[5] R. T. Sharp, Proc. Camb. Phil. Soc. **68**, 571 (1970).
[6] B. R. Judd, *Operator Techniques in Atomic Spectroscopy* (McGraw-Hill, New York, 1963).
[7] A. P. Stone, J. Math. Phys. **11**, 29 (1970).
[8] B. G. Wybourne, J. Math. Phys. **10**, 467 (1969).
[9] J. G. Cleary and B. G. Wybourne, J. Math. Phys. **12**, 45 (1971).
[10] B. R. Judd, J. Phys. (Paris) C **31**, 4 (1970).
[11] S. J. Alisauskas and A. P. Jucys, J. Math. Phys. **10**, 2227 (1969).

On the Existence of Relations among the Eigenvalues of Difference Equations

P. Tsilimigras

Nuclear Research Center "Democritos," Aghia Paraskevi Attikis, Athens, Greece

(Received 24 September 1971; Revised Manuscript Received 30 November 1971)

By translating the eigenvalue problem of a difference equation to one of abstract operators, we give a method of finding relations among the eigenvalues of the difference equations.

Let us consider an operator B on any normed space having discrete eigenvalues $\lambda_1, \lambda_2, \lambda_3, \cdots$ with corresponding eigenfunctions $\varphi_1, \varphi_2, \varphi_3, \cdots$. Suppose that there is an operator U such that

$$BU = UF(B) \tag{1}$$

and with the additional property that the $\varphi_i, i = 1, 2, \ldots,$ do not belong to the null space of U. The function F suffices for our application to be a polynomial.

From the above equation (1) we obtain

$$B(U\varphi_i) = F(\lambda_i)(U\varphi_i) \tag{2}$$

Equation (2) means that if λ_i is an eigenvalue of the operator B, then $F(\lambda_i)$ is also an eigenvalue. By continuing in this way we conclude that if we know one eigenvalue λ_i, we can find a sequence of eigenvalues $F(\lambda_i), F(F(\lambda_i)) \cdots$. This sequence can be finite or infinite.

In the case

$$U^n\varphi_i \neq 0, \quad n \geq 1, \tag{3}$$

the eigenvalue λ_i would generate an infinite sequence of eigenvalues.

In the case

$$U^n\varphi_i = 0 \tag{4}$$

for some n, the eigenvalue λ_i would generate a finite sequence of eigenvalues.

A sufficient condition, which we shall use later, in order that (3) be valid is that the point spectrum of the operator U be the empty set.

In many problems of physics we encounter the eigenvalue problem of a difference equation of the form

$$\sum_{i=1}^{N} f(n + k_i) + a(n)f(n) = Ef(n) \tag{5}$$

in the space of square-summable sequences. For simplicity we consider the $k_i, i = 1, \ldots, N$, to be positive integers.

The above equation can be considered the realization of the eigenvalue problem of the operator[1,2]

$$T = \sum_{i=1}^{N} V^{k_i} + A, \tag{6}$$

where V is the unilateral shift operator on a separable Hilbert space H, over the complex field C, with an orthonormal basis $\{e_n\}_1^\infty$, e.g., $Ve_n = e_{n+1}$ and A is a diagonal operator defined as $Ae_n = a(n)e_n$. The form (6) is very convenient for studying relations among the eigenvalues. As an elementary application we shall prove the following theorem.

Theorem: If A is such that $Ae_n = bn + D$ where b and D are arbitrary constants, then for each eigenvalue λ_i of T there is an infinite sequence of eigenvalues $\{\lambda_i + mb\}_{m=1}^\infty$. In particular, if $b \geq 2\sqrt{N}$, the eigenvalues of T form an infinite sequence and $\lambda_{i+1} = \lambda_i + b$.

Proof: Let V be the above-mentioned unilateral shift operator. Then it follows directly that

$$TV = V(T + b) \tag{7}$$

The spectrum of[3] V is the closed unit disk in the complex plane and is purely continuous except for the points $z: |z| < 1$ which belong to the residual spectrum. The point spectrum of V is the empty set. This means that for each eigenfunction φ_i of T

$$V^n\varphi_i \neq 0, \quad n \geq 1. \tag{8}$$

Relations (7) and (8) imply that to each eigenvalue λ_i of T there belongs an infinite sequence of eigenvalues $\{\lambda_i + mb\}_{m=1}^\infty$ and the first part of the theorem is proved.

For the second part we write the operator T in the form $T = T_0 + A$ where $T_0 = \sum_{i=1}^{N} V^{k_i}$. We may assume $b > |D|$ without loss of generality, since addition of a constant does not change the nature of the spectrum. The obvious relations $bn + D > 0$ $\sum_{n=1}^{\infty} (bn + D)^{-2} < \infty$ imply that the operator A^{-1} exists and is a positive definite Hilbert–Schmidt operator.

Furthermore, since $\| T_0 \| = \sqrt{N}$ and $b \geq 2\sqrt{N}$ the circles

$$C_k \equiv \{\lambda \mid |\lambda - \mu_k| \leq \| T_0 \|\}, \quad k = 1, 2, \cdots, \quad (9)$$

are mutually disjoint. Here μ_k are the eigenvalues of A, e.g., $\mu_k = bk + D$.

We are now able to use the following theorem[4]:

If T_0 is a bounded operator and A a self-adjoint one with A^{-1} a positive-definite Hilbert–Schmidt operator, then $T = T_0 + A$ has a discrete spectrum. If the norm $\| T_0 \|$ is such that the circles defined by the relation (9) are mutually disjoint, then the eigenvalues of T are contained in the circles C_k with one eigenvalue to each circle. (μ_k are the eigenvalues of A.)

The application of this theorem and the previous result (that to each eigenvalue λ_i of T there belong an infinite sequence $\{\lambda_i + mb\}_{m=1}^{\infty}$ of eigenvalues) implies the relation $\lambda_{i+1} = \lambda_i + b$, and the proof of the theorem is complete.

---

1 P. Deliyannis and E. Ifantis, J. Math. Phys. **10**, 421 (1969).
2 P. Tsilimigras, Z. Angew. Math. Mech. **50**, 317 (1970).
3 P. Halmos, *A Hilbert Space Problem Book* (Van Nostrand, Princeton, N.J., 1967), p. 230.
4 J. Osborn, J. Math. & Phys. **45**, 391 (1966).

Proof of an Existence Condition for Solutions of the Unitarity System of Equations

C. Eftimiu

Department of Physics, University of Missouri, St. Louis, Missouri 63121

(Received 15 September 1971; Revised Manuscript Received 21 November 1971

A previously established set of conditions for the existence of solutions of a nonlinear system, expressing the unitarity of $n \times n$ symmetric matrices, is proven by means of the homotopy invariance theorem.

In a previous paper[1] we considered the system of equations

$$2a_{ij} \sin\phi_{ij} = \sum_k a_{ik}a_{jk} \cos(\phi_{ik} - \phi_{jk}), \quad i,j = 1, 2, \ldots, n, \quad (1)$$

which expresses the unitarity condition for $n \times n$ symmetric matrices

$$U = 1 + iT \quad (2)$$

with

$$T_{ij} = a_{ij}e^{i\phi_{ij}}, \quad a_{ij} > 0, \quad (3)$$

and we established inductively a condition for the moduli a which guarantees the existence of at least one solution set ϕ of the system (1). The purpose of the present note is to provide a proof of that condition, valid for arbitrary n.

As in Paper I, we start by splitting the system (1) into two subsystems. The first involves only the phases of diagonal elements,

$$2a_{ii} \sin\phi_{ii} = \sum_k a_{ik}^2, \quad i = 1, 2, \ldots, n, \quad (4)$$

and obviously requires that

$$2a_{ii} \geq \sum_k a_{ik}^2, \quad i = 1, 2, \ldots, n, \quad (5)$$

or, that

$$1 - \left(1 - \sum_k{}' a_{ik}^2\right)^{1/2} \leq a_{ii} \leq 1 + \left(1 - \sum_k{}' a_{ik}^2\right)^{1/2}, \quad i = 1, 2, \ldots, n, \quad (6)$$

where the prime indicates the absence of the term with $k = i$.

The second subsystem reads

$$2a_{ij} \sin\phi_{ij} = \sum_k a_{ik}a_{jk} \cos(\phi_{ik} - \phi_{jk}), \quad j > i = 1, 2, \ldots, n \quad (7)$$

and can be rewritten as

$$F_{ij}(t) \equiv A_{ij}t_{ij} - B_{ij} - (1 + t_{ij}^2)^{1/2} \sum_k{}'' a_{ik}a_{jk} \times \cos(\phi_{ik} - \phi_{jk}) = 0, \quad j > i = 1, 2, \ldots, n, \quad (8)$$

where

$$t_{ij} \equiv \tan\phi_{ij}, \quad (9)$$

$$A_{ij} = a_{ij}(2 - a_{ii} \sin\phi_{ii} - a_{jj} \sin\phi_{jj}), \quad (10)$$

$$B_{ij} = a_{ij}(a_{ii} \cos\phi_{ii} + a_{jj} \cos\phi_{jj}), \quad (11)$$

and the double prime indicates the absence from sum of the terms with $k = i$ and $k = j$.

We will assume now that conditions (5) or (6) hold, and will also impose the restriction

$$0 \leq \phi \leq \tfrac{1}{2}\pi \quad (12)$$

so that the phases of the diagonal elements are uniquely determined from (4). When these are substituted into (10) and (11), the problem reduces to finding out under what additional conditions on the moduli a the system (8) has a solution in the $\frac{1}{2}n(n-1)$ — dimensional positive cube

$$C = \bigotimes_{i<j} \{\phi_{ij} \mid 0 \leq \phi_{ij} \leq \tfrac{1}{2}\pi\}. \quad (13)$$

A direct inspection of the system (8) readily shows that the system (8) cannot have a solution in C unless $A_{ij} > 0$. This inequality imposes on a_{ii} upper bounds which are more restrictive than those indicated in (6), and we will assume that they hold because we are here only interested in the solutions in C. In particular, this means that $\sin\phi_{ii} = 1$ only when a_{ii} assume their minimum values given in (6). Of course, if these more restrictive upper bounds on a_{ii} do not hold, the system (8) may still have a solution in a domain larger than C, as indicated by the numerical example given in I.

Furthermore, since $\| T_0 \| = \sqrt{N}$ and $b \geq 2\sqrt{N}$ the circles

$$C_k \equiv \{\lambda \mid |\lambda - \mu_k| \leq \| T_0 \|\}, \quad k = 1, 2, \cdots, \quad (9)$$

are mutually disjoint. Here μ_k are the eigenvalues of A, e.g., $\mu_k = bk + D$.

We are now able to use the following theorem[4]:

If T_0 is a bounded operator and A a self-adjoint one with A^{-1} a positive-definite Hilbert–Schmidt operator, then $T = T_0 + A$ has a discrete spectrum. If the norm $\| T_0 \|$ is such that the circles defined by the relation (9) are mutually disjoint, then the eigenvalues of T are contained in the circles C_k with one eigenvalue to each circle. (μ_k are the eigenvalues of A.)

The application of this theorem and the previous result (that to each eigenvalue λ_i of T there belong an infinite sequence $\{\lambda_i + mb\}_{m=1}^{\infty}$ of eigenvalues) implies the relation $\lambda_{i+1} = \lambda_i + b$, and the proof of the theorem is complete.

---

1 P. Deliyannis and E. Ifantis, J. Math. Phys. **10**, 421 (1969).
2 P. Tsilimigras, Z. Angew. Math. Mech. **50**, 317 (1970).
3 P. Halmos, *A Hilbert Space Problem Book* (Van Nostrand, Princeton, N.J., 1967), p. 230.
4 J. Osborn, J. Math. & Phys. **45**, 391 (1966).

Proof of an Existence Condition for Solutions of the Unitarity System of Equations

C. Eftimiu

Department of Physics, University of Missouri, St. Louis, Missouri 63121

(Received 15 September 1971; Revised Manuscript Received 21 November 1971

A previously established set of conditions for the existence of solutions of a nonlinear system, expressing the unitarity of $n \times n$ symmetric matrices, is proven by means of the homotopy invariance theorem.

In a previous paper[1] we considered the system of equations

$$2a_{ij} \sin\phi_{ij} = \sum_k a_{ik}a_{jk} \cos(\phi_{ik} - \phi_{jk}), \quad i,j = 1, 2, \ldots, n, \quad (1)$$

which expresses the unitarity condition for $n \times n$ symmetric matrices

$$U = 1 + iT \quad (2)$$

with

$$T_{ij} = a_{ij}e^{i\phi_{ij}}, \quad a_{ij} > 0, \quad (3)$$

and we established inductively a condition for the moduli a which guarantees the existence of at least one solution set ϕ of the system (1). The purpose of the present note is to provide a proof of that condition, valid for arbitrary n.

As in Paper I, we start by splitting the system (1) into two subsystems. The first involves only the phases of diagonal elements,

$$2a_{ii} \sin\phi_{ii} = \sum_k a_{ik}^2, \quad i = 1, 2, \ldots, n, \quad (4)$$

and obviously requires that

$$2a_{ii} \geq \sum_k a_{ik}^2, \quad i = 1, 2, \ldots, n, \quad (5)$$

or, that

$$1 - \left(1 - \sum_k{}' a_{ik}^2\right)^{1/2} \leq a_{ii} \leq 1 + \left(1 - \sum_k{}' a_{ik}^2\right)^{1/2}, \quad i = 1, 2, \ldots, n, \quad (6)$$

where the prime indicates the absence of the term with $k = i$.

The second subsystem reads

$$2a_{ij} \sin\phi_{ij} = \sum_k a_{ik}a_{jk} \cos(\phi_{ik} - \phi_{jk}), \quad j > i = 1, 2, \ldots, n \quad (7)$$

and can be rewritten as

$$F_{ij}(t) \equiv A_{ij}t_{ij} - B_{ij} - (1 + t_{ij}^2)^{1/2} \sum_k{}'' a_{ik}a_{jk} \times \cos(\phi_{ik} - \phi_{jk}) = 0, \quad j > i = 1, 2, \ldots, n, \quad (8)$$

where

$$t_{ij} \equiv \tan\phi_{ij}, \quad (9)$$

$$A_{ij} = a_{ij}(2 - a_{ii} \sin\phi_{ii} - a_{jj} \sin\phi_{jj}), \quad (10)$$

$$B_{ij} = a_{ij}(a_{ii} \cos\phi_{ii} + a_{jj} \cos\phi_{jj}), \quad (11)$$

and the double prime indicates the absence from sum of the terms with $k = i$ and $k = j$.

We will assume now that conditions (5) or (6) hold, and will also impose the restriction

$$0 \leq \phi \leq \tfrac{1}{2}\pi \quad (12)$$

so that the phases of the diagonal elements are uniquely determined from (4). When these are substituted into (10) and (11), the problem reduces to finding out under what additional conditions on the moduli a the system (8) has a solution in the $\frac{1}{2}n(n-1)$ — dimensional positive cube

$$C = \bigotimes_{i<j} \{\phi_{ij} \mid 0 \leq \phi_{ij} \leq \tfrac{1}{2}\pi\}. \quad (13)$$

A direct inspection of the system (8) readily shows that the system (8) cannot have a solution in C unless $A_{ij} > 0$. This inequality imposes on a_{ii} upper bounds which are more restrictive than those indicated in (6), and we will assume that they hold because we are here only interested in the solutions in C. In particular, this means that $\sin\phi_{ii} = 1$ only when a_{ii} assume their minimum values given in (6). Of course, if these more restrictive upper bounds on a_{ii} do not hold, the system (8) may still have a solution in a domain larger than C, as indicated by the numerical example given in I.

Let us now construct the homotopy

$$F(t;\lambda): C \times [0,1] \subset R^{n(n-1)/2+1} \to R^{n(n-1)/2}, \quad (14)$$

where

$$F_{ij}(t;\lambda) \equiv A_{ij}(\lambda)t_{ij} - B_{ij}(\lambda) - (1 + t_{ij}{}^2)^{1/2} \times \sum_k{}'' a_{ik}a_{jk} \cos\lambda(\phi_{ik} - \phi_{jk}),$$

$$j > i = 1, 2, \ldots, n \quad (15)$$

and $A_{ij}(\lambda), B_{ij}(\lambda)$ are those functions of λ which are obtained from (10) and (11), respectively, when the moduli $a_{kk}(k = i, j)$ are replaced by $\lambda a_{kk} + (1 - \lambda)\bar{a}_{kk}$ where $\bar{a}_{kk}$ are lower ends of the ranges for a_{kk} as given in (6).

Clearly,

$$\begin{aligned} &A_{ij}(1) = A_{ij}, \\ &B_{ij}(1) = B_{ij}, \\ &A_{ij}(0) = \mathfrak{A}_{ij} = a_{ij}(2 - \bar{a}_{ii} - \bar{a}_{jj}) \geq A_{ij}, \\ &B_{ij}(0) = 0. \end{aligned} \quad (16)$$

[The last equation follows from the fact that $B_{ij}(0)$ is the expression (11) in which all a_{kk} are replaced by their minimum values, indicated in (6). For these values, as it can be seen from (4), $\sin\phi_{ii} = 1$ and $\cos\phi_{ii} = 0$.]

The system

$$F(t;1) = 0 \quad (17)$$

coincides with the system (8). On the other hand, the system

$$F_{ij}(t;0) \equiv \mathfrak{A}_{ij}t_{ij} - (1 + t_{ij}^2)^{1/2} \sum_k{}'' a_{ik}a_{jk} = 0, \quad j > i = 1, 2, \ldots, n \quad (18)$$

has one, and only one solution in C:

$$t_{ij} = \Gamma_{ij}(1 - \Gamma_{ij}^2)^{-1/2} \quad (19)$$

if, and only if

$$\Gamma_{ij} \equiv \mathfrak{A}_{ij}^{-1} \sum_k{}'' a_{ik}a_{jk} < 1, \quad j > i = 1, 2, \ldots, n. \quad (20)$$

Let us show now that if (20) holds, $\deg(F(;\lambda), C, 0)$ is constant for $\lambda \in [0, 1]$. To this end, according to the homotopy invariance theorem,[2] we only have to check that the system

$$F(t;\lambda) = 0 \quad (21)$$

has no solutions on the boundary $\dot{C}$ of C for all $\lambda \in [0, 1]$. But on the side[3] $\phi_{ij} = 0$ of $\dot{C}$,

$$-F_{ij}(t;\lambda) = B_{ij}(\lambda) + \sum_k{}'' a_{ik}a_{jk} \cos\lambda(\phi_{ik} - \phi_{jk}). \quad (22)$$

In this expression both the term $B_{ij}(\lambda)$ and the sum are nonnegative. For $1 \geq \lambda > 0$, $B_{ij}(\lambda)$ is strictly positive, while for $\lambda = 0$, $B_{ij}(0) = 0$, but the sum is strictly positive; the expression (22) as a whole may not vanish for any $\lambda \in [0, 1]$.

On the side $\phi_{ij} = \frac{1}{2}\pi$,

$$\lim_{t_{ij}\to\infty} F_{ij}(t;\lambda) = -B_{ij}(\lambda) + [A_{ij}(\lambda) - \sum_k{}'' a_{ik}a_{jk} \times \cos\lambda(\phi_{ik} - \phi_{jk})] \lim_{t_{ij}\to\infty} t_{ij}. \quad (23)$$

For $1 \geq \lambda > 0$, this expression may not vanish because, even if the coefficient of the infinite limit vanished, $B_{ij}(\lambda)$ is strictly positive. Again for $\lambda = 0$, $B_{ij}(0) = 0$, but for this value of λ the coefficient of the limit is strictly positive as a consequence of conditions (20).

We conclude that the system $F(t;\lambda) = 0$ has no solution on the boundary of C for any $\lambda \in [0, 1]$. By the homotopy invariance theorem the topological degree is then constant, and since for $\lambda = 0$ the system has [under the conditions (20)] one and only one solution in C, the degree is plus unity for $\lambda = 0$ as well as for any other value in $[0, 1]$, including $\lambda = 1$. This is equivalent to saying that, if conditions (20) hold, the system (8) has at least one solution in C. Naturally, the degree may still be unity if (8) possesses an odd number of solutions, so that nothing can be inferred about the uniqueness of the solution from this argument alone.

Conditions (20) coincide with conditions (28) of I.

ACKNOWLEDGMENTS

I am grateful to Professor Freeman Dyson for his interest in this problem and to Professor Carl Kaysen for his kind hospitality at the Institute for Advanced Study, where this work was done.

---

[1] C. Eftimiu, J. Math. Phys. **12**, 2047 (1971). Hereafter referred to as I.

[2] See, e.g., J. M. Ortega and W. C. Rheinboldt, *Iterative Solutions of Nonlinear Equations in Several Variables* (Academic, New York, 1970), p. 156.

[3] The norm of an element $x \subset R^m$ is here defined as $\|x\| = \max_i |x_i|$. One may thus alternatively view C as the intersection of the ball $\|x\| \leq \frac{1}{2}\pi$ with the nonnegative cone $x \geq 0$. C would then consist of all elements with $\|x\| = 0$ or $\|x\| = \frac{1}{2}\pi$.

Geodesics and Classical Mechanics on Lie Groups

Robert Hermann*
Mathematics Department, Rutgers University, New Brunswick, New Jersey
(Received 23 November 1970)

Hamilton–Jacobi Theory on Lie groups is discussed, and an error in a previous paper is corrected. A method for "analytically continuing" geodesics from compact to noncompact Lie groups is presented.

1. INTRODUCTION

In Ref. 1 I proved certain facts about geodesics of left-invariant pseudo-Riemannian metrics on the Lie group $SO(3, R)$, and I claimed to deduce from these facts that there were such metrics which were not complete. Jensen and Schäffer pointed out to me, in a private communication, that this deduction is in error, and that one can prove directly that these metrics are in fact complete.

This paper has two goals: first, to discuss some general facts about left-invariant dynamical systems on Lie groups, then to discuss in more detail the case where the Lie groups are $SO(3, R)$ and $SO(2, 1)$, and then to clarify the confusion generated by the claim of Ref. 1. For general differential geometric background, Ref. 2 will be used. Certain of the topics to be presented here concerning the Poisson bracket structure on Lie groups are related to the material presented in Ref. 3.

Now, mechanics on manifolds can be developed from both a Lagrangian and Hamiltonian point of view. The Lagrangian approach was emphasized in Ref. 2, although the Hamiltonian approach was also briefly treated there, and further in Ref. 3. We will then begin with the Lagrangian approach involving the tangent bundle, as an introduction, then switch to the Hamiltonian, cotangent bundle viewpoint. Certain of the facts presented here about "Poisson bracket" structures on Lie groups are of much more general interest than the geodesic problem with which we started.

2. THE EULER EQUATIONS FOR THE GEODESICS

The Euler equations referred to in the title of this section are the equations of motion of a rotating rigid body.[2] In Ref. 2 the differential-geometric meaning of these equations was discussed. Here, we will present the results of that discussion.

Let M be a manifold of dimension n. Choose indices and the summation convention as follows:

$$1 \leq i, j \leq n.$$

Let $T(M)$ denote the tangent bundle to M, and let (ω_i) be a basis for differential 1-forms on M. Consider these forms as 1-forms on $T(M)$ without any change of notation, via the pull-back map π^*, where $\pi: T(M) \to M$ is the fiber space projection map. Let y_i denote the real-valued functions on $T(M)$ defined as follows:

$$y_i(v) = \omega_i(v) \text{ for } v \in T(M).$$

Then, the 1-forms (ω_i, dy_i) form a basis for 1-forms on $T(M)$.

Let $t \to \sigma(t)$, $0 \leq t \leq 1$, be a curve in M. Let $t \to \sigma'(t) \in M_{\sigma(t)}$ denote its tangent vector curve. We then have the following relation:

$$\sigma^*(\omega_i) = y_i(\sigma'(t))dt. \tag{2.1}$$

Let $L: T(M) \to R$ be a real valued function on M. In classical mechanics such a function is called a Lagrangian. Let L_i, L_{n+i} denote the functions on $T(M)$ defined so as to satisfy the following relation:

$$dL = L_i\omega_i + L_{n+i}\,dy_i. \tag{2.2}$$

Let (c_{jki}) be the functions on M [hence also on $T(M)$] which satisfy the following relations:

$$d\omega_i = c_{jki}\,\omega_j \wedge \omega_k, \tag{2.3}$$

$$c_{jki} + c_{kji} = 0. \tag{2.4}$$

Definition: A curve $t \to \sigma(t)$ in M is an *extremal of* L if it satisfies the following set of differential equations:

$$\frac{d}{dt}L_{n+i}(\sigma'(t)) - L_i(\sigma'(t)) + L_{n+k}(\sigma'(t))\,y_j(\sigma'(t))\,c_{jik}(\sigma(t)) = 0. \tag{2.5}$$

(See Ref. 2, p. 172 for an explanation of the relation between these equations and the notion of "extremal" as defined in the calculus of variations and classical mechanics.)

Now, Eqs. (2.5) take a simple form if the following conditions are satisfied:

(i) M is the underlying manifold of a Lie group G, with the (ω_i) a basis for the left-invariant 1-forms on M. (2.6)

(ii) $L_i = 0$. (2.7)

Thus, in case (2.6) is satisfied, the (c_{jki}) are constants —in fact they are just the structure constants of the Lie algebra G of G, relative to its basis which is dual to the (ω_i). Explicitly, if (X_i) are the vector fields on M which are dual to the ω_i, i.e., which satisfy the conditions

$$\omega_i(X_j) = \delta_{ij}, \tag{2.8}$$

then

$$[X_i, X_j] = c_{ijk}X_k. \tag{2.9}$$

Condition (2.7) means that L is a function of the y_i alone, i.e. is a function which is invariant under left translations by the Lie group G. Thus, if (2.6) and (2.7) are satisfied, Eqs. (2.5) reduce to the following equations:

$$\frac{d}{dt}L_{n+i}(y(t)) + L_{n+k}(y(t))\,y_j(t)c_{jik} = 0, \tag{2.10}$$

where

$$y_i(t) = y_i(\sigma'(t)). \tag{2.11}$$

In particular, notice that Eqs. (2. 10) are equations which involve the $y_i(t)$ alone, not directly the coordinates of $\sigma(t)$. Thus, the curves $t \to \sigma(t)$ which solve (2. 5) may be found in two stages: First, find the functions $y_i(t)$ which satisfy the first-order differential equations (2. 10), and then find the $t \to \sigma(t)$ in M which satisfy the first-order equations

$$\sigma^*(\omega_i) = y_i(t)dt. \tag{2.12}$$

Equations (2. 10) are the key equations for this program and are called the *Euler equations*. As a special case, we may consider L of the following quadratic type:

$$L = \tfrac{1}{2} g_{ij} y_i y_j, \tag{2.13}$$

where (g_{ij}) is a real, constant symmetric matrix. Then, the Euler equations (2. 10) take the following form:

$$g_{ij} \frac{dy_j}{dt} + g_{kl} y_l y_j c_{jik} = 0. \tag{2.14}$$

In case $G = SO(3,R)$, these equations have interpretations in both geometry and mechanics. Geometrically, the curves $t \to \sigma(t)$ which satisfy (2. 10) and (2. 11) are the geodesics of the following Riemannian metric on M:

$$ds^2 = g_{ij}\, \omega_i \cdot \omega_j. \tag{2.15}$$

($\cdot$ indicates symmetric product of differential forms.) The metric defined by (2. 15) on M is invariant under the left action of $G = SO(3, R)$ on M. In mechanics, Eqs. (2. 10) and (2. 11) are the equations of motion of a rotating, force-free top, with (g_{ij}) the moment of inertia matrix. In particular, it is known from mechanics that the solutions of (2. 14) can be written down explicitly using elliptic functions. (The Jacobian rather than the Weierstrassian elliptic functions are needed for this purpose.) This viewpoint can be reversed, and Eqs. (2. 14) can be used to study the main properties of the Jacobian elliptic functions. This program was carried out partially by Tricomi[4] in Ref. 2, Chap. 17, and in definitive form by Hille.[5] In this paper, we will use Hille's form of the results to study certain properties of geodesics of left-invariant metrics on Lie groups. However, before treating this specific case, we will discuss further general facts, related to the Poisson bracket structure on Lie groups.

3. THE POISSON BRACKET WITH RESPECT TO A MOVING FRAME

Return to the general situation where M is a manifold, with a basis (ω_i) of 1-forms for M. Let $T^d(M) = N$ be the cotangent bundle to M, i.e. an element $\alpha \in T^d(M)$ is a real-valued linear form: $M_p \to R$ on the tangent space to a point p of M. Let (z_i) be the functions on N which are dual to the ω_i, i.e., which satisfy the following condition:

$$\alpha = z_i(\alpha)\omega_i(p) \quad \text{for } \alpha \in M_p^d \tag{3.1}$$

Let θ be the following 1-form on N, called the *contact 1-form* associated with M:

$$\theta = z_i \omega_i \tag{3.2}$$

[Again, we make no notational distinction between a form on M and its pullback to $T^d(M) = N$ via the fiber space projection map.] Set

$$\Omega = d\theta. \tag{3.3}$$

Then, Ω is a closed 2-form on N of maximal rank that determines a Poisson-bracket, Lie algebra structure $\{\ ,\ \}$ on $F(N)$ in the following way[2,3]:

Given $f \in F(N)$, let Y_f be the vector field such that

$$df = Y_f \lrcorner\, \Omega. \tag{3.4}$$

Set

$$\{f_1, f_2\} = Y_{f_1}(f_2). \tag{3.5}$$

Our goal in this section is to compute an explicit formula for the Poisson bracket, when θ is given by formula (3. 1), and

$$d\omega_i = c_{jki}\omega_j \wedge \omega_k, \tag{3.6}$$

where the (c_{jki}) are functions on M.

Given $f \in F(N)$, let f_i, f_{n+i} be the functions such that

$$df = f_i\omega_i + f_{n+i}\, dz_i. \tag{3.7}$$

Let X_i, X_{n+i} be the vector fields on N such that

$$X_i(f) = f_i, \qquad X_{n+i}(f) = f_{n+i} \quad \text{for } f \in F(N). \tag{3.8}$$

Now, using (3. 1) and (3. 2), we have

$$\Omega = dz_i \wedge \omega_i + z_i c_{jki}\, \omega_j \wedge \omega_k. \tag{3.9}$$

Using (3. 3), (3. 6), and (3. 7), we have

$$X_i(f)\omega_i + X_{n+i}(f)dz_i = X_f \lrcorner\, \Omega = X_f(z_i)\omega_i - \omega_i(X_f)dz_i + 2z_i\, c_{jki}\, \omega_j(X_f)\omega_k \quad \text{[using (3. 8)]}.$$

Thus, we have

$$\omega_i(X_f) = -X_{n+i}(f) \tag{3.10}$$

$$X_k(f) = X_f(z_k) + 2z_i\, c_{jki}\, \omega_i(X_f) = X_f(z_k) - 2z_i c_{jki} X_{n+j}(f) \quad \text{[using (3. 9)]}.$$

Hence,

$$X_f = \omega_i(X_f)X_i + X_f(z_k)X_{n+k} = -X_{n+i}(f)X_i + X_k(f) + 2z_i c_{jki} X_{n+j}(f) X_{n+k}. \tag{3.11}$$

Let $f, f^1 \in F(N)$. Using (3. 4), (3. 7), and (3. 10), we have

$$\{f, f^1\} = -X_f(f^1) = f_{n+i} f_i^1 - f_k f_{n+k}^1 + 2z_i c_{kji} f_{n+j} f_{n+k}^1. \tag{3.12}$$

Formula (3. 12) is the main result of this section. It generalizes the classical formula for Poisson bracket which is found in all mechanics books (and to which it reduces in case $d\omega_i = 0$). We now study its properties in case M is the underlying manifold of a Lie group.

4. THE POISSON BRACKET ON THE COTANGENT BUNDLE OF A LIE GROUP

Keep the notations of Sec. 3. Suppose in addition that M is the underlying manifold of a Lie group G.

Suppose also that G, the Lie algebra of G, is identified with the Lie algebra of vector fields on M that are invariant under left translation by elements of G. Suppose (ω_i) are a basis of 1-forms on M that are invariant under left translation by G. Then the (c_{jki}) are constants. Let (X_i) be the vector fields on M that satisfy

$$\omega_i(X_j) = \delta_{ij}.$$

Then, the X_i are a basis for G, and

$$[X_i, X_j] = 2c_{ijk}X_k, \tag{4.1}$$

i.e., the c_{ijk} are the structure constants of the Lie algebra G with respect to the basis (X_i).

Suppose that f, f^1 are functions on $N = T^d(M)$ that are functions of the z_i alone, i.e.,

$$0 = f_i = f_i^1 = X_i(f) = X_i^1(f). \tag{4.2}$$

Then, using (3.11), we have

$$\{f, f^1\} = 2z_i c_{jki} f_{n+j} f^1_{n+k}. \tag{4.3}$$

In particular, we have

$$\{z_j, z_k\} = 2c_{jki}z_i, \tag{4.4}$$

which are the same as relations (4.1).

We can explain this relation in basis-free terms in the following way. The fiber of $N = T^d(M)$ above the identity element e of $M = G$ can be identified with $\mathbf{G}^d$, the dual space to G. N can be identified with the product $M \times \mathbf{G}^d$. The functions $f \in F(M)$ that satisfy (4.2) are the functions on $N = M \times \mathbf{G}^d$ that are functions on $\mathbf{G}^d$ alone. Thus, the set of functions f, f^1 that satisfy (4.2) can be identified with $F(\mathbf{G}^d)$. Now, each $X \in G^d$ defines a linear function

$$f_X : \alpha \to \alpha(X) = f_X(\alpha) \tag{4.5}$$

on $\mathbf{G}^d$. The z_i are the functions that correspond in this case to the basis elements X_i of G. Thus, we have proved the following result.

Theorem 4.1: The Poisson bracket operation on $T^d(G)$ defines a Lie algebra structure on $F(\mathbf{G}^d)$ such that

$$\{f, f'f''\} = \{f, f'\}f'' + f'\{f, f''\} \quad \text{for } f, f', f'' \in F(G^d). \tag{4.6}$$

The linear mapping defined by (4.5) of $\mathbf{G} \to F(\mathbf{G}^d)$ is a Lie algebra homomorphism, and defines G as a Lie subalgebra of $F(\mathbf{G}^d)$.

We can now extend the mapping (4.5). Let $S(\mathbf{G})$ be the symmetric algebra G, i.e. an element of $S(\mathbf{G})$ is an element of a symmetric tensor product G with itself. $S(\mathbf{G})$ is made into a commutative, associative algebra using the symmetric tensor product. G can be identified with a subspace of $S(\mathbf{G})$. The map (4.5) can then be extended to a map $S(\mathbf{G}) \to F(\mathbf{G}^d)$ using the following formula:

$$\text{If } X = X_1 \cdots X_r \in S(\mathbf{G}), \text{ then } f_X = f_{X_1} \cdots f_{X_r}. \tag{4.7}$$

The linear map (4.7) is one-to-one, and an algebra homomorphism, and hence defines $S(\mathbf{G})$ as a linear subspace of $F(\mathbf{G}^d)$. In particular, $S(\mathbf{G})$ has a Lie algebra structure, and we have proved the following result.

Theorem 4.2: If G is a Lie algebra, $S(\mathbf{G})$ has a Lie algebra structure such that

$$[\Delta, \Delta'\Delta''] = [\Delta, \Delta']\Delta'' + \Delta'[\Delta, \Delta''] \quad \text{for } \Delta, \Delta', \Delta'' \in S(\mathbf{G}). \tag{4.8}$$

G, as a linear subspace of $S(\mathbf{G})$, is a Lie subalgebra, with the inherited Lie algebra structure identical with the one with which we began.

Now, let $U(\mathbf{G})$ be the universal associative enveloping algebra[6] of G. It has a filtration $\mathbf{G} = U^1(\mathbf{G}) \subset U^2(\mathbf{G}) \subset \cdots$. It can be proved[6] that the associated graded associative algebra is $S(G)$, i.e.,

$$S^r(\mathbf{G}) = U^r(\mathbf{G})/U^{r-1}(\mathbf{G}), \qquad r = 1, 2, \cdots, \tag{4.9}$$

where $S^r(\mathbf{G})$ denotes the subspace of $S(\mathbf{G})$ consisting of the symmetric tensors of degree r. An element $\Delta \in U(\mathbf{G})$ is called a *Casimir element* if it lies in the center of $U(\mathbf{G})$. The following result now follows from (4.8).

Theorem 4.3: Let Δ be a Casimir element of $U(\mathbf{G})$. Then the corresponding element in $S(\mathbf{G})$ and $F(\mathbf{G}^d)$ lies in the center of the Lie algebra structure defined above on these spaces.

Theorem 4.3 has a consquence for "classical mechanics" on $M = G$. Suppose h is an element of $F(\mathbf{G}^d)$ that is regarded as a Hamiltonian for a mechanical system on G. Given a Casimir element Δ of $U(\mathbf{G})$, denote by f_Δ the corresponding element of $F(\mathbf{G}^d)$. Then, by Theorem 4.3,

$$\{h, f_\Delta\} = 0, \tag{4.10}$$

i.e., f_Δ is an integral of the motion generated by the Hamiltonian h. For example, if $G = SO(3, R)$ and if h is the Hamiltonian corresponding to a rotating, force-free rigid body, then, if Δ is chosen to be the usual second-order Casimir operator, f_Δ is the "integral of total angular momentum."

Now—as pointed out to me by Jensen and Schäffer—these remarks prove the following result, which contradicts the main result of Ref. 1.

Theorem 4.4: Suppose that h is the Hamiltonian of a mechanical system on $M = G$ that is invariant under left translation by elements of G. Suppose that G is a compact, semisimple Lie group. Then the system defined by h is complete, i.e., extremal curves can be indefinitely extended.

Proof: Let Y_h be the vector field on $T^d(G) = G \times \mathbf{G}^d$ defined by h, using formula (3.3). Then Y_h is tangent to G^d, since h is a function of the $(z_1, \cdots, z_n)$ alone. Let Δ be the second-order Casimir element of $U(\mathbf{G})$ defined by the Killing form of G. Let f_Δ be the corresponding element of $F(\mathbf{G}^d)$. Then the level surfaces $f_\Delta^{-1}(a)$, for each $a \in R$, are compact,

and X_h is tangent to them. Hence, the integral curves of X_h can be indefinitely extended, which means that the system is complete.

One may suspect then that the simplest example where the integral curves of X_h cannot be indefinitely extended is that where G is a noncompact, semisimple Lie group, e.g., $SO(2,1)$. We shall then turn to a study of this example.

5. GEODESICS OF LEFT-INVARIANT METRICS ON COMPACT AND NONCOMPACT SEMISIMPLE LIE GROUPS

Let us return to the general situation described in Sec. 2, i.e., M is the underlying manifold of a Lie group G, with (ω_i) a basis for left-invariant 1-forms. Let (X_i) be a dual basis for the Lie algebra $\mathbf{G}$ of G. Let $ds^2 = g_{ij}\omega_i \cdot \omega_j$ be a left-invariant metric on M. If $t \to \sigma(t)$ is a geodesic of this nature, set $g_i = \omega_i(\sigma'(t))$. Then, the $y_i(t)$ satisfy Eqs. (2.14). Set

$$X(t) = y_i(t)X_i. \tag{5.1}$$

Then, $t \to X(t)$ may be regarded as a curve in $\mathbf{G}$. Thus, Eqs. (2.14) translate into the following basis-free equations for the curve $t \to X(t)$:

$$\frac{dX}{dt} = A^{-1}[AX, X], \tag{5.2}$$

where A is a linear transformation $\mathbf{G} \to \mathbf{G}$. Equation (5.2) was also derived in Ref. 2, p. 430, independently of the Lagrangian approach to the Euler equations. If $\mathbf{G}$ is a semisimple Lie algebra, with $B(\ ,\ )$ its Killing form,[6] it was shown in Ref. 2 that A could be chosen to be symmetric with respect to the Killing form, i.e.,

$$B(AX, Y) = B(X, AY) \quad \text{for } X, Y \in \mathbf{G}. \tag{5.3}$$

Let us suppose that G is a compact semisimple Lie algebra,[6] and that K is a symmetric subalgebra[6] with the following Cartan decomposition[6] of $\mathbf{G}$:

$$\mathbf{G} = \mathbf{K} \oplus \mathbf{P}, [\mathbf{P}, \mathbf{P}] \subset \mathbf{K}, \quad [\mathbf{K}, \mathbf{P}] \subset \mathbf{P}. \tag{5.4}$$

Let us also suppose that

$$A(\mathbf{K}) \subset \mathbf{K}, \quad A(\mathbf{P}) \subset \mathbf{P}. \tag{5.5}$$

Now, consider $\mathbf{G}$ as a real subalgebra of its complexification $\mathbf{G}_c = \mathbf{G} \otimes C$. Then, $\mathbf{G}'$ defined as follows is a real subalgebra of $\mathbf{G}_c$ whose complexification is again $\mathbf{G}_c$, i.e., $\mathbf{G}'$ is a noncompact real form of $\mathbf{G}$:

$$\mathbf{G}' = \mathbf{K} + i\mathbf{P}. \tag{5.6}$$

Given $t \to X(t)$ satisfying (5.2), let $X_K(t)$ and $X_P(t)$ denote its projections in $\mathbf{K}$ and P.

Now, let us suppose further that A satisfies the following condition:

$$[AX, X] = 0 \quad \text{for } X \in \mathbf{K}. \tag{5.7}$$

Then, using (5.4) and (5.5), the curves $t \to X_K(t)$, $X_P(t)$ satisfy the following differential equations:

$$\begin{aligned} \frac{dX_K}{dt} &= A^{-1}[AX_P, X_P], \\ \frac{dX_P}{dt} &= A^{-1}[AX_K, X_P] + A^{-1}[X_K, AX_P]. \end{aligned} \tag{5.8}$$

Let us suppose $X(t)$ is defined in a neighborhood of $t = 0$ as a solution of (5.2). Then, using the standard existence theorem[5] for analytic differential equations, it can be extended to be a complex analytic function if the complex variable $u = t + is$ in a neighborhood of $u = 0$. Set

$$Y(t) = iX_K(it) + X_P(it). \tag{5.9}$$

Then, using (5.7), we have

$$\begin{aligned} \frac{dY}{dt} = &- A^{-1}[AX_P(it), X_P(it)] + A^{-1}[AiX_K(it), X_P(it)] \\ &+ A^{-1}[iX_K(it), AX_P(it)]. \end{aligned} \tag{5.10}$$

Notice that these equations are the same as (5.9), except that the Lie bracket $[\ ,\ ]$ for $\mathbf{P}$ has been changed into its negative.

However, we see from (5.6) that this is the Lie bracket for $\mathbf{G}'$. In particular, $Y(t)$ satisfies a differential equation analogous to (5.2), but with the Lie bracket $[\ ,\ ]$ for $\mathbf{G}$ replaced by the Lie bracket $[\ ,\ ]'$ for $\mathbf{G}'$. In particular, if $Y(0) \in \mathbf{G}'$, then $Y(t)$ belongs to $\mathbf{G}'$ for all real t. The resulting equations for $Y(t)$ are the Euler equations for geodesics of a left-invariant metric on G'. Thus, we have given a precise sense to saying that the "differential equations for the geodesics of a left-invariant metric on G are analytically continued to the geodesics of a left-invariant metric on G'." In particular, this gives us the following result, which may be regarded as describing a criterion that the left-invariant metric on G' defined by A be incomplete

Theorem 5.1: Suppose that $u \to X(u)$ is a curve in G_c that is defined and meromorphic in the entire complex u plane and that satisfies the following differential equation and initial conditions:

$$\frac{dX}{du} = A^{-1}[AX, X], \quad X(0) \in \mathbf{P}. \tag{5.11}$$

Also suppose that A is a complex-linear map: $\mathbf{G}_c \to \mathbf{G}_c$ satisfies (5.3), (5.5), and (5.7). Then, if the curve $u \to X(u)$ has at least one pole along the imaginary axis $u = it$, $t \in R$, the left-invariant metric on G' defined by A is incomplete.

This result applies, in particular, to the case $G = SO(3, R)$, $G' = SO(2,1)$. It is well known that the solutions of (5.11) in this case are given by the Jacobian elliptic functions, which do indeed[5] have poles along the imaginary axis.

Finally, I hope that this result will clarify the major error in Ref. 1 which was pointed out by Jensen and Schäffer. I also would like to remark that this example of an incomplete metric on $SO(2,1)$ may be useful as an example for a possible general theory of incomplete homogeneous pseudo-Riemannian manifolds, a topic that would be very important physically because of its connection with the study of singularities in cosmology.

* Supported by the Office of Naval Research, under contract NR-041-378. Reproduction in whole or part is permitted for any purpose by the United States government.

[1] R. Hermann, J. Math. Mech. 13, 667 (1964).
[2] R. Hermann, *Differential Geometry and the Calculus of Variations* (Academic, New York, 1968).
[3] R. Hermann, *Vector Bundles in Mathematical Physics* (Benjamin, New York, 1970), Vol. II.
[4] F. Tricomi, *Differential Equations* (Hefner, New York, 1961).
[5] E. Hille, *Lectures on Ordinary Differential Equations* (Addison-Wesley, Reading, Mass., 1969).
[6] R. Hermann, *Lie Groups for Physicists* (Benjamin, New York, 1966).

Nonpolynomial Lagrangians with Isospin

R. Delbourgo
Physics Department, Imperial College, London, SW7, England
(Received 9 December 1970)

We apply (Laplace, Fourier, ...) transform methods to obtain compact representations of the perturbation S-matrix elements for interaction Lagrangians which are general nonpolynomial functions of isospin multiplets. In order to illustrate the power and simplicity of the method for coping with the isospin complications, we have treated several examples in detail, and these include the commonly used parametrizations of the unitary chiral $SU(2) \otimes SU(2)$ transformations.

1. INTRODUCTION

Many of the nonpolynomial Lagrangians one encounters in physical examples involve internal symmetry groups like $SU(3) \otimes SU(3)$ where nonpolynomial functions of the field and generators make an appearance. The simplest instances of this are functions like $\exp(\gamma_5 \lambda \boldsymbol{\tau} \cdot \boldsymbol{\phi})$ or $(1 + \lambda^2 \boldsymbol{\phi}^2)^{-2}$ which depend on an $SU(2)$ isotriplet of fields $\boldsymbol{\phi}$. Perturbation calculations with such interaction Hamiltonian functions can then become exceedingly difficult in comparison with calculations involving one scalar field only. We describe below a way of simplifying these computations by using transform (Laplace, Fourier, ...) methods whereby the isospin complexities are extracted out and the integrals to be evaluated are similar to the integrals one would meet for a single scalar field.

Section 2 summarizes the rules of calculation when there is just one field ϕ and the potential $V(\phi)$ is some general function of it; the results to second order in V are listed for some typical functions. Section 3 generalizes the method of transforms used in Sec. 2 to the more physical situation of a multiplet of fields $\boldsymbol{\phi}$ when V is a scalar function of them. The case when V is a matrix function of at most an isotriplet of fields (e.g., $\exp\lambda \boldsymbol{\tau} \cdot \boldsymbol{\phi}$) is studied in Sec. 4 and examples are given for some commonly met parametrizations of chiral models. We shall not consider derivative interactions of mesons in this paper.

2. MATRIX ELEMENTS FOR NONPOLYNOMIAL INTERACTIONS

In this section we shall give a brief resume of some techniques[1,2] for evaluating Green's functions in a Lagrangian theory of a single scalar field ϕ with some specified interaction Hamiltonian $V(\phi)$ which is in general a nonpolynomial functional of ϕ. One works in the interaction picture with the free propagator

$$\langle 0 \,|\, T[\phi(x)\phi(x')] \,|\, 0 \rangle \equiv \langle \phi, \phi' \rangle \equiv \Delta(x - x'), \tag{1}$$

and then, to order V^N in the perturbation development, the n-point Green's function is described by the x-space matrix element:

$$S^N(x) \equiv S_{n_1 n_2 \cdots n_N}(x_1, x_2, \ldots, x_N) \tag{2}$$

with $n = \sum n_i$ and where n_i denotes the number of external lines meeting at point x_i.

There are many equivalent ways of expressing $S^N(x)$ in terms of V, but the most convenient for our purposes is the form[1]

$$S^N(x) = \exp\left(\tfrac{1}{2}\sum_{ij} \Delta_{ij} \frac{\partial^2}{\partial\phi_i \partial\phi_j}\right) \prod_k \left(\frac{\partial}{\partial\phi_k}\right)^{n_k} V(\phi_k)\Big|_{\phi=0}, \tag{3}$$

wherein $\Delta_{ij} \equiv \Delta(x_i - x_j)$ and $\phi_i \equiv \phi(x_i)$ are regarded as c-numbers. The summation in Eq. (3) is over $i \neq j$ if V is normally ordered. The simplest way of evaluating S is to get an integral representation for it using transform methods.[3] Thus if we express V in terms of its Laplace transform $\mathcal{U}$:

$$V(\phi) \equiv \int_0^\infty \mathcal{U}(\zeta) e^{-\zeta\phi} d\zeta, \tag{4}$$

or its Fourier transform $\tilde{V}$:

$$V(\phi) \equiv \int_{-\infty}^{\infty} \tilde{V}(\xi) e^{-i\xi\phi} d\xi, \tag{4'}$$

it is a trivial matter to arrive at the expressions

$$S^N(x) = (-1)^n \prod_{k=1}^N \left(\int_0^\infty d\zeta_k \zeta_k^{n_k} \mathcal{U}(\zeta_k)\right) \exp\left(\tfrac{1}{2}\sum_{ij} \zeta_i \zeta_j \Delta_{ij}\right)$$

or (5)

$$S^N(x) = (-i)^n \prod_{k=1}^N \left(\int_{-\infty}^\infty d\xi_k \xi_k^{n_k} \tilde{V}(\xi_k)\right) \exp\left(-\tfrac{1}{2}\sum_{ij} \xi_i \xi_j \Delta_{ij}\right) \tag{5'}$$

using the fundamental result that

$$\langle \exp(\lambda\phi), \exp(\lambda'\phi') \rangle = \exp(\lambda\lambda'\Delta). \tag{6}$$

Strictly speaking, of course, Eqs. (4) and the ensuing representations (5) have validity only for certain types of function V. However, as ϕ always occurs multiplied by a dimensional minor coupling constant λ in nonpolynomial interactions, we can usually apply these transforms for a certain complex domain of λ and then analytically continue the representations (5) if necessary to the physical value of λ. In so doing, one will, in general, encounter singularities in the region of auxiliary (ζ or ξ) variable integration, and one of the chief problems of the theory is to define these such that the $S^N(x)$ are real in the Euclidean region—it is here that ambiguity problems may arise.[4] However such sources of difficulty are not our concern in this paper; rather we shall be especially interested in the complexities which arise when one must deal[5] with a whole multiplet of fields $\boldsymbol{\phi}$ as we do later on. Let us therefore pass over this par-

* Supported by the Office of Naval Research, under contract NR-041-378. Reproduction in whole or part is permitted for any purpose by the United States government.

[1] R. Hermann, J. Math. Mech. 13, 667 (1964).
[2] R. Hermann, *Differential Geometry and the Calculus of Variations* (Academic, New York, 1968).
[3] R. Hermann, *Vector Bundles in Mathematical Physics* (Benjamin, New York, 1970), Vol. II.
[4] F. Tricomi, *Differential Equations* (Hefner, New York, 1961).
[5] E. Hille, *Lectures on Ordinary Differential Equations* (Addison-Wesley, Reading, Mass., 1969).
[6] R. Hermann, *Lie Groups for Physicists* (Benjamin, New York, 1966).

Nonpolynomial Lagrangians with Isospin

R. Delbourgo
Physics Department, Imperial College, London, SW7, England
(Received 9 December 1970)

We apply (Laplace, Fourier, ...) transform methods to obtain compact representations of the perturbation S-matrix elements for interaction Lagrangians which are general nonpolynomial functions of isospin multiplets. In order to illustrate the power and simplicity of the method for coping with the isospin complications, we have treated several examples in detail, and these include the commonly used parametrizations of the unitary chiral $SU(2) \otimes SU(2)$ transformations.

1. INTRODUCTION

Many of the nonpolynomial Lagrangians one encounters in physical examples involve internal symmetry groups like $SU(3) \otimes SU(3)$ where nonpolynomial functions of the field and generators make an appearance. The simplest instances of this are functions like $\exp(\gamma_5 \lambda \boldsymbol{\tau} \cdot \boldsymbol{\phi})$ or $(1 + \lambda^2 \boldsymbol{\phi}^2)^{-2}$ which depend on an $SU(2)$ isotriplet of fields $\boldsymbol{\phi}$. Perturbation calculations with such interaction Hamiltonian functions can then become exceedingly difficult in comparison with calculations involving one scalar field only. We describe below a way of simplifying these computations by using transform (Laplace, Fourier, ...) methods whereby the isospin complexities are extracted out and the integrals to be evaluated are similar to the integrals one would meet for a single scalar field.

Section 2 summarizes the rules of calculation when there is just one field ϕ and the potential $V(\phi)$ is some general function of it; the results to second order in V are listed for some typical functions. Section 3 generalizes the method of transforms used in Sec. 2 to the more physical situation of a multiplet of fields $\boldsymbol{\phi}$ when V is a scalar function of them. The case when V is a matrix function of at most an isotriplet of fields (e.g., $\exp\lambda \boldsymbol{\tau} \cdot \boldsymbol{\phi}$) is studied in Sec. 4 and examples are given for some commonly met parametrizations of chiral models. We shall not consider derivative interactions of mesons in this paper.

2. MATRIX ELEMENTS FOR NONPOLYNOMIAL INTERACTIONS

In this section we shall give a brief resume of some techniques[1,2] for evaluating Green's functions in a Lagrangian theory of a single scalar field ϕ with some specified interaction Hamiltonian $V(\phi)$ which is in general a nonpolynomial functional of ϕ. One works in the interaction picture with the free propagator

$$\langle 0 \,|\, T[\phi(x)\phi(x')] \,|\, 0 \rangle \equiv \langle \phi, \phi' \rangle \equiv \Delta(x - x'), \tag{1}$$

and then, to order V^N in the perturbation development, the n-point Green's function is described by the x-space matrix element:

$$S^N(x) \equiv S_{n_1 n_2 \cdots n_N}(x_1, x_2, \ldots, x_N) \tag{2}$$

with $n = \sum n_i$ and where n_i denotes the number of external lines meeting at point x_i.

There are many equivalent ways of expressing $S^N(x)$ in terms of V, but the most convenient for our purposes is the form[1]

$$S^N(x) = \exp\left(\tfrac{1}{2}\sum_{ij} \Delta_{ij} \frac{\partial^2}{\partial\phi_i \partial\phi_j}\right) \prod_k \left(\frac{\partial}{\partial\phi_k}\right)^{n_k} V(\phi_k)\Big|_{\phi=0}, \tag{3}$$

wherein $\Delta_{ij} \equiv \Delta(x_i - x_j)$ and $\phi_i \equiv \phi(x_i)$ are regarded as c-numbers. The summation in Eq. (3) is over $i \neq j$ if V is normally ordered. The simplest way of evaluating S is to get an integral representation for it using transform methods.[3] Thus if we express V in terms of its Laplace transform $\mathcal{U}$:

$$V(\phi) \equiv \int_0^\infty \mathcal{U}(\zeta) e^{-\zeta\phi} d\zeta, \tag{4}$$

or its Fourier transform $\tilde{V}$:

$$V(\phi) \equiv \int_{-\infty}^{\infty} \tilde{V}(\xi) e^{-i\xi\phi} d\xi, \tag{4'}$$

it is a trivial matter to arrive at the expressions

$$S^N(x) = (-1)^n \prod_{k=1}^N \left(\int_0^\infty d\zeta_k \zeta_k^{n_k} \mathcal{U}(\zeta_k)\right) \exp\left(\tfrac{1}{2}\sum_{ij} \zeta_i \zeta_j \Delta_{ij}\right)$$

or (5)

$$S^N(x) = (-i)^n \prod_{k=1}^N \left(\int_{-\infty}^\infty d\xi_k \xi_k^{n_k} \tilde{V}(\xi_k)\right) \exp\left(-\tfrac{1}{2}\sum_{ij} \xi_i \xi_j \Delta_{ij}\right) \tag{5'}$$

using the fundamental result that

$$\langle \exp(\lambda\phi), \exp(\lambda'\phi') \rangle = \exp(\lambda\lambda'\Delta). \tag{6}$$

Strictly speaking, of course, Eqs. (4) and the ensuing representations (5) have validity only for certain types of function V. However, as ϕ always occurs multiplied by a dimensional minor coupling constant λ in nonpolynomial interactions, we can usually apply these transforms for a certain complex domain of λ and then analytically continue the representations (5) if necessary to the physical value of λ. In so doing, one will, in general, encounter singularities in the region of auxiliary (ζ or ξ) variable integration, and one of the chief problems of the theory is to define these such that the $S^N(x)$ are real in the Euclidean region—it is here that ambiguity problems may arise.[4] However such sources of difficulty are not our concern in this paper; rather we shall be especially interested in the complexities which arise when one must deal[5] with a whole multiplet of fields $\boldsymbol{\phi}$ as we do later on. Let us therefore pass over this par-

ticular ambiguity problem by mentioning that prescriptions, like taking Cauchy principal values, etc., can be devised for defining contours in the ζ and ξ planes to render S^N real for all $x \neq 0$ (i.e., Δ finite). The passage to $x \to 0 (\Delta \to \infty)$ can be a second source of ambiguity as it has to do with the definition of distributions $\phi^n(x)$ which enter whenever a power series expansion of $V(\phi)$ in powers of ϕ is made.[6] For resolving this difficulty, incidental though it is to the object of our exercise, we shall always interpret $S(x)$ as a classical function whenever its Fourier transform $\tilde{S}(p)$ exists—otherwise $S(x)$ requires renormalization[3] in the traditional way if its behavior in Δ as $\Delta \to \infty$ is polynomial.

It is not necessary in practise to evaluate every $S_{n_1 n_2 n_3 \cdots n_N}(x)$ because of the fortunate circumstance that differentiation with respect to Δ increases the number of external lines. Specifically,

$$\frac{\partial}{\partial \Delta_{12}} S_{n_1 n_2 n_3 \cdots n_N}(x) = S_{n_1+1\, n_2+1\, n_3 \cdots n_N}(x), \quad \text{etc.,} \tag{7}$$

can be proved from representations (3) or (5). Thus it is sufficient to evaluate just the matrix elements $S_{n0\ldots0}(x)$, a considerable saving of labor. In fact most of our attention will be focussed on the second order functions $S^2(x_1 - x_2) \equiv S_{n\,0}(x_1, x_2)$ as these are the only elements one can readily calculate in terms of familiar functions. (Higher functions $S^3, S^4, \cdots$, even for the simplest nonpolynomial interactions, bring in confluent hypergeometric functions of three or more variables whose properties are not well known.)

The power of the transform method for arriving at compact representations of the Green's functions is best illustrated by working out $S_{n0}(x)$ for some typical interactions.

[1a] Polynomial and exponential:

$$V(\phi) = \phi^\nu e^{-\lambda\phi}, \tag{8}$$

where we take ν integer and $\lambda > 0$ for the present. Therefore, $\mathcal{V}(\zeta) = \delta^\nu(\zeta - \lambda)$ and the auxiliary ζ integrations are easily carried out to give

$$S_{n0} = (-1)^n \left(\Delta \frac{\partial}{\partial \zeta}\right)^\nu (\zeta^{n+\nu} e^{\lambda\zeta\Delta})\Big|_{\zeta=\lambda}. \tag{9}$$

Thus for $\nu = 0$ or $\lambda = 0$ we recover well-known results. Indeed the continuation to noninteger ν may be regarded as a definition of the distribution ϕ^ν.

[1b] Rational:

$$V(\phi) = \phi^\nu (1 + \lambda\phi)^{-\mu}, \tag{10}$$

where we take $\lambda > 0$ and ν, μ positive integers with $\nu > \mu$ at first. Thus

$$\mathcal{V}(\zeta) = \frac{1}{\Gamma(\mu)} \left(-\frac{\partial}{\partial\lambda}\right)^{\mu-1} \left(\frac{\partial}{\partial\zeta}\right)^{\nu-\mu+1} \left(\frac{e^{-\zeta/\lambda}}{\lambda}\theta(\zeta + 0)\right).$$

One of the auxiliary ζ integrations is easily carried out to leave us with the matrix elements

$$S_{n0} = \frac{(-1)^n}{\Gamma(\mu)} \lambda^n \left(\frac{\partial}{\partial\lambda^2}\right)^{\mu-1} (\lambda^2)^{\mu-1} \int_0^\infty d\zeta \frac{(\lambda^2\zeta)^n (\Delta\zeta)^\nu e^{-\zeta}}{(1 - \lambda^2\Delta\zeta)^\mu}, \tag{11}$$

all of which can be expressed as derivatives of the basic function

$$\int_0^\infty d\zeta \frac{e^{-\zeta}}{1 - \lambda^2\Delta\zeta} = \frac{1}{\lambda^2\Delta} \mathrm{Ei}\left(\frac{1}{\lambda^2\Delta}\right).$$

As this function has infinitely many sheets, we can define it on the principal sheet by taking the principal value of the ζ integral.

[1c] Gaussian:

$$V(\phi) = \left(\frac{\partial}{\partial\phi}\right)^\nu e^{-\lambda^2\phi^2/2}. \tag{12}$$

In this case it is more appropriate to use the Gaussian transform:

$$\left(\frac{\partial}{\partial\phi}\right)^\nu e^{-\lambda^2\phi^2/2} = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} d\eta \cdot e^{-\eta^2/2 + \eta\lambda\phi} \cdot (\lambda\eta)^\nu,$$

$$S_{2n\,0} = \frac{1}{2\pi} \int d\eta d\eta' (\lambda\eta)^{\nu+2n} (\lambda\eta')^\nu e^{-\eta^2/2 - \eta'^2/2 + \eta\eta'\lambda^2\Delta}$$

$$= (\tfrac{1}{2}\lambda^2)^n \left(\frac{\partial}{\partial\Delta}\right)^\nu \frac{(2n)!}{n!} [1 - \lambda^4\Delta^2]^{-1/2-n}, \tag{13}$$

whose real part must be taken. Thus beyond the branch points at $\Delta = \pm 1/\lambda^2$ we are obliged to define $S_{2n\,0} \equiv 0$ by averaging on each side of the cut.

If we formally expand (8), (10), and (12) in powers of ϕ and use the fact that

$$\langle \phi^n, \phi'^n \rangle = n!\Delta^n,$$

the power series in Δ which follows, of course, agrees term by term with the expansions of (9), (11), and (13). Besides the obvious fact that the transform technique automatically sums these series, there is the advantage that it also interprets the series as an asymptotic one [e.g. in Eq. (11)] even when the series is formally divergent—the ambiguities[4] of such summations are then reflected in the possible choice of auxiliary variable contours which avoid the singularities in the range of integration. However, although it would be fair to say that the results we have listed above can be obtained by other means without a great deal of difficulty, to our knowledge the results given in the next sections are not easily obtained in any other way[5].

3. SCALAR INTERACTIONS AND MULTIPLETS OF FIELDS

Consider now the situation where one has an R-dimensional multiplet of fields $\boldsymbol{\phi} = \phi_a$, $a = 1, \ldots, R$, and an interaction $V(\boldsymbol{\phi})$ which is a scalar function of them. The N'th-order matrix elements will be described by tensors $S_{a_1 b_1 \ldots, a_2 b_2 \ldots, \ldots, a_N b_N \ldots}(x_1, x_2, \ldots, x_N)$, where the n_i labels $a_i, b_i, \ldots$ refer to the field components entering at vertex x_i. The corresponding generalization of (3) is

$$S^N(x) = \exp\left(\tfrac{1}{2}\sum_{ij} \frac{\partial}{\partial\boldsymbol{\phi}_i} \Delta_{ij} \frac{\partial}{\partial\boldsymbol{\phi}_j}\right) \prod_k \left(\frac{\partial}{\partial\phi_{a_k}} \frac{\partial}{\partial\phi_{b_k}} \cdots\right) V(\phi_k)\Big|_{\phi=0}, \tag{14}$$

and since $\langle \phi_a, \phi'_b \rangle = \delta_{ab}\Delta$, the index of the exponential reduces to $\frac{1}{2}\sum_{ij} \Delta_{ij} \partial^2/\partial\phi_{ai}\partial\phi_{aj}$. The transform tech-

nique is again indicated for alleviating the complexities of "isospin." Generalize (4) and (4′) to

$$V(\boldsymbol{\phi}) = \int_0^\infty \mathcal{U}(\boldsymbol{\zeta}) e^{-\boldsymbol{\zeta}\cdot\boldsymbol{\phi}} d^R\zeta \tag{15}$$

and

$$V(\boldsymbol{\phi}) = \int_{-\infty}^{\infty} \tilde{V}(\boldsymbol{\xi}) e^{-i\boldsymbol{\xi}\cdot\boldsymbol{\phi}} d^R\xi \tag{15'}$$

so that the Green's functions reduce to

$$S^N(x) = (-1)^n \prod_{k=1}^N \left(\int_0^\infty d^R\zeta_k \zeta_{a_k} \zeta_{b_k} \cdots \mathcal{U}(\zeta_k) \right) \times \exp\left(\tfrac{1}{2} \sum_{ij} \zeta_i \zeta_j \Delta_{ij} \right) \tag{16}$$

or

$$(-i)^n \prod_{k=1}^N \left(\int_{-\infty}^{\infty} d^R\xi_k \xi_{a_k} \xi_{b_k} \cdots \tilde{V}(\xi_k) \right) \exp\left(-\tfrac{1}{2} \sum_{ij} \xi_i \xi_j \Delta_{ij} \right). \tag{16'}$$

Because V is a scalar, $\mathcal{U} = \mathcal{U}(|\boldsymbol{\zeta}|)$ and $\tilde{V} = \tilde{V}(|\boldsymbol{\xi}|)$. To make further progress, break up S into scalar functions $S^{(I)}$:

$$S_{a_1 b_1 \cdots a_2 b_2 \cdots a_N b_N \cdots} = \sum_I K^{(I)}_{a_1 b_1 \cdots a_N b_N} S^{(I)}, \tag{17}$$

where the number I is related to the number of external lines and the way they are attached to the vertices. Because of the supposed linear independence of the K^I, it is possible to discover the $S^{(I)}$ by taking I different traces over the tensor labels. After tracing, the integral representations (16) reduce to the final form

$$\mathcal{S}_{m_1 m_2 \cdots m_N} = \prod_{k=1}^N \left(\int d^R\zeta_k \, |\boldsymbol{\zeta}_k|^{m_k} \mathcal{U}(|\boldsymbol{\zeta}_k|) \right) \exp\left(\tfrac{1}{2} \sum \zeta_i \zeta_j \Delta_{ij} \right) \tag{18}$$

and derivatives thereof with respect to the Δ_{ij}. *The whole problem then boils down to the evaluation of certain scalar integrals like (18).*

To illustrate, consider the third-order four-point function described tensorially by $S_{a_1 b_1 a_2 a_3}(x_1, x_2, x_3)$. It decomposes kinematically as

$$S_{a_1 b_1 a_2 a_3} = \delta_{a_1 b_1} \delta_{a_2 a_3} S^{(0)} + (\delta_{a_1 a_2} \delta_{b_1 a_3} + \delta_{a_1 a_3} \delta_{b_1 a_2}) S^{(1)} \tag{19}$$

so that

$$R^2 S^{(0)} + 2RS^{(1)} = \mathcal{S}_{bbaa} = \prod_{k=1}^3 [d^R\zeta_k \mathcal{U}(|\boldsymbol{\zeta}_k|)] \exp(\,) |\boldsymbol{\zeta}_1|^2 \boldsymbol{\zeta}_2 \cdot \boldsymbol{\zeta}_3 = \frac{\partial}{\partial \Delta_{23}} \mathcal{S}_{200}$$

and, from S_{abab},

$$RS^{(0)} + R(R+1)S^{(1)} = \prod_{k=1}^3 \left(\int d^R\zeta_k \mathcal{U}(|\boldsymbol{\zeta}_k|) \right) \exp(\,) \boldsymbol{\zeta}_1 \cdot \boldsymbol{\zeta}_3 \boldsymbol{\zeta}_1 \cdot \boldsymbol{\zeta}_2 = \frac{\partial^2}{\partial \Delta_{13} \partial \Delta_{12}} \mathcal{S}_{000},$$

which can be solved for $S^{(0)}$ and $S^{(1)}$ in terms of the $\mathcal{S}$, etc. Thus, if we are interested in the second-order functions, the relevant integrals to be computed are of the type

$$\mathcal{S}_{mm'} = \int d^R\zeta d^R\zeta' |\boldsymbol{\zeta}|^m |\boldsymbol{\zeta}'|^{m'} e^{\boldsymbol{\zeta}\cdot\boldsymbol{\zeta}'\Delta} \mathcal{U}(|\boldsymbol{\zeta}|) \mathcal{U}(|\boldsymbol{\zeta}'|), \tag{20}$$

which can be done without too much trouble as they reduce to a triple integral (one angular, two radial)

$$\mathcal{S}_{mm'} \propto \int d\zeta d\zeta' \zeta^{m+R-1} \zeta'^{m'+R-1} \mathcal{U}(\zeta) \mathcal{U}(\zeta') \times \int d\theta (\sin\theta)^{R-2} e^{\zeta\zeta'\Delta\cos\theta}. \tag{21}$$

If R is odd, the angular integrations leave us with exponential forms $e^{\zeta\zeta'\Delta}$ and derivatives thereof, but if R is even, Bessel functions arise. We give examples below for $R = 2$ and 3 which demonstrate this explicitly.

[2a] Two-dimensional multiplets $\boldsymbol{\phi} = (\phi_1, \phi_2)$:

Take $V(\boldsymbol{\phi}) = e^{-\lambda\phi}$, where $\phi \equiv (\phi^2)^{1/2}$, and use Fourier transforms

$$\tilde{V}(\boldsymbol{\xi}) = \frac{1}{4\pi^2} \int e^{i\boldsymbol{\xi}\cdot\boldsymbol{\phi}} V(\boldsymbol{\phi}) d^2\xi$$

so that, for $\lambda > 0$,

$$\tilde{V}(\boldsymbol{\xi}) = \frac{1}{4\pi^2} \int \phi d\phi d\theta e^{(i\xi\cos\theta - \lambda)\phi} = -\frac{1}{2\pi} \frac{\partial}{\partial\lambda} \frac{1}{(\lambda^2 + \xi^2)^{1/2}}. \tag{22}$$

The vacuum matrix element, the so-called "superpropagator," is given by

$$\mathcal{S}_{00} = \frac{1}{2\pi} \frac{\partial^2}{\partial\lambda\partial\lambda'} \int_0^\infty \frac{\xi d\xi}{(\lambda^2 + \xi^2)^{1/2}} \frac{\xi' d\xi'}{(\lambda'^2 + \xi'^2)^{1/2}} \times \int_0^\pi e^{-\xi\xi'\Delta\cos\theta} d\theta.$$

Let us take Δ pure imaginary so that angular integration gives $2\pi J_0(i\xi\xi'\Delta)$ that can again be integrated wrt ξ' to leave us with the final form

$$\langle e^{-\lambda\phi}, e^{-\lambda'\phi'} \rangle = -\frac{i}{\Delta} \frac{\partial^2}{\partial\lambda\partial\lambda'} \int_0^\infty \frac{e^{-i\xi\lambda\lambda'\Delta}}{(1 + \xi^2)^{1/2}} d\xi,$$

which is an inhomogeneous cylinder function. In fact, if we symmetrize in λ and λ',

$$\langle \cosh\lambda\phi, \cosh\lambda'\phi' \rangle = \frac{\pi}{2\Delta} \frac{\partial^2}{\partial\lambda\partial\lambda'} L_0(\lambda\lambda'\Delta) = 1 + \frac{\pi\lambda\lambda'\Delta}{2} L_0(\lambda\lambda'\Delta), \tag{23}$$

where the appropriate cylinder function[7] is just the Struve function L:

$$L_\nu(z) \equiv I_{-\nu}(z) - \frac{(\frac{1}{2}z)^\nu}{2\Gamma(\frac{1}{2})\Gamma(\nu + \frac{1}{2})} \int_0^\infty \frac{\sin(z\xi)}{(1 + \xi^2)^{1/2 - \nu}} d\xi. \tag{24}$$

$L_0(z)$ is an entire function of z as can be seen from the series development

$$L_\nu(z) = \sum_{m=0}^\infty \frac{(\frac{1}{2}z)^{\nu+2m+1}}{\Gamma(m + \frac{3}{2})\Gamma(\nu + m + \frac{3}{2})} \tag{24'}$$

and indeed an alternative derivation of (23) is to use the power series expansion of $\cosh\lambda\phi$,

$$\langle (\boldsymbol{\phi}^2)^n, (\boldsymbol{\phi}'^2)^n \rangle = [\Gamma(n+1)]^2 (2\Delta)^{2n}, \tag{25}$$

and to identify the resulting series in Δ as a Struve function.

[3a] Three-dimensional multiplets $\boldsymbol{\phi} = (\phi_1, \phi_2, \phi_3)$:

Let $V(\phi) = (1 + \lambda\phi)^{-1}$, where $\phi \equiv (\boldsymbol{\phi}^2)^{1/2}$. It is easy to check that for this case the Laplace transform is

$$\mathcal{V}(\boldsymbol{\zeta}) = 2e^{-\zeta/\lambda}/\pi\lambda^2\zeta.$$

The vacuum expectation value will be given by[8]

$$\left\langle \frac{1}{1+\lambda\phi}, \frac{1}{1+\lambda'\phi'} \right\rangle = \int \zeta^2 d\zeta\, \zeta'^2 d\zeta' \frac{e^{-\zeta/\lambda}}{\lambda^2\zeta} \frac{e^{-\zeta'/\lambda'}}{\lambda'^2\zeta'} \int_0^1 d(\cos\theta) e^{\zeta\zeta'\Delta\cos\theta} = \int_0^\infty \frac{\zeta e^{-\zeta} d\zeta}{1 - \lambda\lambda'\Delta\zeta} \tag{26}$$

More complicated examples can be built up from (26); thus

$$\left\langle \frac{1}{1-\lambda^2\phi^2}, \frac{1}{1-\lambda^2\phi'^2} \right\rangle = \int_0^\infty \frac{\zeta e^{-\zeta} d\zeta}{1 - \lambda^4\Delta^2\zeta^2}, \tag{27}$$

which incidentally shows that in three dimensions

$$\langle (\boldsymbol{\phi}^2)^n, (\boldsymbol{\phi}'^2)^n \rangle = \Gamma(2n+2)\Delta^{2n} \tag{28}$$

and can be compared with the two-dimensional result (25).

One may discover the generalizations of (25) and (28) to an arbitrary R dimensional multiplet using Fourier transform methods. When $V(\phi) = (\boldsymbol{\phi}^2)^n$,

$$\tilde{V}(\boldsymbol{\xi}) = \frac{(-1)^n \Gamma(n + \frac{1}{2}R)\delta^{2n}(\xi)}{2\Gamma(n+\frac{1}{2})(\pi^{1/2}\xi)^{R-1}},$$

$$\therefore \langle (\boldsymbol{\phi}^2)^n, (\boldsymbol{\phi}'^2)^n \rangle = \sqrt{\pi}\,[\Gamma(n+\tfrac{1}{2}R)/\Gamma(n+\tfrac{1}{2})]^2[\Delta^{2n}/\Gamma(\tfrac{1}{2}R)\Gamma(\tfrac{1}{2}R - \tfrac{1}{2})] \times \int d\xi\, d\xi'\, d\theta\, (\sin\theta)^{R-2} e^{-\xi\xi'\cos\theta}\delta^{2n}(\xi)\delta^{2n}(\xi') = \frac{\Gamma(n+\frac{1}{2}R)}{\Gamma(\frac{1}{2}R)}\Gamma(n+1)(2\Delta)^{2n}. \tag{29}$$

Formula (24) is a useful one whenever series expansions in powers of Δ_{ij} are made for their subsequent transformation into Sommerfeld–Watson contour integrals. It also shows the fundamental distinction between odd- and even-dimensional spaces in the series they engender.

4. MATRIX INTERACTIONS

We finally consider the problem when $\boldsymbol{\phi}$ enters as a matrix in the interaction. Thus suppose $V = V(\Phi)$ is a matrix function of $\Phi = \sum_{a=1}^R \Gamma_a\phi_a$, where $\boldsymbol{\Gamma}$ are a specified set of $m \times m$ matrices with stated anticommutation properties

$$\{\Gamma_a, \Gamma_b\} = d_{abc}\Gamma_c. \tag{30}$$

We are interested in the S-matrix elements which will be N direct matrix products. The difficulty we now meet is that

$$\langle e^{\boldsymbol{\phi}\cdot\boldsymbol{\Gamma}}, e^{\boldsymbol{\phi}\cdot\boldsymbol{\Gamma}} \rangle \neq e^{\boldsymbol{\Gamma}\cdot\boldsymbol{\Gamma}'\Delta},$$

except for the trivial case when all the $\boldsymbol{\Gamma}$ commute. We have not been able to solve this problem in complete generality except for the case when $R = 2$ or 3 so that $\boldsymbol{\Gamma}$ refers to $SU(2)$ matrices—which is good enough for treating $SU(2) \otimes SU(2)$ chiral interactions such as $e^{\lambda\gamma_5\boldsymbol{\tau}\cdot\boldsymbol{\phi}}$ or $(1 + \lambda\gamma_5\boldsymbol{\tau}\cdot\boldsymbol{\phi})(1 - \lambda\gamma_5\boldsymbol{\tau}\cdot\boldsymbol{\phi})^{-1}$. Let us then demonstrate the method by some examples.

[2b] Suppose Φ is a symmetric Hermitian 2×2 matrix so that $\Phi = \sigma_1\phi_1 + \sigma_3\phi_3$, and take the case where

$$V = e^{i\lambda\Phi} = \cos\lambda\phi + i\boldsymbol{\sigma}\cdot\hat{\boldsymbol{\phi}}\sin\lambda\phi = \left(1 - \frac{i}{\lambda}\boldsymbol{\sigma}\cdot\frac{\partial}{\partial\boldsymbol{\phi}}\right)\cos\lambda\phi. \tag{31}$$

Now in general one can establish the lemma

$$\left\langle \frac{\partial}{\partial\phi_a}F(\boldsymbol{\phi}), \frac{\partial}{\partial\phi'_b}F(\boldsymbol{\phi}') \right\rangle = \int d^R\zeta\, d^R\zeta'\, \zeta_a\zeta'_b\mathfrak{F}(\zeta)\mathfrak{F}(\zeta')e^{\boldsymbol{\zeta}\cdot\boldsymbol{\zeta}'\Delta} = \frac{1}{R}\delta_{ab}\int d\zeta\, d\zeta'\, (\boldsymbol{\zeta}\cdot\boldsymbol{\zeta}')e^{\boldsymbol{\zeta}\cdot\boldsymbol{\zeta}'\Delta}\mathfrak{F}(\zeta)\mathfrak{F}(\zeta') = \frac{1}{R}\delta_{ab}\frac{\partial}{\partial\Delta}\langle F(\boldsymbol{\phi}), F(\boldsymbol{\phi}')\rangle. \tag{32}$$

Therefore, we have

$$\langle e^{i\lambda\Phi}, e^{i\lambda\Phi'} \rangle = \left(1 \otimes 1 - \frac{1}{2\lambda^2}\boldsymbol{\sigma}\otimes\boldsymbol{\sigma}\frac{\partial}{\partial\Delta}\right)\langle \cos\lambda\phi,\ \cos\lambda\phi' \rangle = 1\otimes 1\left(1 - \frac{\pi\lambda^2\Delta}{2}L_0(-\lambda^2\Delta)\right) + (\sigma_1\otimes\sigma_1 + \sigma_3\otimes\sigma_3)\frac{\partial}{\partial(\lambda^2\Delta)}\frac{\pi\lambda^2\Delta}{4}L_0(-\lambda^2\Delta), \tag{33}$$

making use of the continuation of formula (23). The $\Delta \to \infty$ asymptotic expansion of the Struve function is provided by $I_0(-\lambda^2\Delta)$ and produces exponential terms $e^{-\lambda^2\Delta}$ similarly to the one-dimensional case; thus there is an essential singularity when $x \to 0$.

[3b] Let $V = e^{i\lambda\Phi}$ again, with $\Phi = \sigma_1\phi_1 + \sigma_2\phi_2 + \sigma_3\phi_3$. The steps are as before except that there are three dimensions, and in place of (23) we have

$$\langle \cos\lambda\phi,\ \cos\lambda\phi' \rangle = \frac{\partial}{\partial\Delta}[\Delta\cosh(\lambda^2\Delta)], \tag{34}$$

$$\therefore \langle e^{i\lambda\Phi}, e^{i\lambda\Phi'} \rangle = \left[1\otimes 1 - \tfrac{1}{3}\boldsymbol{\sigma}\otimes\boldsymbol{\sigma}\frac{\partial}{\partial(\lambda^2\Delta)}\right] \times [\cosh(\lambda^2\Delta) + \lambda^2\Delta\sinh(\lambda^2\Delta)]. \tag{35}$$

[3c] Suppose that we take the alternative parametrization $V(\phi) = (1 - i\lambda\Phi)/(1 + i\lambda\Phi)$. By using the integral representation

$$V(\Phi) = \int_0^\infty d\eta\, e^{-\eta}[2e^{-i\lambda\Phi\eta} - 1]$$

and formula (35), the superpropagator reduces to

$$\langle V(\Phi), V(\Phi') \rangle = \int_0^\infty d\eta d\eta' e^{-(\eta+\eta')}(4\langle e^{-i\eta\lambda\Phi}, e^{-i\eta'\lambda\Phi'}\rangle - 3.1\otimes 1) = \int_0^\infty d\eta d\eta' e^{-(\eta+\eta')}\left[1\otimes 1\left(-3 + 4\frac{\partial}{\partial\Delta}\Delta\cosh(\lambda^2\Delta\eta\eta')\right) - \tfrac{4}{3}\boldsymbol{\sigma}\otimes\boldsymbol{\sigma}\left(1 + \frac{\partial}{\partial\Delta}\Delta\right)\sinh(\lambda^2\Delta\eta\eta')\right] = 1\otimes 1\,.\,\frac{\partial}{\partial\Delta}\int_0^\infty \frac{\Delta(1 + 3\lambda^4\Delta^2\zeta^2)}{1 - \lambda^4\Delta^2\zeta^2}e^{-\zeta}d\zeta$$

$$-\tfrac{4}{3}\sigma\otimes\sigma\left(1+\frac{\partial}{\partial\Delta}\Delta\right)\int_0^\infty\frac{\lambda^2\Delta\zeta e^{-\zeta}}{1-\lambda^4\Delta^2\zeta^2}d\zeta$$

and so on.

The basic reason why we have been able to make good progress with the above computations is that $SU(2)$ matrices Γ of degree j satisfy polynomial equations $P_j(\Gamma\cdot\hat{\phi})=0$ so that expansions in powers of Φ always terminate in their tensorial complexity.

For larger groups like $SU(3)$, where the anticommutation structure constants are not so simple,[9] an alternative procedure will need to be devised. This will be relevant to gravitational interactions[10] in exponential coordinates where $g_{\mu\nu}=[\exp(\eta+\kappa h)]_{\mu\nu}$ is a 4×4 symmetric matrix, with η standing for the Minkowski metric and h for the gravitational spin-2 field.

Note Added in Proof: The cases of gravity and $SU(3)$ have now been satisfactorily resolved by an entirely different technique. See "On Matrix Superpropagators I and II" by J. Ashmore R. Delbourgo, J. Math. Phys. (to be published).

1 E. S. Fradkin, Nucl. Phys. **49**, 624 (1963); G. V. Efimov, Zh. Eksp. Teor. Fiz. **74**, 2107 (1963)[Sov. Phys. JETP **17**, 1417 (1963)].

2 R. Delbourgo, A. Salam, and J. Strathdee, Phys. Rev. **187**, 1999 (1969); A. Salam and J. Strathdee, Phys. Rev. D 1,3296 (1970).

3 G. V. Efimov, Nucl. Phys. **74**, 657 (1965); H. M. Fried, Phys. Rev. **174**, 1725 (1968); R. Delbourgo, K. Koller, and A. Salam, Ann. Phys. (N.Y.) **16**, 19 (1971).

4 H. M. Fried, Nuovo Cimento **52A**, 1333 (1967); M. K. Volkov, Ann. Phys. (N.Y.) **49**, 202 (1968); J. Giambiagi and J. Tiomno, Lettere Nuovo Cimento **2**, 674 (1969); B. W. Lee and B. Zumino, Nucl. Phys. **13B**, 671 (1969).

5 The only works on nonpolynomial methods which give *detailed* calculations for Lagrangians with isospin that we are aware of are by A. Hunt, K. Koller, and Q. Shafi, Phys. Rev. (N.Y.) to be published (N.Y.) and J. Sultoon, Ph. D. thesis (Imperial College, 1970).

6 I. M. Gel'fand and G. Shilov, *Generalized Functions*, Vol. I (Academic, New York); A. Salam and J. Strathdee Ref. 2); M. Volkov and J. Giambiagi and J. Tiomno, Ref. 4. See also H. Lehmann and K. Pohlmeyer, Commun. Math. Phys. **20**, 101 (1971).

7 A. Erdelyi *et al.*, *Higher Transcendental Functions*, Bateman Manuscript Project (McGraw-Hill, New York, 1953), Vol. 2, Sec. 7.5.4.

8 Note that as $\Delta\to\infty$, (26) behaves as Δ^{-1} in contradistinction to the one-dimensional case which from (11) behaves as $(\ln\Delta)/\Delta$.

9 To see how difficult the problem becomes, even for $SU(3)$, consult A. Macfarlane, A. Sudbery, and P. H. Weisz, Commun. Math. Phys. **11**, 77 (1966).

10 R. Delbourgo and A. Hunt, Lettere Nuovo Cimento **4**, 1010 (1970); C. Isham, A. Salam and J. Strathdee, Phys. Rev. **3**, 1805 (1971).

A Class of Mean Field Models

H. Silver and N. E. Frankel

School of Physics, University of Melbourne, Parkville, Victoria, 3052, Australia.

and

B. W. Ninham

Department of Applied Mathematics, University of New South Wales, Kensington, N.S.W., Australia.

(Received 29 October 1970)

A model of isotropically interacting ν-dimensional classical spins with an infinite range potential of the molecular field-type is solved. The partition function is represented as the integral of $e^{-\beta H_N}$ over an appropriate weight function, which, for given ν, is the Pearson random walk probability distribution in ν dimensions. A molecular field-type phase transition is obtained for all ν.

I. INTRODUCTION

Stanley[1] has shown that the Hamiltonian

$$H_N^{(\nu)}=-J\sum_{<ij>}\mathbf{S}_i\cdot\mathbf{S}_j, \tag{1}$$

where $\mathbf{S}_i=(S_i(1),S_i(2),\text{----},S_i(\nu))$ are ν-dimensional vectors of magnitude $\sqrt{\nu}$ and $[S_i(\alpha),S_i(\beta)]_-=0$ (for $\nu>1$ the magnitude of the spin is infinite); hence, for all ν, the model describes classical spins, spans an infinite class of models ranging from the Ising model with $\nu=1$, through the classical planar model ($\nu=2$), the classical Heisenberg model ($\nu=3$), and tending towards the Berlin–Kac spherical model as $\nu\to\infty$.

Stanley's exact solutions have been confined to nearest neighbor one-dimensional chains and hence do not exhibit a phase transition. It is therefore of some interest to study this model with an infinite range potential which will yield a phase transition. The simplest infinite range potential is obtained when every spin interacts equally with every other spin, that is a molecular field-type potential.

We study the statistical mechanics of a system whose Hamiltonian is

$$H_N^{(\nu)}=-(2J/N)\sum_{i\le i<j\le N}\mathbf{S}_i\cdot\mathbf{S}_j-\mu_0\mathbf{B}\cdot\sum_{i=1}^{N}\mathbf{S}_i, \tag{2}$$

where $\mathbf{B}$ is an external field and μ_0 is the magnetic moment associated with a single spin. Since $\|\mathbf{S}_i\|=\sqrt{\nu}$,

$$\left(\sum_{i=1}^{N}\mathbf{S}_i\right)^2=N\nu+2\sum_{1\le i<j\le N}\mathbf{S}_i\cdot\mathbf{S}_j \tag{3}$$

and

$$H_N^{(\nu)}=J\nu-\frac{J}{N}\left(\sum_{i=1}^{N}\mathbf{S}_i\right)^2-\mu_0\mathbf{B}\cdot\sum_{i=1}^{N}\mathbf{S}_i. \tag{4}$$

This Hamiltonian characterizes the totality of molecular field potentials previously considered; for $\nu=1$ it is the Curie–Weiss Ising model. The work of the paper is arranged as follows:

In Sec. II the partition function Q_N is evaluated by reducing the summation over all the spins $\mathbf{S}_i$, $i=1,2\ldots,N$, to integration over the single variable $\mathbf{S}=\sum_{i=1}^{N}\mathbf{S}_i$. In this way the partition function is expressed as a functional of a Pearson random walk probability distribution. This representation explicitly displays the connection between the molecular field-type phase transition and the underlying random walk. The distribution has a well-known integral representation and the multiple integral obtained for Q_N is evaluated by the method of steepest descents (generalized to such multidimensional integrals). The mechanism of the phase transition (in the case $\mathbf{B}=0$) is then the

$$-\tfrac{4}{3}\sigma\otimes\sigma\left(1+\frac{\partial}{\partial\Delta}\Delta\right)\int_0^\infty\frac{\lambda^2\Delta\zeta e^{-\zeta}}{1-\lambda^4\Delta^2\zeta^2}d\zeta$$

and so on.

The basic reason why we have been able to make good progress with the above computations is that $SU(2)$ matrices Γ of degree j satisfy polynomial equations $P_j(\Gamma\cdot\hat{\phi})=0$ so that expansions in powers of Φ always terminate in their tensorial complexity.

For larger groups like $SU(3)$, where the anticommutation structure constants are not so simple,[9] an alternative procedure will need to be devised. This will be relevant to gravitational interactions[10] in exponential coordinates where $g_{\mu\nu}=[\exp(\eta+\kappa h)]_{\mu\nu}$ is a 4×4 symmetric matrix, with η standing for the Minkowski metric and h for the gravitational spin-2 field.

Note Added in Proof: The cases of gravity and $SU(3)$ have now been satisfactorily resolved by an entirely different technique. See "On Matrix Superpropagators I and II" by J. Ashmore R. Delbourgo, J. Math. Phys. (to be published).

1 E. S. Fradkin, Nucl. Phys. **49**, 624 (1963); G. V. Efimov, Zh. Eksp. Teor. Fiz. **74**, 2107 (1963)[Sov. Phys. JETP **17**, 1417 (1963)].

2 R. Delbourgo, A. Salam, and J. Strathdee, Phys. Rev. **187**, 1999 (1969); A. Salam and J. Strathdee, Phys. Rev. D 1,3296 (1970).

3 G. V. Efimov, Nucl. Phys. **74**, 657 (1965); H. M. Fried, Phys. Rev. **174**, 1725 (1968); R. Delbourgo, K. Koller, and A. Salam, Ann. Phys. (N.Y.) **16**, 19 (1971).

4 H. M. Fried, Nuovo Cimento **52A**, 1333 (1967); M. K. Volkov, Ann. Phys. (N.Y.) **49**, 202 (1968); J. Giambiagi and J. Tiomno, Lettere Nuovo Cimento **2**, 674 (1969); B. W. Lee and B. Zumino, Nucl. Phys. **13B**, 671 (1969).

5 The only works on nonpolynomial methods which give *detailed* calculations for Lagrangians with isospin that we are aware of are by A. Hunt, K. Koller, and Q. Shafi, Phys. Rev. (N.Y.) to be published (N.Y.) and J. Sultoon, Ph. D. thesis (Imperial College, 1970).

6 I. M. Gel'fand and G. Shilov, *Generalized Functions*, Vol. I (Academic, New York); A. Salam and J. Strathdee Ref. 2); M. Volkov and J. Giambiagi and J. Tiomno, Ref. 4. See also H. Lehmann and K. Pohlmeyer, Commun. Math. Phys. **20**, 101 (1971).

7 A. Erdelyi *et al.*, *Higher Transcendental Functions*, Bateman Manuscript Project (McGraw-Hill, New York, 1953), Vol. 2, Sec. 7.5.4.

8 Note that as $\Delta\to\infty$, (26) behaves as Δ^{-1} in contradistinction to the one-dimensional case which from (11) behaves as $(\ln\Delta)/\Delta$.

9 To see how difficult the problem becomes, even for $SU(3)$, consult A. Macfarlane, A. Sudbery, and P. H. Weisz, Commun. Math. Phys. **11**, 77 (1966).

10 R. Delbourgo and A. Hunt, Lettere Nuovo Cimento **4**, 1010 (1970); C. Isham, A. Salam and J. Strathdee, Phys. Rev. **3**, 1805 (1971).

A Class of Mean Field Models

H. Silver and N. E. Frankel

School of Physics, University of Melbourne, Parkville, Victoria, 3052, Australia.

and

B. W. Ninham

Department of Applied Mathematics, University of New South Wales, Kensington, N.S.W., Australia.

(Received 29 October 1970)

A model of isotropically interacting ν-dimensional classical spins with an infinite range potential of the molecular field-type is solved. The partition function is represented as the integral of $e^{-\beta H_N}$ over an appropriate weight function, which, for given ν, is the Pearson random walk probability distribution in ν dimensions. A molecular field-type phase transition is obtained for all ν.

I. INTRODUCTION

Stanley[1] has shown that the Hamiltonian

$$H_N^{(\nu)}=-J\sum_{<ij>}\mathbf{S}_i\cdot\mathbf{S}_j, \tag{1}$$

where $\mathbf{S}_i=(S_i(1),S_i(2),\text{----},S_i(\nu))$ are ν-dimensional vectors of magnitude $\sqrt{\nu}$ and $[S_i(\alpha),S_i(\beta)]_-=0$ (for $\nu>1$ the magnitude of the spin is infinite); hence, for all ν, the model describes classical spins, spans an infinite class of models ranging from the Ising model with $\nu=1$, through the classical planar model ($\nu=2$), the classical Heisenberg model ($\nu=3$), and tending towards the Berlin–Kac spherical model as $\nu\to\infty$.

Stanley's exact solutions have been confined to nearest neighbor one-dimensional chains and hence do not exhibit a phase transition. It is therefore of some interest to study this model with an infinite range potential which will yield a phase transition. The simplest infinite range potential is obtained when every spin interacts equally with every other spin, that is a molecular field-type potential.

We study the statistical mechanics of a system whose Hamiltonian is

$$H_N^{(\nu)}=-(2J/N)\sum_{i\le i<j\le N}\mathbf{S}_i\cdot\mathbf{S}_j-\mu_0\mathbf{B}\cdot\sum_{i=1}^{N}\mathbf{S}_i, \tag{2}$$

where $\mathbf{B}$ is an external field and μ_0 is the magnetic moment associated with a single spin. Since $\|\mathbf{S}_i\|=\sqrt{\nu}$,

$$\left(\sum_{i=1}^{N}\mathbf{S}_i\right)^2=N\nu+2\sum_{1\le i<j\le N}\mathbf{S}_i\cdot\mathbf{S}_j \tag{3}$$

and

$$H_N^{(\nu)}=J\nu-\frac{J}{N}\left(\sum_{i=1}^{N}\mathbf{S}_i\right)^2-\mu_0\mathbf{B}\cdot\sum_{i=1}^{N}\mathbf{S}_i. \tag{4}$$

This Hamiltonian characterizes the totality of molecular field potentials previously considered; for $\nu=1$ it is the Curie–Weiss Ising model. The work of the paper is arranged as follows:

In Sec. II the partition function Q_N is evaluated by reducing the summation over all the spins $\mathbf{S}_i$, $i=1,2\ldots,N$, to integration over the single variable $\mathbf{S}=\sum_{i=1}^{N}\mathbf{S}_i$. In this way the partition function is expressed as a functional of a Pearson random walk probability distribution. This representation explicitly displays the connection between the molecular field-type phase transition and the underlying random walk. The distribution has a well-known integral representation and the multiple integral obtained for Q_N is evaluated by the method of steepest descents (generalized to such multidimensional integrals). The mechanism of the phase transition (in the case $\mathbf{B}=0$) is then the

"sticking" of the saddle point for $T > T_c$, the critical temperature.

In Sec. III expressions for the thermodynamic functions of interest (magnetization, internal energy, specific heat for the case $\mathbf{B} = 0$, and susceptibility for $\mathbf{B} \neq 0$) are developed along with expansions about $T = 0$ and $T = T_c$.

Section IV presents a discussion of these results.

II. PARTITION FUNCTION

The normalized partition function corresponding to the Hamiltonian (4) is

$$Q_N^{(\nu)}(K_1, \mathbf{K}_2) = Z_N^{(\nu)}(0,0)^{-1} \times Z_N^{(\nu)}(K_1, \mathbf{K}_2), \tag{5}$$

where

$$Z_N^{(\nu)}(K_1, \mathbf{K}_2) = e^{-\nu K_1} \int_{-\infty}^{\infty} \cdots \int_{\|\mathbf{S}_i\| = \sqrt{\nu}} \exp\left[\frac{K_1}{N}\left(\sum_{i=1}^{N} S_i\right)^2 + \mathbf{K}_2 \cdot \sum_{i=1}^{N} \mathbf{S}_i\right] \times d\mathbf{S}_i \cdots d\mathbf{S}_N, \tag{6}$$

where $K_1 = \beta J$ and $\mathbf{K}_2 = \beta \mu_0 \mathbf{B}$. Clearly, we have

$$Z_N^{(\nu)}(0,0) = \left(\int_{\|\mathbf{S}\| = \sqrt{\nu}} d\mathbf{S}\right)^N$$

so that

$$Z_N^{(\nu)}(0,0) = [2\pi^{\nu/2} \nu^{(\nu-1)/2} / \Gamma(\nu/2)]^N. \tag{7}$$

The integral is just the surface are of a ν-ball of radius $\sqrt{\nu}$. The integrand in Eq. (5) is a function of $\mathbf{S} = \sum_{i=1}^{N} \mathbf{S}_i$ only and we can write

$$Q_N^{(\nu)}(K_1, \mathbf{K}_2) = e^{-\nu K_1} \int_{-\infty}^{\infty} P_N^{(\nu)}(\mathbf{S}) e^{(K_1/N)\mathbf{S}^2 + \mathbf{K}_2 \cdot \mathbf{S}} \, d\mathbf{S}, \tag{8}$$

where $P_N^{(\nu)}(\mathbf{S})d\mathbf{S}$ is the probability that $\sum_{i=1}^{N} \mathbf{S}_i \in (\mathbf{S}, \mathbf{S} + d\mathbf{S})$ given that $\|\mathbf{S}_i\| = \sqrt{\nu}$. $Z_N^{(\nu)}(0,0)$ has been divided out as it is the measure of the sample space on which we define $P_N^{(\nu)}(\mathbf{S})$ [for example, for $\nu = 1$, $Z_N^{(\nu)}(0,0) = 2^N$, the total number of walks of N steps]. $P_N^{(\nu)}(\mathbf{S})$ is the probability distribution for Pearson's walk,[2] whose integral representation was first given, for $\nu = 2$ by Kluyver[3] and for $\nu = 3$ by Rayleigh.[4] Now, because of the symmetrical nature of the walks,

$$P_N^{(\nu)}(\mathbf{S}) = P_N^{(\nu)}(S) = \frac{\partial}{\partial S} W_N^{(\nu)}(S),$$

where $W_N^{(\nu)}(S)$ is the probability that $\|\sum_{i=1}^{N} \mathbf{S}_i\| < S$. $P_N^{(\nu)}(S)$ is normalized by

$$\int_{-\infty}^{\infty} P_N^{(\nu)}(S) dS = 1.$$

The distribution $W_N^{(\nu)}(S)$ is given by Watson[5] and we find

$$P_N^{(\nu)}(S) = \frac{[\Gamma(\nu/2)]^N}{\pi^{\nu/2} S^{\nu-1}} \int_0^{\infty} (\tfrac{1}{2}St)^{\nu/2} J_{\nu/2-1}(St) \left(\frac{J_{\nu/2-1}(\sqrt{\nu}t)}{(\frac{1}{2}\sqrt{\nu}t)^{\nu/2-1}}\right)^N dt$$

$$= \frac{[\Gamma(\nu/2)]^N}{\sqrt{\nu}\,\pi^{\nu/2} S^{\nu-1}} \int_0^{\infty} \left(\tfrac{1}{2}\frac{S}{\sqrt{\nu}}u\right)^{\nu/2} J_{\nu/2-1}\left(\frac{S}{\sqrt{\nu}}u\right) \times \left(\frac{J_{\nu/2-1}(u)}{(\frac{1}{2}u)^{\nu/2-1}}\right)^N du, \tag{9}$$

where $J_\alpha(z)$ is a Bessel function.

For $N \gg 1$ it is easy to prove, using Watson's lemma, that

$$P_N^{(\nu)}(S) \sim [1/(2N\pi)^{\nu/2}] e^{-S^2/2N} \tag{10}$$

provided $S/N\sqrt{\nu} < 1$. Clearly $P_N^{(\nu)}(S) = 0$ if $S > \sqrt{\nu} N$, the maximum extension of N walks each of length $\sqrt{\nu}$. The Gaussian form of the limiting distribution, Eq. (10), is a general consequence of the Markovian nature of the random walk problem. The polar integrals in Eq. (8) are performed using Eq. (A10) of Appendix A, and we find

$$Q_N^{(\nu)}(K_1, \mathbf{K}_2) = \tfrac{1}{2} e^{-\nu K_1} \int_{-N\sqrt{\nu}}^{N\sqrt{\nu}} P_N^{(\nu)}(S) e^{K_1 S^2/N} 2\pi^{\nu/2} S^{\nu-1} \times \frac{I_{\nu/2-1}(K_2 S)}{(\frac{1}{2}K_2 S)^{\nu/2-1}} \, dS, \tag{11}$$

where $I_\alpha(z)$ is a Bessel function of imaginary argument. By Eq. (9)

$$Q_N^{(\nu)}(K_1, \mathbf{K}_2) = \frac{[\Gamma(\nu/2)]^N e^{-\nu K_1}}{\sqrt{\nu}} \int_{-N\sqrt{\nu}}^{N\sqrt{\nu}} e^{K_1 S^2/N} \frac{I_{\nu/2-1}(K_2 S)}{(\frac{1}{2}K_2 S)^{\nu/2-1}} \left[\int_0^{\infty} \left(\frac{1}{2}\frac{S}{\sqrt{\nu}}u\right)^{\nu/2} \times J_{\nu/2-1}\left(\frac{S}{\sqrt{\nu}}u\right) \left(\frac{J_{\nu/2-1}(u)}{(\frac{1}{2}u)^{\nu/2-1}}\right)^N du\right] dS. \tag{12}$$

With the change of variable, $S = \sqrt{\nu} N x$, we have[6]

$$Q_N^{(\nu)}(K_1, \mathbf{K}_2) = N[\Gamma(\nu/2)]^N e^{-\nu K_1} \times \int_{-1}^{1} e^{\nu N K_1 x^2} \frac{I_{\nu/2-1}(\sqrt{\nu} N K_2 x)}{(\frac{1}{2}\sqrt{\nu} N K_2 x)^{\nu/2-1}} \left[\int_0^{\infty} (\tfrac{1}{2}Nxu)^{\nu/2} \times J_{\nu/2-1}(Nxu) \left(\frac{J_{\nu/2-1}(u)}{(\frac{1}{2}u)^{\nu/2-1}}\right)^N du\right] dx. \tag{13}$$

The further reduction of this expression proceeds as follows. The integral can be recast into an easily manageable form if we substitute the integral representations[7]

$$J_\alpha(z) = \frac{(\frac{1}{2}z)^\alpha}{\Gamma(\alpha + \frac{1}{2})\Gamma(\frac{1}{2})} \int_{-\pi/2}^{\pi/2} e^{iz\sin\phi} \cos^{2\alpha}\phi \, d\phi,$$

$$I_\alpha(z) = \frac{(\frac{1}{2}z)^\alpha}{\Gamma(\alpha + \frac{1}{2})\Gamma(\frac{1}{2})} \int_{-\pi/2}^{\pi/2} e^{z\sin\phi} \cos^{2\alpha}\phi \, d\phi. \tag{14}$$

This gives

$$Q_N^{(\nu)}(K_1, \mathbf{K}_2) = e^{-\nu K_1} [\Gamma(\nu/2)]^N \frac{N^\nu}{2^{\nu-1}\pi[\Gamma(\frac{1}{2}\nu - \frac{1}{2})]^2} \times \int_{-1}^{1} dx \int_0^{\infty} du \int_{-\pi/2}^{\pi/2} d\phi_1 \int_{-\pi/2}^{\pi/2} d\phi_2 f(x, u, \phi_1, \phi_2) \times e^{\nu N g(x,u,\phi_1,\phi_2)}, \tag{15}$$

where

$$f(x, u, \phi_1, \phi_2) = x^{\nu-1} u^{\nu-1} (\cos\phi_1 \cos\phi_2)^{\nu-2},$$

$$g(x, u, \phi_1, \phi_2) = K_1 x^2 + (1/\sqrt{\nu}) K_2 x \sin\phi_1 + (1/\nu) ixu \sin\phi_2 + (1/\nu) \times \log[J_{\nu/2-1}(u)/(\tfrac{1}{2}u)^{\nu/2-1}]. \tag{16}$$

For large N the major contribution to the multiple integral comes from around the saddle point of $g(x,u,\phi_1,\phi_2)$. This saddle point is determined by

$$\frac{\partial g}{\partial x} = \frac{\partial g}{\partial u} = \frac{\partial g}{\partial \phi_1} = \frac{\partial g}{\partial \phi_2} = 0,$$
$$(x = x_s, u = u_s, \phi_1 = \phi_{1_s}, \phi_2 = \phi_{2_s}). \quad (17)$$

Explicitly, we have

$$\begin{aligned}
\frac{\partial g}{\partial x} &= 2K_1 x + \frac{1}{\sqrt{\nu}} K_2 \sin\phi_1 + \frac{1}{\nu} iu \sin\phi_2,\\
\frac{\partial g}{\partial u} &= \frac{1}{\nu} ix \sin\phi_2 - \frac{1}{\nu}\frac{J_{\nu/2}(u)}{J_{\nu/2-1}(u)},\\
\frac{\partial g}{\partial \phi_1} &= \frac{1}{\sqrt{\nu}} K_2 x \cos\phi_1,\\
\frac{\partial g}{\partial \phi_2} &= \frac{1}{\nu} ixu \cos\phi_2.
\end{aligned} \quad (18)$$

If the trivial solution $x_s = 0$ is excluded, the last two equations give $\cos\phi_{i_s} = 0$, so that $|\sin\phi_{i_s}| = 1$. Without loss of generality[8] we may take $\sin\phi_{i_s} = 1$. The first two of Eqs. (18) then reduce to

$$\begin{aligned}
2\nu K_1 x_s + \sqrt{\nu} K_2 + iu_s &= 0,\\
ix_s - J_{\nu/2}(u_s)/J_{\nu/2-1}(u_s) &= 0.
\end{aligned} \quad (19)$$

Combination of these two equations yields the transcendental equation satisfied by the saddle point x_s, namely[9]

$$x_s = -iJ_{\nu/2}[i(2\nu K_1 x_s + \sqrt{\nu} K_2)]/J_{\nu/2-1}[i(2\nu K_1 x_s + \sqrt{\nu} K_2)],$$

whence

$$x_s = I_{\nu/2}(2\nu K_1 x_s + \sqrt{\nu} K_2)]/I_{\nu/2-1}(2\nu K_1 x_s + \sqrt{\nu} K_2)] \quad (20)$$

and

$$x_s = \frac{1}{2\nu K_1}\frac{\partial}{\partial x_s}\log\left(\frac{I_{\nu/2-1}(2\nu K_1 x_s + \sqrt{\nu} K_2)}{[\tfrac{1}{2}(2\nu K_1 x_s + \sqrt{\nu} K_2]^{\nu/2-1}}\right). \quad (21)$$

The method of steepest descents applied to Eq. (15) yields

$$Q_N^{(\nu)}(K_1, \mathbf{K}_2) \sim e^{-\nu K_1}[\Gamma(\nu/2)]^N C_{N,\nu}(x_s) e^{\nu N G(x_s)}, \quad (22)$$

where

$$\begin{aligned}
G(x_s) &= g(x_s, u_s, \phi_{1_s}, \phi_{2_s})\\
&= -K_1 x_s^2 + \frac{1}{\nu}\log\left(\frac{I_{\nu/2-1}(2\nu K_1 x_s + \sqrt{\nu} K_2)}{[\tfrac{1}{2}(2\nu K_1 x_s + \sqrt{\nu} K_2)]^{\nu/2-1}}\right)
\end{aligned} \quad (23)$$

and

$$\frac{\partial G}{\partial x_s} = 0. \quad (24)$$

Since $\cos\phi_{i_s} = 0$, $f(x_s, u_s, \phi_{1_s}, \phi_{2_s}) = 0$, so that explicit determination of $C_{N,\nu}(x_s)$ is difficult and in any event unnecessary since $\log[C_{N,\nu}(x_s)]/N \to 0$, as $N \to \infty$. This follows because $f(x_s, u_s, \phi_{1_s}, \phi_{2_s})$ is not of the order e^N.

III. THERMODYNAMICS

The mean magnetic moment per particle of the system, $\mathbf{M}$, is given by

$$\mathbf{M} = \left\langle \frac{\mu_0}{N}\sum_{i=1}^{N} \mathbf{S}_i \right\rangle, \quad (25)$$

and from Eq. (5) we have

$$\mathbf{M} = Z_N^{-1}\int_{\|\mathbf{S}_i\| = \sqrt{\nu}}^{\infty}\cdots\int e^{-\beta H_N(\mathbf{S}_i)}\left(\frac{\mu_0}{N}\sum_{i=1}^{N}\mathbf{S}_i\right) d\mathbf{S}_1 \ldots d\mathbf{S}_N \quad (26)$$

$$= \mu_0 \frac{\int_{-\infty}^{\infty} N^{-1} S P_N^{(\nu)}(S) e^{(K_2/N)\mathbf{S}^2 + \mathbf{K}_2\cdot\mathbf{S}}\, d\mathbf{S}}{\int_{-\infty}^{\infty} P_N^{(\nu)}(S) e^{(K_1/N)\mathbf{S}^2 + \mathbf{K}_2\cdot\mathbf{S}} d\mathbf{S}}. \quad (27)$$

It is easily shown that

$$\mathbf{M} = \sqrt{\nu}\,\mu_0 x_s \hat{\mathbf{B}}, \quad (28)$$

where $\hat{\mathbf{B}}$ is a vector of unit length parallel to $\mathbf{B}$. Now

$$2\nu K_1 x_s + \sqrt{\nu} K_2 = \beta(2\nu J x_s + \sqrt{\nu}\,\mu_0 B); \quad (29)$$

by comparison of Eqs. (20) and (28) it is clear that we can interpret $2Jx_s$ as the "mean field" of Weiss.[10] For $\mathbf{B} = 0$, we have $K_2 = 0$, and Eq. (20) reduces to

$$x_s = I_{\nu/2}(2\nu K x_s)/I_{\nu/2-1}(2\nu K x_s). \quad (30)$$

For $K < \frac{1}{2}$, Eq. (30) has only the trivial solution $x_s = 0$, while for $K > \frac{1}{2}$ there is a nonzero solution. Thus a spontaneous magnetization exists for $T < T_c = 2J/k$.

For $\nu = 1$ and $\nu = 3$, from Eq. (20) we recover the well known molecular field equations

$$x_s = \tanh(2K_1 x_s + K_2), \quad (31)$$

$$x_s = \coth(6K_1 x_s + \sqrt{3}K_2) - 1/6K_1 x_s + \sqrt{3}K_2. \quad (32)$$

Equation (31) is the classical Weiss equation for the molecular field, while the right-hand side of Eq. (32) is a Langevin function of argument $6K_1 x_s + \sqrt{3}K_2$. In Appendix C we show that for $\nu \to \infty$ (and $\mathbf{B} = 0$) we recover the spherical model result [Appendix B, Eq. (B12)] which gives the saddle point as

$$2Kx_s^3 - (2K - 1)x_s = 0. \quad (33)$$

The Helmholtz free energy per spin is given by

$$-\beta\psi = \lim_{N\to\infty}\log(Q_N)/N \quad (34)$$

and by Eq. (22)

$$-\beta\psi/\nu = (1/\nu)\log\Gamma(\nu/2) + G(x_s). \quad (35)$$

In the case $\mathbf{B} = 0$, when $T > T_c$, $x_s = 0$ and $-\beta\psi = 0$. On the other hand, for $T < T_c$, $-\beta\psi \neq 0$. Thus T_c is the critical temperature of the system. If we let U and C_v denote the internal energy per particle and specific heat per particle, respectively, then

$$U = -J\frac{d}{dK}(-\beta\psi), \qquad C_v = kK^2\frac{d^2}{dK^2}(-\beta\psi). \quad (36)$$

Now

$$\begin{aligned}
\frac{d}{dK}(-\beta\psi) &= \frac{\partial}{\partial K}(-\beta\psi) + \frac{\partial}{\partial x_s}(-\beta\psi)\frac{dx_s}{dK}\\
&= \frac{\partial}{\partial K}(-\beta\psi)
\end{aligned} \quad (37)$$

since, for $T < T_c$, $(\partial/\partial x_s)(-\beta\psi) = \partial G/\partial x_s = 0$, while, for $T > T_c$, $dx_s/dK = 0$. We find

$$\frac{U}{\nu} = -Jx_s^2, \qquad \frac{C_v}{\nu} = 2kK^2 x_s \frac{dx_s}{dK}. \tag{38}$$

Clearly $U = C_v = 0$ for $T > T_c$, again the usual feature characteristic of molecular field theories.

For $T \lesssim T_c$, $x_s \cong 0$, and from Appendix D we find

$$x_s \sim \sqrt{[(2+\nu)/\nu](1 - T/T_c)}, \tag{39}$$

$$U/\nu \cong -J[(2+\nu)/\nu](1 - T/T_c) \tag{40}$$

$$\to 0, \tag{41}$$

and

$$C_v/\nu \to (k/2)(2+\nu)/\nu, \qquad \text{as } T \to T_c - 0. \tag{42}$$

For $T \cong 0$, $K \gg 1$ and from Appendix D we find, for $\nu = 1$,

$$x_s \sim 1 - 2e^{-4K}, \tag{43}$$

$$U \cong -J(1 - 4e^{-4K}) \tag{44}$$

$$\to -J, \tag{45}$$

and

$$C_v \to 0, \qquad \text{as } T \to 0+, \tag{46}$$

while, for $\nu > 1$,

$$x_s \sim 1 - \tfrac{1}{2}[(\nu-1)/\nu]T/T_c, \tag{47}$$

$$U/\nu \cong -J(1 - [(\nu-1)/\nu]T/T_c) \tag{48}$$

$$\to -J, \tag{49}$$

and

$$C_v/\nu \to (k/2)(\nu-1)/\nu, \qquad \text{as } T \to 0+. \tag{50}$$

These tend to the correct spherical model results (Appendix B) in the limit $\nu \to \infty$. In Fig. 1 the normalized specific heat, $C_v/\nu k$ is plotted as a function of temperature T, for various spin dimensionalities.

The susceptibility χ is given by

$$\chi = N\frac{\partial M}{\partial B}. \tag{51}$$

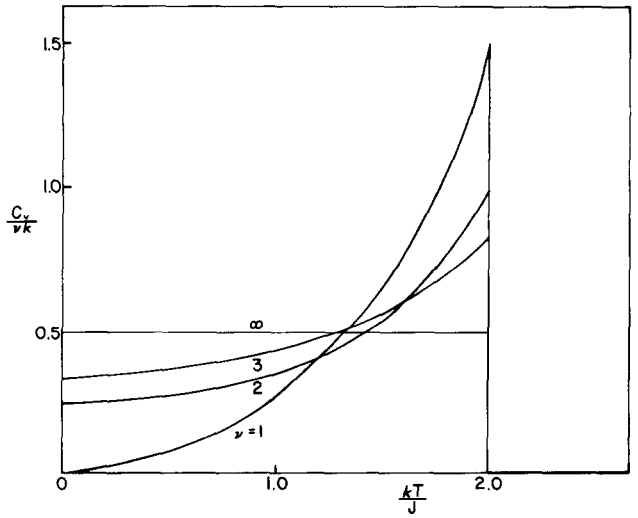


FIG 1. Sketch of the normalized specific heat $C_v/\nu k$ as a function of temperature T for various spin dimensionalities.

For $T > T_c$ and with a weak external field B, $x_s \cong 0$ and we use the small argument expansion (Appendix D, Eq. (D2)] to find

$$x_s = \tfrac{1}{2}\frac{2\nu K_1 x_s + \sqrt{\nu}K_2}{\nu/2}\left[1 - \frac{(2\nu K_1 x_s + \sqrt{\nu}K_2)^2}{(\nu/2)(\nu/2-1)} + \cdots\right] \tag{52}$$

so that

$$x_s \cong 2K_1 x_s + K_2/\sqrt{\nu}, \tag{53}$$

and we find

$$x_s = \mu_0 B/\sqrt{\nu}k(T - T_c), \tag{54}$$

where T_c is the critical temperature for zero external field, that is, $T_c = 2J/k$. Then, by Eqs. (28) and (51),

$$\chi = N\mu_0^2/k(T - T_c) \tag{55}$$

and for every spin dimensionality the model obeys the Curie-Weiss law.

IV. DISCUSSION

For a given value of ν the Curie-Weiss model is characterized by a Pearson walk in ν dimensions. This walk has a probability distribution $P_N^{(\nu)}(S)$ which for $\nu = 1$ is discrete, while for $\nu > 1$ is a nonregular function of S for $|S| < \sqrt{\nu}\,N$. The Curie-Weiss spherical model is characterized by a random walk which has a probability distribution $P_N(S)$ regular for $|S| < N$.

The Curie-Weiss Ising model ($\nu = 1$) is qualitatively different from all the models with $\nu > 1$. This is because for all $\nu > 1$ the implicit absolute value of the spin is infinite.

A general feature of all mean field models is the zero specific heat for $T > T_c$ (in the case $\mathbf{B} = 0$). This "insensitiveness" of the specific heat to the dimensionality of the spin is a general consequence of the Gaussian form [Eq. (10)] of the limiting probability distribution. This, in turn, is due to the Markovian nature of the random walk problem. An additional observation consequent on this is the following. Any random walk with finite correlations is still asymptotically Gaussian.[11] It seems clear that any modification of the Hamiltonian which generates by perturbation theory a modified walk with finite correlations will lead to a partition function with the same structure as the unperturbed case (for example, a jump discontinuity in the specific heat).

The relationship between the partition function and a particular kind of random walk, namely the Pearson walk, is explicitly displayed in this problem. Recently, in a series of papers, Domb[11,12] has noted the close relationship between dominant terms in the coefficients in high temperature series expansions for the spin-$\frac{1}{2}$ Ising model, classical planar model and classical Heisenberg model and the geometrical properties of self-avoiding walks on the given lattice. It may then prove instructive to study the form of Eq. (11) for the partition function, where $P_N(S)$ now represents the distribution for a random walk other than the Pearson walk. The only self-avoiding walk problem that has been exactly solved is the so-called "Manhattan walk" of Kasteleyn[13] and we can take $P_N(S)$

for this walk and then evaluate Q_N. In this case, the presence of infinite range correlations between the walks means that the problem is no longer Markovian, and so the limiting distribution is no longer Gaussian and we expect qualitatively different thermodynamic behavior for this model. Alternatively we can take $P_N(S)$ for a walk on a defective lattice and investigate how defects alter the nature of the phase transition. Clearly, however, the Hamiltonian corresponding to the partition function represented by the integral in Eq. (11) must be discerned if we are to attach physical significance to such studies.

ACKNOWLEDGMENT

One of us (H.S.) gratefully acknowledges the financial assistance of the Commonwealth Government through their Commonwealth Post Graduate Award.

APPENDIX A: POLAR INTEGRATIONS IN ν-DIMENSIONS

We want to integrate over the angular coordinates in the integral

$$I = \int_{-\infty}^{\infty} f(\mathbf{S})d\mathbf{S}, \tag{A1}$$

where $\mathbf{S} = (S_1, S_2, \ldots, S_\nu)$ is a ν-vector. To do this, we write

$$I = \int_0^\infty ds\left(\int_{S=\|\mathbf{S}\|} f(\mathbf{S})d\mathbf{S}\right) \tag{A2}$$

and note that

$$\int_{S=\|\mathbf{S}\|} f(\mathbf{S})d\mathbf{S} = \int_{-\infty}^{\infty} \delta(S - \|\mathbf{S}\|) f(\mathbf{S})d\mathbf{S} \tag{A3}$$

$$= 2S\int_{-\infty}^{\infty} \delta(S^2 - \|\mathbf{S}\|^2) f(\mathbf{S})d\mathbf{S}. \tag{A4}$$

Now we introduce the delta function integral representation,

$$\delta(x) = \frac{1}{2\pi i}\int_{-i\infty}^{i\infty} e^{zx}dz, \tag{A5}$$

and find on interchanging orders of integration

$$\int_{S=\|\mathbf{S}\|} f(\mathbf{S})d\mathbf{S} = 2S\frac{1}{2\pi i}\int_{\epsilon-i\infty}^{\epsilon+i\infty} e^{zS^2}\left(\int_{-\infty}^{\infty}\cdots\int f(S_1, \ldots, S_\nu) \times e^{-z(S_1^2+\cdots+S_\nu^2)}dS_1\ldots dS_\nu\right), \tag{A6}$$

where $\epsilon > 0$ is chosen so that the line $\mathrm{Re}z = \epsilon$ is to the right of all singularities of the integrand.
In the paper we have to evaluate (A1) with $f(\mathbf{S}) = g(s)e^{\mathbf{K}\cdot\mathbf{S}}$. The inner integrals in Eq. (A6) are then easily performed and we find

$$\int_{S=\|\mathbf{S}\|} f(\mathbf{S})d\mathbf{S} = 2Sg(\mathbf{S})\frac{1}{2\pi i}\int_{\epsilon-i\infty}^{\epsilon+i\infty} e^{zS^2}\left(\prod_{i=1}^{\nu}\sqrt{\frac{\pi}{z}}\, e^{K_i^2/4z}\right)dz \tag{A7}$$

$$= 2\pi^{\nu/2}S^{\nu-1}g(S)\frac{1}{2\pi i}\int_{\epsilon-i\infty}^{\epsilon+i\infty} e^{z+(KS)^2/4z} z^{-[\nu/2-1]-1}dz \tag{A8}$$

$$= 2\pi^{\nu/2}S^{\nu-1}g(S)\frac{I_{\nu/2-1}(KS)}{(\frac{1}{2}KS)^{\nu/2-1}}. \tag{A9}$$

Finally,

$$\int_{-\infty}^{\infty} e^{\mathbf{K}\cdot\mathbf{S}}g(\mathbf{S})d\mathbf{S} = 2\pi^{\nu/2}\int_0^\infty S^{\nu-1}g(S)\frac{I_{\nu/2-1}(KS)}{(\frac{1}{2}KS)^{\nu/2-1}}dS. \tag{A10}$$

APPENDIX B: THE SPHERICAL MODEL WITH MEAN FIELD

The Curie–Weiss spherical model has the Hamiltonian

$$H_N = -\frac{2J}{N}\sum_{1\le i<j\le N} S_iS_j - \mu_0 B\sum_{i=1}^N S_i, \tag{B1}$$

where the S_i are constrained by $\sum_{i=1}^N S_i^2 = N$, so that Eq. (B1) can be written

$$H_N = J - \frac{J}{N}\left(\sum_{i=1}^N S_i\right)^2 - \mu_0 B\sum_{i=1}^N S_i. \tag{B2}$$

The partition function is given by

$$Q_N(K_1, K_2) = Z_N^{-1}e^{-K_1}\int_{-\infty}^{\infty}\cdots\int_{\sum_{i=1}^N S_i^2=N} \exp\left[\frac{K_1}{N}\left(\sum_{i=1}^N S_i\right)^2 + K_2\sum_{i=1}^N S_i\right] \times dS_1 \ldots dS_N, \tag{B3}$$

where

$$Z_N = 2\pi^{N/2}N^{(N-1)/2}/\Gamma(N/2) \tag{B4}$$

is the surface area of the ν-ball. This partition function has been evaluated by Baker[14] and by Thompson and Lieb[15] but we re-evaluate it here by explicitly displaying the random walk underlying(B3). Since the integrand is a function of $S = \sum_{i=1}^N S_i$ only,

$$Q_N(K_1, K_2) = e^{-K_1}\int_{-\infty}^{\infty} P_N(S)e^{(K_1/N)S^2+K_2S}dS, \tag{B5}$$

where $P_N(S)dS$ is the probability that $\sum_{i=1}^N S_i \in (S, S + dS)$ given $\sum_{i=1}^N S_i^2 = N$. Clearly

$$P_N(S) = \int\cdots\int_{\sum_{i=1}^N S_i=S,\ \sum_{i=1}^N S_i^2=N} dS_1 \ldots dS_N. \tag{B6}$$

If we make an orthogonal change of variables to $Y_1, Y_2, \ldots, Y_N$ with

$$Y_1 = N^{-1/2}\sum_{i=1}^N S_i,$$

we find very simply that

$$P_N(S) = \frac{\Gamma(N/2)}{N\sqrt{\pi}\,\Gamma((N-1)/2)}\left(1 - \left[\frac{S}{N}\right]^2\right)^{(N-3)/2}, \quad \text{for } |S| < N,$$

$$= 0, \quad \text{for } |S| > N, \tag{B7}$$

provided $N \geq 3$. We can easily show that

$$P_N(S) \sim (1/\sqrt{2\pi N})e^{-S^2/2N}, \quad \text{as } N\to\infty, \tag{B8}$$

provided $|S| < N$. We find

$$Q_N(K_1, K_2) \sim e^{-K_1}\frac{\Gamma(N/2)}{\sqrt{\pi}\,\Gamma((N-1)/2)}\int_{-1}^{1} e^{NG(x)}dx, \tag{B9}$$

$$G(x) = \tfrac{1}{2}\log(1 - x^2) + K_1x^2 + K_2x.$$

(B9) is easily evaluated by the method of steepest descents:

$$Q_N(K_1, K_2) \sim e^{-K_1}[\Gamma(N/2)/\sqrt{\pi}\,\Gamma((N-1)/2)]e^{NG(x_s)}, \tag{B10}$$

where x_s satisfies the cubic equation

$$2K_1x_s^3 + K_2x_s^2 - (2K_1 - 1)x_s - K_2 = 0. \quad \text{(B11)}$$

For $B = 0$, $K_2 = 0$ and we obtain the simpler equation

$$2Kx_s^3 - (2K - 1)x_s = 0 \quad \text{(B12)}$$

so that

$$x_s = 0, \quad \text{or } x_s = \sqrt{1 - K_c/K} = \sqrt{1 - T/T_c},$$

where $K_c = \frac{1}{2}$, $T_c = 2J/k$. Since $x_s = \langle(\sum_{i=1}^{N} S_i)/N\rangle$ the magnetization per spin is given by

$$M = \mu_0 x_s. \quad \text{(B13)}$$

For $B = 0$, the thermodynamic functions are easily evaluated and we find

$$\begin{aligned} -\beta\psi &= -\tfrac{1}{2}\log(K/K_c) + (K - K_c), && K > K_c, \\ &= 0, && K < K_c, \\ U &= -J(1 - K/K_c), && K > K_c, \\ &= 0, && K < K_c, \\ C_v &= k/2, && K > K_c, \\ &= 0, && K < K_c. \end{aligned} \quad \text{(B14)}$$

For $T > T_c$ and $B \cong 0$, $x_s \cong 0$ and from Eq. (B11) we have

$$-(2K_1 - 1)x_s - K_2 \cong 0$$

so that

$$x_s \cong \mu_0 B/k(T - T_c)$$

and the susceptibility χ is given by

$$\chi = N\mu_0^2/k(T - T_c), \quad \text{(B15)}$$

where T_c is the critical temperature in zero field. Thus the Curie–Weiss spherical model also obeys the Curie–Weiss law.

APPENDIX C: THE LIMIT $\nu \to \infty$ (B = 0)

By Eq. (35)

$$-\beta\psi/\nu = (1/\nu)\log\Gamma(\nu/2) + G(x_s), \quad \text{(C1)}$$

where

$$G(x_s) = -Kx_s^2 + (1/\nu)\log[I_{\nu/2-1}(2\nu Kx_s)/(\nu Kx_s)^{\nu/2-1}]. \quad \text{(C2)}$$

Then using Stirling's approximation and the asymptotic expansion of the Bessel function of large order and large argument,[16] we find

$$-\beta\psi/\nu \sim -Kx_s^2 - \tfrac{1}{2} + \tfrac{1}{2}\log 2 + \tfrac{1}{2}(\sqrt{1 + 16K^2x_s^2} - \log(1 + \sqrt{1 + 16K^2x_s^2})), \quad \text{as } \nu \to \infty. \quad \text{(C3)}$$

x_s is given by $\partial G/\partial x_s = (\partial/\partial x_s)(-\beta\psi/\nu) = 0$, and we find from Eq. (C3)

$$x_s \sim 4Kx_s/1 + \sqrt{1 + 16K^2x_s^2}, \quad \text{as } \nu \to \infty, \quad \text{(C4)}$$

which reduces to $x_s = 0$, or $x_s = \sqrt{1 - K_c/K}$, the same as Eq. (B12) in the Appendix B. Finally, substituting $\sqrt{1 - K_c/K}$ for x_s in Eq. (C8), we find

$$\begin{aligned} -\beta\psi/\nu &\sim -\tfrac{1}{2}\log(K/K_c) + (K - K_c), && K > K_c, \\ &\sim 0, && K < K_c, \end{aligned} \quad \text{(C5)}$$

which is $-\beta\psi$ for the Curie–Weiss spherical model [Eq. (B14)]. This completes the proof that, for $B = 0$, we recover the spherical model in the limit $\nu \to \infty$.

APPENDIX D: EXPANSIONS ABOUT $T = 0$ AND $T = T_c$

The transcendental equation satisfied by x_s for $\mathbf{B} = 0$ is

$$x_s = I_{\nu/2}(2\nu Kx_s)/I_{\nu/2-1}(2\nu Kx_s). \quad \text{(D1)}$$

For $T \lessapprox T_c$, $x_s \cong 0$ and we use the Bessel function expansion for small argument

$$I_\alpha(z) = \sum_{n=0}^{\infty} \frac{(\frac{1}{2}z)^{2n+\alpha}}{n!(n + \alpha)!}, \quad \text{(D2)}$$

so that

$$\begin{aligned} \frac{I_\alpha(z)}{I_{\alpha-1}(z)} &= \frac{[(\frac{1}{2}z)^\alpha/\alpha!][1 + (\frac{1}{2}z)^2/(1 + \alpha) + \cdots]}{[(\frac{1}{2}z)^{\alpha-1}/(\alpha - 1)!][1 + (\frac{1}{2}z)^2/\alpha + \cdots]} \\ &\cong (z/2\alpha)[1 + (\tfrac{1}{2}z)^2/(1 + \alpha) + \cdots] \\ &\quad \times [1 - (\tfrac{1}{2}z)^2/\alpha + \cdots] \\ &= (z/2\alpha)[1 - (\tfrac{1}{2}z)^2/\alpha(1 + \alpha) + \cdots]; \end{aligned} \quad \text{(D3)}$$

and so around T_c Eq. (D1) reduces to

$$x_s = \frac{2\nu Kx_s}{2 \cdot \nu/2}\left(1 - \frac{(\frac{1}{2} \cdot 2\nu Kx_s)^2}{(\nu/2)(1 + \nu/2)} + \cdots\right),$$

and we find

$$x_s \cong \sqrt{[(2 + \nu)/\nu](1 - T/T_c)}. \quad \text{(D4)}$$

For $T \cong 0$, $K \gg 1$ and we use the Bessel function expansion for large argument,

$$I_{\nu/2}(z) \sim \frac{e^z}{\sqrt{2\pi z}}\left[\sum_{n=0}^{\infty} L_n^{(\nu+2)} \cdot (-z)^{-n} + i^{-(\nu+1)}e^{-2z}\sum_{n=0}^{\infty} L_n^{(\nu+2)} \cdot z^{-n}\right], \quad \text{(D5)}$$

where the coefficients $L_n^{(\nu)}$ are defined by means of the recursion relation:

$$L_n^{(\nu)} = \frac{(\nu - 2)^2 - (2n - 1)^2}{8n} L_{n-1}^{(\nu)}, \quad L_0^{(\nu)} = 1.$$

The second series in Eq. (D5) is necessary for the $\nu = 1$ and $\nu = 3$ cases as then $L_n^{(\nu)} = 0$ for $n > 0$.

For $\nu = 1$ we find

$$\begin{aligned} I_{1/2}(z)/I_{-1/2}(z) &\sim (1 - e^{-2z})/(1 + e^{-2z}) \\ &\cong 1 - 2e^{-2z}, \end{aligned} \quad \text{(D6)}$$

and so Eq. (D1) around $T = 0$ reduces to

$$x_s \cong 1 - 2e^{-4Kx_s} \cong 1 - 2e^{-4K} \tag{D7}$$

since $x_s \cong 1$ for $T \cong 0$.

For $\nu > 1$ we find

$$\frac{I_{\nu/2}(z)}{I_{\nu/2-1}(z)} \sim \frac{1 - L_1^{(\nu+2)}/z + \text{-----}}{1 - L_1^{(\nu)}/z + \text{-----}} \cong 1 - \frac{L_1^{(\nu+2)} - L_1^{(\nu)}}{z} = 1 - \frac{\nu - 1}{2} \cdot \frac{1}{z} \tag{D8}$$

(this result also holds for $\nu = 3$). Around $T = 0$, Eq. (D1) reduces to

$$x_s \cong 1 - [(\nu - 1)/\nu](1/4K). \tag{D9}$$

1 H. E. Stanley, Phys. Rev. **176,** 718 (1968); **179,** 570 (1969); Phys. Rev. Letters **20,** 589 (1968); J. Appl. Phys. **40,** 1272 (1969); J. Phys. Soc. Japan Suppl. **26,** 102 (1969).

2 K. Pearson, Nature **72,** 294, 342 (1905). Pearson formulated the problem of random flights in two dimensions as follows: "A man starts from a point 0 and walks l yards in a straight line; he then turns through any angle whatever and walks another l yards in a second straight line. He repeats this process N times. I require the probability that after these N stretches he is at a distance between R and $R + dR$ from his starting point 0".

3 J. C. Kluyver, Koninkl. Ned. Akad. Wetenschap., Proc. **8,** 341 (1906).

4 Lord Rayleigh, Phil. Mag. **37,** 321 (1919).

5 G. N. Watson, *Theory of Bessel Functions* (Cambridge U.P., London, 1944), p. 419.

6 A simpler integral of this form is evaluated by Lieb and Thompson by finding upper and lower bounds: E. H. Lieb and C. J. Thompson, J. Math. Phys. **10,** 1403 (1969).

7 Reference 5, pp. 48, 79. The choice of these particular representations is determined by the requirement that the integrand in Eq. (15) be a sharply peaked function of x, u, ϕ_1, and ϕ_2.

8 One of the other solutions gives the same Eq. (20) while the other two correspond to Eq. (20) with $K_2 \to -K_2$, that is, a reversed magnetic field.

9 For the Hamiltonian (4), the ground state energy is given by $H_N^{(\nu)} = -(N-1)J\nu - \mu_0 BN\sqrt{\nu}$. Clearly the magnetic field must be scaled ($B = \sqrt{\nu}\mathcal{B}$) if it is to appear in the result when we take the limit $\nu \to \infty$. We could have chosen the spins to be of unit length ($\|S_i\| = 1$). The equation for the saddle point x_s would then have been

$$x_s = I_{\nu/2}(2K_1x_s + K_2)/I_{\nu/2-1}(2K_1X_s + K_2).$$

In this case, however, for $\mathbf{B} = 0$, the critical temperature is no longer independent of ν: $T_c = 2J/k\nu$. To take the limit $\nu \to \infty$ we now have to scale J by $J = \nu\mathcal{J}$ and the same asymptotic results are obtained.

10 P. Weiss, J. Phys. (4th series) **6,** 661 (1907).

11 M. N. Barber and B. W. Ninham, *Random and Restricted Walks* (Gordon and Breach, New York, 1970), Chaps. 6 and 13.

12 C. Domb, J. Phys. C: Solid State Phys. **3,** 256 (1970); Advan. Chem. Phys. **15,** 229 (1969).

13 P. W. Kasteleyn, Physica **29,** 1329 (1963).

14 G. Baker, Phys. Rev. **126,** 2071 (1962).

15 E. H. Lieb and C. J. Thompson, J. Math. Phys. **10,** 1403 (1969).

16 Ref. 5, Chap. 8.

Exact and Simultaneous Measurements

Platon C. Deliyannis

Department of Mathematics, Illinois Institute of Technology, Chicago, Illinois 60616

(Received 1 November 1971)

The observables which can be determined through their states of exact measurement are characterized, and a criterion for the set of states in which two observables are both measured exactly is established so that the two observables will be compatible.

INTRODUCTION

An interesting and, one might perhaps venture to say, important problem concerning observables is the following: How many measurements are required to determine a given observable u, and of what kind? The term "measurement" is taken here in a rather weak sense; we simply mean a process of obtaining the expectation value of u in some state m. In a strong sense a measurement could be considered as a process of determining the probability distribution $p_{m,u}$ of u in the state m. In this latter sense it is well known that if we measure u in every state, then u is completely known. The problem of determining u through "weak measurements" has not been solved yet; Gudder[1] has obtained partial results of importance, which we discuss in Sec. 2. There are, however, certain states for which "weak" and "strong" measurements cannot be distinguished—those in which u is measured exactly—because, then, the probability distribution $p_{m,u}$ is just the Dirac measure at the expectation value of the observable. Naturally one is immediately confronted with the difficulty that no such state may exist for a particular u, so that certain restrictions on u are in order. These, and how to decide on the validity of the required conditions, are the content of the first part of the present paper (Theorems 1 and 2).

In the second part the question of simultaneous measurements is studied in the same spirit. Physically, two observables u, v are simultaneously measurable if somehow their values can be obtained through a single measurement in so many states that no other measurements are required. This seems to suggest that the two observables u, v are so closely related to a third one, say w, that it suffices to measure w in order to obtain the values of u, v. In the current abstract mathematical model we shall be studying, the formulation is very different, but equivalent logically to the above result, i.e., that u, v are simultaneously measurable iff they are functions of some w. This is done, however, without reference to measurements. It is this connection to measurements which we study in the second part (Theorems 3 and 4).

1. PRELIMINARIES

The set $\mathcal{L}$ of all events is assumed to have the following structure:

(i) A partial order $\leqq$; there exist $0, I \in \mathcal{L}$ with $0 \leqq A \leqq I$ for all $A \in \mathcal{L}$.

(ii) A complementation $'$; we have $(A')' = A$ for all

and so Eq. (D1) around $T = 0$ reduces to

$$x_s \cong 1 - 2e^{-4Kx_s} \cong 1 - 2e^{-4K} \tag{D7}$$

since $x_s \cong 1$ for $T \cong 0$.

For $\nu > 1$ we find

$$\frac{I_{\nu/2}(z)}{I_{\nu/2-1}(z)} \sim \frac{1 - L_1^{(\nu+2)}/z + \text{-----}}{1 - L_1^{(\nu)}/z + \text{-----}} \cong 1 - \frac{L_1^{(\nu+2)} - L_1^{(\nu)}}{z} = 1 - \frac{\nu - 1}{2} \cdot \frac{1}{z} \tag{D8}$$

(this result also holds for $\nu = 3$). Around $T = 0$, Eq. (D1) reduces to

$$x_s \cong 1 - [(\nu - 1)/\nu](1/4K). \tag{D9}$$

1 H. E. Stanley, Phys. Rev. **176,** 718 (1968); **179,** 570 (1969); Phys. Rev. Letters **20,** 589 (1968); J. Appl. Phys. **40,** 1272 (1969); J. Phys. Soc. Japan Suppl. **26,** 102 (1969).

2 K. Pearson, Nature **72,** 294, 342 (1905). Pearson formulated the problem of random flights in two dimensions as follows: "A man starts from a point 0 and walks l yards in a straight line; he then turns through any angle whatever and walks another l yards in a second straight line. He repeats this process N times. I require the probability that after these N stretches he is at a distance between R and $R + dR$ from his starting point 0".

3 J. C. Kluyver, Koninkl. Ned. Akad. Wetenschap., Proc. **8,** 341 (1906).

4 Lord Rayleigh, Phil. Mag. **37,** 321 (1919).

5 G. N. Watson, *Theory of Bessel Functions* (Cambridge U.P., London, 1944), p. 419.

6 A simpler integral of this form is evaluated by Lieb and Thompson by finding upper and lower bounds: E. H. Lieb and C. J. Thompson, J. Math. Phys. **10,** 1403 (1969).

7 Reference 5, pp. 48, 79. The choice of these particular representations is determined by the requirement that the integrand in Eq. (15) be a sharply peaked function of x, u, ϕ_1, and ϕ_2.

8 One of the other solutions gives the same Eq. (20) while the other two correspond to Eq. (20) with $K_2 \to -K_2$, that is, a reversed magnetic field.

9 For the Hamiltonian (4), the ground state energy is given by $H_N^{(\nu)} = -(N-1)J\nu - \mu_0 BN\sqrt{\nu}$. Clearly the magnetic field must be scaled ($B = \sqrt{\nu}\mathcal{B}$) if it is to appear in the result when we take the limit $\nu \to \infty$. We could have chosen the spins to be of unit length ($\|S_i\| = 1$). The equation for the saddle point x_s would then have been

$$x_s = I_{\nu/2}(2K_1x_s + K_2)/I_{\nu/2-1}(2K_1X_s + K_2).$$

In this case, however, for $\mathbf{B} = 0$, the critical temperature is no longer independent of ν: $T_c = 2J/k\nu$. To take the limit $\nu \to \infty$ we now have to scale J by $J = \nu\mathcal{J}$ and the same asymptotic results are obtained.

10 P. Weiss, J. Phys. (4th series) **6,** 661 (1907).

11 M. N. Barber and B. W. Ninham, *Random and Restricted Walks* (Gordon and Breach, New York, 1970), Chaps. 6 and 13.

12 C. Domb, J. Phys. C: Solid State Phys. **3,** 256 (1970); Advan. Chem. Phys. **15,** 229 (1969).

13 P. W. Kasteleyn, Physica **29,** 1329 (1963).

14 G. Baker, Phys. Rev. **126,** 2071 (1962).

15 E. H. Lieb and C. J. Thompson, J. Math. Phys. **10,** 1403 (1969).

16 Ref. 5, Chap. 8.

Exact and Simultaneous Measurements

Platon C. Deliyannis

Department of Mathematics, Illinois Institute of Technology, Chicago, Illinois 60616

(Received 1 November 1971)

The observables which can be determined through their states of exact measurement are characterized, and a criterion for the set of states in which two observables are both measured exactly is established so that the two observables will be compatible.

INTRODUCTION

An interesting and, one might perhaps venture to say, important problem concerning observables is the following: How many measurements are required to determine a given observable u, and of what kind? The term "measurement" is taken here in a rather weak sense; we simply mean a process of obtaining the expectation value of u in some state m. In a strong sense a measurement could be considered as a process of determining the probability distribution $p_{m,u}$ of u in the state m. In this latter sense it is well known that if we measure u in every state, then u is completely known. The problem of determining u through "weak measurements" has not been solved yet; Gudder[1] has obtained partial results of importance, which we discuss in Sec. 2. There are, however, certain states for which "weak" and "strong" measurements cannot be distinguished—those in which u is measured exactly—because, then, the probability distribution $p_{m,u}$ is just the Dirac measure at the expectation value of the observable. Naturally one is immediately confronted with the difficulty that no such state may exist for a particular u, so that certain restrictions on u are in order. These, and how to decide on the validity of the required conditions, are the content of the first part of the present paper (Theorems 1 and 2).

In the second part the question of simultaneous measurements is studied in the same spirit. Physically, two observables u, v are simultaneously measurable if somehow their values can be obtained through a single measurement in so many states that no other measurements are required. This seems to suggest that the two observables u, v are so closely related to a third one, say w, that it suffices to measure w in order to obtain the values of u, v. In the current abstract mathematical model we shall be studying, the formulation is very different, but equivalent logically to the above result, i.e., that u, v are simultaneously measurable iff they are functions of some w. This is done, however, without reference to measurements. It is this connection to measurements which we study in the second part (Theorems 3 and 4).

1. PRELIMINARIES

The set $\mathcal{L}$ of all events is assumed to have the following structure:

(i) A partial order $\leqq$; there exist $0, I \in \mathcal{L}$ with $0 \leqq A \leqq I$ for all $A \in \mathcal{L}$.

(ii) A complementation $'$; we have $(A')' = A$ for all

$A, A \leqq B$ iff $B' \leqq A', 0' = I$ and $I' = 0$. With $\wedge, \vee$ denoting infimum and supremum with respect to $\leqq$, we also have $A \wedge A' = 0$, $A \vee A' = I$.

We shall say that A, B are *disjoint* (write $A \perp B$) iff $A \leqq B'$ or equivalently $B \leqq A'$.

(iii) An operation $\sum$; any sequence of pairwise disjoint events $A_i \in \mathfrak{L}$ has a supremum $\sum A_i$ ($A_1 + A_2 + \cdots + A_n$ for a finite sequence).

(iv) Orthomodularity; if $A \leqq B$, then $B = A + (B \wedge A')$, and we shall write $B - A$ for $B \wedge A'$.

(v) Separability; any set of nonzero disjoint events is at most countable.

Two events A, B are said to *commute* if $A = A_1 + C$, $B = B_1 + C$, where $A_1 \perp B_1$.

A *state* is a map $m: \mathfrak{L} \to$ interval $[0, 1]$ such that for any pairwise disjoint $A_i \in \mathfrak{L}$ we have $m(\sum A_i) = \sum m(A_i)$. Write $\mathfrak{M}$ for the set of all states.

For most of our arguments we shall need the following property of $\mathfrak{L}$ which we therefore assume:

(JPZ) $\mathfrak{L}$ is a lattice, and, if $mA = mB = 1$, then $m(A \wedge B) = 1$ also.

Finally we assume that enough states exist.

(vi) For each $A \neq 0$ there exists a state m such that $m(A) = 1$.

An observable is a σ-homomorphism of the Borel sets on the line R into $\mathfrak{L}$, i.e., a map $E \to u(E)$ such that $u(\phi) = 0$, $u(R) = I$, $u(E) \leqq u(F)$ for $E \subseteq F$ and $u(\cup E_i) = \sum u(E_i)$ provided that the E_i are pairwise disjoint. The range of an observable is always a Boolean σ-algebra with respect to the operations in $\mathfrak{L}$. For any Borel function f the observable $f(u)$ is the map $E \to u(f^{-1}E)$. The *spectrum* $\sigma(u)$ is the intersection of all closed E such that $u(R - E) = 0$; the *point spectrum* $\sigma_p(u)$ is the set of all $\lambda \in R$ such that $u(\{\lambda\}) \neq 0$. The observable u is *bounded* iff $\sigma(u)$ is compact. The event A will be identified to the observable

$$E \to \begin{cases} A & \text{if } 1 \in E,\ 0 \notin E \\ A' & \text{if } 0 \in E,\ 1 \notin E \\ I & \text{if } 0 \in E,\ 1 \in E \\ 0 & \text{if } 0 \notin E,\ 1 \notin E \end{cases}.$$

An observable is *discrete* if its range is an atomic Boolean algebra; note that the atoms of the range need not be atoms of $\mathfrak{L}$.

Two observables u, v are *compatible* iff for any Borel sets E, F the events $u(E), v(F)$ commute.

For any $m \in \mathfrak{M}$, the measure $p_{m,u}: E \to m(u(E))$ is the *probability distribution* of u in the state m; it is always supported by $\sigma(u)$. The integral $m(u) = \int_R \lambda dp_{m,u}(\lambda)$ (whenever it exists) is the *expectation* of u in the state m. It is known that $m(u)$ exists for all states m iff u is bounded.[2] The *dispersion* of u in the state m is of course $m(u^2) - [m(u)]^2$, and u is *measured exactly* in m iff its dispersion is zero, which happens iff $p_{m,u}$ is supported by a single point.

For the purposes of the present paper the term *subsystem* will mean a subset of $\mathfrak{L}$ closed under the operations $', \wedge, \vee$ and $\sum$.

We shall make use of certain well-known facts which we summarize in the following proposition:

Proposition: The points λ of $\sigma_p(u)$ and the atoms A of the range of u are in a one-to-one correspondence given by $A = u(\{\lambda\})$. For $m \in \mathfrak{M}$, $\lambda = m(u)$, and $A = u(\{\lambda\})$ we have that u is measured exactly in m iff A is an atom of the range of u and $m(A) = 1$.

2. EXACT MEASUREMENTS

It is convenient to have the terminology introduced by the following definition:

Definition: The *maximal observation* of an observable u is the pair consisting of the set $\mathfrak{M}_u$ of all states in which u is measured exactly and the map $m \to m(u)$ from $\mathfrak{M}_u$ to the reals.

Theorem 1: An observable is completely determined by its maximal observation iff it is discrete. More precisely: if u is discrete, then the range $\mathfrak{B}$ of u consists of all events which are measured exactly in each state in $\mathfrak{M}_u$ and $u(E) = \sup\{B \in \mathfrak{B} \mid m(B) = 1$ for some $m \in \mathfrak{M}_u$ for which $m(u) \in E\}$. On the other hand, if u is not discrete, there exists a $v \neq u$ such that u and v have the same maximal observation.

Proof: Let $\mathfrak{B}$ be the range of u and write $I = B_a + B_c$, where B_a is the union of all atoms in $\mathfrak{B}$ [which are countably many by (v)] and $B_c = B_a'$; note that $B_a = u(\sigma_p(u))$. To each atom B_i of $\mathfrak{B}$ there corresponds a unique $\lambda_i \in \sigma_p(u)$ such that $B_i = u(\{\lambda_i\})$ and we have that $m \in \mathfrak{M}_u$ iff $m(B_i) = 1$ for some i; in such a case $m(u) = \lambda_i$.

Now let $v = \sum \lambda_i B_i$, i.e., let $v(E) = \sum\{B_i \mid \lambda_i \in E\}$ for each Borel set E; it is clear that $v = f(u)$, where $f(\lambda) = \lambda\chi_S(\lambda)$, S being the point spectrum $\sigma_p(u)$. Let a be any nonzero number and consider the observable $w = v + a(u - v) = g(u)$, where $g(\lambda) = \lambda[\chi_S(\lambda) + a\chi_T(\lambda)]$, $T = R - S$. Setting $(1/a)E = \{x \mid ax \in E\}$, we see that

$$g^{-1}E = \begin{cases} [\sigma_p(u) \cap E] \cup [(1/a)E - \sigma_p(u)] & \text{if } 0 \notin E \\ [\sigma_p(u) \cap E] \cup [(1/a)E - \sigma_p(u)] \cup [R - \sigma(u)] & \text{if } 0 \in E \end{cases}$$

so that $w(E) = u(g^{-1}(E)) = [B_a \cap E] \cup [B_c \cap u((1/a)E)]$. To find the atoms in the range of w, we consider the case $E = \{\lambda_0\}$; then $(1/a)E = \{(1/a)\lambda_0\}$, so $u((1/a)E) = 0$ unless $(1/a)\lambda_0 \in \sigma_p(u)$; but in this case $u((1/a)E) \leqq B_a$ and thus $w(\{\lambda_0\}) = B_a \cap u(\{\lambda_0\})$. Therefore, $w(\{\lambda_0\}) \neq 0$ iff $u(\{\lambda_0\}) \neq 0$, and are in all cases equal. This implies that $\mathfrak{M}_w = \mathfrak{M}_u$, and, as for any m in this set we have $m(u) = m(v) = \lambda_i$ for some i, we obtain the same maximal observation for u and w. However, for $B_c \neq 0$ we have $w \neq u$, because otherwise we would have $a(m(u) - m(v)) = 0$ for all m, which would imply $a = 0$ contrary to the hypothesis. So we have established that if u is uniquely determined by its maximal observation, it has to be discrete.

We now assume that $\mathfrak{B}$ is atomic, i.e., $u = \sum \lambda_i B_i$. We shall show that $\mathfrak{B}$ consists of those events which are measured exactly in each of the states of $\mathfrak{M}_u$. To this end we first note that if $0 < B < B_i$ for some i, then there is a state $m \in \mathfrak{M}_u$ in which B is not measured exactly. Let $C = B_i - B \neq 0$ and consider [by (vi)] states m_1, m_2 in which B, C have value 1, respectively. Let $m = \frac{1}{2}(m_1 + m_2)$; then $m(B) = \frac{1}{2}$ since $m_2(B) = 0$, but $m(B_i) = 1$ since $m_1(B_i) = m_2(B_i) = 1$, i.e., $m \in \mathfrak{M}_u$.

Consider an even B measured exactly in each state $m \in \mathfrak{M}_u$. Suppose $m(B) = 1$; if $m(B_i) = 1$ too, then by (JPZ) we have $m(B \wedge B_i) = 1$, while if $m(B_i) = 0$ we have $m(B \wedge B_i) = 0$. If $m(B) = 0$, we also have $m(B \wedge B_i) = 0$. Thus $B \wedge B_i$ is measured exactly in each state of $\mathfrak{M}_u$, and by our previous remark we have that either $B \wedge B_i = 0$ or $B \wedge B_i = B_i$; this last will occur iff for some state $m \in \mathfrak{M}_u$ we have $m(B) = m(B_i) = 1$. Thus we see that $B \geqq B_i$ for all those i for which some state $m \in \mathfrak{M}_u$ exists such that $m(B) = m(B_i) = 1$. Now, if B_j is such that for no state $m \in \mathfrak{M}_u$ we have $m(B_j) = m(B) = 1$, there must exist a state $m \in \mathfrak{M}_u$ with $m(B_j) = 1, m(B) = 0$; then $m(B') = 1$, and we obtain $B_j \leqq B'$. So the atoms of $\mathfrak{B}$ fall in two classes: those $\leqq B$ and those $\leqq B'$, which implies at once that B is a union of atoms of $\mathfrak{B}$ and thus $B \in \mathfrak{B}$.

So we have established that $\mathfrak{M}_u$ determines $\mathfrak{B}$ and hence its atoms B_i; since $\lambda_i = m(u)$ for those $m \in \mathfrak{M}_u$ for which $m(B_i) = 1$, we have $u(E) = \sum\{B_i \mid \lambda_i \in E\} = \sup\{B \in \mathfrak{B} \mid m(B) = 1$ for some $m \in \mathfrak{M}_u$ for which $m(u) \in E\}$.

Theorem 2: The set of states in which u is measured exactly determines whether or not u is discrete.

Proof: Consider the set of events which are measured exactly in every state $m \in \mathfrak{M}_u$. For each $m \in \mathfrak{M}_u$ we consider the set $\mathfrak{L}_m = \{A \in \mathfrak{L} \mid m(A) = 1\}$; clearly $\mathfrak{L}_m \neq 0$ since $A_m = u(\{m(u)\}) \in \mathfrak{L}_m$. Given any $A \in \mathfrak{L}_m$, we repeat the argument in the proof of Theorem 1 to obtain $A \geqq A_m$, i.e., that A_m is the smallest element of $\mathfrak{L}_m$ and hence determined by $\mathfrak{M}_u$. On the other hand, every atom of the range of u has the form A_m for suitable m; thus u will be discrete iff the union of all distinct A_m is I.

Corollary: If $m(u) = m(v)$ and $m(u^2) = m(v^2)$ whenever one side of each equation is defined, and, if u or v is discrete, then $u = v$.

Remarks: As we have already mentioned, Gudder has shown that in a quite full logic, if we have $m(u) = m(v)$ for all $m \in \mathfrak{M}$, then $u = v$ provided that u or v is bounded and has countable spectrum. We first note that this last requirement on u is more than what his argument requires: in fact the theorem holds for discrete observables, for which the spectrum can include a whole interval since it is the closure of the point spectrum [Lemma (3, 1) of Ref. 1]. The main theorem which implies the above is Theorem 4.5 in Ref. 1, which does not require a countable spectrum: If u, v are bounded, $m(u) = m(v)$ for all $m \in \mathfrak{M}$, and $\lambda = \sup\sigma(u)$, $\mu = \sup\sigma(v)$, then $\lambda = \mu$ and $u(\{\lambda\}) = v(\{\mu\})$. From this Gudder easily obtains $u(\{\lambda\}) = v(\{\lambda\})$ for all $\lambda \in R$. But this implies that u, v have the same point spectrum S and that the atoms in their respective ranges are the same events. Thus u will be discrete iff v is, and in such a case we have $u(E) = \sum\{u(\{\lambda\}) \mid \lambda \in S \cap E\} = \sum\{v(\{\lambda\}) \mid \lambda \in S \cap E\} = v(E)$ for all E, i.e., $u = v$.

So we have a stronger form of the corollary valid: For u bounded and discrete the map $m \to m(u)$ determines u completely without knowledge of $m(u^2)$. There seems, however, no way of deciding which are the states in which we can measure u exactly, although they are determined. So there seems no way of deducing Gudder's theorem from Theorems 1, 2. Vice versa, too; knowledge of a maximal observation does not appear to allow us to determine the map $m \to m(u)$ on $\mathfrak{M}$, and thereby let us deduce theorem 1 from the result of Gudder.

Finally let us note that the corollary is in one respect more general than Gudder's theorem in that it does not restrict the observables to be bounded; this appears to be an essential restriction for the proof of Theorem 4.5 in Ref. 1, so that no relaxation of $m(u^2) = m(v^2)$ seems apparent.

3. EXACT SIMULTANEOUS MEASUREMENTS

The proof of the next theorem is essentially contained in Theorem 1, but we wish to state it separately because of its physical content.

Theorem 3: Let u be a discrete observable; then v is a function of u iff v is measured exactly in each state in which u is.

Proof: If $v = f(u)$, then $m(v) = \int_R f(\lambda) dp_{m,u}(\lambda)$ so that the desired conclusion is immediate. So assume now that v is measured exactly in each $m \in \mathfrak{M}_u$, so that the same holds for each $v(E)$ in its range. Then by the argument in the proof of Theorem 1 we have $v(E)$ in the range of u. But in general z is a function of w iff the range of z is contained in the range of w (see, e.g., Ref. 3); thus v is a function of u.

We shall now establish a criterion for compatibility of two observables based on the supply of states in which both are measured exactly. It is perhaps interesting to note the similarity this condition bears to the corresponding condition in the totally different axiomatic scheme introduced by Segal.[4]

Definition: Let u, v be observables with ranges $\mathfrak{B}$ and $\mathfrak{C}$, respectively; the *subsystem generated* by u, v is the smallest subsystem $\mathfrak{L}(u, v)$ containing $\mathfrak{B} \cup \mathfrak{C}$.

Theorem 4: Let u, v be observables and $\mathfrak{M}_0 = \mathfrak{M}_u \cap \mathfrak{M}_v$. If u, v are discrete and compatible, then $\mathfrak{M}_0$ separate the elements of $\mathfrak{L}(u, v)$. in the sense that if $A, B \in \mathfrak{L}(u, v)$ and $A \neq B$, then $m(A) \neq m(B)$ for some $m \in \mathfrak{M}_0$. Conversely, if $\mathfrak{M}_0$ separates $\mathfrak{L}(u, v)$, then u and v are compatible.

Proof: Assume u, v discrete and compatible; then $\mathfrak{L}(u, v)$ is the Boolean σ-algebra having as atoms all nonzero $B_i \wedge C_j$, where $\{B_i\}$ and $\{C_j\}$ are the atoms of $\mathfrak{B}$ and $\mathfrak{C}$, respectively. To see this, consider the set $\mathfrak{D}$ of all nonzero $D_{ij} = B_i \wedge C_j$, which are necessarily in $\mathfrak{L}(u, v)$ and note that they are pairwise disjoint; further we have $\sum_{i,j} D_{ij} = I$ since $\sum B_i = \sum C_j = I$ and the distributive laws hold by compatibility of u, v (see Ref. 3). For the same reason the set of all possible suprema of subsets of $\mathfrak{D}$ forms a Boolean σ-algebra contained in $\mathfrak{L}(u, v)$, and hence has to coincide with $\mathfrak{L}(u, v)$; clearly then the D_{ij} are the atoms of $\mathfrak{L}(u, v)$. By (JPZ) we see that the states in which all elements of $\mathfrak{L}(u, v)$ are measured exactly form $\mathfrak{M}_0$; on the other hand, property (vi) guarantees that these states separate the elements of $\mathfrak{L}(u, v)$.

Conversely now, suppose that $\mathfrak{M}_0$ separates the elements of $\mathfrak{L}(u, v)$, and consider $\mathfrak{L}_0 = \{A \in \mathfrak{L} \mid m(A) = 0$

or 1 for each $m \in \mathfrak{M}_0\} \supseteq \mathcal{B} \cup \mathcal{C}$. Trivially $\mathcal{L}_0$ is closed under complements, disjoint unions and by (JPZ) it is also closed under the lattice operations; thus $\mathcal{L}_0 \supseteq \mathcal{L}(u,v)$ so that all elements of $\mathcal{L}(u,v)$ are measured exactly in each state of $\mathfrak{M}_0$. The rest follows from the next theorem which is of interest in itself.

Theorem 5: A subsystem $\mathcal{L}_1$ of $\mathcal{L}$ such that, for some set $\mathfrak{M}_0$ of states, the following two conditions hold, is a Boolean σ-algebra:

(i) Every $A \in \mathcal{L}_1$ is measured exactly in each state of $\mathfrak{M}_0$;

(ii) the states in $\mathfrak{M}_0$ separate the elements of $\mathcal{L}_1$.

Proof: Let $h(A) = \{m \in \mathfrak{M}_0 \mid m(A) = 1\}$ for $A \in \mathcal{L}_1$. Clearly $A \leqq B$ implies $h(A) \subseteq h(B)$ and $h(A') = \mathfrak{M}_0 - h(A)$ (set difference). Also $h(A \wedge B) = \{m \in \mathfrak{M}_0 | m(A \wedge B) = 1\} \subseteq \{m \in \mathfrak{M}_0 | m(A) = 1$ and $m(B) = 1\} = h(A) \cap h(B)$; by (JPZ) the reverse also holds: $m \in h(A) \cap h(B)$ implies $m(A) = m(B) = 1$; hence $m(A \wedge B) = 1$, i.e., $m \in h(A \wedge B)$. So we have $h(A \wedge B) = h(A) \cap h(B)$. Now let $h(A) \subseteq h(B)$, or $h(A) = h(A) \cap h(B)$, i.e., $h(A) = h(A \wedge B)$; then $m(A) = m(A \wedge B)$ for all $m \in \mathfrak{M}_0$ since the values of any m on any C in $\mathcal{L}_1$ are 0 or 1, and, as $\mathfrak{M}_0$ separates the elements of $\mathcal{L}_1$, we have $A = A \wedge B$, i.e., $A \leqq B$.

We therefore have that h is one-to-one. Since h also preserves pairwise disjoint unions and for $A \leqq B$ we have $h(A - B) = h(A) - h(B)$ (set difference), we see that the range of h is a Boolean σ-algebra of subsets of $\mathfrak{M}_0$. But h is one-to-one and preserves all relevant operations; hence $\mathcal{L}_1$ is also a Boolean σ-algebra.

[1] S. P. Gudder, Pacific J. Math. **19**, 81 (1966).
[2] S. P. Gudder, Trans. Amer. Math. Soc. **119**, 428 (1965).
[3] A. Ramsey, J. Math. Mech. 15, 227 (1966).
[4] I. E. Segal, A Mathematical Approach to Elementary Particles and Their Fields (University of Chicago Lecture Notes, 1955).

Spectrum of Local Hamiltonians in the Yukawa$_2$ Field Theory

Jonathan Dimock* †

Department of Physics, Harvard University, Cambridge, Massachusetts 02138

(Received 15 July 1971)

It is proved that the local Hamiltonians constructed for the two-dimensional Yukawa model by Glimm and Jaffe have the same continuous spectrum as the free Hamiltonian. The proof depends on the construction of asymptotic creation and annihilation operators.

1. INTRODUCTION

The spatially cutoff Hamiltonian in the two-dimensional Yukawa model is given formally by the expression

$$H(g) = H_0 + \lambda \int :\bar{\psi}(x)\psi(x):\varphi(x)g(x)dx - \tfrac{1}{2}\delta m^2 \int :\varphi(x)^2:g(x)^2dx - E(g), \quad (1)$$

where $g \in C_0^\infty$ is assumed equal to one on a large set and δm^2 and $E(g)$ are the infinite renormalization constants. Glimm and Jaffe have proved that $H(g)$ can be defined as a positive self-adjoint operator,[1] and that it gives rise to locally correct dynamics.[2] They have also proved that $H(g)$ has a vacuum state $\Omega(g)$ (known to be unique for small values of λ), and that in the spectral interval $[0, \min\{m_b, m_f\})$ the spectrum is pure discrete with finite multiplicity.[1] Here m_b and m_f are the bare boson mass and bare fermion mass, respectively. In this paper, we use estimates developed elsewhere[3] to show that $H(g)$ has the same continuous spectrum as H_0, i.e., that $[\min\{m_b, m_f\}, \infty) \subset$ spectrum $H(g)$.

Such a spectrum has the following interpretation proposed by Glimm and Jaffe in connection with the $P(\varphi)_2$ model.[4] The states associated with the continuum are those which at large times (in the Schrödinger picture) represent free particles not in the region of interaction. The discrete spectrum corresponds to states bound into the region of interaction (supp$g$), with reduced mass because of the interaction. One expects that in the infinite volume limit $g \to 1$, the density of bound states becomes infinite giving rise to a new continuum beginning at one of the physical masses.

We remark that the mathematical problem of constructing the theory in the infinite volume limit is quite difficult and involves a change of Hilbert spaces. The present results should be regarded as one step toward the eventual study of the spectrum of a physical Hamiltonian with no cutoffs.

The method we use to obtain the continuum involves the construction of asymptotic creation and annihilation operators. We prove the existence of strong limits of the form

$$\beta_\epsilon^\#(f) = \lim_{t \to \infty} e^{-iH(g)t} e^{iH_0 t} b_\epsilon^\#(f) e^{-iH_0 t} e^{iH(g)t}, \quad (2)$$

where $b_\epsilon^\#(f) = \int b_\epsilon^\#(k) f(k) dk$, $f \in L^2$, is one of the standard creation or annihilation operators for bosons ($\epsilon = 0$), fermions ($\epsilon = +1$), or antifermions ($\epsilon = -1$). Then by applying the asymptotic creation operators $\beta_\epsilon^*(f)$ to the vacuum, we build asymptotic states on which $H(g)$ acts like a free Hamiltonian. The first detailed application of this method was given by Kato and Mugibayashi.[5] Høegh-Krohn developed and refined the technique and gave a series of extensions to more and more singular interactions, including the local $P(\varphi)_2$ Hamiltonian.[6-8] The $(\varphi^4)_2$ Hamiltonian was also studied independently by Kato and Mugibayashi.[9] The present paper marks the first application to interactions with infinite renormalizations.

The Hamiltonian $H(g)$ is constructed in Ref. 1 as the limit of positive self-adjoint Hamiltonians

$$H(g, \kappa) = H_0 + H_{I,g,\kappa} + C_{g,\kappa}, \quad (3)$$

where $H_{I,g,\kappa}$ is an approximation to

or 1 for each $m \in \mathfrak{M}_0\} \supseteq \mathcal{B} \cup \mathcal{C}$. Trivially $\mathcal{L}_0$ is closed under complements, disjoint unions and by (JPZ) it is also closed under the lattice operations; thus $\mathcal{L}_0 \supseteq \mathcal{L}(u,v)$ so that all elements of $\mathcal{L}(u,v)$ are measured exactly in each state of $\mathfrak{M}_0$. The rest follows from the next theorem which is of interest in itself.

Theorem 5: A subsystem $\mathcal{L}_1$ of $\mathcal{L}$ such that, for some set $\mathfrak{M}_0$ of states, the following two conditions hold, is a Boolean σ-algebra:

(i) Every $A \in \mathcal{L}_1$ is measured exactly in each state of $\mathfrak{M}_0$;

(ii) the states in $\mathfrak{M}_0$ separate the elements of $\mathcal{L}_1$.

Proof: Let $h(A) = \{m \in \mathfrak{M}_0 \mid m(A) = 1\}$ for $A \in \mathcal{L}_1$. Clearly $A \leqq B$ implies $h(A) \subseteq h(B)$ and $h(A') = \mathfrak{M}_0 - h(A)$ (set difference). Also $h(A \wedge B) = \{m \in \mathfrak{M}_0 | m(A \wedge B) = 1\} \subseteq \{m \in \mathfrak{M}_0 | m(A) = 1$ and $m(B) = 1\} = h(A) \cap h(B)$; by (JPZ) the reverse also holds: $m \in h(A) \cap h(B)$ implies $m(A) = m(B) = 1$; hence $m(A \wedge B) = 1$, i.e., $m \in h(A \wedge B)$. So we have $h(A \wedge B) = h(A) \cap h(B)$. Now let $h(A) \subseteq h(B)$, or $h(A) = h(A) \cap h(B)$, i.e., $h(A) = h(A \wedge B)$; then $m(A) = m(A \wedge B)$ for all $m \in \mathfrak{M}_0$ since the values of any m on any C in $\mathcal{L}_1$ are 0 or 1, and, as $\mathfrak{M}_0$ separates the elements of $\mathcal{L}_1$, we have $A = A \wedge B$, i.e., $A \leqq B$.

We therefore have that h is one-to-one. Since h also preserves pairwise disjoint unions and for $A \leqq B$ we have $h(A - B) = h(A) - h(B)$ (set difference), we see that the range of h is a Boolean σ-algebra of subsets of $\mathfrak{M}_0$. But h is one-to-one and preserves all relevant operations; hence $\mathcal{L}_1$ is also a Boolean σ-algebra.

[1] S. P. Gudder, Pacific J. Math. **19**, 81 (1966).
[2] S. P. Gudder, Trans. Amer. Math. Soc. **119**, 428 (1965).
[3] A. Ramsey, J. Math. Mech. 15, 227 (1966).
[4] I. E. Segal, A Mathematical Approach to Elementary Particles and Their Fields (University of Chicago Lecture Notes, 1955).

Spectrum of Local Hamiltonians in the Yukawa$_2$ Field Theory

Jonathan Dimock* †

Department of Physics, Harvard University, Cambridge, Massachusetts 02138

(Received 15 July 1971)

It is proved that the local Hamiltonians constructed for the two-dimensional Yukawa model by Glimm and Jaffe have the same continuous spectrum as the free Hamiltonian. The proof depends on the construction of asymptotic creation and annihilation operators.

1. INTRODUCTION

The spatially cutoff Hamiltonian in the two-dimensional Yukawa model is given formally by the expression

$$H(g) = H_0 + \lambda \int :\bar{\psi}(x)\psi(x):\varphi(x)g(x)dx - \tfrac{1}{2}\delta m^2 \int :\varphi(x)^2:g(x)^2dx - E(g), \quad (1)$$

where $g \in C_0^\infty$ is assumed equal to one on a large set and δm^2 and $E(g)$ are the infinite renormalization constants. Glimm and Jaffe have proved that $H(g)$ can be defined as a positive self-adjoint operator,[1] and that it gives rise to locally correct dynamics.[2] They have also proved that $H(g)$ has a vacuum state $\Omega(g)$ (known to be unique for small values of λ), and that in the spectral interval $[0, \min\{m_b, m_f\})$ the spectrum is pure discrete with finite multiplicity.[1] Here m_b and m_f are the bare boson mass and bare fermion mass, respectively. In this paper, we use estimates developed elsewhere[3] to show that $H(g)$ has the same continuous spectrum as H_0, i.e., that $[\min\{m_b, m_f\}, \infty) \subset$ spectrum $H(g)$.

Such a spectrum has the following interpretation proposed by Glimm and Jaffe in connection with the $P(\varphi)_2$ model.[4] The states associated with the continuum are those which at large times (in the Schrödinger picture) represent free particles not in the region of interaction. The discrete spectrum corresponds to states bound into the region of interaction (supp$g$), with reduced mass because of the interaction. One expects that in the infinite volume limit $g \to 1$, the density of bound states becomes infinite giving rise to a new continuum beginning at one of the physical masses.

We remark that the mathematical problem of constructing the theory in the infinite volume limit is quite difficult and involves a change of Hilbert spaces. The present results should be regarded as one step toward the eventual study of the spectrum of a physical Hamiltonian with no cutoffs.

The method we use to obtain the continuum involves the construction of asymptotic creation and annihilation operators. We prove the existence of strong limits of the form

$$\beta_\epsilon^\#(f) = \lim_{t \to \infty} e^{-iH(g)t} e^{iH_0 t} b_\epsilon^\#(f) e^{-iH_0 t} e^{iH(g)t}, \quad (2)$$

where $b_\epsilon^\#(f) = \int b_\epsilon^\#(k) f(k) dk$, $f \in L^2$, is one of the standard creation or annihilation operators for bosons ($\epsilon = 0$), fermions ($\epsilon = +1$), or antifermions ($\epsilon = -1$). Then by applying the asymptotic creation operators $\beta_\epsilon^*(f)$ to the vacuum, we build asymptotic states on which $H(g)$ acts like a free Hamiltonian. The first detailed application of this method was given by Kato and Mugibayashi.[5] Høegh-Krohn developed and refined the technique and gave a series of extensions to more and more singular interactions, including the local $P(\varphi)_2$ Hamiltonian.[6-8] The $(\varphi^4)_2$ Hamiltonian was also studied independently by Kato and Mugibayashi.[9] The present paper marks the first application to interactions with infinite renormalizations.

The Hamiltonian $H(g)$ is constructed in Ref. 1 as the limit of positive self-adjoint Hamiltonians

$$H(g, \kappa) = H_0 + H_{I,g,\kappa} + C_{g,\kappa}, \quad (3)$$

where $H_{I,g,\kappa}$ is an approximation to

$$\lambda \int : \overline{\psi}(x)\psi(x): \varphi(x)g(x)dx$$

in which a cutoff function has been added in momentum space, and $C_{g,\kappa}$ has logarithmically divergent renormalization constants $\delta m^2(\kappa)$ and $E(g,\kappa)$. The convergence is in the norm resolvent sense. For a large class of cutoff functions

$$\|R_\kappa(\zeta) - R(\zeta)\| \to 0 \quad \text{as } \kappa \to \infty, \tag{4}$$

where $R_\kappa(\zeta) = [H(g,\kappa) - \zeta]^{-1}$ and $R(\zeta) = [H(g) - \zeta]^{-1}$.

We now detail the input to the proof that the $t \to \infty$ limit exists. As we shall see, one needs information about the operator $R_\kappa X_\kappa^\epsilon(k) R_\kappa$, where

$$X_\kappa^\epsilon(k) = [b_\epsilon(k), H_{I,g,\kappa} + C_{g,\kappa}]. \tag{5}$$

In Ref. 3 uniform bounds on $\|R_\kappa X_\kappa^\epsilon(k) R_\kappa\|$ were obtained by explicitly carrying out the renormalization cancellations. It was proved that for any $\delta > 0$, there exists a constant independent of κ such that

$$\|R_\kappa X_\kappa^\epsilon(k) R_\kappa\| \le \text{const} \begin{cases} \mu^{-1+\delta}, & \epsilon = 0 \\ \mu^{-1/2+\delta}, & \epsilon = \pm 1 \end{cases}, \tag{6}$$

$$\|R_\kappa^3 X_\kappa^\epsilon(k) R_\kappa^3\| \le \text{const} \begin{cases} \mu^{-(7/4)+\delta}, & \epsilon = 0 \\ \mu^{-(5/4)+\delta}, & \epsilon = \pm 1 \end{cases} \tag{7}$$

Here $\mu = \mu(k)$ is either $\mu_b(k) = (k^2 + m_b^2)^{1/2}$ or $\mu_f = (k^2 + m_f^2)^{1/2}$.

The operators $X_\kappa^\epsilon(k)$ are sums of Wick monomials and for $\epsilon = \pm 1$, it is convenient to distinguish two types of terms, $X_\kappa^\epsilon(k) = X_\kappa^{\epsilon,+}(k) + X_\kappa^{\epsilon,-}(k)$, according to whether $\nu(k) = (\mu_f + k)^{1/2}$ or $\nu(-k)$ appears in the kernel. One case or the other always occurs, these factors arising from the spinors

$$u(p) = \begin{bmatrix} \nu(p) \\ -\nu(-p) \end{bmatrix} \text{ and } v(p) = \begin{bmatrix} \nu(-p) \\ \nu(p) \end{bmatrix}.$$

For example one can see how the separation takes place in the following explicit expression for $\epsilon = +1$,

$$\begin{aligned} X_\kappa^{\epsilon=+1}(k) = \lambda \int \tilde{g}(k' + k + p)[&-(4\pi)^{-1}(2)^{-1/2} \\ &\times (\mu_b(k')\mu_f(k)\mu_f(p))^{-1/2}(\nu(k)\nu(-p) \\ &+ \nu(-k)\nu(p))]\chi_\kappa(k', k, p)(b_0^*(k') \\ &+ b_0(-k'))b_{-1}^*(p)dk'dp + \lambda \int \tilde{g}(k' + k + p) \\ &\times [-(4\pi)^{-1}(2)^{-1/2}(\mu_b(k')\mu_f(k)\mu_f(p))^{-1/2} \\ &\times (\nu(k)\nu(p) + \nu(-k)\nu(-p))]\chi_\kappa(k', k, p)(b_0^*(k') \\ &+ b_0(-k'))b_{+1}(-p)dk'dp \end{aligned}$$

where $\chi_\kappa(k', k, p)$ is the cutoff function. Now define

$$\mathcal{J}_\kappa^{\epsilon,\pm}(k) = \begin{cases} \mu_b(k)^{1/2} X_\kappa^0(k), & \epsilon = 0 \\ \mu_f(k)^{1/2}\nu(\pm k)^{-1} X_\kappa^{\epsilon,\pm}(k), & \epsilon = \pm 1 \end{cases}. \tag{8}$$

These operators have the uniform bounds

$$\|R_\kappa \mathcal{J}_\kappa^{\epsilon,\pm}(k) R_\kappa\| \le \text{const } \mu^{-(1/2)+\delta} \tag{9}$$

$$\|R_\kappa^3 \mathcal{J}_\kappa^{\epsilon,\pm}(k) R_\kappa^3\| \le \text{const } \mu^{-(5/4)+\delta}. \tag{10}$$

These follow from (6) and (7) for $\epsilon = 0$. For $\epsilon \pm 1$, slight modifications of the original proofs yield the result. [Basically, the kernels of $X^{\epsilon,\pm}(\epsilon = \pm 1)$ have factors $\nu(\pm k)\mu_f(k)^{-1/2}$ formerly bounded by a constant which do not appear in $\mathcal{J}^{\epsilon,\pm}$.]

We shall make explicit use of the localization of the Hamiltonian. In terms of the $\mathcal{J}^\epsilon$, this is reflected in the fact that for a suitable choice of cutoff functions, there exists an interval $[-M, M]$ containing supp g such that the Fourier transform of $\langle \theta, R_\kappa \mathcal{J}_\kappa^{\epsilon,\pm}(k) R_\kappa \chi \rangle$ vanishes outside $[-M, M]$ for all κ ($\kappa > 1$) and all θ, χ. (A possible choice is $\chi_\kappa(k, p_1, p_2) = \tilde{\tau}(k/\kappa)\tilde{\rho}(p_1/\kappa) \times \tilde{\rho}(p_2/\kappa)$ with $\tilde{\tau}(p) = \tilde{\tau}(-p)$, $\tilde{\rho}(p) = \tilde{\rho}(-p)$, $\tilde{\tau}(0) = \tilde{\rho}(0) = 1$, and the Fourier transforms τ, ρ in C_0^∞.) The demonstration of this fact was given explicitly for $\epsilon = 0$ in Ref. 3 (Proposition 7.1), and the proof is similar for $\epsilon = \pm 1$.

2. ASYMPTOTIC CREATION AND ANNIHILATION OPERATORS

In obtaining the limit (2), we shall make use of some of the standard estimation techniques which have been developed for the two-dimensional field theories by Glimm and Jaffe. A good general reference is Ref. 4. In particular it is known that $D(N^{1/2}) \supset D(H^{1/2})$ and $\|N^{1/2}R^{1/2}\| \le \text{const}$ ("first-order estimate"). Also by standard estimates, $D(b_\epsilon(f)) \supset D(N^{1/2})$ and $\|b_\epsilon(f)(N+1)^{-1/2}\| \le \text{const } \|f\|_2$, and similarly for $b_\epsilon^*(f)$. Thus the expressions $b_{\epsilon,t}^\#(f)$ given by

$$\begin{aligned} b_{\epsilon,t}(f) &= e^{-iHt} b_\epsilon(e^{-i\mu_\epsilon t} f) e^{iHt}, \\ b_{\epsilon,t}^*(f) &= e^{-iHt} b_\epsilon^*(e^{i\mu_\epsilon t} f) e^{iHt} \end{aligned} \tag{11}$$

are well-defined operators on $D(H^{1/2})$, and

$$\|b_{\epsilon,t}^\#(f)\psi\| \le \text{const } \|f\|_2 \|(H + I)^{1/2}\psi\|. \tag{12}$$

Theorem 1: For $\psi \in D(H)$ and $f \in L^2$, the vectors $b_{\epsilon,t}^\#(f)\psi$ have strong limits as $t \to \infty$.

Remark: We could equally well consider $t \to -\infty$.

Proof: Throughout the proof we suppress ϵ. We first consider the annihilation operators. Because the estimate (12) is uniform in t, it is sufficient to consider ψ in a core for H, say $C^\infty(H)$, and f in a dense set in L^2, say C_0^∞, and vanishing in a neighborhood of the origin, called D_0.

We approximate $b_t(f)$ by $b_{t,\kappa}(f)$ by replacing H by H_κ. We also approximate our vector $\psi \in C^\infty(H)$, which we may take to have the form $\psi = R(\zeta)^n\theta$ with $\zeta < 0$ and some n large, by a vector $\psi_\kappa = R_\kappa(\zeta)^n\theta$ in $C^n(H_\kappa) \subset D(b_{t,\kappa}(f))$. The strategy of the proof is the following. We first show

$$\|b_t(f)\psi - b_{t,\kappa}(f)\psi_\kappa\| \to 0 \quad \text{as } \kappa \to \infty. \tag{13}$$

Then we show

$$\|(b_{t,\kappa}(f) - b_{t',\kappa}(f))\psi_\kappa\| \to 0 \quad \text{as } t, t' \to \infty. \tag{14}$$

uniformly in κ. By an $\epsilon/3$ argument, these imply $\|(b_t(f) - b_{t'}(f))\psi\| \to 0$ as $t, t' \to \infty$.

For (13) we make the expansion

$$\begin{aligned} &\|b_t(f)\psi - b_{t,\kappa}(f)\psi_\kappa\| \\ &\quad \le \|\delta(e^{-iHt})b(e^{-i\mu t}f)R^n e^{iHt}\theta\| \\ &\quad + \|e^{-iH_\kappa t} b(e^{-i\mu t}f)\delta R R^{n-1} e^{iHt}\theta\| \end{aligned}$$

$$+ \| e^{-iH_\kappa t} b(e^{-i\mu t} f) R_\kappa \delta(R^{n-1}) e^{iHt}\theta\| + \| e^{-iH_\kappa t} b(e^{-i\mu t} f) R_\kappa^n \delta(e^{iHt})\theta\|, \quad (15)$$

where $\delta(e^{-iHt}) = e^{-iHt} - e^{-iH_\kappa t}$, etc. The third term converges by resolvent convergence, $\| b(e^{-i\mu t} f)(N+1)^{-1/2}\| \le \text{const} \| f \|_2$, and the uniform estimate $\| (N+1)^{1/2} R_\kappa^{1/2}\| \le \text{const}$. Resolvent convergence also implies that $e^{-iH_\kappa t}$ converges strongly to e^{-iHt}, so that the first and fourth terms in (15) converge. The convergence of the second term follows similarly, using

$$\| N^{1/2}(R_\kappa - R)\| \to 0 \quad \text{as } \kappa \to \infty, \quad (16)$$

which we establish in a lemma at the end of this section. This completes the proof of (13).

For (14) instead of proving strong convergence directly, it is sufficient to prove uniform weak convergence on a dense set in the unit ball. Suppose for every $\delta > 0$ there exists a T such that if $t, t' > T$, then

$$|\langle \chi, (b_{t,\kappa}(f) - b_{t',\kappa}(f))\psi_\kappa\rangle| \le \| \chi\| \, \delta \quad (17)$$

for all $\chi \in C^\infty(H_\kappa)$. By the Riesz representation theorem we have for $t, t' > T$,

$$\| (b_{t,\kappa}(f) - b_{t',\kappa}(f))\psi_\kappa\| \le \delta. \quad (18)$$

If (17) is true for all κ, then so is (18) which gives the uniformity in κ.

Recalling that $\psi_\kappa = R_\kappa(\xi)^n\theta$, the functions we want to study are

$$\langle \chi, b_{s,\kappa}(f)\psi_\kappa\rangle = \langle e^{iH_\kappa s}\chi, b(e^{-i\mu s} f) R_\kappa(\xi)^n e^{iH_\kappa s}\theta\rangle.$$

Using the standard number operator estimates and the first order estimate, it is relatively straightforward to show that these are differentiable functions of s, and integrating the derivatives gives ($V_\kappa = H_{I,\kappa} + C_\kappa$)

$$\langle \chi, (b_{t,\kappa}(f) - b_{t',\kappa}(f))\psi_\kappa\rangle = \int_{t'}^{t} \langle e^{iH_\kappa s}\chi, [iV_\kappa, b(e^{-i\mu s} f)] R_\kappa(\xi)^n e^{iH_\kappa s}\theta\rangle ds. \quad (19)$$

We must estimate the right-hand side of this equation. We let $e^{iH_\kappa s} \to e^{i(H_\kappa - \zeta)s}$ (no change) and note that $e^{i(H_\kappa - \zeta)s}\chi = -i(d/ds) e^{i(H_\kappa - \zeta)s} R_\kappa(\zeta)\chi$. Then integrating by parts we see that the right-hand side of (19) is equal to

$$-i\int_{t'}^{t} \langle e^{iH_\kappa s} R_\kappa(\xi)\chi, [iV_\kappa, b(-i\mu e^{-i\mu s} f)] R_\kappa(\xi)^n e^{iH_\kappa s}\theta\rangle ds - i\int_{t'}^{t} \langle e^{iH_\kappa s} R_\kappa(\xi)\chi, [iV_\kappa, b(e^{-i\mu s} f)] iR_\kappa(\xi)^{n-1} e^{iH_\kappa s}\theta\rangle ds + i\langle e^{iH_\kappa s} R_\kappa(\xi)\chi, [iV_\kappa, b(e^{-i\mu s} f)] R_\kappa(\xi)^n e^{iH_\kappa s}\theta\rangle\Big|_{s=t'}^{s=t}. \quad (20)$$

This step is justified by snowing that the various derivatives exist in the proper topologies and we omit the elementary details. The effect of this manipulation is to add a factor $R_\kappa(\xi)$ on χ. For the first two terms in (20) we follow the same procedure and obtain two more factors of the resolvent on χ and more surface terms. Altogether we obtain eight integrated terms of the form (up to a constant)

$$\int_{t'}^{t} \langle e^{iH_\kappa s}\chi, R_\kappa^3 [V_\kappa, b(e^{-i\mu s} f_\mu)] R_\kappa^3 e^{iH_\kappa s}\theta'\rangle ds, \quad (21)$$

where $f_\mu = \mu^j f$, $j = 0, 1, 2$, or 3, lies in D_0 and $\theta' = R_\kappa^k\theta$ for some k. (We assume $n \ge 6$). Then there are seven surface terms of the form

$$\langle e^{iH_\kappa s}\chi', R_\kappa [V_\kappa, b(e^{-i\mu s} f_\mu)] R_\kappa e^{iH_\kappa s}\theta'\rangle\Big|_{s=t'}^{s=t}, \quad (22)$$

where $\chi' = R_\kappa^l \chi$, $l = 0, 1, 2$.

The terms (21) have the form $\int_{t'}^{t} A(s)ds$. We obtain the convergence required by (17) by showing that for any integer m, there exists a constant independent of κ such that

$$|A(s)| \le \text{const} \| \chi\| \, |s|^{-m}. \quad (23)$$

To this end we note that the integrand $A(s)$ can be written as

$$A(s) = \int e^{-i\mu s} f_\mu(k) \langle e^{iH_\kappa s}\chi, R_\kappa^3 X_\kappa(k) R_\kappa^3 e^{iH_\kappa s}\theta'\rangle dk = \int e^{-i\mu s} \Big(\sum_{\pm} f_\mu^{\pm}(k)\gamma^{\pm}(k)\Big) dk, \quad (24)$$

where

$$\gamma^{\pm}(k) = \langle e^{iH_\kappa s}\chi, R_\kappa^3 \mathcal{J}_\kappa^{\pm}(k) R_\kappa^3 e^{iH_\kappa s}\theta'\rangle$$
$$f_\mu^{\pm}(k) = \begin{cases} \frac{1}{2}\mu_b^{-1/2} f_\mu(k), & \epsilon = 0 \\ \nu(\pm k)\mu_f^{-1/2} f_\mu(k), & \epsilon = \pm 1 \end{cases}. \quad (25)$$

As mentioned previously, with the proper choice of cutoff functions, the support of $\hat{\gamma}^{\pm}(x) = (2\pi)^{-1/2}\int e^{ikx} \times \gamma^{\pm}(k)dk$ is contained in an interval $[-M, M]$ for all κ, s. Then $\gamma^{\pm}(k)$ is infinitely differentiable and using the κ, s uniform bound $\| \hat{\gamma}^{\pm}\|_\infty \le \| \gamma^{\pm}\|_1 \le \text{const} \| \chi\|$ which follows from (10), we obtain the κ, s uniform bound

$$\Big| \frac{d^r}{dk^r}\gamma^{\pm}(k)\Big| = (2\pi)^{-1/2} \Big| \int_{-M}^{M} e^{-ikx}(-ix)^r \hat{\gamma}^{\pm}(x)dx\Big| \le \text{const } M^{r+1} \| \hat{\gamma}^{\pm}\|_\infty \le \text{const} \| \chi\|. \quad (26)$$

Furthermore $f_\mu^{\pm} \in D_0$ is infinitely differentiable, and all its derivatives are in D_0. Now in (24) we write

$$e^{-i\mu s} = s^{-m}\Big(\frac{i\mu}{k}\frac{d}{dk}\Big)^m e^{-i\mu s}$$

and integrate by parts m times. This is justified since the integrands always have compact support away from the origin. This gives the factor $|s|^{-m}$, and since in (26) we have the necessary uniform bounds on the derivatives, we obtain (23).

Now consider the surface terms (22). These are the difference of terms of the form

$$\int e^{-i\mu t} f_\mu(k) \langle e^{iH_\kappa t}\chi', R_\kappa X_\kappa(k) R_\kappa e^{iH_\kappa t}\theta'\rangle dk = \sum_{\pm} B^{\pm}(f_\mu^{\pm}, t), \quad (27)$$

where $f_\mu^{\pm} \in \mathcal{S}$ (Schwartz space) is as before and

$$B^{\pm}(h, t) = \int e^{-i\mu t} h(k)\lambda^{\pm}(k, t)dk,$$
$$\lambda^{\pm}(k, t) = \langle e^{iH_\kappa t}\chi', R_\kappa \mathcal{J}_\kappa^{\pm}(k) R_\kappa e^{iH_\kappa t}\theta'\rangle. \quad (28)$$

All these terms converge to zero with the proper uniformity. In fact we show that for all $h \in \mathcal{S}$, $|B^{\pm}(h, t)| \to 0$ as $t \to \infty$ uniformly in κ and χ on the unit ball. Using $\| \chi'\| \le \text{const} \| \chi\|$ and (9), we have the κ, t uniform bound

$$|B^{\pm}(h,t)| \leq \| \lambda^{\pm}(\cdot,t)\|_{\infty} \| h\|_1 \leq \text{const} \| h\|_1 \| \chi\|. \quad (29)$$

Since the L^1 norm is a Schwartz norm, this shows that it is sufficient to prove convergence for h in a dense set in $\mathcal{S}$ in the $\mathcal{S}$ topology. We choose the set of all $\mathcal{S}$ functions whose Fourier transforms are C_0^{∞}. We now write

$$\begin{aligned} |B^{\pm}(h,t)| &= |\int o(k,t)\tau^{\pm}(k,t)dk| \\ &= |\int \hat{o}(x,t)\hat{\tau}^{\pm}(-x,t)dx| \\ &\leq \| \hat{o}(\cdot,t)\|_{\infty} \| \hat{\tau}^{\pm}(\cdot,t)\|_1, \end{aligned} \quad (30)$$

where for some large integer r,

$$\begin{aligned} o(k,t) &= e^{-i\mu t}\mu^{-2r}, \\ \tau^{\pm}(k,t) &= \mu^{2r}h(k)\lambda^{\pm}(k,t). \end{aligned} \quad (31)$$

Since $h \in \mathcal{S}$, $\| \hat{\tau}^{\pm}(\cdot,t)\|_{\infty} \leq \| \tau^{\pm}(\cdot,t)\|_1 \leq \text{const} \| \chi\|$. Furthermore

$$\hat{\tau}^{\pm}(x,t) = (2\pi)^{-1/2}\int ((-\Delta + m^2)^r\hat{h})(x-y)\hat{\lambda}^{\pm}(y,t)dy \quad (32)$$

vanishes outside an interval independent of κ and t since both $(-\Delta + m^2)^r\hat{h}$ and $\hat{\lambda}^{\pm}do$. Therefore we obtain the κ, t uniform bound

$$\| \hat{\tau}^{\pm}(\cdot,t)\|_1 \leq \text{const} \| \hat{\tau}^{\pm}(\cdot,t)\|_{\infty} \leq \text{const} \| \chi\|. \quad (33)$$

Finally $\hat{o}(x,t) = (2\pi)^{-1/2}\int e^{ikx - i\mu t}\mu^{-2r}$ is a smooth solution of the Klein–Gordon equation and has the standard rate of decrease in two dimensions,

$$\| \hat{o}(\cdot,t)\|_{\infty} \leq \text{const} \ |t|^{-1/2}. \quad (34)$$

See Jost[10] or S. Nelson.[11] Thus combining (30), (33), and (34) we have for a constant independent of κ,

$$|B^{\pm}(h,t)| \leq \text{const} \| \chi\| \ |t|^{-1/2}, \quad \hat{h} \in C_0^{\infty}. \quad (35)$$

Thus the surface terms (22) converge to zero and (17) is proved.

This completes the proof of the theorem for the annihilation operators. For the creation operators we make the replacements $e^{-i\mu t} \to e^{i\mu t}$ and $X \to X^*$. Neither of these makes any substantial difference and the whole proof goes through as before.

We now give the proof of (16).

Lemma: As $\kappa \to \infty$, $\| N^{1/2}(R_{\kappa} - R)\| \to 0$.

Proof: We introduce the operator $P_j = (1 + j^{-1}N)^{-1}$, which satisfies $\| N^r P_j\| \leq j^r (0 \leq r \leq 1)$. Since $\| N^{1/2}P_j(R_{\kappa} - R)\| \to 0$ as $\kappa \to \infty$ by resolvent convergence, we need only show that as $j \to \infty$, $N^{1/2}P_jR_{\kappa}$ converges to $N^{1/2}R_{\kappa}$ uniformly in κ for $\kappa \leq \infty$. This follows from the uniform quadratic estimate proved in Ref. 3 which says $\| NR_{\kappa}\| \leq \text{const}$ $(\kappa \leq \infty)$. Thus as $j \to \infty$,

$$\begin{aligned} \| N^{1/2}(1 - P_j)R_{\kappa}\| &= j^{-1}\| N^{1/2}P_jNR_{\kappa}\| \\ &\leq \text{const} \ j^{-1/2} \\ &\to 0. \end{aligned}$$

We remark that the same proof gives that $\| N_{\tau}^{1/2}(R_{\kappa} - R)\| \to 0$ as $\kappa \to \infty$, provided that $\tau < \frac{1}{2}$. Here $N_{\tau} = \sum_{\epsilon} \int \mu_{\epsilon}(k)^{\tau} b_{\epsilon}(k)^* b_{\epsilon}(k)dk$. A similar proof extends the result to $\tau < 1$ (A. Jaffe, private communication).

3. SPECTRUM OF THE HAMILTONIAN

By Theorem 1 we may define operators $\beta_{\epsilon}(f)$ and $\beta_{\epsilon}^*(f)$ on $D(H)$ by

$$\beta_{\epsilon}^{\#}(f)\psi = \lim_{t \to \infty} b_{\epsilon,t}^{\#}(f)\psi. \quad (36)$$

We now establish properties of these operators.

Proposition 1: For $\psi, \varphi \in D(H)$, $\langle \beta_{\epsilon}^*(\bar{f})\psi, \varphi\rangle = \langle \psi, \beta_{\epsilon}(f)\varphi\rangle$ and thus $\beta_{\epsilon}^*(\bar{f}) \subset \beta_{\epsilon}(f)^*$ and $\beta_{\epsilon}(f) \subset \beta_{\epsilon}^*(\bar{f})^*$.

Proof: For finite t we have

$$\langle b_{\epsilon}^*(e^{i\mu_{\epsilon}t}\bar{f})e^{iHt}\psi, e^{iHt}\varphi\rangle = \langle e^{iHt}\psi, b_{\epsilon}(e^{-i\mu_{\epsilon}t}f)e^{iHt}\varphi\rangle.$$

Letting $t \to \infty$, we obtain the result.

Proposition 2: On $D(H) \times D(H)$, the canonical (anti) commutation relations hold, that is,

$$\begin{aligned} [\beta_0(f), \beta_0^*(g)] &= \langle \bar{f}, g\rangle_{L^2}, \\ \{\beta_{\pm 1}(f), \beta_{\pm 1}^*(g)\} &= \langle \bar{f}, g\rangle_{L^2}, \end{aligned} \quad (37)$$

with all other canonical (anti) commutators equal to zero.

Proof: This follows from the strong convergence and the corresponding relations at finite t.

Proposition 3: The asymptotic operators $\beta_{\epsilon}^{\#}(f)$ have the same commutation relations with H as the $b_{\epsilon}^{\#}(f)$ do with H_0. In particular on $D(H)$, for all $f \in L^2$,

$$\begin{aligned} e^{iHs}\beta_{\epsilon}(f)e^{-iHs} &= \beta_{\epsilon}(e^{-i\mu_{\epsilon}s}f), \\ e^{iHs}\beta_{\epsilon}^*(f)e^{-iHs} &= \beta_{\epsilon}^*(e^{i\mu_{\epsilon}s}f). \end{aligned} \quad (38)$$

Also for $f \in \mathcal{S}$ we have on $D(H) \times D(H)$,

$$\begin{aligned} [H, \beta_{\epsilon}(f)] &= \beta_{\epsilon}(-\mu_{\epsilon}f), \\ [H, \beta_{\epsilon}^*(f)] &= \beta_{\epsilon}^*(\mu_{\epsilon}f). \end{aligned} \quad (39)$$

Proof: For finite t we have the operator identity on $D(H)$:

$$\begin{aligned} e^{iHs}b_{\epsilon,t}(f)e^{-iHs} &= b_{\epsilon,t-s}(e^{-i\mu_{\epsilon}s}f), \\ e^{iHs}b_{\epsilon,t}^*(f)e^{-iHs} &= b_{\epsilon,t-s}^*(e^{i\mu_{\epsilon}s}f). \end{aligned}$$

As $t \to \infty$ we obtain (38). Differentiating (38) on $D(H) \times D(H)$ and setting $s = 0$ gives (39).

Proposition 4: If θ is an eigenstate of H, then $\beta_{\epsilon}(f)\theta = 0$ for all $f \in L^2$ and all ϵ.

Proof: First consider vectors in the Fock space with a finite number of particles and whose wavefunctions in the n-particle subspaces are of the form $h_1 \otimes \cdots \otimes h_n$ (symmetrized or antisymmetrized), with $h_i \in \mathcal{S}$. For any such vector ψ and any $f \in \mathcal{S}$,

$$\| b_{\epsilon}(e^{-i\mu_{\epsilon}t}f)\psi\| \to 0 \quad \text{as } t \to \infty, \quad (40)$$

for if one writes out this norm explicitly, one finds that all terms are proportional to something of the form

$$|\int e^{-i\mu_\epsilon t} f(k) h_i(k) dk|,$$

which goes to zero as $t \to \infty$, as can be seen using (34). Finite combinations of these vectors form a dense set of vectors satisfying (40). This domain is a core for $(N+1)^{1/2}$ and since we have the t uniform bound

$$\| b_\epsilon(e^{-i\mu_\epsilon t} f)\psi \| \leq \text{const} \, \| f \|_2 \| (N+1)^{1/2} \psi \|,$$

we can extend (40) to all $\psi \in D((N+1)^{1/2})$, $f \in L_2$. Now suppose θ is an eigenstate of H, $e^{iHt}\theta = e^{iEt}\theta$. Then since $D(H) \subset D((N+1)^{1/2})$,

$$\| e^{-iHt} b_\epsilon(e^{-i\mu_\epsilon t} f) e^{iHt}\theta \| = \| b_\epsilon(e^{-i\mu_\epsilon t} f)\theta \| \to 0 \quad \text{as } t \to \infty; \tag{41}$$

but the left-hand side of (41) converges to $\| \beta_\epsilon(f)\theta \|$, hence $\beta_\epsilon(f)\theta = 0$.

Proposition 5: If θ is an eigenstate of H, then $\| \beta_\epsilon^*(f)\theta \| = \| f \|_2 \| \theta \|$.

Proof: This follows from Propositions 2 and 4.

We now can prove the main result.

Theorem 2: $[\min\{m_b, m_f\}, \infty) \subset$ spectrum H.

Proof: We must show that for any λ in this interval and any $\delta > 0$, there exists a $\psi \in D(H)$, such that $\| (H - \lambda)\psi \| \leq \delta \| \psi \|$. To find such a ψ, consider vectors of the form $\beta_\epsilon^*(f)\Omega$ with $f \in C_0^\infty$ and Ω a vacuum for H, i.e., $H\Omega = 0$, $\|\Omega\| = 1$. For $\chi \in D(H)$, we have by Proposition 3, $\langle H\chi, \beta_\epsilon^*(f)\Omega \rangle = \langle \chi, \beta_\epsilon^*(\mu_\epsilon f)\Omega \rangle$. Thus $\beta_\epsilon^*(f)\Omega \in D(H)$ and

$$H\beta_\epsilon^*(f)\Omega = \beta_\epsilon^*(\mu_\epsilon f)\Omega. \tag{42}$$

Now take ϵ corresponding to the smaller mass, and narrow support for f so that if $k \in \text{supp} f$, $|\mu_\epsilon - \lambda| < \delta$. Then using Proposition 5, we have

$$\begin{aligned} \| (H - \lambda)\beta_\epsilon^*(f)\Omega \| &= \| \beta_\epsilon^*((\mu_\epsilon - \lambda) f)\Omega \| \\ &= \| (\mu_\epsilon - \lambda) f \|_2 \\ &\leq \delta \| f \|_2 \\ &= \delta \| \beta_\epsilon^*(f)\Omega \|, \end{aligned} \tag{43}$$

which proves the theorem.

We conclude with some general remarks. We have obtained one particle asymptotic states by applying the $\beta_\epsilon^*(f)$ to the vacuum. Similarly, one might generate n-particle states by applying operators of the form $\beta_{\epsilon_1}^*(f_1) \cdots \beta_{\epsilon_n}^*(f_n)$ to the vacuum.[4] Summing the n-particle subspaces gives an asymptotic Fock space $\mathfrak{F}_{\text{out}}$. Such a construction has been outlined by Høegh-Krohn[7,8] for other models. One finds that the Hamiltonian acts on $\mathfrak{F}_{\text{out}}$ like a free Hamiltonian, schematically $H = \sum_\epsilon \int \mu_\epsilon(k) \beta_\epsilon^*(k) \beta_\epsilon(k) dk$, just as we have shown for the one-particle states in (42). Using operators defined as limits of the $b_{\epsilon,t}^\#(f)$ as $t \to -\infty$, one might also obtain another Fock space $\mathfrak{F}_{\text{in}}$. Scalar products of states in $\mathfrak{F}_{\text{in}}$ with states in $\mathfrak{F}_{\text{out}}$ would give an S matrix $S(g)$. Such a construction is of limited interest at this time, since the physical S matrix has to be constructed after taking the limit $g \to 1$, and it is not clear in what sense (if any) $S(g)$ might converge to S. Working in the infinite volume limit, one could hope to obtain an S matrix using the Haag-Ruelle collision theory. This theory depends on a detailed knowledge of the spectrum of the Hamiltonian. In this sense the present results are a step toward the construction of a physical S matrix.

ACKNOWLEDGMENTS

I would like to thank A. Jaffe for proposing the problem and for many helpful suggestions. I also thank J. Glimm and R. Schrader for useful remarks.

---

* Present address: Courant Institute of Mathematical Sciences, New York University, New York 10012.

† Supported in part by Air Force Office of Scientific Research, Contract F44620-70-C-0030 and by National Science Foundation Grant GP-31239X.

1 J. Glimm and A. Jaffe, Ann. Phys. (N.Y.) **60**, 321 (1970).
2 J. Glimm and A. Jaffe, J. Functional Anal. **7**, 323 (1971).
3 J. Dimock, "Estimates, Renormalized Currents, and Field Equations for the Yukawa$_2$ Field Theory," Ann. Phys. (N.Y.) (to be published).
4 J. Glimm and A. Jaffe, "Field Theory Models," in *1970 Les Houches Lectures*, edited by C. Dewitt and R. Stora (Gordon and Breach, New York, 1971).
5 Y. Kato and N. Mugibayashi, Progr. Theoret. Phys. **30**, 103 (1963).
6 R. Høegh-Krohn, J. Math. Phys. **9**, 2075 (1968).
7 R. Høegh-Krohn, Commun. Math. Phys. **18**, 109 (1970).
8 R. Høegh-Krohn, Commun. Math. Phys. **21**, 256 (1971).
9 Y. Kato and N. Mugibayashi, Progr. Theoret. Phys. (Kyoto) **45**, 628 (1971).
10 R. Jost, Helv. Phys. Acta **39**, 21 (1966).
11 S. Nelson, Trans. Amer. Math. Soc. **154**, 227 (1971).

On the Mathematical Theory of Electromagnetic Radiation from Flanged Waveguides

A. Q. Howard*
Institute for Telecommunication Sciences, Boulder, Colorado
(Received 12 August 1971; Revised Manuscript Received 22 October 1971)

A mathematical technique particularly suited to describing the electromagnetic radiation from open-ended waveguide structures is discussed. As an example of the utility of the method, the exact solutions, for both wave polarizations corresponding to a uniform line source embedded in a dielectric filled slot in a ground plane, are given. The edge condition is used to estimate and partially correct for the truncation error resulting from approximating the infinite system of linear equations by a finite one. The truncation corrected field expressions are shown to recover the proper local field behavior in the vicinity of the aperture perimeter. Some numerical results, for both aperture and radiation fields are given and in the latter case compared with the Kirchhoff approximation to the radiation field.

1. INTRODUCTION

The rather straightforward but sometimes mathematically poorly conditioned method of mode matching as discussed, for example, by Wexler[1] and more recently by Masterman and Clarricoats[2] has in the past been applied to a variety of electromagnetic boundary value problems arising from mathematical analyses of waveguide configurations in which severe or abrupt discontinuities exist in the waveguide walls.[3] We shall consider an important type of discontinuity, the termination of a guide or cavity in an open ended flange plane.

The method to be used complements solutions obtained by Weinstein,[4] who used the Wiener–Hopf technique to treat unflanged cases. Mittra and Lee[5] have discussed the relationship between the Wiener-Hopf and mode matching technique as well as giving much of the formal theory for both. The solutions should also establish greater confidence in the application of the geometrical theory of diffraction to narrow waveguide radiators.[6] The analysis is in many respects a generalization of the work of Nussenzveig.[7]

The method to be used to partially correct for truncation error[8] is an adaptation of "the mixed method" employed by Do Amaral and Bautista Vidal,[9] who treated the intermediate region case $0.1 < k_0 a < 1.7$, where $2a$ is the width of the guide and k_0 is the wavenumber.

Although the numerical results we present refer to a situation different in two respects from Nussenzveig's work, our theory in contrast is developed for both wave polarizations and also admits the possibilities of a finite depth well which is filled with a uniform lossy medium.

In order to incorporate the truncation correction, the infinite system of linear equations is transformed into an equivalent infinite set for which the algebraic dependence of the amplitudes as a function of mode index is known asymptotically. This information, which is related to the edge condition, is used to consider the truncation error.

The matrix elements of the scatter matrix[10] for either polarization are shown to reduce to a canonical integral which is proportional to a finite sine or cosine transform of the Hankel function of order zero. An asymptotic series for the canonical integral is obtained. This series, coupled with the fact that the scatter matrix is nearly diagonal dominant[11] and extremely well conditioned, makes it numerically possible, although as shown not necessary, to solve the resulting system of equations for orders in excess of fifty.

The paper is organized into three sections. In Sec. 2 the formal exact solution for the simplest nontrivial problem in the class of flanged waveguide cavities is obtained. Then the resulting matrix equations are formulated in an equivalent set for the aperture amplitudes. To implement the solution of the infinite system, an analysis of the matrix elements is made. The symmetry of the scatter matrix is exploited to generate a matrix of order N with $2N-1$ canonical integrals. In general these integrals must be evaluated numerically. To facilitate computation, a series which is asymptotic in mode index is obtained for the canonical integral. Section 2 concludes with the determination of the asymptotic evaluation of the aperture amplitudes for large mode index.

The third and final section deals with the formulation of the truncation corrected equations, the truncation error estimate, and some graphical results for the radiation and aperture fields. A comparison of the resulting radiation fields with the Kirchhoff predicted fields shows good agreement except near resonance and in low elevation angle regions.

2. THE FORMULATION: THE DISCRETE AND CONTINUOUS WAVE REPRESENTATIONS

In this section we will obtain the formal solution in all space, and for both wave polarizations, to the electromagnetic fields in a parallel plate flanged slot geometry depicted in Fig 1.

A uniform electric or magnetic line source is located a distance h below the slot aperture, and for simplicity is located in the central (y, z) plane midway between the vertical walls. A Cartesian coordinate system is defined such that the plane $z = 0$ corresponds to the flange interface. The positive z direction is normal and points outward from the flange. The x coordinate is measured positive to the right from the origin contained in the intersection of the interface and central plane. The y axis, which points into the paper, completes the right-handed triad (x, y, z). The slot half-width is denoted by a and the slot depth by d. Because of symmetry, the fields are independent of the y coordinate. Also, because of the geometry, it can be shown that the problem is scalar.[12]

For vertically polarized waves, the primitive field quantity is the y component of the magnetic intensity $\varphi(x, z) = H_y(x, z)$, which gives rise to TM waves (transverse to z). Similarly, horizontally polarized or TE waves are obtained by defining the field primitive to be $\varphi(x, z) = E_y(x, z)$. The time harmonic factor is taken to be $e^{-i\omega t}$ and the free space wavenumber (for $z > 0$) is $k_0 = 2\pi/\text{wavelength}$. The cavity region ($-h < z < 0$ and $|x| < a$) is filled with a uniform and in general lossy dielectric specified by a dielectric constant ϵ and a conductivity σ. Thus the wave-

number k_1 for the interior region is given by $k_1 = k_0(\epsilon/\epsilon_0 + i\sigma/\omega\epsilon_0)^{1/2}$. Notice that the units are MKS and the permeability is taken to be μ_0 in all regions.

In the interior region, we thus seek a solution to Maxwell's equations φ_{int} which satisfies the wave equation

$$(\nabla^2 + k_1^2)\varphi_{\text{int}}(x, z) = -\delta(z + h)\delta(x), \tag{2.1}$$

where δ is the Dirac delta function. Before we discuss the wave representations, let us consider the source term in (2.1). In the absence of any boundaries, the solution to (2.1) which corresponds to a diverging cylinder wave is given by

$$\varphi(x, z) = \tfrac{1}{4}iH_0^{(1)}(k_0[(z + h)^2 + x^2]^{1/2}),$$

where $H_0^{(1)}$ is the zeroth-order Hankel function of the first kind. Although the general solution to (2.1) is the polarization dependent Green's function for this particular geometry, it is important to also note the close relation of the right-hand side of (2.1) to a uniform distribution of Hertzian vertical dipoles for TM waves and Hertzian horizontal dipoles in the TE case. For the latter, the interpretation is fairly obvious. The relation of the TM magnetic line current and an equivalent uniform vertical dipole density has been discussed by Wait.[13]

A discrete spectrum, that is, a modal series, is suitable for the representation of φ_{int}. In order that the source condition (2.1) be satisfied, it is convenient to give a piecewise definition of φ_{int}:

$$\varphi_{\text{int}} = \sum_{n=0}^{\infty} B_n(e^{ik_1\gamma_n z} + \hat{\Gamma}_n e^{-ik_1\gamma_n z})\cos k_1\beta_n x \quad \text{for } z > -h \tag{2.2a}$$

$$\varphi_{\text{int}} = \sum_{n=0}^{\infty} A_n(e^{-ik_1\gamma_n z} + \Gamma_n e^{ik_1\gamma_n z})\cos k_1\beta_n x \quad \text{for } z < -h \tag{2.2b}$$

where $n = 0, 1, 2, \cdots$ and $A_n, B_n, \Gamma_n, \hat{\Gamma}_n, \gamma_n, \beta_n$ are as yet unspecified constants. For (2.2) to satisfy (2.1) in its homogeneous form, we must have

$$\beta_n^2 + \gamma_n^2 = 1. \tag{2.3}$$

The perfectly conducting walls at $x = \pm a$ impose the conditions

$$\begin{Bmatrix} E_z \\ E_y \end{Bmatrix} = 0 \quad \text{for} \begin{Bmatrix} \text{TM} \\ \text{TE} \end{Bmatrix} \text{waves.}$$

These conditions are satisfied by defining β_n to be

$$\beta_n = \begin{cases} n\pi/(k_1 a) & \text{TM waves} \quad (2.4) \\ (2n + 1)\pi/(2k_1 a) & \text{TE waves} \quad (2.5) \end{cases}$$

At the possible risk of slight confusion, unless otherwise stated, the symbol β_n will be used for either polarization. Note that γ_n is now defined to within a sign. The sign is chosen by requiring the waves to decay as they move away from the source. Thus

$$\gamma_n = (1 - \beta_n^2)^{1/2}, \quad \text{Im}(\gamma_n) \geq 0. \tag{2.6}$$

Similar to the determination of the β_n's, the perfectly conducting floor of the slot ($z = -d$) leads directly to the expression for Γ_n:

$$\Gamma_n = \pm e^{2ik_1\gamma_n d} \quad \text{for} \begin{pmatrix} \text{TM} \\ \text{TE} \end{pmatrix} \text{waves.} \tag{2.7}$$

In the following we will use the elementary orthogonality condition

$$\int_{-a}^{+a} \cos k_1\beta_n x \cos k_1\beta_m x\, dx = 2\delta_{nm}a/\epsilon_n, \tag{2.8}$$

where δ_{nm} is the Kronecker delta and ϵ_n is the Newmann factor defined by

$$\delta_{nm} = \begin{cases} 1, n = m \\ 0, n \neq m \end{cases}, \quad \epsilon_n = \begin{cases} 1, n = 0 \\ 2, n = 1, 2, 3, \cdots \end{cases}. \tag{2.9}$$

Two conditions, the continuity of the field in the plane $z = -h$ and that the derivative of φ_{int} with respect to z be discontinuous as dictated by the right-hand side of (2.1), lead to the following relation between the unknowns B_n and $\hat{\Gamma}_n$:

$$B_n = [(H_n^2\Gamma_n + 1)\epsilon_n]/[4i\gamma_n k_1 a H_n(1 - \Gamma_n\hat{\Gamma}_n)], \tag{2.10}$$

where ϵ_n is given by (2.9) and H_n is defined

$$H_n = e^{-ik_1\gamma_n h}. \tag{2.11}$$

The coefficient $\hat{\Gamma}_n$ is, of course, a sum of reflection coefficients of all modes into the nth mode. The reflections arise because of the aperture discontinuity, that is, the truncation of the waveguide. If $\text{Im}(\gamma_n k_0 h) \gg 1$, $\hat{\Gamma}_n$ reduces to the reflection coefficient of the nth mode into the most nearly resonant mode, i.e., the mode which most and nearly satisfies $n\pi = \text{Re}(k_1 a)$.

In the external region ($z > 0$), it is convenient to represent the primitive field $\varphi_{\text{int}}(x, z)$ as a continuous spectrum of inhomogeneous plane waves:

$$\varphi_{\text{ext}}^{(j)}(x, z) = \int_0^{\infty} \cos\mu x\, e^{i\beta(\mu)z} g^{(j)}(\mu)d\mu, \tag{2.12}$$

where $\beta(\mu) = (k_0^2 - \mu^2)^{1/2}$.

Here the superscript is a polarization index where (1) corresponds to TM and (2) to TE and $g^{(j)}(\mu)$ is to be determined. In order that the radiation condition be fulfilled for $z \to +\infty$, the branch of the complex square root is determined by the condition $\text{Im}\beta(\mu) \geq 0$. There are several reasons for the choice of the

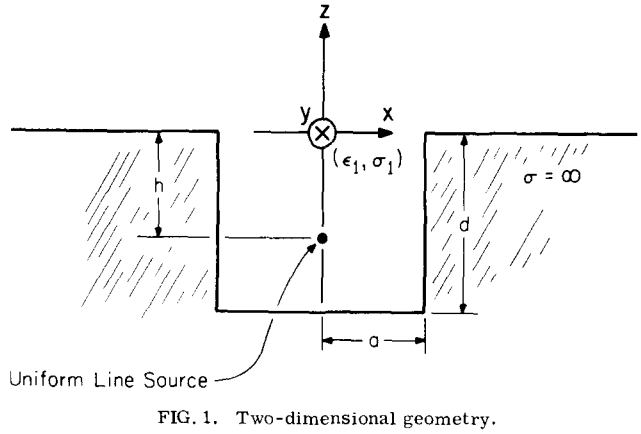


FIG. 1. Two-dimensional geometry.

representation (2.12). For one, it is symmetric in x, and it is transformable in x over the infinite interval. That is,

$$\int_0^\infty \cos\mu s \cos\mu' s \, ds = \tfrac{1}{2}\pi\delta(\mu - \mu') \quad \text{for } \mu, \mu' > 0. \tag{2.13}$$

Furthermore, it is closely related to the plane wave representation of the zeroth-order Hankel function. This property will be exploited later on in the calculation of the radiation pattern, and also in the simplification of the matrix elements. It also satisfies the wave equation in the source free region.

The conditions of continuity of tangential fields in the plane $z = 0$ lead to the following systems.

For TM waves, these conditions give

$$\frac{1}{k_0^2}\frac{\partial\varphi_{\text{ext}}^{(1)}}{\partial z}\Big|_{z=0^+} = \begin{cases} \dfrac{1}{k_1^2}\dfrac{\partial\varphi_{\text{int}}^{(1)}}{\partial z}\Big|_{z=0^-}, & |x| < a, \\ 0, & |x| > a, \end{cases} \tag{2.14}$$

$$\varphi_{\text{ext}}^{(1)}\Big|_{z=0^+} = \varphi_{\text{int}}^{(1)}\Big|_{z=0^-}, \quad |x| < a. \tag{2.15}$$

Similarly the corresponding conditions for the TE case yield

$$\varphi_{\text{ext}}^{(2)}\Big|_{z=0^+} = \begin{cases} \varphi_{\text{int}}^{(2)}\Big|_{z=0^-}, & |x| < a, \\ 0 & |x| > a, \end{cases} \tag{2.16}$$

$$\frac{\partial\varphi_{\text{ext}}^{(2)}}{\partial z}\Big|_{z=0^+} = \frac{\partial\varphi_{\text{int}}^{(2)}}{\partial z}\Big|_{z=0^-}, \quad |x| < a. \tag{2.17}$$

Since the left-hand sides of (2.14) and (2.16) are defined for all x, these equations can be used to isolate $g^{(j)}(\mu)$ for $j = 1$ and 2, respectively. Similarly the mode amplitude B_n can be written in terms of $g^{(j)}(\mu)$ by applying (2.8) to (2.15) for $j = 1$ and (2.17) for $j = 2$.

Consequently, if we insert the appropriate representations into (2.14) we can use (2.13) to solve for $g^{(j)}(\mu)$:

$$g^{(1)}(\mu) = \frac{k_0^2}{k_1\pi\beta(\mu)}\sum_{n=0}^{\infty}\gamma_n B_n(1 - \hat{\Gamma}_n) \times \int_{-a}^{+a}\cos\mu s \cos k_1\beta_n s \, ds. \tag{2.18}$$

From this point on, it will be necessary on occasion to interchange the operations of integration and summation. For example, the justification for the interchange of the operations of summation and integration in (2.18) follows from Lebesgue's bounded convergence theorem.[14] That is, since the summation in (2.18) is a representation of E_x in the aperture, and E_x is singular at $x = \pm a$, the series is *not* uniformly convergent on the interval $(-a, +a)$. Hence we must use the bounded convergence theorem in this case. Note that the interchange of summation and integration in going from (2.16) to (2.20) relies only on the uniform convergence property of a Fourier series of a continuous function.[15] Since similar results apply to the other interchanges which follow, this point need not be discussed again. Thus (2.8) can be used to isolate the product $B_m(1 + \hat{\Gamma}_m)$ occurring in the right-hand side of (2.15) obtaining

$$\frac{2a}{\epsilon_m}B_m(1 + \hat{\Gamma}_m) = \int_0^\infty g^{(1)}(\mu)d\mu \times \int_{-a}^{a}\cos\mu s' \cos k_1\beta_m s' \, ds'. \tag{2.19}$$

Similar operations yield the following two equations for the TE polarization. The condition (2.16) yields the information

$$g^{(2)}(\mu) = \frac{1}{\pi}\sum_{n=0}^{\infty}B_n(1 + \hat{\Gamma}_n)\int_{-a}^{a}\cos\mu s \cos k_1\beta_n s \, ds, \tag{2.20}$$

while (2.17) reduces to

$$\frac{2k_1 a}{\epsilon_m}\gamma_m B_m(1 - \hat{\Gamma}_m) = \int_0^\infty \beta(\mu)g^{(2)}(\mu)d\mu \times \int_{-a}^{a}\cos\mu s' \cos\beta_m k_1 s' \, ds'. \tag{2.21}$$

The pairs of equations (2.18), (2.19) and (2.20), (2.21) may be cast into either integral equations for $g^{(j)}(\mu)$ or matrix equations for B_m. We choose the latter alternative. In this regard it is important to define the quantities $D_m^{\pm}$. The aperture amplitudes are defined as

$$D_m^{\pm} = B_m(1 \pm \hat{\Gamma}_m). \tag{2.22}$$

Note that $D_m^{\pm}$ will not have polarization indices. The coefficients $D_m^{\pm}$ are algebraic in their dependence on n. This will be verified in Sec. 3 where the edge singularities are discussed. In contrast to the algebraic decay of the aperture amplitudes $D_m^{\pm}$ with mode number, B_m decreases exponentially, and $\hat{\Gamma}_m$ increases exponentially with mode number. Thus, if matrix equations were developed for either of these quantities, severe numerical problems would result. The coefficients $D_m^{\pm}$ are linearly related through the source excitation condition (2.10). This condition gives directly

$$D_m^{+} = F_m^{(1)} - [(1 + \Gamma_m)/(1 - \Gamma_m)]D_m^{-}, \tag{2.23}$$

where

$$F_m^{(1)} = [(H_m^2\Gamma_m + 1)\epsilon_m]/[2i\gamma_m k_1 a H_m(1 - \Gamma_m)]. \tag{2.24}$$

By eliminating $g^{(1)}(\mu)$ in (2.18) and (2.19) and incorporating the definition (2.22) in the result we get

$$D_m^{+} = \frac{\epsilon_m k_0^2}{2k_1 a\pi}\sum_{n=0}^{\infty}I_{mn}^{(1)}\gamma_n D_n^{-}, \tag{2.25}$$

where the matrix element $I_{mn}^{(1)}$ is given by

$$I_{mn}^{(1)} = \int_0^\infty \frac{d\mu}{\beta(\mu)}\int_{-a}^{a}ds\int_{-a}^{a}\cos\mu s \cos\mu s' \cos\beta_n k_1 s \times \cos\beta_m k_1 s' \, ds'. \tag{2.26}$$

Similarly, eliminating $g^{(2)}(\mu)$ from (2.20) and (2.21) results in the following matrix equation for the TE mode amplitudes:

$$D_m^{-} = \frac{\epsilon_m}{2k_1 a\gamma_m\pi}\sum_{n=0}^{\infty}I_{mn}^{(2)}D_n^{+}, \tag{2.27}$$

where

$$I_{mn}^{(2)} = \int_0^\infty d\mu\beta(\mu)\int_{-a}^{a}ds\int_{-a}^{a}\cos\mu s \cos\mu s' \cos\beta_n k_1 s \times \cos\beta_m k_1 s' \, ds'. \tag{2.28}$$

Finally, the matrix equations (2.25) and (2.27) can be written in terms of D_n^- and D_n^+, respectively, through use of (2.23). That is,

$$\sum_{n=0}^{\infty} L_{mn}^{(1)} D_n^- = F_m^{(1)}, \quad \text{for TM waves,} \tag{2.29}$$

where

$$L_{mn}^{(1)} = \frac{\epsilon_m k_0^2 \gamma_n}{2k_1 a\pi} I_{mn}^{(1)} + \frac{1+\Gamma_m}{1-\Gamma_m}\delta_{mn} \tag{2.30}$$

and $F_m^{(1)}$ is given by (2.24).

For the TE case similar steps lead to

$$\sum_{n=0}^{\infty} L_{mn}^{(2)} D_n^+ = F_m^{(2)}, \tag{2.31}$$

where

$$L_{mn}^{(2)} = \frac{\epsilon_m}{2k_1 a\gamma_m \pi} I_{mn}^{(2)} + \frac{1-\Gamma_m}{1+\Gamma_m}\delta_{mn}$$

and

$$F_m^{(2)} = \frac{(H_m^2\Gamma_m + 1)\epsilon_m}{2i\gamma_m k_1 a H_m (1+\Gamma_m)}. \tag{2.32}$$

Let us consider $I_{mn}^{(1)}$, the TM element defined by (2.26). A special form of the plane wave representation of the Hankel function of order zero is needed. The most general form is written

$$H_0^{(1)}(k_0 R) = \frac{1}{\pi}\int_C \frac{e^{ixt + i\beta(t)z}}{\beta(t)}\,dt, \qquad z > 0, \tag{2.33}$$

where $\mathrm{Im}[\beta(t)] \geq 0$ and $R = (x^2 + z^2)^{1/2}$.

The contour C is depicted in Fig. 2. Making the substitution $t = k_0 \sin\varphi$, we find that the right-hand side of (2.33) can be written

$$\frac{1}{\pi}\int_{C'} e^{ik_0 R\cos(\varphi-\psi)}\,d\varphi, \qquad \text{where } \psi = \tan^{-1}(x/z),$$

where the contour C', which is a simple open curve, comes in from $i\infty$ in the complex φ plane in the strip $\psi + \pi/2 < \mathrm{Re}(\varphi) < 3\pi/2 + \psi$ and goes to $-i\infty$ in the strip $\psi - \pi/2 < \mathrm{Re}(\varphi) < \psi + \pi/2$. By choosing $\mathrm{Re}(\varphi) = 3\pi/2 + \psi$ in the former and $\mathrm{Re}(\varphi) = \pi/2 + \psi$ in the later, we find that (2.33) is equivalent to a commonly used integral representation of the zero order Hankel function first given by Sommerfeld.[16]

Note that, in (2.33), the branch points $t = \pm k_0$ must be avoided as shown in Fig. 2 in order that the radiation condition for $z > 0$ be fulfilled. That is,

$$\mathrm{Im}[\beta(\mu)] \geq 0.$$

With this convention we see that

$$\int_C \frac{e^{i\mu x}}{\beta(\mu)}\,d\mu = \int_C \frac{\cos\mu x}{\beta(\mu)}\,d\mu = \pi H_0^{(1)}(k_0|x|). \tag{2.34}$$

The absolute value sign within the Hankel function in (2.34) is needed to insure equality for $x < 0$. The result (2.34) is physically reasonable since $H_0^{(1)}$ is the free space Green's function for our two-dimensional problem and it must display symmetry in the $x = 0$ plane. Via (2.34) the infinite integration in (2.26) can be performed with the result

$$I_{mn}^{(1)} = \tfrac{1}{2}\pi \int_{-a}^{a} ds \int_{-a}^{a} \cos\beta_m k_1 s' \cos\beta_n k_1 s\, H_0^{(1)}(k_0|s + s'|)ds'. \tag{2.35}$$

If we change to sum and difference coordinates

$$u = s + s', \qquad v = s - s',$$

we find that

$$I_{mn}^{(1)} = \tfrac{1}{4}\pi\left[\int_{A_1}\int q_{mn}(u,v)dudv + \int_{A_2}\int q_{mn}(u,v)dudv\right], \tag{2.36}$$

where

$$q_{mn}(u,v) = (\cos\{\beta_m k_1[(u-v)/2)\}\cos\{\beta_n k_1[(u+v)/2]\}) \times H_0^{(1)}(k_0|u|).$$

The areas A_1 and A_2 are depicted in Fig. 3 and thus may be parameterized:

$$\int_{A_1}\int dudv = \int_0^{-2a} du \int_{-u-2a}^{u+2a} dv, \quad \int_{A_2}\int dudv = \int_0^{2a} du \int_{u-2a}^{-u+2a} dv.$$

The result of elementary integration over v and trigonometric reduction and the substitution $u =$ at is

$$I_{mn}^{(1)} = \frac{-(-1)^{m+n}a^2}{m^2 - n^2}\int_0^2 H_0^{(1)}(\rho t)\,[m\sin(m\pi t) - n\sin(n\pi t)]\,dt \quad \text{where } \rho = k_0 a \tag{2.37}$$

and as limiting cases of (2.37)

$$I_{nn}^{(1)} = -\tfrac{1}{2}\pi a^2 \int_0^2 t H_0^{(1)}(\rho t)[\cos(n\pi t) + \mathrm{sinc}(n\pi t)]dt, \tag{2.38}$$

$$I_{00}^{(1)} = -\pi a^2 \int_0^2 t H_0^{(1)}(\rho t)dt,$$

where $\mathrm{sinc}(x) = \sin(x)/x$.

To simplify notation, it is convenient to define the function $h_m(\rho)$:

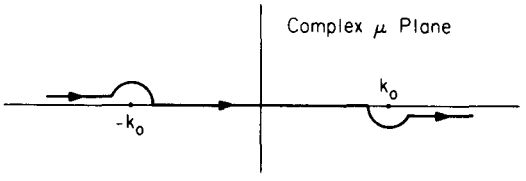


FIG. 2. The contour C.

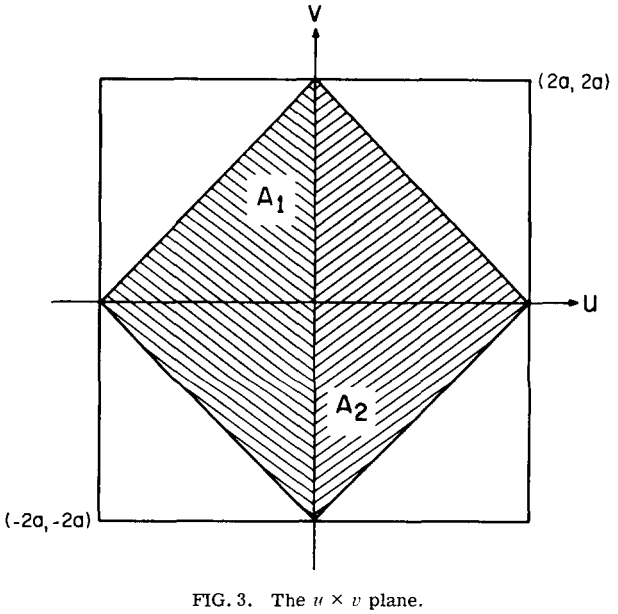


FIG. 3. The $u \times v$ plane.

$$h_m(\rho) = \int_0^2 H_0^{(1)}(\rho t)\sin(m\pi t)dt. \tag{2.39}$$

In terms of the canonical integral $h_m(\rho)$, the nondiagonal elements $I_{mn}^{(1)}$ is written

$$I_{mn}^{(1)} = -\{[(-1)^{m+n}a^2]/(m^2-n^2)\}[mh_m(\rho) - nh_n(\rho)]. \tag{2.40}$$

It is interesting to note that, to within notational differences, the matrix element corresponding to $I_{mn}^{(1)}$ as obtained by Nussenzveig[7] agrees with ours.

It is not possible to carry out the infinite integration over μ first in the case of the TE matrix element. This is because the integral μ is not absolutely convergent and hence the order of integration is important. What can be done is to perform the two integrations over the finite limits. Upon doing this and making the substitution $\mu = at$, (2.28) becomes

$$I_{mn}^{(2)} = 2\alpha_n\alpha_m(-1)^{n+m}\int_{-\infty}^{+\infty}\frac{(\rho^2-t^2)^{1/2}\cos^2 t}{(\alpha_n^2-t^2)(\alpha_m^2-t^2)}dt, \tag{2.41}$$

where $\alpha_n = \frac{1}{2}(2n+1)\pi$, $\rho = k_0 a$.

In order to utilize the integral representation of the zeroth-order Hankel function (2.34) in (2.41), and also to enable us to use partial fraction techniques such that each term of the partial fraction decomposition is a convergent integral, we rationalize the numerator.

Then, as usual, we express

$$\frac{\rho^2-t^2}{(\alpha_n^2-t^2)(\alpha_m^2-t^2)} = \sum_{i=1}^{4} A_i\frac{1}{\lambda_i - t}, \tag{2.42}$$

where

$$\lambda_1 = \alpha_n, \qquad A_1 = -A_2 = \frac{\rho^2-\alpha_n^2}{2\alpha_n(\alpha_m^2-\alpha_n^2)},$$
$$\lambda_2 = -\alpha_n,$$
$$\lambda_3 = \alpha_m$$
$$\lambda_4 = -\alpha_m, \qquad A_3 = -A_4 = -\frac{\rho^2-\alpha_m^2}{2\alpha_m(\alpha_m^2-\alpha_n^2)}. \tag{2.43}$$

Thus combining (2.41) and (2.42) gives

$$I_{mn}^{(2)} = \sum_{j=1}^{4} A_j' I(\lambda_j), \tag{2.44}$$

where

$$A_j' = 2\alpha_n\alpha_m(-1)^{m+n}A_j \tag{2.45}$$

and

$$I(\lambda) = \int_{-\infty}^{+\infty}\frac{\cos^2 t\,dt}{(\rho^2-t^2)^{1/2}(\lambda-t)}. \tag{2.46}$$

By inspection of (2.46) we see that

$$I(-\lambda) = -I(\lambda). \tag{2.47}$$

Thus, using this symmetry along with the symmetries given in (2.43), we obtain

$$I_{mn}^{(2)} = 2A_1' I(\lambda_1) + 2A_3' I(\lambda_3). \tag{2.48}$$

The integral $I(\lambda)$, defined by (2.46), can be written as a linear combination of the canonical integral defined by (2.39), where $m \to m + \frac{1}{2}$. This identification may be made by considering $I(\lambda)$ as a Fourier convolution integral, as discussed for example by Goldberg.[17] We have

$$I(\lambda) = \int_{-\infty}^{+\infty} f(t-\lambda)g(t)dt \equiv f * g \tag{2.49}$$

$$= \frac{1}{2\pi}\int_{-\infty}^{+\infty} e^{-i\omega\lambda}\hat{f}(\omega)\hat{g}(\omega)d\omega, \tag{2.50}$$

where

$$f(t) = 1/t, \qquad g(t) = \cos^2 t/(\rho^2-t^2)^{1/2} \tag{2.51}$$

and $\hat{f}, \hat{g}$ are the Fourier transforms of f and g, i.e.,

$$\hat{f}(\omega) = \int_{-\infty}^{+\infty} e^{i\omega t}f(t)dt.$$

The Fourier transform of $1/t$ is related to the Heaviside unit step function

$$u(\omega) = \begin{cases} 1, & \omega > 0 \\ 0, & \omega < 0 \end{cases}$$

by

$$\hat{f}(\omega) = 2\pi i u(\omega) - \pi i. \tag{2.52}$$

Thus from (2.50), (2.51), and (2.52) we get

$$I(\lambda) = i\int_{-\infty}^{+\infty} e^{-i\omega\lambda}[u(\omega) - \tfrac{1}{2}]d\omega \int_{-\infty}^{+\infty} e^{i\omega t}\frac{\cos^2 t}{(\rho^2-t^2)^{1/2}}dt$$

or

$$I(\lambda) = i\int_0^{\infty} e^{-i\omega\lambda}d\omega\int_{-\infty}^{+\infty} e^{i\omega t}\frac{\cos^2 t}{(\rho^2-t^2)^{1/2}}dt - i\int_{-\infty}^{+\infty}\frac{\delta(t-\lambda)\cos^2 t}{(\rho^2-t^2)^{1/2}}dt. \tag{2.53}$$

Since λ is an odd-half integer multiple of π, the second term in (2.53) does not contribute. Hence

$$I(\lambda) = \tfrac{1}{2}i\int_0^{\infty} e^{-i\omega\lambda}d\omega\int_{-\infty}^{+\infty}\frac{\cos\omega t}{(\rho^2-t^2)^{1/2}}(1+\cos 2t)dt. \tag{2.54}$$

The form of the integral formula (2.34) to be used in (2.54) is

$$\int_{-\infty}^{+\infty}\frac{\cos\mu x}{(\rho^2-\mu^2)^{1/2}}d\mu = \pi H_0^{(1)}(\rho x), \qquad x > 0. \tag{2.55}$$

To use (2.55), we express the cosine products in the integrand as a sum of cosines, and interrupt the integration over ω such that the cosine arguments are positive for all values of the integration variable t, to obtain

$$I(\lambda) = \tfrac{1}{4}\pi i\Big(2\int_0^{\infty} e^{-i\omega\lambda}H_0^{(1)}(\rho\omega)d\omega + \int_0^{\infty} e^{-i\omega\lambda}H_0^{(1)}(\rho(\omega+2))d\omega + \int_0^{2} e^{-i\omega\lambda}H_0^{(1)}(\rho(2-\omega))d\omega + \int_2^{\infty} e^{-i\omega\lambda}H_0^{(1)}(\rho(\omega-2))d\omega\Big),$$

which simplifies to

$$I(\lambda) = \tfrac{1}{2}\pi\int_0^2 \sin(\lambda t)H_0^{(1)}(\rho t)dt.$$

Thus, according to (2.48), $I_{mn}^{(2)}$ becomes

$$I_{mn}^{(2)} = \frac{\pi(-1)^{m+n}}{\alpha_m^2-\alpha_n^2}[\alpha_m(\rho^2-\alpha_n^2)h_{m+1/2}(\rho) - \alpha_n(\rho^2-\alpha_m^2)h_{m+1/2}(\rho)], \text{ where } \alpha_n = (2n+1)\pi/2. \tag{2.56}$$

In (2.56) we have used the notation given by (2.39). The diagonal elements are given by

$$I_{nn}^{(2)} = \frac{\pi(\rho^2 - \alpha_n^2)}{2\alpha_n} \int_0^2 (\sin\alpha_n t - t\cos\alpha_n t) H_0^{(1)}(\rho t) dt. \tag{2.57}$$

This completes the analysis of the integral forms of the matrix elements. In anticipation of the work to follow, it is important to note that, for either polarization, $\lim_{n\to\infty} I_{mn}^{(j)} = 0$ by the Riemann–Lebesgue theorem as, for example, discussed by Widder.[18] This is a necessary condition for $I_{mn}^{(j)}$ to be diagonally dominant.

Continuing on with the process of rendering the equations of the form (2.29) amenable to digital computation, we obtain an asymptotic series for the canonical integral $h_m(\rho)$ defined by (2.39). We write the integral as

$$h_n(\rho) = \int_0^\infty H_0^{(1)}(\rho t)\, \sin n\pi t\, dt - \int_2^\infty H_0^{(1)}(\rho t)\, \sin n\pi t\, dt. \tag{2.58}$$

The infinite integration can be carried out in closed form[19]:

$$\int_0^\infty H_0^{(1)}(\rho t)\, \sin(n\pi t) dt = \frac{1 + (2i/\pi)\ln\{n\pi/\rho - [(n\pi/\rho)^2 - 1]^{1/2}\}}{[(n\pi)^2 - \rho^2]^{1/2}} \quad \text{for } n\pi > \rho. \tag{2.59}$$

The remaining term can be approximately evaluated by replacing $H_0^{(1)}$ within the integrand by the first four terms of its asymptotic expansion,[20]

$$H_0^{(1)}(z) \sim \sqrt{\frac{2}{\pi}}\, e^{i(z-\pi/4)} \sum_{k=0}^{3} \frac{a_k}{z^{k+1/2}}, \tag{2.60}$$

where $a_0 = 1$, $a_1 = -i/8$, $a_2 = 9/128$, and $a_3 = 75i/1024$.

The integrations of the terms in (2.60) are expressible as incomplete gamma functions. Since we are assuming $n\pi$ to be large these functions may be approximated also by their asymptotic series. Hence one finds to the orders indicated

$$\int_2^\infty H_0^{(1)}(\rho t)\sin(n\pi t) dt \sim n\sqrt{\frac{\pi}{\rho}}\, \frac{\exp\{i[2(\rho - n) - \tfrac{1}{4}\pi]\}}{(n\pi)^2 - \rho^2} \times \left\{ \sum_{k=0}^{3} \frac{a_k}{(2\rho)^k} \left[\left(1 + \frac{i(k + \tfrac{1}{2})\rho}{(n\pi)^2 - \rho^2}\right) + o\left(\frac{1}{\rho^4}\right) \right] \right\} + o\left(\frac{1}{n^5}\right). \tag{2.61}$$

Thus (2.59) and (2.61) approximate $h_n(\rho)$ in the regime $1 < \rho < n\pi$.

Finally we conclude this section with the asymptotic evaluation of the mode amplitudes for large index. These results will be used to correct the truncated systems in the following section. We then will be prepared to estimate the error incurred in truncating the infinite system (2.29). According to Meixner,[21] the field component $E_x(x, 0)$ in the aperture plane has the singularity

$$E_x(x, 0) \underset{|x|\to a}{\sim} A_0/(a^2 - x^2)^{1/3}, \tag{2.62}$$

where we have accounted for the obvious symmetry in the x coordinate. Equivalently, the edge singularity (2.62) can be obtained from Oberhettinger's[22] eigenfunction solution to the field in the presence of a perfectly conducting wedge. It should be noted that the constant A_0 in (2.62) depends upon the excitation, i.e., the form of the incident wave as well as the value of $k_0 a$. If we Fourier analyze the right-hand side of (2.62), we find

$$\frac{A_0}{(a^2 - x^2)^{1/3}} = A_0 \sum_{n=0}^{\infty} a_n \cos\left(\frac{n\pi x}{a}\right). \tag{2.63}$$

It is not difficult to show that the coefficient a_n is given explicitly by

$$a_n = [J_{1/6}(n\pi)\pi^{1/2}\Gamma(\tfrac{2}{3})]/[a^{2/3}(n\pi/2)^{1/6}]. \tag{2.64}$$

In (2.64), $J_{1/6}$ is the Bessel function of the first kind of order 1/6. This completes Sec. 2. We are now ready to discuss the approximate solution to the infinite system.

3. THE TRUNCATED EQUATIONS, RADIATION FIELDS, AND SOME NUMERICAL RESULTS

We now consider only the TM case. Similar results apply for the TE case. As a means of solving the system (2.29), we truncate it at some finite order N. To estimate the error in so doing, we consider a percent error defined as

$$P_n^{(N)} = 100.\, |(D_n^- - \tilde{D}_n^-)/D_n^-|, \tag{3.1}$$

where

$$\sum_{n=0}^{N-1} L_{mn}\tilde{D}_n^- = F_m, \tag{3.2}$$

$$\sum_{n=0}^{N-1} L_{mn}D_n^- = F_m + e_m, \tag{3.3}$$

and

$$e_m = -\sum_{n=N}^{\infty} L_{mn}D_n^-. \tag{3.4}$$

Note that since we will be discussing only the TM case, the polarization superscript is deleted.

The singularity (2.62) in the component E_x at the edges implies that

$$\lim_{n\to\infty} \gamma_n D_n^- = a_n A_0, \tag{3.5}$$

where a_n is given by (2.64) and A_0 as discussed after (2.62) is a constant which depends upon the incident wave form. This is a direct consequence of a theorem due to Erdélyi.[23] From (3.5) then it follows that

$$D_n^- \approx \frac{(-1)^n \alpha^{(N)}}{n^{5/3}} + o(n^{-7/3}), \quad \text{where } \alpha^{(N)} = n^{5/3}(-1)^n \tilde{D}_n^-,\ n = N - 1. \tag{3.6}$$

and the error term $o(n^{-7/3})$ arises from terms of order $o(r^{1/3})$ in the expansion of the field components perpendicular to the edges.

Explicitly, using (3.6) and (3.4) and with the definitions (2.30) and (2.40), we have

$$e_m \approx \frac{\alpha^{(N)}\epsilon_m(k_0 a)^2(-1)^m}{2\pi(k_1 a)} \sum_{n=N}^{\infty} \gamma_n \frac{[mh_m(\rho) - nh_n(\rho)]}{(m^2 - n^2)n^{5/3}}, \quad m = 0, 1, \ldots, N - 1. \tag{3.7}$$

Note that $h_m(\rho)$, $m = 0, 1, \ldots, N-1$, must be calculated to solve for the first iteration coefficients $\tilde{D}_n^-$, while, for moderate and large values of N, $h_n(\rho)$, $n = N, N+1, \cdots$, can be calculated using the asymptotic series (2.61) for $h_n(\rho)$. Hence the error term e_m given by (3.7) is readily calculated.

Let us now consider the improvement and also indicate the convergence of corrected equations for the aperture amplitudes D_n^-. Since the aperture fields are a sensitive indicator of the effect of truncation, we illustrate our results in terms of them. Specifically we consider the magnetic intensity component $\varphi_{\text{int}}(x, z) = H_y(x, z)_{z=0^-}$ in the aperture. Thus from (2.2a) we have

$$H_y(x, 0^-) = \sum_{n=0}^{\infty} D_n^+ \cos(k_1 \beta_n x), \qquad |x| < a. \tag{3.8}$$

Furthermore, in order that we can most easily interpret our results, we also make the simplifying assumptions that $k_1 = k_0$ and $d \to \infty$. A letter by the author[8] displayed the convergence of the truncated system (3.2) as a function of N. Figures 4, 6, 8 show the aperture field (3.8) for $k_0 h = 1$ and $k_0 a = 0.2$, 3, and 5. Figure 10 differs from 8 only in that $k_0 h = 30$. In each figure we are comparing the absolute value of the aperture field (3.8) where the amplitudes $\tilde{D}_n^-$ are calculated in the first case by (3.2) and in the second by (3.3). In the latter case a total of 200 modes are summed to approximate e_m. Although the results obtained show that the minimum values of the graphs of the truncated and corrected aperture fields are almost identical, we have displaced the truncated fields downward slightly so that the two fields may be better compared.

Figures 4, 6, 8, 10 demonstrate that the truncation correction factor e_m is a small correction to the matrix Eq. (3.3). This, of course, was implicitly assumed from the outset and is now seen to be a good assumption.

The normalized radiation patterns as shown in Figs. 5, 7, 9, 11 correspond to the aperture fields in Figs. 4, 6, 8, 10. The pattern plots are the normalized power plots of $|E_\theta|^2/|\max(E_\theta)|^2$ where $\theta = \tan^{-1}(x/z)$ in reference to Fig. 1.

The radiation field is obtained by substituting the Fourier coefficient $g^{(1)}(\mu)$ defined by (2.18) into (2.12), getting

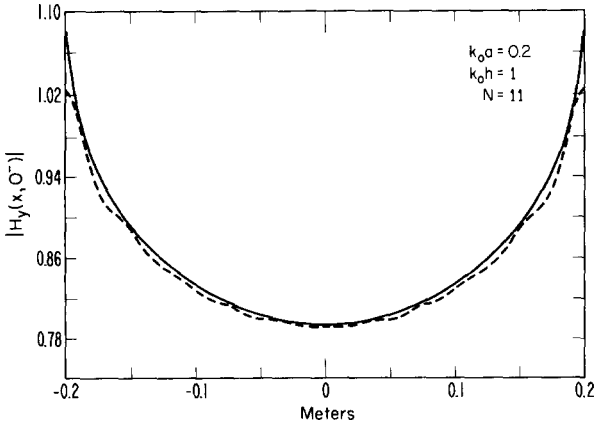


FIG. 4. Aperture field.

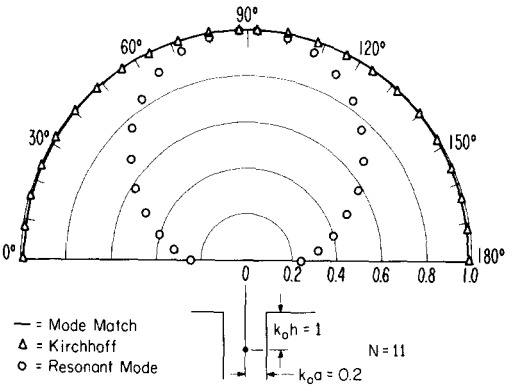


FIG. 5. Radiation patterns.

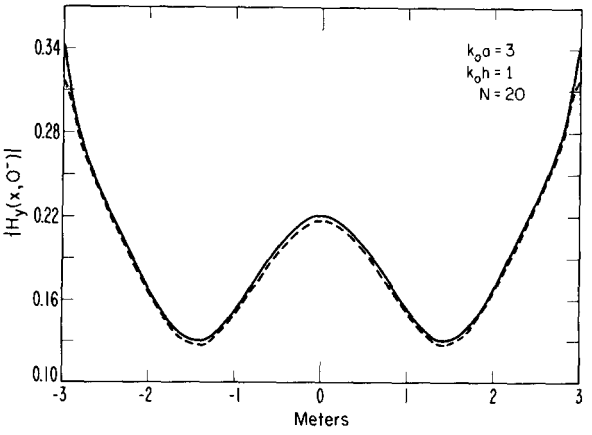


FIG. 6. Aperture field.

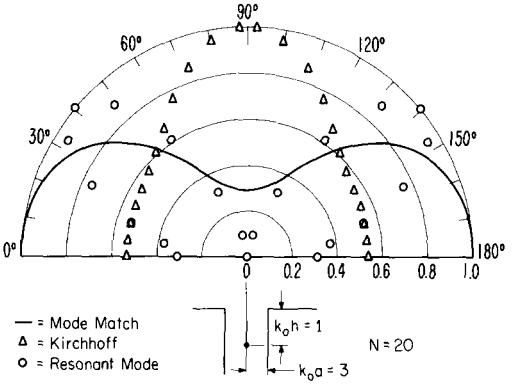


FIG. 7. Radiation patterns.

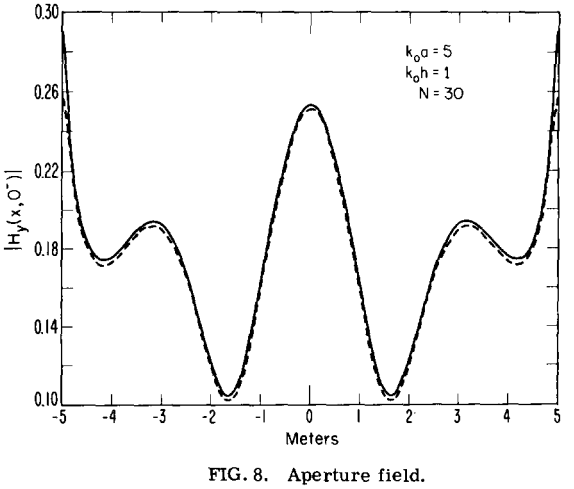


FIG. 8. Aperture field.

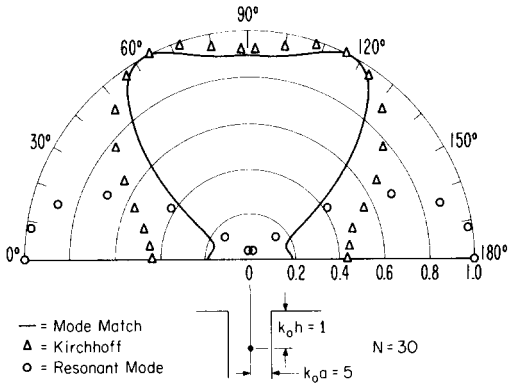


FIG. 9. Radiation patterns.

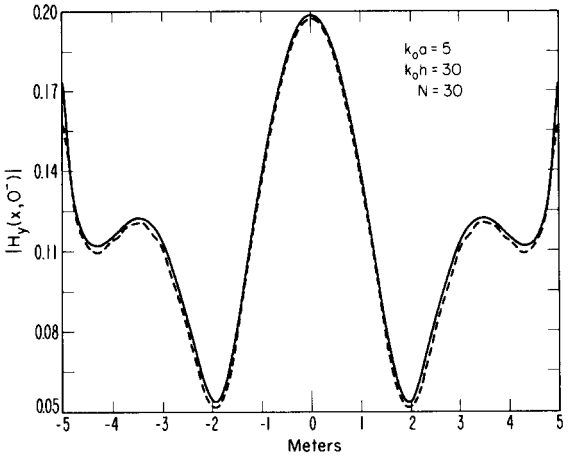


FIG. 10. Aperture field.

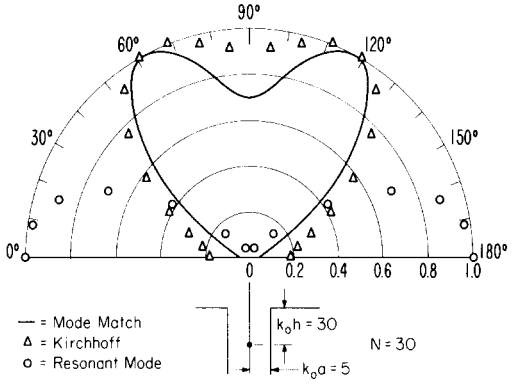


FIG. 11. Radiation patterns.

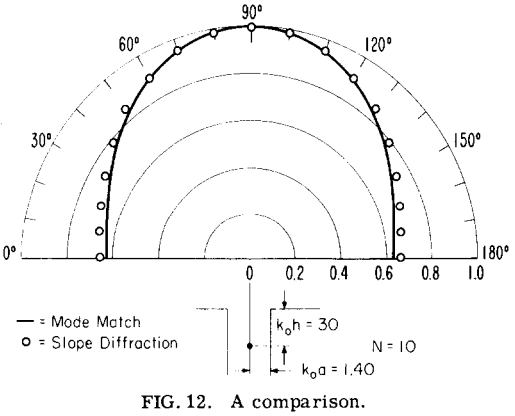


FIG. 12. A comparison.

$$\varphi_{\text{ext}}(x,z) = \frac{k_0^2}{k_1\pi}\sum_{n=0}^{\infty}\gamma_n D_n^- Q_n(x,z), \tag{3.9}$$

where, upon performing the integration over μ, we obtain

$$Q_n(x,z) = \tfrac{1}{4}\pi\int_{-a}^{+a}\cos(\beta_n k_1 s)[H_0^{(1)}(k_0R^+) + H_0^{(1)}(k_0R^-)]ds \tag{3.10}$$

and where $R^{\pm} = [(x \pm s)^2 + z^2]^{1/2}$.

In the radiation zone, i.e., for $k_0R^{\pm} \gg 1$, one finds that

$$\varphi_{\text{ext}} \equiv H_y(x,z) \sim (-k_0/k_1)(k_0a)H_0^{(1)}(k_0R) \times \sum_{n=0}^{\infty}\frac{\gamma_n D_n^-(-1)^n\Theta\sin\Theta}{(n\pi)^2 - \Theta^2} \tag{3.11}$$

where $\Theta = k_0a\sin\theta$.

It can be shown that in the radiation zone

$$E_\theta = (i/\omega\epsilon_0)\hat{i}_\theta\cdot(\bar{\nabla}\times\bar{H}) \sim (\mu_0/\epsilon_0)^{1/2}H_y, \tag{3.12}$$

where $\hat{i}_\theta$ is the unit vector in the direction of increasing θ.

Calculations of radiation from the open end of waveguide antennas are usually based upon approximate methods originally devised by Kirchhoff. Although these methods are highly versatile, it is not clear that they should be applicable to configurations in which the waveguide width is on the order of one wavelength. Thus it is of interest to compare our results with those obtained with the Kirchhoff approximation. Within the present formalism, the most direct route to the Kirchhoff radiation pattern is to note that, in this approximation, the aperture amplitude D_n^- to be used in (3.11), is given by assuming the parallel plate region extends to $z = +\infty$, and hence, in the notation (2.2a), $\hat{\Gamma}_n = 0$. This gives

$$D_n^{(kir)^-} \equiv \epsilon^{ik_0\gamma_n h}. \tag{3.13}$$

In Figs. 5, 7, 9, 11, the solid curve corresponds to the normalized $|E_\theta|^2$ pattern obtained from (3.12); the triangular symbols correspond to the same quantity with the coefficients defined by (3.13), while the circle symbols correspond to the same quantity obtained from (3.11) in which only the term which most nearly satisfies

$$\text{Re}(k_1a) = n\pi \tag{3.14}$$

in the sum is retained. This is the resonant mode contribution to the radiation field.

The discrepancies between the matrix solution (the solid curve) and the Kirchhoff solution (triangles) in the lit region stem from the fact that the more rigorous solution predicts a maximum value of the coefficient D_n^- for the n which most nearly satisfies (3.14), while the Kirchhoff predicted values satisfy (3.13). For narrow apertures, in which only the TEM mode ($n = 0$) propagates, we would then expect good agreement since both theories would yield D_0^- a maximum. This narrow aperture limit is shown in Fig. 5. Figure 7 was chosen near resonance $k_0a = 3 \sim \pi$. The figure demonstrates the inability of Kirchhoff theory to account for resonance. Figures 8 and 10

demonstrate the minor effects, except for the total aperture amplitude, of the position of the line source in the central plane. In all the radiation patterns, the solutions in the low angle ($|\theta| \lesssim \pi/2$) regions show discrepancy. This is a consequence of the inability of the Kirchhoff predicted fields to properly describe the edge diffracted components of the radiation. That is, $D_n^{(kir)}$ as given by (3.13) displays exponential behavior for large n while D_n^- according to (3.5) shows an algebraic dependence. These radiation conclusions are in agreement with the analysis of the unflanged cases, treated in an elegant fashion by Weinstein.[4]

To conclude our discussion of radiation patterns, we present a comparison of our results with a slope diffraction analysis obtained by Rudduck and Wu.[24] Their configuration consists of a guide width of $k_0 a = 1.40$ with a TEM illumination of the aperture. Since this width prohibits propagation of all higher-order modes, we take $k_0 h$ to be large (= 30.0) for the purpose of comparison. The results are displayed in Fig. 12.

4. CONCLUDING REMARKS

We have shown that it is possible to treat the radiation from an open-ended structure terminating in a flange plane in a formally exact and straightforward manner. In particular we have shown that the resultant infinite system of equations can closely be approximated by a finite system provided one accounts for truncation. With notable exceptions, namely near resonance and in the low observation angle regions, Kirchhoff theory is shown to correctly predict radiation patterns from such configurations. Since the nature of the formulation required that rather large systems of linear equations be considered, the matrix elements were analytically reduced to simple form. Thus the elements were factored into canonical integrals which depended on only one matrix index. This enabled an $N \times N$ matrix to be generated from $2N$ integrals. An asymptotic series in mode index was developed to provide an efficient means of obtaining the truncation correction terms.

Finally, it should be noted that our technique can be applied to similar but more complicated structures. For example, preliminary results on the analogous analysis for a circular flanged guide in which the flange boundary condition is relaxed to account for a more general impedance condition have been obtained.[25]

ACKNOWLEDGMENTS

The author acknowledges Dr. J. R. Wait for important comments and suggestions, and gives thanks to Dr. G. Hufford for helpful discussion.

---

* The material herein is excerpted from a chapter of the author's dissertation to be submitted to the Department of Physics at the University of Colorado.

1 A. Wexler, IEEE Trans. Microwave Theory Tech. **17**, 416 (1969).
2 P. H. Mastermann and P. J. B. Clarricoats, Proc. IEE **118**, 51 (1971).
3 A. Wexler, IEEE Trans. Microwave Theory Tech. **15**, 508 (1967).
4 L. Weinstein, *The Theory of Diffraction and the Factorization Method* (Golem, Boulder, Colorado, 1969), Chap. 1.
5 R. Mittra and S. W. Lee, *Analytical Techniques in the Theory of Guided Waves* (Macmillan, New York, 1971).
6 H. Y. Lee, L. B. Felsen, and J. B. Keller, SIAM J. Appl. Math. **16**, 268 (1968).
7 H. M. Nussenzveig, Phil. Trans. Roy. Soc. London **A252**, 1 (1959).
8 A. Q. Howard, Electron. Lett. **7**, Nos. 5/6, 129 (1971).
9 C. M. Do Amaral and J. W. Bautista Vidal, Appl. Sci. Res. B **11**, 1 (1964).
10 The scatter matrix $L_{mn}^{(j)}$ for the problem under consideration is defined by (2.30) or (2.31) as $j = 1$ or 2.
11 The property of diagonal dominance is discussed by Ref. 1.
12 J. A. Stratton, *Electromagnetic Theory* (McGraw-Hill, New York, 1941), pp. 349–51.
13 J. R. Wait, Radio Sci. **68D**, 81 (1964).
14 P. R. Halmos, *Measure Theory* (Van Nostrand, New York, 1950), p. 110.
15 R. V. Churchill, *Fourier Series and Boundary Value Problems* (McGraw-Hill, New York, 1963), p. 104.
16 A. Sommerfeld, Math. Ann. **47**, 335 (1896).
17 R. R. Goldberg, *Fourier Transforms* (Cambridge U. P., London, 1961), pp. 18–20.
18 D. V. Widder, *Advanced Calculus* (Prentice-Hall, Engelwood Cliffs, N.J., 1961), 2nd ed., p. 408.
19 A. D. Wheelon, *Tables of Summable Series and Integrals Involving Bessel Functions* (Holden-Day, San Francisco, 1968).
20 M. Abramowitz and I. Stegun, *Handbook of Mathematical Functions* (U.S. Govt. Printing Office, Washington, D.C., 1965), p. 364.
21 J. Meixner, NYU Res. Report No. EM-72 (1954).
22 F. Oberhettinger, Commun. Pure Appl. Math. **7**, 551 (1954).
23 A. Erdelyi, *Asymptotic Expansions* (Dover, New York, 1956).
24 R. C. Rudduck and B. C. F. Wu, IEEE Trans. Antennas Propagation **17**, No. 6 (1969).
25 A. Q. Howard, ITS Tech. Memo. OT-TM 46, Boulder, Colorado (1971).

On Uniqueness of the Kerr–Newman Black Holes*

Robert Wald†
Joseph Henry Laboratories, Princeton University, Princeton, New Jersey 08540
(Received 21 June 1971; Revised Manuscript Received 6 October 1971)

It is proven that the Kerr–Newman space-times with $e^2 + a^2 < m^2$ are the only electrovac black hole solutions of Einstein's equations which can be obtained by analytic variation of the space-time geometry starting from the Schwarzschild solution.

1. INTRODUCTION

In a recent paper,[1] the author has described a new theorem which leads one to believe that the final state of general (nonspherical) gravitational collapse is a Kerr–Newman black hole.[2–5] The present paper presents the detailed statement and proof of this theorem.

It is widely believed that the complete gravitational collapse of a body results in the production of a black hole as opposed to a "naked singularity." (The physical arguments leading to this conjecture are summarized in Ref. 1.) Thus, it is of great interest to find all the solutions of Einstein's equations which describe black holes, since, if this conjecture is

demonstrate the minor effects, except for the total aperture amplitude, of the position of the line source in the central plane. In all the radiation patterns, the solutions in the low angle ($|\theta| \lesssim \pi/2$) regions show discrepancy. This is a consequence of the inability of the Kirchhoff predicted fields to properly describe the edge diffracted components of the radiation. That is, $D_n^{(kir)}$ as given by (3.13) displays exponential behavior for large n while D_n^- according to (3.5) shows an algebraic dependence. These radiation conclusions are in agreement with the analysis of the unflanged cases, treated in an elegant fashion by Weinstein.[4]

To conclude our discussion of radiation patterns, we present a comparison of our results with a slope diffraction analysis obtained by Rudduck and Wu.[24] Their configuration consists of a guide width of $k_0 a = 1.40$ with a TEM illumination of the aperture. Since this width prohibits propagation of all higher-order modes, we take $k_0 h$ to be large (= 30.0) for the purpose of comparison. The results are displayed in Fig. 12.

4. CONCLUDING REMARKS

We have shown that it is possible to treat the radiation from an open-ended structure terminating in a flange plane in a formally exact and straightforward manner. In particular we have shown that the resultant infinite system of equations can closely be approximated by a finite system provided one accounts for truncation. With notable exceptions, namely near resonance and in the low observation angle regions, Kirchhoff theory is shown to correctly predict radiation patterns from such configurations. Since the nature of the formulation required that rather large systems of linear equations be considered, the matrix elements were analytically reduced to simple form. Thus the elements were factored into canonical integrals which depended on only one matrix index. This enabled an $N \times N$ matrix to be generated from $2N$ integrals. An asymptotic series in mode index was developed to provide an efficient means of obtaining the truncation correction terms.

Finally, it should be noted that our technique can be applied to similar but more complicated structures. For example, preliminary results on the analogous analysis for a circular flanged guide in which the flange boundary condition is relaxed to account for a more general impedance condition have been obtained.[25]

ACKNOWLEDGMENTS

The author acknowledges Dr. J. R. Wait for important comments and suggestions, and gives thanks to Dr. G. Hufford for helpful discussion.

---

* The material herein is excerpted from a chapter of the author's dissertation to be submitted to the Department of Physics at the University of Colorado.

1 A. Wexler, IEEE Trans. Microwave Theory Tech. **17**, 416 (1969).
2 P. H. Mastermann and P. J. B. Clarricoats, Proc. IEE **118**, 51 (1971).
3 A. Wexler, IEEE Trans. Microwave Theory Tech. **15**, 508 (1967).
4 L. Weinstein, *The Theory of Diffraction and the Factorization Method* (Golem, Boulder, Colorado, 1969), Chap. 1.
5 R. Mittra and S. W. Lee, *Analytical Techniques in the Theory of Guided Waves* (Macmillan, New York, 1971).
6 H. Y. Lee, L. B. Felsen, and J. B. Keller, SIAM J. Appl. Math. **16**, 268 (1968).
7 H. M. Nussenzveig, Phil. Trans. Roy. Soc. London **A252**, 1 (1959).
8 A. Q. Howard, Electron. Lett. **7**, Nos. 5/6, 129 (1971).
9 C. M. Do Amaral and J. W. Bautista Vidal, Appl. Sci. Res. B **11**, 1 (1964).
10 The scatter matrix $L_{mn}^{(j)}$ for the problem under consideration is defined by (2.30) or (2.31) as $j = 1$ or 2.
11 The property of diagonal dominance is discussed by Ref. 1.
12 J. A. Stratton, *Electromagnetic Theory* (McGraw-Hill, New York, 1941), pp. 349–51.
13 J. R. Wait, Radio Sci. **68D**, 81 (1964).
14 P. R. Halmos, *Measure Theory* (Van Nostrand, New York, 1950), p. 110.
15 R. V. Churchill, *Fourier Series and Boundary Value Problems* (McGraw-Hill, New York, 1963), p. 104.
16 A. Sommerfeld, Math. Ann. **47**, 335 (1896).
17 R. R. Goldberg, *Fourier Transforms* (Cambridge U. P., London, 1961), pp. 18–20.
18 D. V. Widder, *Advanced Calculus* (Prentice-Hall, Engelwood Cliffs, N.J., 1961), 2nd ed., p. 408.
19 A. D. Wheelon, *Tables of Summable Series and Integrals Involving Bessel Functions* (Holden-Day, San Francisco, 1968).
20 M. Abramowitz and I. Stegun, *Handbook of Mathematical Functions* (U.S. Govt. Printing Office, Washington, D.C., 1965), p. 364.
21 J. Meixner, NYU Res. Report No. EM-72 (1954).
22 F. Oberhettinger, Commun. Pure Appl. Math. **7**, 551 (1954).
23 A. Erdelyi, *Asymptotic Expansions* (Dover, New York, 1956).
24 R. C. Rudduck and B. C. F. Wu, IEEE Trans. Antennas Propagation **17**, No. 6 (1969).
25 A. Q. Howard, ITS Tech. Memo. OT-TM 46, Boulder, Colorado (1971).

On Uniqueness of the Kerr–Newman Black Holes*

Robert Wald†
Joseph Henry Laboratories, Princeton University, Princeton, New Jersey 08540
(Received 21 June 1971; Revised Manuscript Received 6 October 1971)

It is proven that the Kerr–Newman space-times with $e^2 + a^2 < m^2$ are the only electrovac black hole solutions of Einstein's equations which can be obtained by analytic variation of the space-time geometry starting from the Schwarzschild solution.

1. INTRODUCTION

In a recent paper,[1] the author has described a new theorem which leads one to believe that the final state of general (nonspherical) gravitational collapse is a Kerr–Newman black hole.[2–5] The present paper presents the detailed statement and proof of this theorem.

It is widely believed that the complete gravitational collapse of a body results in the production of a black hole as opposed to a "naked singularity." (The physical arguments leading to this conjecture are summarized in Ref. 1.) Thus, it is of great interest to find all the solutions of Einstein's equations which describe black holes, since, if this conjecture is

correct, one would then have a complete and detailed knowledge of all the possible final states of gravitational collapse. The only known black hole solutions are the three-parameter family of Kerr–Newman space-times.[2–5] Theorems proven by Carter[6] and Israel[7] have led to the conjecture that they are the only black hole solutions. Israel proved that the Reissner–Nordström black holes (i.e., Kerr–Newman black holes with no angular momentum) are the only static, asymptotically flat, electrovac space-times with closed, simply connected surfaces of constant g_{00} which have a nonsingular event horizon $g_{00} = 0$. Carter considered axisymmetric black holes with no electromagnetic fields and proved that perturbations of these solutions which preserve the vacuum, axisymmetric, black hole properties are uniquely determined by the changes in two parameters associated with these solutions. Thus, Carter's result gives good reason to believe that the two-parameter Kerr family may be the only axisymmetric, vacuum, black hole solutions. In addition, independent investigations[8] of Ipser and myself have shown that the Kerr–Newman electromagnetic field is the only well behaved axisymmetric, electromagnetic perturbation of the Kerr black hole with no magnetic monopole moment.

In the present paper it is proven that the Kerr–Newman black holes are the only black hole solutions obtainable by analytic variation of the space-time geometry starting from the Schwarzschild geometry. (See Sec. 2 for a precise statement of what is meant by this.) The important restrictive assumptions of "static" (as opposed to stationary) in Israel's work and "axial symmetry" and "no electromagnetic fields" in Carter's work are not made here. Furthermore, as argued in Ref. 1, it is highly plausible that—at least for small deviations from spherical symmetry—the final space-time geometries resulting from collapse can be analytically developed from the Schwarzschild geometry (known to be the final state of spherically symmetric collapse). Thus, the theorem proven here leads one to believe that—at least for small deviations from spherical symmetry—the final state of gravitational collapse is a Kerr–Newman black hole. (Of course, it should be emphasized that not only is there no compelling reason to believe that this analytic variation assumption is valid for collapse with large deviations from spherical symmetry, but there is not even any compelling reason to believe that such collapse would result in a black hole, since it is only for infinitesimal deviations from spherical symmetry that collapse to a black hole is reasonably well established.[9] Thus, a great deal remains to be proven!)

The results of this paper may be described in physical terms as follows: Consider an electrovac black hole which differs slightly from a Schwarzschild black hole. The results of first-order perturbation theory (see Lemmas 1 and 2 of Sec. IV) show that in the first approximation this black hole can differ from the Schwarzschild black hole in mass, angular momentum, and charge, but it cannot have any higher multipole moments (e.g., no mass quadrupole moment or magnetic dipole moment, etc.). However, in the second and higher approximations, when account is taken of the first-order changes in angular momentum and charge, the black hole can have higher multipole moments. It is proven here that when all orders of approximation are taken into account, all the higher multipole moments must be fixed relative to the mass, angular momentum, and charge in precisely the same manner as for Kerr–Newman black holes.

Hartle and Thorne[10] have given an expression for the exterior field metric of a nearly spherical star (or black hole) in which the effect of mass quadrupole moment is calculated to lowest nonvanishing order and the effect of angular momentum is calculated to second order. Examination of this expression suggests that for black holes one will have singular behavior at the horizon unless the quadrupole moment is fixed in terms of the angular momentum in the same manner as for Kerr black holes. Aside from putting this conclusion on a rigorous footing, the present paper may be viewed as an improvement of the Hartle and Thorne result in the following significant ways: (1) Axial symmetry and reflection symmetry (assumed by Hartle and Thorne) are not assumed here; (2) electromagnetic fields are permitted to occur here; (3) the Hartle and Thorne result is generalized to all orders and it is thus shown that the Kerr–Newman solutions are the only black hole solutions that can be obtained by a perturbation series expansion about the Schwarzschild solution.

Another method for generalizing perturbation results to analytic variation (and, in fact, to even a somewhat wider class of functions than analytic) has been developed independently by Carter (private communication) and applied to his axisymmetric black hole perturbation analysis. Carter has recently shown that this method can also be applied to the problem treated here (where the gauge arbitrariness causes a number of difficulties).

After this paper was submitted for publication, Hawking announced the proof of a new theorem which states that a stationary black hole must be either static or axisymmetric. As a consequence, a result on black hole uniqueness stronger than the one presented here in that it excludes various possibilities for black holes not analytically connected to Schwarzschild could be obtained by combining the theorems of Hawking, Carter,[6] Israel,[7] and Wald and Ipser.[8] The proof of the theorem presented here is entirely independent of (and considerably simpler than) these combined theorems.

In Sec. 2, the analytic family $\mathfrak{A}$ of electrovac black hole space-times containing the Schwarzschild solution is defined and the theorem is stated. In Sec. 3 we derive conditions (summarized at the end of the section) on the nth-order perturbations of the metric and electromagnetic field tensor. Two lemmas are presented in Sec. 4 and the theorem is proven in Sec. 5.

2. STATEMENT OF THE THEOREM

We define a one-parameter analytic family of electrovac black hole space-times containing the Schwarzschild solution of mass m_0 to be a family of space-time metrics $g(\alpha)$ and electromagnetic field tensors $F(\alpha)$ satisfying the following five conditions:

(I) For each α, the pair $g(\alpha), F(\alpha)$ is a solution of the Einstein–Maxwell equations

$$G_{\mu\nu}(g(\alpha)) = 8\pi T_{\mu\nu}(\alpha) = 2(F_{\mu\beta}F_{\nu}{}^{\beta} - \tfrac{1}{4}g_{\mu\nu}F_{\rho\sigma}F^{\rho\sigma}) \quad (1)$$

$$F^{\mu\nu}{}_{;\nu} = 0, \qquad {}^{*}F^{\mu\nu}{}_{;\nu} = 0, \quad (2)$$

where ";" denotes covariant derivative and $^*F$ denotes the dual of F. For $\alpha = 0$, the metric $g(0)$ is the Schwarzschild metric of mass m_0, $m_0 > 0$, and $F(0) = 0$.

(II) Each $g(\alpha)$ is asymptotically flat and stationary. [Stationary means $g(\alpha)$ has a Killing vector which is timelike at large distances.] This requires that for each α there exist "Schwarzschild-like coordinates" denoted t, r, θ, ϕ having the following properties: (a) $\partial/\partial t$ is a Killing vector which is timelike for large r; and (b) the metric is asymptotically Minkowskian and the field tensor components are $0(1/r^2)$ as $r \to \infty$ in the coordinates t, X, Y, Z, where $X = r\sin\theta\cos\phi$, $Y = r\sin\theta\sin\phi$, and $Z = r\cos\theta$.

The Schwarzschild-like coordinates t, r, θ, ϕ will be denoted x^μ, or simply x, in the following and the components of $g(\alpha)$ in this coordinate system will be denoted ${}^{s}g_{\mu\nu}(\alpha, x)$ (s standing for "Schwarzschild-like coordinates"). For $\alpha = 0$ we take these coordinates to be the standard coordinates of the Schwarzschild metric $g(0)$, i.e.,

$$ds^2(\alpha = 0) = -(1 - {}^{2m_0}/r)dt^2 + (1 - {}^{2m_0}/r)^{-1}dr^2 + r^2(d\theta^2 + \sin^2\theta d\phi^2).$$

(III) The exterior region, i.e., the domain of outer communications, and the horizon of each $g(\alpha)$ is nonsingular. (The domain of outer communications is defined as the set of points lying on timelike curves which escape to arbitrarily large distances in both future and past directions; the horizon is defined as its boundary.) This requires the existence of analytic extension coordinates (i.e., "generalized Kruskal coordinates") which cover the exterior region and horizon of $g(\alpha)$. (The Schwarzschild-like coordinates need not cover this entire region, e.g., they may break down on the horizon.)

The analytic extension coordinates will be denoted z^μ, or simply z, in the following, and the components of $g(\alpha)$ in this coordinate system will be denoted ${}^{A}g_{\mu\nu}(\alpha, z)$ (A standing for "analytic extension coordinates"). We shall assume that, for fixed α, ${}^{A}g_{\mu\nu}(\alpha, z)$ varies analytically in z, although, strictly, it is only necessary to assume it is C^2 in z. For $\alpha = 0$ we take these coordinates to be the standard Kruskal coordinates[11] of the Schwarzschild metric $g(0)$.

(IV) The metric and field tensor components in the A coordinate system vary analytically with α, i.e., ${}^{A}g_{\mu\nu}(\alpha, z)$ and ${}^{A}F_{\mu\nu}(\alpha, z)$ are analytic functions of α. We shall assume that ${}^{A}g_{\mu\nu}(\alpha, z)$ and ${}^{A}F_{\mu\nu}(\alpha, z)$ are jointly analytic in (α, z), although, strictly, it is only necessary to assume separate analyticity in α and existence and continuity of all partial derivatives of second order in spacial derivatives and arbitrary order in α (so that we may commute α derivatives with the spacial derivatives occurring in Einstein's equations). (In Ref. 1 only separate analyticity in α and z was postulated because it was thought to imply joint analyticity by Hartog's theorem; however, Hartog's theorem requires the functions to be defined on an open complex domain.)

(V) The transformation from the analytic coordinates to the Schwarzschild-like coordinates $x^\mu = x^\mu(\alpha, z)$ varies analytically with α. Again, we assume that, in the region where the Schwarzschild-like coordinates are defined, we have joint analyticity of $x^\mu(\alpha, z)$. [Strictly, only continuity of the partial derivatives of $x^\mu(\alpha, z)$ of third order in z and arbitrary order in α is required.] This implies, using (IV), that, where defined, ${}^{s}g_{\mu\nu}(\alpha, x)$ and ${}^{s}F_{\mu\nu}(\alpha, x)$ are analytic in (α, x). In addition, we explicitly postulate that the α-partial derivatives of ${}^{s}g_{\mu\nu}$ and ${}^{s}F_{\mu\nu}$ are respectively $O(1/r)$ and $O(1/r^2)$ as $r \to \infty$. [This does not automatically follow from (II) as examples like $\alpha r/(\alpha^2 + r^2)$ and $(1/r)\sin\alpha r$ show.]

The first three conditions require each $g(\alpha)$ to be an electrovac black hole, i.e., an asymptotically flat, stationary, vacuum except for electromagnetic fields, solution of Einstein's equations with a nonsingular exterior region and horizon. (IV) and (V) require the space–time properties of the black holes $g(\alpha)$ to vary analytically with α.

Note added in proof: It would probably be more reasonable physically to consider and analytically vary only the exterior region and future part of the event horizon, since the past horizon would not be produced in gravitational collapse. This would affect the results only in that it would allow the Kerr–Newman solutions with $e^2 + a^2 = m^2$ to be in $\mathfrak{A}$.

One may wonder why we introduce two coordinate systems (s and A), i.e., why not simply start with the Schwarzschild metric expressed, say, in Kruskal coordinates and look for all black hole solutions obtainable by analytically varying these metric components, without ever worrying about the s system. The trouble with doing this is that we would then have no mathematically convenient way of expressing the condition that the solution be stationary and asymptotically flat. Similarly, we could analytically vary the metric components starting with the Schwarzschild metric in standard Schwarzschild coordinates and not worry about the A coordinates; but then we would have no mathematically convenient way of insuring that the horizon be nonsingular. There would be no problem if it were reasonable to demand that the s and A systems be compatible, i.e., that for all black holes one could find a single coordinate system satisfying both properties (II) and (III) simultaneously; however, this is not reasonable, since it is easy to see that for the Schwarzschild metric any coordinate system that uses the static Killing vector as a coordinate (as required for an s system) cannot cover the entire horizon (as required for an A system). We overcome the above difficulties by working with both systems and requiring that as we analytically vary the metric components in the A system [property (IV)], the coordinate transformation from the A system to the s system also varies analytically [property (V)]. [In other words, we require that (or, more precisely, we seek only solutions for which) the timelike Killing vector and the location of the asymptotically flat region of the space–time change smoothly as the metric components are smoothly varied; this need not automatically happen since as one smoothly varies the metric components one might suddenly get a completely different global interpretation of the space–time; but in such a case the

properties of the space-time change drastically (even though the metric components in the A system vary analytically), so we are justified in including property (V) as part of the definition of black hole space-time analytically developable from Schwarzschild.] By using both coordinate systems we can conveniently express all the black hole requirements. Property (V) then allows us to relate perturbations in the A system to those in the s system.

The (total) analytic family $\mathfrak{A}$ of black hole space-times containing the Schwarzschild solution is defined to be the set of all space-times belonging to some one-parameter analytic family defined above. We prove the following theorem:

Theorem: The analytic family $\mathfrak{A}$ of electrovac black hole space-times containing the Schwarzschild solution is completely spanned by the Kerr-Newman space-times with $e^2 + a^2 < m^2$.

The meaning of the theorem is heuristically illustrated in Fig. 1.

It is not difficult to verify that the Kerr-Newman space-times,

$$ds^2 = -\frac{r^2 + a^2 + e^2 - 2mr}{r^2 + a^2\cos^2\theta}(dt - a\sin^2\theta d\phi)^2 + \frac{\sin^2\theta}{r^2 + a^2\cos^2\theta}[adt - (r^2 + a^2)d\phi]^2 + (r^2 + a^2\cos^2\theta)\left(\frac{dr^2}{r^2 + a^2 + e^2 - 2mr} + d\theta^2\right) \tag{3}$$

$$F = \frac{2e}{(r^2 + a^2\cos^2\theta)^2} \times (r^2 - a^2\cos^2\theta)dr \wedge (dt - a\sin^2\theta d\phi) - \frac{4ear\sin\theta\cos\theta}{(r^2 + a^2\cos^2\theta)^2} d\theta \wedge [adt - (r^2 + a^2)d\phi] \tag{4}$$

belong to $\mathfrak{A}$ for $e^2 + a^2 < m^2$. Namely, to show that an arbitrary Kerr-Newman solution characterized by the parameters m_1, a_1, e_1 with ${e_1}^2 + {a_1}^2 < {m_1}^2$ is in $\mathfrak{A}$, construct a one-parameter family $g(\alpha), F(\alpha)$ of Kerr-Newman solutions by taking $m(\alpha) = m_0 + \alpha(m_1 - m_0)$, $a(\alpha) = \alpha a_1$, $e(\alpha) = \alpha e_1$, so that $g(1)$, $F(1)$ is the given solution. From the properties of the Kerr-Newman space-times,[5] one may verify that $g(\alpha), F(\alpha)$ satisfy conditions (1)–(5) of the definition of one-parameter analytic family. The given solution is in this one-parameter family, and thus is in $\mathfrak{A}$.

On the other hand, the Kerr-Newman space-times with $e^2 + a^2 = m^2$ (where the structure of the horizon changes) or $e^2 + a^2 > m^2$ (where the horizon disappears altogether) do not belong to $\mathfrak{A}$. For ${e_1}^2 + {a_1}^2 \geqq {m_1}^2$, a one-parameter family $g(\alpha), F(\alpha)$ constructed as above would fail to satisfy (IV) and (V) [and even (III) for ${e_1}^2 + {a_1}^2 > {m_1}^2$].

We now prove the nontrivial part of the theorem namely, that the Kerr-Newman space-times are the only members of $\mathfrak{A}$, i.e., that any one-parameter analytic family $g(\alpha), F(\alpha)$ satisfying (I)–(V) is composed of only Kerr-Newman space-times. We proceed by obtaining conditions on all the derivatives $g^{(n)}$ and $F^{(n)}$ of $g(\alpha)$ and $F(\alpha)$ with respect to α at $\alpha = 0$.

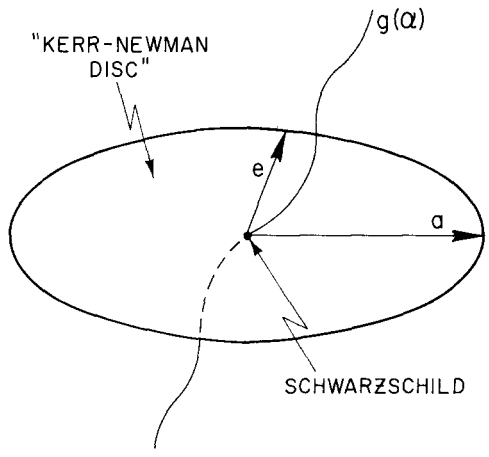


FIG 1. "Black hole solution space." Each point in the figure represents an electrovac black hole solution of Einstein's equations. The dependence on mass is suppressed and e and a are, respectively, the values of the charge and angular momentum per unit mass of the black hole. The vertical direction represents all other quantities besides mass, charge, and angular momentum upon which a black hole might depend, e.g., higher multipole moments. The disc represents the Kerr-Newman black holes and the curve $g(\alpha)$ represents an arbitrary one-parameter analytic family satisfying Conditions 1–5. The theorem states that $g(\alpha)$ must lie entirely in the "Kerr-Newman disc."

3. CONDITIONS ON $g^{(n)}$ AND $F^{(n)}$

Properties (IV) and (V) require the components ${}^s g_{\mu\nu}(\alpha, x)$ and ${}^s F_{\mu\nu}(\alpha, x)$ to be analytic functions of α. Hence, we may expand them in a power series in α:

$${}^s g_{\mu\nu}(\alpha, x) = \sum_{n=0}^{\infty} \frac{\alpha^n}{n!}\, {}^s g^{(n)}_{\mu\nu}(x), \tag{5a}$$

$${}^s F_{\mu\nu}(\alpha, x) = \sum_{n=0}^{\infty} \frac{\alpha^n}{n!}\, {}^s F^{(n)}_{\mu\nu}(x), \tag{5b}$$

where

$${}^s g^{(n)}_{\mu\nu}(x) \equiv \frac{\partial^n\, {}^s g_{\mu\nu}}{\partial\alpha^n}(\alpha, x)\,\Big|_{\alpha=0}\ x \text{ fixed}, \tag{6a}$$

$${}^s F^{(n)}_{\mu\nu}(x) \equiv \frac{\partial^n\, {}^s F_{\mu\nu}}{\partial\alpha^n}(\alpha, x)\,\Big|_{\alpha=0}\ x \text{ fixed}. \tag{6b}$$

Note that ${}^s g^{(0)}_{\mu\nu} = {}^s g_{\mu\nu}(\alpha = 0)$ is the Schwarzschild metric and ${}^s F^{(0)}_{\mu\nu} = 0$. We denote by $g^{(n)}$ and $F^{(n)}$ the tensors defined on the Schwarzschild manifold $g(0)$ whose components in the standard Schwarzschild coordinates of this manifold are just ${}^s g^{(n)}_{\mu\nu}$ and ${}^s F^{(n)}_{\mu\nu}$. *We now translate Conditions 1–5 on $g^{(n)}, F(\alpha)$ into conditions on the "nth-order quantities" $g^{(n)}$ and $F^{(n)}$ and show that, given lower orders, $g^{(n)}$ and $F^{(n)}$ are determined uniquely up to the addition of tensors of a particular form. From this "uniqueness theorem" on $g^{(n)}$ and $F^{(n)}$, it will be shown that each $g(\alpha), F(\alpha)$ is indeed a Kerr-Newman solution.*

Property (II) and the additional explicit assumption stated in (V) immediately yield the following two conditions on $g^{(n)}$ and $F^{(n)}$ for $n \geqq 1$: (1) ${}^s g^{(n)}_{\mu\nu}(t, r, \theta, \phi)$ and ${}^s F^{(n)}_{\mu\nu}(t, r, \theta, \phi)$ must be independent of t; and (2) in the coordinates t, X, Y, Z, the components of $g^{(n)}$ must be $O(1/r)$ and the components of $F^{(n)}$ must be $O(1/r^2)$ as $r \to \infty$.

Next, we obtain equations for $g^{(n)}$ and $F^{(n)}$ (called the "nth-order equations") by differentiating the

Einstein–Maxwell equations (1), (2) n times with respect to α (keeping x fixed) and setting $\alpha = 0$. When we differentiate (1) once with respect to α and set $\alpha = 0$, the right-hand side of (1) gives no contribution because F appears quadratically and $F(\alpha = 0) = 0$. We thus get a linear, homogeneous system of equations for the "first order perturbation" $g^{(1)}$ which (by definition) are the linearized Einstein field equations in the Schwarzschild background $g(0)$. These equations, which have been extensively treated in perturbation analyses of the Schwarzschild metric,[12–15] will be denoted as

$$\mathcal{G}_{\mu\nu}(g^{(1)}) = 0. \tag{7}$$

Similarly, we get equations for $F^{(1)}$ by differentiating the "Maxwell half" of the Einstein–Maxwell equations once with respect to α (keeping x fixed) and setting $\alpha = 0$. The equations thus obtained for the "test field" $F^{(1)}$ are simply the Maxwell equations in the Schwarzschild background, which will be denoted as

$$\mathcal{M}_{\mu}(F^{(1)}) = 0. \tag{8}$$

To obtain the essential form of the nth-order equations, we make use of the following general principle.

Higher Order Equations Principle: nth-order equations involve nth-order quantities in precisely the same manner as first-order equations involve first-order quantities.

The proof of this statement is given in Appendix A.

By the above principle, the nth order equations for $g^{(n)}$ and $F^{(n)}$ must be of the form

$$\mathcal{G}_{\mu\nu}(g^{(n)}) = (\text{lower order terms}), \tag{9}$$

$$\mathcal{M}_{\mu}(F^{(n)}) = (\text{lower order terms}), \tag{10}$$

where $\mathcal{G}_{\mu\nu}$ and $\mathcal{M}_{\mu}$ are the same operators as those in Eqs. (7) and (8), and "lower order terms" means terms involving only $g^{(i)}$ and $F^{(j)}$ with $i, j < n$.

The final conditions on $g^{(n)}$ and $F^{(n)}$ are obtained from the existence of the analytic coordinates z. The analyticity of ${}^{A}g_{\mu\nu}(\alpha, z)$ and ${}^{A}F_{\mu\nu}(\alpha, z)$ in (α, z) [property (IV)] requires that all the partial derivatives of ${}^{A}g_{\mu\nu}(\alpha, z)$ and ${}^{A}F_{\mu\nu}(\alpha, z)$ with respect to α at $\alpha = 0$ must be analytic functions of z throughout the exterior region and horizon of the Schwarzschild metric $g(0)$, where the partial derivatives

$${}^{A}g^{[n]}_{\mu\nu}(z) \equiv \frac{\partial^n}{\partial\alpha^n}\, {}^{A}g_{\mu\nu}(\alpha, z)\,\Big|_{\substack{z \text{ fixed}\\ \alpha=0}},$$

$${}^{A}F^{[n]}_{\mu\nu}(z) \equiv \frac{\partial^n}{\partial\alpha^n}\, {}^{A}F_{\mu\nu}(\alpha, z)\,\Big|_{\substack{z \text{ fixed}\\ \alpha=0}}$$

are to be taken holding z (not x) fixed. [The square bracket around the superscript n denotes partial derivatives taken with z fixed, as opposed to the round bracket, Eq. (6), which denotes partial derivatives taken with x held fixed.] In particular, ${}^{A}g^{[n]}_{\mu\nu}$ and ${}^{A}F^{[n]}_{\mu\nu}$ must be finite at the horizon. In order to make use of this condition, we must relate ${}^{A}g^{[n]}_{\mu\nu}$ and ${}^{A}F^{[n]}_{\mu\nu}$ to the tensors $g^{(n)}$ and $F^{(n)}$ defined by their Schwarzschild components, Eq. (6). We first obtain expressions for ${}^{A}F^{[1]}_{\mu\nu}$ and ${}^{A}g^{[1]}_{\mu\nu}$.

By the tensor transformation law, we have

$${}^{A}F_{\mu\nu}(\alpha, z) = {}^{S}F_{\rho\sigma}(\alpha, x(\alpha, z))\, \frac{\partial x^{\rho}}{\partial z^{\mu}}\, \frac{\partial x^{\sigma}}{\partial z^{\nu}}. \tag{11}$$

From the fact that $F(\alpha = 0) = 0$ and that when $\alpha = 0$ the analytic coordinates reduce to the Kruskal coordinates of the Schwarzschild metric $g(0)$ (see Condition 3), it is not difficult to verify that, differentiating (11) once with respect to α, holding z fixed, we get

$${}^{A}F^{[1]}_{\mu\nu} = {}^{K}F^{(1)}_{\mu\nu}, \tag{12}$$

where ${}^{K}F_{\mu\nu}{}^{(1)}$ denotes the components of the tensor $F^{(1)}$ in the Kruskal coordinates of $g(0)$.

The expression for ${}^{A}g^{[1]}_{\mu\nu}$ is slightly more complicated on account of the fact that $g(\alpha)$ does not vanish at $\alpha = 0$. However, it is not difficult to verify that, defining $f^{\mu}(\alpha, z)$ by

$$f^{\mu}(\alpha, z) = z^{\mu}(\alpha, x(\alpha, z)) - z^{\mu}(0, x(0, z)) \tag{13}$$

(so that f^{μ} is the difference between the actual and zeroth-order analytic coordinates), we have

$${}^{A}g^{[1]}_{\mu\nu} = {}^{K}g^{(1)}_{\mu\nu} - f^{[1]}_{\mu;\nu} - f^{[1]}_{\nu;\mu}, \tag{14}$$

where ${}^{K}g^{(1)}_{\mu\nu}$ denotes the components of $g^{(1)}$ in Kruskal coordinates,

$$(f^{[1]})^{\mu} \equiv \frac{\partial f^{\mu}(\alpha, z)}{\partial\alpha}\,\Big|_{\substack{z \text{ fixed}\\ \alpha=0}}$$

and ";" denotes covariant derivative in the Schwarzschild metric $g(0)$. In other words, the first-order change in the analytic coordinate components of the metric is the sum of the first-order change in the metric, ${}^{K}g^{(0)}_{\mu\nu} \to {}^{K}g^{(0)}_{\mu\nu} + \alpha {}^{K}g^{(1)}_{\mu\nu}$, and the first-order (gauge transformation) effect of the coordinate transformation $z^{\mu} \to z^{\mu} + \alpha f^{[1]\mu}$.

By the higher-order equations principle, it follows that the equations obtained by differentiating (11) n times with respect to α (holding z fixed) and setting $\alpha = 0$ must be of the form

$${}^{A}F^{[n]}_{\mu\nu} = {}^{K}F^{(n)}_{\mu\nu} + (\text{lower-order terms}), \tag{15}$$

where "lower-order terms" denotes terms involving only $F^{(i)}$ and $f^{[j]}$ with $i, j < n$. Similarly,

$${}^{A}g^{[n]}_{\mu\nu} = {}^{K}g^{(n)}_{\mu\nu} - f^{[n]}_{\mu;\nu} - f^{[n]}_{\nu;\mu} + (\text{lower-order terms}), \tag{16}$$

where "lower-order terms" here denotes terms involving $g^{(i)}$ and $f^{[j]}$ with $i, j < n$.

We summarize the conditions on $g^{(n)}$ and $F^{(n)}$, $n \geqq 1$, which we have obtained above.

(C1) ${}^{S}g^{(n)}_{\mu\nu}(t, r, \theta, \phi)$ and ${}^{S}F^{(n)}_{\mu\nu}(t, r, \theta, \phi)$ must not depend on t.

(C2) In the coordinates t, X, Y, Z, the components of $g^{(n)}$ must be $O(1/r)$ and the components of $F^{(n)}$ must be $O(1/r^2)$ as $r \to \infty$.

(C3) $g^{(n)}$ satisfies Eq. (9), where $\mathcal{G}_{\mu\nu}$ is the linearized Einstein field equations operator in the Schwarzschild background. $F^{(n)}$ satisfies Eq. (10), where $\mathcal{M}_{\mu}$ is the Maxwell equations operator in the Schwarzschild background.

(C4) ${}^A g^{[n]}_{\mu\nu}$ and ${}^A F^{[n]}_{\mu\nu}$ must be finite at the horizon. These quantities are related to $g^{(n)}$ and $F^{(n)}$ via Eqs. (15) and (16).

4. TWO LEMMAS

The proof of the theorem is based on the two lemmas presented below. The first lemma has been effectively proven by Vishveshwara.[14]

Lemma 1 (Vishveshwara): Let h be a symmetric tensor defined on the Schwarzschild manifold satisfying the following four conditions.

(G1) ${}^s h_{\mu\nu}(t, r, \theta, \phi)$ is independent of t.

(G2) In coordinates t, X, Y, Z, components of h are $O(1/r)$ as $r \to \infty$.

(G3) h satisfies the linearized Einstein equations $\mathcal{G}_{\mu\nu}(h) = 0$.

(G4) There exists a gauge transformation ζ^μ such that the Kruskal components $({}^K h_{\mu\nu} - \zeta_{\mu;\nu} - \zeta_{\nu;\mu})$ are finite at the horizon.

(In other words, h is a well behaved, stationary perturbation of the Schwarzschild metric.) Then,

$$ {}^s h_{\mu\nu} = C_0 T_0 + C_x T_{1x} + C_y T_{1y} + C_z T_{1z} + \xi_{\mu;\nu} + \xi_{\nu;\mu}, \tag{17} $$

where C_0, C_x, C_y, C_z are arbitrary constants; T_0, T_{1x}, T_{1y}, T_{1z} are tensors defined as follows:

$$ T_0 = \frac{1}{r}\begin{bmatrix} 1 & 0 & 0 & 0 \\ & (1-2m/r)^{-2} & 0 & 0 \\ & & 0 & 0 \\ & (sym.) & & 0 \end{bmatrix}, $$

$$ T_{1x} = \frac{1}{r}\begin{bmatrix} 0 & 0 & -\sin\phi & -\sin\theta\cos\theta\cos\phi \\ & 0 & 0 & 0 \\ & & 0 & 0 \\ & (sym.) & & 0 \end{bmatrix} \tag{18} $$

$$ T_{1y} = \frac{1}{r}\begin{bmatrix} 0 & 0 & \cos\phi & -\sin\theta\cos\theta\sin\phi \\ & 0 & 0 & 0 \\ & & 0 & 0 \\ & (sym.) & & 0 \end{bmatrix}, $$

$$ T_{1z} = \frac{1}{r}\begin{bmatrix} 0 & 0 & 0 & \sin^2\theta \\ & 0 & 0 & 0 \\ & & 0 & 0 \\ & (sym.) & & 0 \end{bmatrix} $$

(the rows and the columns being in the order t, r, θ, ϕ); and ξ^μ is a gauge transformation which preserves $\partial/\partial t$ as a Killing vector and preserves the asymptotic behavior (G2) of h as $r \to \infty$, but is otherwise arbitrary.

In Appendix B, Vishveshwara's results are described and the necessary steps are given to prove the above lemma from his analysis.

Lemma 2: Let $\mathcal{F}$ be an antisymmetric tensor defined on the Schwarzschild manifold satisfying the following four conditions.

(F1) ${}^s \mathcal{F}_{\mu\nu}(t, r, \theta, \phi)$ is independent of t.

(F2) In coordinates t, X, Y, Z, the components of $\mathcal{F}$ are $O(1/r^2)$ as $r \to \infty$.

(F3) $\mathcal{F}$ satisfies the source-free Maxwell's equations $\mathfrak{M}_\mu(\mathcal{F}) = 0$.

(F4) The Kruskal components ${}^K \mathcal{F}_{\mu\nu}$ are finite at the horizon.

(In other words, $\mathcal{F}$ is a well behaved, static, test electromagnetic field in the Schwarzschild background.) Then, if magnetic monopoles are excluded,

$$ \mathcal{F} = (C/r^2) dt \wedge dr, \tag{19} $$

where C is an arbitrary constant. (If magnetic monopoles are permitted, duality rotations of this field tensor are also acceptable.)

Proof: An arbitrary "static" (i.e., independent of t) source free electromagnetic field tensor in the Schwarzschild background can be derived from an electrostatic potential V (yielding arbitrary $\mathcal{F}_{0i} = -\partial V/\partial x^i$ and a magnetic scalar potential[16] ψ (yielding arbitrary ${}^*\mathcal{F}_{0i} = -\partial\psi/\partial x^i$ and hence arbitrary $\mathcal{F}_{jk}$). Maxwell's equations require that V satisfy[7,17]

$$ 0 = \frac{1}{r^2}\frac{\partial}{\partial r}\left(r^2 \frac{\partial V}{dr}\right) + \frac{1}{1-2m/r}\frac{1}{r^2\sin\theta} \times \left(\frac{\partial}{\partial\theta}\sin\theta\frac{\partial V}{\partial\theta} + \frac{1}{\sin\theta}\frac{\partial^2 V}{\partial\phi^2}\right). \tag{20} $$

By the tensor transformation law,

$$ \begin{aligned} {}^K\mathcal{F}_{v\theta} &= {}^s\mathcal{F}_{r\theta}\frac{\partial r}{\partial v} + {}^s\mathcal{F}_{t\theta}\frac{\partial t}{\partial v} \\ &= {}^s\mathcal{F}_{r\theta}\left(-\frac{8m^2}{r} v e^{-r/2m}\right) + {}^s\mathcal{F}_{t\theta}\left(\frac{4mu}{u^2 - v^2}\right), \end{aligned} \tag{21} $$

where u and v are the standard Kruskal coordinates.[11] Setting $v = 0$ and letting $u \to 0$ (i.e., setting $t = 0$ and letting $r \to 2m$), we get that the finiteness of ${}^K\mathcal{F}_{v\theta}$ implies that ${}^s\mathcal{F}_{t\theta} \to 0$ as $r \to 2m$, i.e., $\partial V/\partial\theta \to 0$ as $r \to 2m$. Similarly, finiteness of ${}^K\mathcal{F}_{v\phi}$ requires $\partial V/\partial\phi \to 0$ as $r \to 2m$. But examination of the explicit solutions[7,17] of (20) shows that all the nonspherically symmetric solutions which go to zero as $r \to 2m$ also blow up as $r \to \infty$. Hence, only the spherically symmetric solutions are acceptable, and the most general acceptable solution of (20) is thus,

$$ V = c_0 + c_1/r, \tag{22} $$

which is just the electrostatic monopole field of the Reissner-Nordström solution.

In fact, this result can also be obtained without direct examination of the explicit solutions of (20). Namely, a solution of (20) in some region must take its maximum and minimum values on the boundary of the region[18] and hence is uniquely determined in the region by its boundary values. For solutions which are well behaved as $r \to \infty$, we may require $V \to 0$ as $r \to \infty$. The solution in the region $r > 2m$ is then uniquely determined by its value at the boundary $r = 2m$. Since, as we have shown above, for acceptable solutions V must be constant at $r = 2m$, and since the solutions $V = c_1/r$ yield all possible constant values at $r = 2m$, they must be the only acceptable solutions.

The equations and boundary conditions for ψ are identical to those for V,[16] so we also have

$$\psi = c_0' + c_1'/r. \tag{23}$$

This solution, however, gives rise to a magnetic monopole field, so we exclude it on physical grounds. [Not rejecting this solution would affect our final results only in that it would permit duality rotations of the field tensor (4) associated with the Kerr–Newman metric (3).] Thus, the most general acceptable field tensor is

$$\mathcal{F} = 2d(Vdt) = -(2c_1/r^2)dr \wedge dt, \tag{24}$$

which concludes the proof of the lemma.

5. PROOF OF THE THEOREM

The lemmas of the previous section show that, given lower orders, the arbitrariness of $g^{(n)}$ and $F^{(n)}$ is quite limited. Namely, consider two analytic, one-parameter families $g(\alpha), F(\alpha)$ and $g'(\alpha), F'(\alpha)$ which are identical in all orders $j < n$, i.e., $g^{(j)} = f'^{[j]}$, $F^{(j)} = F'^{(j)}$, and $f^{[j]} = f'^{[j]}$ [where f is defined by Eq. (13)] for all $j < n$. Let ${}^s h_{\mu\nu}^{(n)} = {}^s g_{\mu\nu}^{(n)} - {}^s g'^{(n)}_{\mu\nu}$. From (C1) and (C2) it is easy to see that $h^{(n)}$ satisfies the hypotheses (G1) and (G2) of Lemma 1. If we take the difference of Eqs. (9) for $g^{(n)}$ and $g'^{(n)}$, and $g'^{(n)}$,

$$\mathcal{G}_{\mu\nu}(g^{(n)}) = (\text{lower-order terms}), \tag{25a}$$

$$\mathcal{G}_{\mu\nu}(g'^{(n)}) = (\text{lower-order terms})' \tag{25b}$$

and recall that $\mathcal{G}_{\mu\nu}$ is linear and that by hypothesis the primed and unprimed lower-order terms are equal, we get

$$\mathcal{G}_{\mu\nu}(h^{(n)}) = 0, \tag{26}$$

i.e., ${}^s h_{\mu\nu}^{(n)}$ satisfies (G3). Similarly, if we take the difference of the expressions (16) for ${}^A g_{\mu\nu}^{[n]}$ and ${}^A g'^{[n]}_{\mu\nu}$,

$${}^A g_{\mu\nu}^{[n]} = {}^K g_{\mu\nu}^{(n)} = f_{\mu;\nu}^{[n]} = f_{\nu;\mu}^{[n]} + (\text{lower-order terms}), \tag{27a}$$

$${}^A g'^{[n]}_{\mu\nu} = {}^K g'^{(n)}_{\mu\nu} = f'^{[n]}_{\mu;\nu} = f'^{[n]}_{\nu;\mu} + (\text{lower-order terms})' \tag{27b}$$

noting that the difference of the two analytic functions ${}^A g_{\mu\nu}^{[n]}$ and ${}^A g'^{[n]}_{\mu\nu}$ must also be analytic, we obtain that ${}^K h_{\mu\nu}^{(n)} + (f^{[n]} - f'^{[n]})_{\mu;\nu} + (f^{[n]} - f'^{[n]})_{\nu;\mu}$ must be finite at the horizon, i.e., (G4) is also satisfied. So, ${}^s h_{\mu\nu}^{(n)}$ satisfies the hypotheses of Lemma 1 and is thus given by Eq. (17).

Similarly, ${}^s \mathcal{F}_{\mu\nu}^{(n)} = {}^s F_{\mu\nu}^{(n)} - {}^s F'^{(n)}_{\mu\nu}$ satisfies conditions (F1)–(F4) of Lemma 2 and hence must be of the form (19).

Furthermore, if $g(\alpha)$ and $g'(\alpha)$ are identical in all orders $j < n$ as above and additionally ${}^s g_{\mu\nu}^{(n)} = {}^s g'^{(n)}_{\mu\nu}$, then taking the difference of Eqs. (27) we get that $(f^{[n]} - f'^{[n]})_{\mu;\nu} + (f^{[n]} - f'^{[n]})_{\nu;\mu}$ must be analytic in a region containing the exterior region and horizon of the Schwarzschild metric $g(0)$. By a direct calculation[19] this in turn implies that $(f_\mu^{[n]} - f_\mu'^{[n]})$ must be analytic in this region, i.e., given $g^{(n)}$ and all $g'^{(j)}$, $F^{(j)}$ and $f^{[j]}$ with $j < n$, $f^{[n]}$ is uniquely determined up to the addition of a function which is analytic in a region containing the domain of outer communications and horizon of the Schwarzschild manifold.

Thus, freedom in choosing the nth-order quantities is quite limited: $g^{(1)}$ must be of the form (17) and $F^{(1)}$ must be of the form (19). With $g^{(1)}$ chosen, $f^{[1]}$ is determined up to the addition of an analytic function. With $g^{(1)}, F^{(1)}$, and $f^{[1]}$ chosen, $g^{(2)}$ is determined uniquely up to the addition of a tensor of the form (17) and $F^{(2)}$ is determined uniquely up to the addition of a tensor of the form (19), etc. We now claim that the arbitrariness which *does* exist stems only from coordinate arbitrariness and, in the illustrative terminology of Fig. 1, freedom of choice of path and path parameter in the "Kerr–Newman disc", i.e., we claim that $g(\alpha), F(\alpha)$ must be a family of Kerr–Newman space-times. Specifically, the arbitrariness of $f^{[n]}$ arises from the freedom of choice of analytic coordinates z^μ; the arbitrariness $\xi_{\mu;\nu} + \xi_{\nu;\mu}$ of (17) arises from freedom of choice of the x^μ coordinates; the freedom of the parameters C_{1x} and C_{1y} of (17) may also be attributed to the x^μ coordinate freedom (namely, the choice of Z-axis); the freedom of C_0 and C_{1z} of (17) and C of (19) arises, respectively, from the freedom of choosing $m(\alpha)$, $a(\alpha)$, and $e(\alpha)$, i.e., the path in the "Kerr–Newman disc." We will show that an arbitrary one-parameter analytic family $g(\alpha), F(\alpha)$ satisfying (I)–(V) is composed of only Kerr–Newman solutions by constructing all possible families $g_K(\alpha), F_K(\alpha)$ of Kerr–Newman solutions and showing that $g(\alpha), F(\alpha)$ must coincide with some $g_K(\alpha), F_K(\alpha)$.

To construct an arbitrary Kerr–Newman family $g_K(\alpha), F_K(\alpha)$, we start with the Kerr–Newman solution, Eqs. (3), (4), noting that the Kerr–Newman metric is of the form

$$ds^2 = (\text{Schwarzschild metric of mass } m) - 2maT_{1z} + (\text{terms quadratic or higher order in } a \text{ and } e), \tag{28}$$

where T_{1z} is defined by Eq. (18). Next, we perform an arbitrary rotation. Noting that T_{1x}, T_{1y}, T_{1z} transforms as a vector under rotations since $-(T_{1x} + iT_{1y})/2$, T_{1z}, and $(T_{1x} - iT_{1y})/2$ are $l = 1$ magnetic parity tensor spherical harmonics,[13] we get that the metric in the new coordinates is of the form

$$ds^2 = (\text{Schwarzschild metric of mass } m) - 2ma(l_x T_{1x} + l_y T_{1y} + l_z T_{1z}) + (\text{terms quadratic or higher order in } a \text{ and } e), \tag{29}$$

where l_x, l_y, l_z are the direction cosines of the old z axis. We write $a_x = al_x$, $a_y = al_y$, $a_z = al_z$ and replace a where it occurs in the higher order terms by $(a_x^2 + a_y^2 + a_z^2)^{1/2}$. Next, we replace m, a_x, a_y, a_z, e, by arbitrary analytic functions $m(\alpha), a_x(\alpha), a_y(\alpha), a_z(\alpha), e(\alpha)$, with $m(0) = m_0$, $a_x(0) = a_y(0) = a_z(0) = e(0) = 0$, i.e.,

$$m(\alpha) = m_0 + \sum_{n=1}^{\infty} \frac{\alpha^n}{n!} C_m^{(n)}, \qquad e(\alpha) = \sum_{n=1}^{\infty} \frac{\alpha^n}{n!} C_e^{(n)},$$

$$a_x(\alpha) = \sum_{n=1}^{\infty} \frac{\alpha^n}{n!} C_{ax}^{(n)}, \qquad a_y(\alpha) = \sum_{n=1}^{\infty} \frac{\alpha^n}{n!} C_{ay}^{(n)}, \tag{30}$$

$$a_z(\alpha) = \sum_{n=1}^{\infty} \frac{\alpha^n}{n!} C_{az}^{(n)},$$

where $C_m^{(n)}, C_{ax}^{(n)}, C_{ay}^{(n)}, C_{az}^{(n)}, C_e^{(n)}$ are arbitrary constants (such that these series converge). Next, we make an arbitrary coordinate transformation of the

Schwarzschild-like coordinates $x^\mu \to x'^\mu$ expressed as $x^\mu = x'^\mu + \xi^\mu(\alpha, x'^\beta)$, subject only to the conditions that $\xi^\mu(\alpha)$ preserve the Killing vector $\partial/\partial t$ and the asymptotic properties of the metric as $r \to \infty$, and that $\xi^\mu(\alpha = 0) = 0$.

Finally, we make an arbitrary analytic in α and z coordinate transformation of the analytic coordinates[4,5] associated with the Kerr–Newman spacetime to get new analytic coordinates z'^μ, related to the old ones by $z^\mu = z'^\mu + y^\mu(\alpha, z'^\beta)$, where y^μ is an arbitrary analytic function of α and z'^β for z'^β in a region containing the domain of outer communications and the horizon of $g(\alpha)$, with $y^\mu(\alpha = 0) = 0$.

Each family $g_K(\alpha), F_K(\alpha)$ constructed as above is composed of only Kerr–Newman solutions and satisfies properties (I)–(V). It is also not difficult to verify that the first order quantities $g_K^{(1)}$ and $F_K^{(1)}$ are of the form

$$g_K^{(1)} = 2C_m^{(1)}T_0 - 2m_0(C_{ax}^{(1)}T_{1x} + C_{ay}^{(1)}T_{1y} + C_{az}^{(1)}T_{1z}) + \xi^{(1)}_{\mu;\nu} + \xi^{(1)}_{\nu;\mu}, \quad (31)$$

$$F_K^{(1)} = (2C_e^{(1)}/r^2)dr \wedge dt. \quad (32)$$

Thus, $g_K^{(1)}$ is just of the form (17) and $F_K^{(1)}$ is of the form (19). Following the same reasoning as employed in the discussion of the "higher-order equations principle" (see Appendix A), it is easy to see that $g_K^{(n)}$ and $F_K^{(n)}$ are of the form

$$g_K^{(n)} = 2C_m^{(n)}T_0 - 2m_0(C_{ax}^{(n)}T_{1x} + C_{ay}^{(n)}T_{1y} + C_{az}^{(n)}T_{1z}) + \xi^{(n)}_{\mu;\nu} + \xi^{(n)}_{\nu;\mu} + (\text{lower-order terms}), \quad (33)$$

$$F_K^{(n)} = (2C_e^{(n)}/r^2)dr \wedge dt + (\text{lower-order terms}), \quad (34)$$

where "lower-order terms" means terms involving only $\xi^{(j)}, C_m^{(j)}, C_{ax}^{(j)}, C_{ay}^{(j)}, C_{az}^{(j)}, C_e^{(j)}$ with $j < n$.

We now prove the following statement by induction: Given an arbitrary one-parameter analytic family $g(\alpha), F(\alpha)$ satisfying properties (I)–(V) [and hence (C1)–(C4)], for each n there is a $g_K(\alpha), F_K(\alpha)$ such that for all $j \leqq n$, $g^{(j)} = g_K^{(j)}$, $F^{(j)} = F_K^{(j)}$, and $f^{[j]} = f_K^{[j]}$. Namely, for $n = 1$, the lemmas of Sec. IV directly show that $g^{(1)}$ must be of the form (17) and $F^{(1)}$ must be of the form (19). From (31) and (32) it is clear that we can find a g_K and F_K (by appropriately picking $C_m^{(1)}, C_{ax}^{(1)}, C_{ay}^{(1)}, C_{az}^{(1)}, C_e^{(1)}, \xi_\mu^{(1)}$) with $g_K^{(1)} = g^{(1)}$ and $F_K^{(1)} = F^{(1)}$. Also, having obtained $g_K^{(1)} = g^{(1)}$, $f^{[1]}$ and any $f_K^{[1]}$ associated with $g_K^{(1)}$ can differ at most by an analytic function; our freedom in choosing $y^{[1]}$ ensures that we can get $f_K^{[1]} = f^{[1]}$. Similarly, assuming there is a $g_K(\alpha), F_K(\alpha)$ such that $g^{(j)} = g_K^{(j)}$, $F^{(j)} = F_K^{(j)}$, and $f^{[j]} = f_K^{[j]}$ for all $j \leqq n$, we have already shown that the differences $(g^{(n+1)} - g_K^{(n+1)})$ and $(F^{(n+1)} - F_K^{(n+1)})$ must, respectively, be of the form (17) and (19). But from (33) and (34) it is clear that we can suitably choose $C_m^{(n+1)}, C_{ax}^{(n+1)}, C_{ay}^{(n+1)}, C_{az}^{(n+1)}, C_e^{(n+1)}, \xi_\mu^{(n+1)}$, (leaving lower orders unchanged) to get a $g_K(\alpha), F_K(\alpha)$ with $g_K^{(j)} = g^{(j)}$ and $F_K^{(j)} = F^{(j)}$ for all $j \leqq n + 1$. Similarly, the freedom of choice of $y^{[n+1]}$ allows us to get $f_K^{[n+1]} = f^{[n+1]}$ without affecting the lower orders.

Thus, given a one-parameter family $g(\alpha), F(\alpha)$ which satisfies properties (I)–(V), we can find a $g_K(\alpha)$, $F_K(\alpha)$ which agrees with $g(\alpha), F(\alpha)$ to arbitrarily high order. This implies that $g(\alpha), F(\alpha)$ itself must be a family of Kerr–Newman solutions. To see this, we argue as follows: Given an arbitrary one-parameter analytic family $g(\alpha), F(\alpha)$, for each n we can define $\xi_\mu^{(n)}, C_m^{(n)}, C_{ax}^{(n)}, C_{ay}^{(n)}, C_{az}^{(n)}, C_e^{(n)}$ and $y_\mu^{[n]}$ from the form of the Kerr–Newman family $g_K(\alpha), F_K(\alpha)$ which agrees with $g(\alpha), F(\alpha)$ to order n. We define[20] $m(\alpha) = m_0 + \sum_{n=1}^{\infty} C_m^{(n)}(\alpha^n/n!)$, etc. Then, by reversing the steps performed in constructing the family $g_K(\alpha), F_K(\alpha)$, it is clear that for any α_0 we can bring $g(\alpha_0), F(\alpha_0)$ into the standard Kerr–Newman form (3) and (4) by coordinate transformations and parameter substitutions. Thus, $g(\alpha), F(\alpha)$ is a family of Kerr–Newman solutions, which completes the proof of the theorem.

Note that the theorem treats electrovac black holes, i.e., only electromagnetic fields are allowed to be present. However, the theorem may be generalized without difficulty to allow for the possible presence of other types of fields provided that there exist no well behaved, static, test fields of these types in the Schwarzschild background. The analysis of Bekenstein[21] shows that this is the case for scalar fields and nonzero mass vector fields. Hence, the Kerr–Newman space-times (which have only electromagnetic fields) remain the only black hole solutions which are analytically developable from the Schwarzschild solution even if the possible presence of these other fields is not excluded *a priori*.

APPENDIX A: HIGHER-ORDER EQUATIONS PRINCIPLE

Consider a set of quantities $U_i(\alpha)$ (e.g., ${}^s g_{\mu\nu}$, ${}^s F_{\mu\nu}$, and their x derivatives) which are n times differentiable with respect to the parameter α and satisfy an equation

$$E(U_i(\alpha)) = 0, \quad (A1)$$

where E is an n times differentiable function of the U_i which has no explicit α dependence. The α derivatives

$$U_i^{(n)} \equiv \frac{d^n U_i(\alpha)}{d\alpha^n}\Big|_{\alpha=0}$$

are called nth order quantities and the equation obtained by differentiating (A1) n times with respect to α at $\alpha = 0$ is called the nth-order equation.

The first-order equation is, thus,

$$\frac{\partial E}{\partial U_i}\frac{dU_i}{d\alpha} = 0, \quad (A2)$$

which is linear and homogeneous in the first-order quantities.

Higher-order equations principle: The nth-order equation involves nth-order quantities in precisely the same manner as the first-order equation involves first-order quantities, i.e., the nth-order equation is of the form

$$\frac{\partial E}{\partial U_i}\frac{d^n U_i}{d\alpha^n} = (\text{terms of lower order than } n), \quad (A3)$$

where "terms of lower order than n" means terms involving only combinations of the $U_i^{(j)}$ with $j < n$.

Proof (by induction): The statement is trivial for $n = 1$. Assume the statement is true for $n = k$, i.e., that

$$\frac{\partial E}{\partial U_i}\frac{d^k U_i}{d\alpha^k} = (\text{terms of lower order than } k). \quad \text{(A4)}$$

The $(k + 1)$th-order equation is obtained by differentiating the kth-order equation one more time with respect to α at $\alpha = 0$. Hence, the $(k + 1)$th-order equation is

$$\frac{\partial^2 E}{\partial U_j \partial U_i}\frac{dU_j}{d\alpha}\frac{d^k U_i}{d\alpha^k} + \frac{\partial E}{\partial U_i}\frac{d^{k+1}U_i}{d\alpha^{k+1}} = \frac{d}{d\alpha}(\text{terms of lower order than } k), \quad \text{(A5)}$$

which is of the form

$$\frac{\partial E}{\partial U_i}\frac{d^{k+1}U_i}{d\alpha^{k+1}} = [\text{terms of lower order than } (k+1)]. \quad \text{(A6)}$$

This completes the proof.

[Another way of seeing this result is to note that nth-order quantities appear in nth-order equations only linearly and in combination with zeroth order quantities (since other combinations are of higher order). It is not difficult to convince oneself that this combination of nth- and zeroth-order quantities must be independent of n.]

APPENDIX B: VISHVESHWARA'S PERTURBATION ANALYSIS

We briefly outline here the pertinent results of Vishveshwara[14] and provide the steps necessary to bring his results into the form of Lemma 1.

Consider the linearized Einstein field equations in the Schwarzschild background, which we have denoted as [(see Eq. (7)]

$$\mathcal{G}_{\mu\nu}(h) = 0. \quad \text{(B1)}$$

Since the Schwarzschild metric is spherically symmetric, the operator $\mathcal{G}_{\mu\nu}$ must be rotationally invariant. Therefore, in the same manner and for the same reasons as one expands scalar functions occurring in rotationally invariant problems in ordinary spherical harmonics $Y_{lM}(\theta, \phi)$ and vector functions occurring in rotationally invariant problems in vector spherical harmonics, we expand ${}^s h_{\mu\nu}$ in tensor spherical harmonics.[13] We obtain from (B1) equations for the expansion coefficients, which in general are functions of r and t. These equations were first obtained by Regge and Wheeler[12] after making a choice of gauge which considerably simplified the problem. A number of errors appeared in the equations given by Regge and Wheeler, but a corrected set of equations has been given by Edelstein and Vishveshwara.[15]

For the case in which we are interested, namely ${}^s h_{\mu\nu}$ not a function of t, the equations simplify and reduce to a single ordinary differential equation in r for each "angular momentum" l and parity type. For $l \geq 2$, Vishveshwara[14] has shown that the Kruskal components of the solution of the radial equation which is well behaved as $r \to \infty$ must blow up as $r \to 2m$, and that this bad behavior cannot be remedied by a gauge transformation. Note that Vishveshwara only considers axially symmetric perturbations in his paper, but the equations depend only on l, not the "magnetic quantum number" M so that the above result holds for arbitrary perturbations with $l \geqq 2$.

On the other hand, with $l = 0, 1$ there are stationary perturbations which are well behaved. With a suitable choice of gauge, these perturbations may be written

$$l = 0, \qquad {}^s h_{\mu\nu} = C_0 T_0 \quad \text{(B2a)}$$

$$l = 1 \text{ ("electric")}, \qquad {}^s h_{\mu\nu} = 0, \quad \text{(B2b)}$$

$$l = 1 \text{ ("magnetic")}, \qquad {}^s h_{\mu\nu} = C_x T_{1x} + C_y T_{1y} + C_z T_{1z}, \quad \text{(B2c)}$$

where C_0, C_x, C_y, C_z are arbitrary constants and T_0, T_{1x}, T_{1y}, T_{1z} are defined by Eq. (18). [Physically, the $l = 0$ (spherically symmetric) perturbation generates a change in mass. The $l = 1$, "electric" perturbation generates a shift in the center of mass and can thus be eliminated by a gauge transformation.[13] The $l = 1$, "magnetic" perturbation generates a change in angular momentum.]

Thus, in our chosen gauge the most general time-independent solution of (B1) which is well behaved as $r \to \infty$ and is such that its Kruskal components plus a gauge transformation remain finite as $r \to 2m$ is just

$${}^s h_{\mu\nu} = C_0 T_0 + C_x T_{1x} + C_y T_{1y} + C_z T_{1z}. \quad \text{(B3)}$$

It is clear that if we "unfix" our gauge, we find that the general solution of (B1) satisfying the above conditions is

$${}^s h_{\mu\nu} = C_0 T_0 + C_x T_{1x} + C_y T_{1y} + C_z T_{1z} + \xi_{\mu;\nu} + \xi_{\nu;\mu}, \quad \text{(B4)}$$

where ξ^μ is a gauge transformation which is independent of t and preserves the good asymptotic behavior of ${}^s h_{\mu\nu}$ as $r \to \infty$ but is otherwise arbitrary. This is precisely the statement of Lemma 1 given in Sec. IV.

---

* Supported in part by NSF grant GP7669 and USAF Office of Scientific Research Grant AF49(638)1545.

† NSF predoctoral fellow.

1 R. M. Wald, Phys. Rev. Letters **26**, 1653 (1971).
2 R. P. Kerr, Phys. Rev. Letters **11**, 237 (1963).
3 E. T. Newman, E. Couch, R. Chinnapared, A. Exton, A. Prakash, and R. Torrence, J. Math. Phys. **6**, 918 (1965).
4 R. H. Boyer and R. W. Lindquist, J. Math. Phys. **8**, 265 (1967).
5 B. Carter, Phys. Rev. **174**, 1559 (1968).
6 B. Carter, Phys. Rev. Letters **26**, 331 (1971).
7 W. Israel, Phys. Rev. **164**, 1776 (1967); Commun Math. Phys. **8**, 245 (1968).
8 R. M. Wald, Ph. D. thesis (Princeton University, 1972) (unpublished); J. Ipser, Phys. Rev. Letters **27**, 529 (1971).
9 See Ref. 17, A. G. Doroshkevich, Ya. B. Zel'dovich, and L. D. Novikov, Zh. Eksp. Teor. Fiz. **49**, 170 (1965); [Sov. Phys. JETP **22**, 122 (1966)]. V. de la Cruz, J. E. Chase, and W. Israel, Phys. Rev. Letters **24**, 423 (1970), and R. Price (to be published).
10 J. B. Hartle and K. S. Thorne, Astrophys. J. **153**, 807 (1968).
11 M. D. Kruskal, Phys. Rev. **119**, 1743 (1960).

[12] T. Regge and J. A. Wheeler, Phys. Rev. **108**, 1063 (1957).
[13] F. Zerilli, Phys. Rev. D **2**, 2141 (1970).
[14] C. V. Vishveshwara, Phys. Rev. D **1**, 2870 (1970).
[15] L. A. Edelstein and C. V. Vishveshwara, Phys. Rev. D **1**, 3514 (1970).
[16] J. L. Anderson and J. M. Cohen, Astrophys. Space Sci. **9**, 146 (1970).
[17] J. M. Cohen and R. M. Wald, J. Math. Phys. **12**, 1845 (1971).
[18] See, e.g., R. Courant and D. Hilbert, *Methods of Mathematical Physics* (Interscience, New York, 1962), Vol II, p. 321.
[19] To verify explicitly that analyticity of the Kruskal components $f_{\mu;\nu} + f_{\nu;\mu}$ implies analyticity of f_μ, set $w^0 = u + v$, $w^1 = u - v$, [where u, v are the standard Kruskal coordinates of the Schwarzschild metric $g(0)$], and set $\psi_0 = f_u + f_v$, $\psi_1 = f_u - f_v$. Analyticity of $(f_{\mu;\nu} + f_{\nu;\mu})$ directly implies analyticity of $\partial/\partial w^0(\psi_0/A)$, $\partial/\partial w^1(\psi_1/A)$, and $(\partial\psi_0/\partial w^1 + \partial\psi_1/\partial w^0)$, where $A(w^0, w^1)$ is analytic in w^0 and w^1 in a region containing the exterior region and horizon of $g(0)$. [In terms of the Schwarzschild radial coordinate, r, $A = (32m/r)\exp(-r/2m)$.] This implies analyticity of ψ_0, ψ_1 and hence of f_u and f_v. With this established it is easy to verify that f_θ and f_φ must also be analytic.
[20] The series for $m(\alpha)$ must converge because the series for $g(\alpha)$ converges (by hypothesis) and $g(\alpha)$ must have finite mass. Similarly, the series for $a_i(\alpha)$ and $e(\alpha)$ must converge. For the case of $\xi_\mu(\alpha)$ one can show that the effect of the sequence of coordinate transformations $\sum_{n=1}^{N} \xi_\mu^{(n)}\alpha^n/n!$, $(N = 1, 2, \ldots)$, on $g(\alpha)$ must converge, but this does not imply in an immediately obvious fashion that the series itself must converge and, in fact, the remote possibility that it does not converge has not as yet been rigorously ruled out.
[21] J. Bekenstein (to be published).

Light-Cone Singularities and Lorentz Poles

S. Ferrara
Laboratori Nazionali del CNEN, Frascati, Italy
and
G. Rossi*
Istituto di Fisica dell'Università, Roma, and Istituto di Fisica dell'Università, L'Aquila, Italy
(Received 8 October 1971)

The causal Meyer–Suura structure functions are projected into irreducible representations of the Lorentz group. A clarification of the connection between light-cone singularities and Lorentz poles is obtained: We find that in general a light-cone singularity of the type $1/(-x^2 + i\epsilon x_0)^\alpha$, in the operator product of the hadronic electromagnetic current, is built up from a sequence of Lorentz poles at $\lambda_n = 1 + \alpha - n$ whose residues are polynomials of order n in the virtual photon square mass.

1. INTRODUCTION

In this paper we study the connection between scaling properties (light-cone singularities) and Regge-like behavior (J-plane singularities) of the off-mass-shell Compton amplitude. Recently many authors[1] suggested that presence of light-cone singularities in the commutator of the hadronic electromagnetic current, which are responsible for the scaling properties of the structure functions in the Bjorken limit, could imply the existence of fixed poles in the off-mass-shell forward Compton amplitude. To study this phenomenon, it is necessary to develop the harmonic analysis with respect to the Lorentz group in configuration space, showing explicitly how Lorentz poles contribute to build up light-cone singularities.

In Ref. 2 the Wick-rotated imaginary part of the off-mass-shell forward Compton amplitude has been subjected to an $O(4)$ analysis and the connection between its Lorentz pole content and light-cone singularities has been investigated by performing a Sommerfeld–Watson transform. The authors were able to relate the behavior of the $O(4)$ partial waves at small distances to the scaling properties of the Compton amplitude. These techniques were used to study a wide class of light-cone singularities suggested by ladder models.

In this work we discuss two kinds of expansion of the Compton amplitude; the first one, which is relevant in the Bjorken limit, is given by an integral over all possible light-cone singularities. This expansion is more transparent in momentum space where it appears as an expansion in terms of homogeneous functions of the variables q^2, ν, i.e., over irreducible representations of the group of projective transformations on the complex variables q^2, ν.[3] The second one, relevant in the Regge limit, is obtained by projecting over the irreducible representations of the Lorentz group.

We find in general that an infinite number of Lorentz poles "conspire" to build up a light-cone singularity, more precisely, a term like $1/(-x^2 + i\epsilon x_0)^\alpha$ related to the sequence of Lorentz poles which are located at $\lambda_n = 1 + \alpha - n$ and whose residues are polynomials of order n in the virtual photon mass. The possible nonpolynomiality of the residues should be interpreted as an indication that an infinite sequence of light-cone singularities contribute in Regge limit.

We point out that these poles have nothing to do with conventional Regge poles, i.e., with the behavior of the "structure functions" $f_i(x\cdot p)$ in configuration space for large $x\cdot p$, but have a pure kinematical origin, in the sense that they reflect the presence of a "power type" singularity $(x_+^2)^{-\alpha}$ of the current commutator near the light cone. In particular, if these poles occur at integer points, because of the Regge signature factor, they cannot contribute to the imaginary part of the off-mass-shell Compton amplitude. However they do contribute for nonintegral values, so these singularities could be physically relevant as a manifestation of noncanonical light-cone singularities.

As is well known, such singularities would correspond to renormalization effects of dimension in the operator product of the two currents.

Finally we remark that the techniques we develop may also be useful for studying dynamical situations suggested by some ladder models, in which these kinds of singularities are realized.

Sections 2 and 3 are devoted to studying the properties of the integral transforms we are led to introduce in order to derive the previously mentioned results. In particular the connection of the expan-

[12] T. Regge and J. A. Wheeler, Phys. Rev. **108**, 1063 (1957).
[13] F. Zerilli, Phys. Rev. D **2**, 2141 (1970).
[14] C. V. Vishveshwara, Phys. Rev. D **1**, 2870 (1970).
[15] L. A. Edelstein and C. V. Vishveshwara, Phys. Rev. D **1**, 3514 (1970).
[16] J. L. Anderson and J. M. Cohen, Astrophys. Space Sci. **9**, 146 (1970).
[17] J. M. Cohen and R. M. Wald, J. Math. Phys. **12**, 1845 (1971).
[18] See, e.g., R. Courant and D. Hilbert, *Methods of Mathematical Physics* (Interscience, New York, 1962), Vol II, p. 321.
[19] To verify explicitly that analyticity of the Kruskal components $f_{\mu;\nu} + f_{\nu;\mu}$ implies analyticity of f_μ, set $w^0 = u + v$, $w^1 = u - v$, [where u, v are the standard Kruskal coordinates of the Schwarzschild metric $g(0)$], and set $\psi_0 = f_u + f_v$, $\psi_1 = f_u - f_v$. Analyticity of $(f_{\mu;\nu} + f_{\nu;\mu})$ directly implies analyticity of $\partial/\partial w^0(\psi_0/A)$, $\partial/\partial w^1(\psi_1/A)$, and $(\partial\psi_0/\partial w^1 + \partial\psi_1/\partial w^0)$, where $A(w^0, w^1)$ is analytic in w^0 and w^1 in a region containing the exterior region and horizon of $g(0)$. [In terms of the Schwarzschild radial coordinate, r, $A = (32m/r)\exp(-r/2m)$.] This implies analyticity of ψ_0, ψ_1 and hence of f_u and f_v. With this established it is easy to verify that f_θ and f_φ must also be analytic.
[20] The series for $m(\alpha)$ must converge because the series for $g(\alpha)$ converges (by hypothesis) and $g(\alpha)$ must have finite mass. Similarly, the series for $a_i(\alpha)$ and $e(\alpha)$ must converge. For the case of $\xi_\mu(\alpha)$ one can show that the effect of the sequence of coordinate transformations $\sum_{n=1}^{N} \xi_\mu^{(n)}\alpha^n/n!$, $(N = 1, 2, \ldots)$, on $g(\alpha)$ must converge, but this does not imply in an immediately obvious fashion that the series itself must converge and, in fact, the remote possibility that it does not converge has not as yet been rigorously ruled out.
[21] J. Bekenstein (to be published).

Light-Cone Singularities and Lorentz Poles

S. Ferrara
Laboratori Nazionali del CNEN, Frascati, Italy
and
G. Rossi*
Istituto di Fisica dell'Università, Roma, and Istituto di Fisica dell'Università, L'Aquila, Italy
(Received 8 October 1971)

The causal Meyer–Suura structure functions are projected into irreducible representations of the Lorentz group. A clarification of the connection between light-cone singularities and Lorentz poles is obtained: We find that in general a light-cone singularity of the type $1/(-x^2 + i\epsilon x_0)^\alpha$, in the operator product of the hadronic electromagnetic current, is built up from a sequence of Lorentz poles at $\lambda_n = 1 + \alpha - n$ whose residues are polynomials of order n in the virtual photon square mass.

1. INTRODUCTION

In this paper we study the connection between scaling properties (light-cone singularities) and Regge-like behavior (J-plane singularities) of the off-mass-shell Compton amplitude. Recently many authors[1] suggested that presence of light-cone singularities in the commutator of the hadronic electromagnetic current, which are responsible for the scaling properties of the structure functions in the Bjorken limit, could imply the existence of fixed poles in the off-mass-shell forward Compton amplitude. To study this phenomenon, it is necessary to develop the harmonic analysis with respect to the Lorentz group in configuration space, showing explicitly how Lorentz poles contribute to build up light-cone singularities.

In Ref. 2 the Wick-rotated imaginary part of the off-mass-shell forward Compton amplitude has been subjected to an $O(4)$ analysis and the connection between its Lorentz pole content and light-cone singularities has been investigated by performing a Sommerfeld–Watson transform. The authors were able to relate the behavior of the $O(4)$ partial waves at small distances to the scaling properties of the Compton amplitude. These techniques were used to study a wide class of light-cone singularities suggested by ladder models.

In this work we discuss two kinds of expansion of the Compton amplitude; the first one, which is relevant in the Bjorken limit, is given by an integral over all possible light-cone singularities. This expansion is more transparent in momentum space where it appears as an expansion in terms of homogeneous functions of the variables q^2, ν, i.e., over irreducible representations of the group of projective transformations on the complex variables q^2, ν.[3] The second one, relevant in the Regge limit, is obtained by projecting over the irreducible representations of the Lorentz group.

We find in general that an infinite number of Lorentz poles "conspire" to build up a light-cone singularity, more precisely, a term like $1/(-x^2 + i\epsilon x_0)^\alpha$ related to the sequence of Lorentz poles which are located at $\lambda_n = 1 + \alpha - n$ and whose residues are polynomials of order n in the virtual photon mass. The possible nonpolynomiality of the residues should be interpreted as an indication that an infinite sequence of light-cone singularities contribute in Regge limit.

We point out that these poles have nothing to do with conventional Regge poles, i.e., with the behavior of the "structure functions" $f_i(x\cdot p)$ in configuration space for large $x\cdot p$, but have a pure kinematical origin, in the sense that they reflect the presence of a "power type" singularity $(x_+^2)^{-\alpha}$ of the current commutator near the light cone. In particular, if these poles occur at integer points, because of the Regge signature factor, they cannot contribute to the imaginary part of the off-mass-shell Compton amplitude. However they do contribute for nonintegral values, so these singularities could be physically relevant as a manifestation of noncanonical light-cone singularities.

As is well known, such singularities would correspond to renormalization effects of dimension in the operator product of the two currents.

Finally we remark that the techniques we develop may also be useful for studying dynamical situations suggested by some ladder models, in which these kinds of singularities are realized.

Sections 2 and 3 are devoted to studying the properties of the integral transforms we are led to introduce in order to derive the previously mentioned results. In particular the connection of the expan-

sions in momentum and configuration space (related by a Fourier transform) is given. In Sec. 4 the decomposition of a light-cone singularity in terms of Lorentz poles is carried out. The proofs of the main formulas we use are collected in the Appendix.

2. CONFORMAL TRANSFORM

Let us consider the functions[4]

$$V_1(q^2, \nu) = (1/q^2)[W_1(q^2, \nu) + (\nu^2/q^2)W_2(q^2, \nu)], \quad (2.1)$$

$$V_2(q^2, \nu) = -(1/q^2)W_2(q^2, \nu), \quad (2.2)$$

linear combinations of the structure functions $W_1(q^2, \nu)$, $W_2(q^2, \nu)$ defined by the equations

$$\begin{aligned} W_{\mu\nu}(q,p) &= \frac{1}{(2\pi)^4}\int d^4x e^{-iq\cdot x}\langle p|[J_\mu^{el}(x), J_\nu^{el}(0)]|p\rangle \\ &= -\left(g_{\mu\nu} - \frac{q_\mu q_\nu}{q^2}\right)W_1(q^2, \nu) + \left(p_\mu - \frac{\nu}{q^2}q_\mu\right) \\ &\quad \times \left(p_\nu - \frac{\nu}{q^2}q_\nu\right)W_2(q^2, \nu), \end{aligned} \quad (2.3)$$

where $J_\mu^{el}(x)$ is the hadronic electromagnetic current. For the Fourier transforms ($k = 1, 2$)

$$\begin{aligned} W_k^F(x^2, x\cdot p) &= \int d^4q e^{iq\cdot x}W_k(q^2, \nu), \\ V_k^F(x^2, x\cdot p) &= \int d^4q e^{iq\cdot x}V_k^F(q^2, \nu), \end{aligned} \quad (2.4)$$

the following relations hold:

$$\begin{aligned} W_1^F(x^2, x\cdot p) &= -\Box V_1^F(x^2, x\cdot p) - P_\mu P_\nu \partial^\mu\partial^\nu V_2^F(x^2, x\cdot p), \\ W_2'^F(x^2, x\cdot p) &= \Box V_2(x^2, x\cdot p). \end{aligned} \quad (2.5)$$

Experimental data suggest the following asymptotic behavior:

$$\lim_{\substack{\nu\to\infty \\ \omega=-q^2/2\nu \text{ fixed}}} V_k(q^2, \nu) \sim \nu^{\alpha_k - 2}F_k(\omega) \quad (2.6)$$

with $\alpha_1 = 1$, $\alpha_2 = 0$.

A typical contribution $\nu^{\alpha-2}F(\omega)$ (we shall omit the index k from now on) to the structure functions in the Bjorken limit corresponds in the Fourier transform to a term of the type $[x^2]^{-\alpha}f(x\cdot p)$ near the light cone (where we indicate by the symbol $[x^2]^{-\alpha}$ the discontinuity of $1/(-x^2 + i0)^\alpha$). The scaling behavior in momentum space has an interesting geometrical interpretation. Let us consider a function $V(q^2, \nu)$ defined on the (four-dimensional) complex affine plane q^2, ν. This space is homogeneous with respect to the group $SL(2, C)$ of projective transformations, i.e., it is equivalent to the quotient space $SL(2, C)/Z$ where Z is the group of matrices of the form $\left(\begin{smallmatrix} 1 & z \\ 0 & 1 \end{smallmatrix}\right)$. A representation of this group is defined on these functions as follows[5]:

$$T_g V(q^2, \nu) = V(\alpha q^2 + \gamma\nu, \beta q^2 + \delta\nu), \quad (2.7)$$

where $g = \left(\begin{smallmatrix} \alpha & \beta \\ \gamma & \delta \end{smallmatrix}\right)$ with $\alpha\delta - \beta\gamma = 1$.

We observe that the homogeneous functions play a special role in this space, as they form an irreducible subspace for the representation. The theory of harmonic analysis on homogeneous spaces gives the following expansion formula:

$$V(q^2, \nu) = \frac{1}{(2\pi)^2 i}\sum_{n=-\infty}^{+\infty}\int_{-i\infty}^{i\infty} d\alpha\, V(q^2, \nu; n, \alpha), \quad (2.8)$$

where $n_1 = (n + \alpha)/2$, $n_2 = (-n + \alpha)/2$, and $V(q^2, \nu; n, \alpha)$ is the Mellin transform of $V(q^2, \nu)$ defined as follows:

$$V(q^2, \nu; n, \alpha) = \tfrac{1}{2}\int d\eta d\bar\eta \eta^{-n_1}\bar\eta^{-n_2}V(\eta q^2, \eta\nu). \quad (2.9)$$

Equation (2.8) is only valid for L^2 functions, while Eq. (2.9) can be used to analytically continue the Mellin transform in α so that, for non-square-summable functions the expansion formula reads

$$V(q^2, \nu) = \frac{1}{(2\pi)^2 i}\sum_{n=-\infty}^{+\infty}\int_C d\alpha\, V(q^2, \nu; n, \alpha), \quad (2.10)$$

where C is a suitable path in the complex α plane.

If we introduce the Fourier transform as

$$V^F(x^2, u) = \int d^4q e^{iq\cdot x}V(q^2, \nu), \qquad u = x\cdot p, \quad (2.11)$$

then from the previous analysis we obtain in configuration space the expansion formula

$$V^F(x^2, u) = \frac{1}{(2\pi)^2 i}\sum_n \int_C d\alpha\, V^F(x^2, u; n, \alpha), \quad (2.12)$$

where

$$V^F(x^2, u) = \tfrac{1}{2}\int d\eta d\bar\eta \eta^{n_1}\bar\eta^{n_2 - 2}V^F(\eta x^2, u). \quad (2.13)$$

If we perform the sum over n, we get

$$V^F(x^2, u) = \frac{1}{2\pi i}\int_C d\alpha \hat V^F(x^2, u; \alpha), \quad (2.14)$$

with the definition

$$\hat V^F(x^2, u; \alpha) = \int_0^\infty dt\, t^{\alpha-1}V^F(tx^2, u). \quad (2.15)$$

Using the causality of the Fourier-transformed structure functions, we can write

$$V^F(x^2, u) = \frac{1}{2\pi i}\int_C d\alpha[x^2]^{-\alpha}f(u; \alpha), \quad (2.16)$$

where we have introduced the "conformal transform" of $V^F(x^2, u)$:

$$f(u; \alpha) = -\frac{1}{2i\sin\pi\alpha}\int_0^\infty d\sigma\, \sigma^{\alpha-1}V^F(\sigma, u) \quad (2.17)$$

and

$$[x^2]^\alpha = \text{disc}\,[1/(-x^2 + i0)^\alpha] = -2i\sin\pi\alpha(x_+^2)^{-\alpha}.$$

We have called the transform defined by Eq. (2.17) "conformal transform" to emphasize that a term like $[x^2]^{-\alpha}f(x\cdot p; \alpha)$ is related in the corresponding light-cone operatorial expansions to an infinite set of tensor operators $O_{\alpha_1\ldots\alpha_n}(0)$ classified according to a ladder of irreducible representations of the conformal algebra whose spectrum is given by the eigenvalues of the nonvanishing Casimir operator:

$$\begin{aligned} 2M_{\mu\nu}M^{\mu\nu} + 2P_\lambda\cdot K^\lambda - 2D^2 + 8iD \\ = 4n(n - \alpha - 1) + 2\alpha^2 - 8, \end{aligned}$$

and $M_{\mu\nu}, P_\lambda, K_\lambda, D$ are the generators of the conformal algebra.[6]

In momentum space we can write the expansion

$$V(q^2, \nu) = \frac{1}{2\pi i} \int_C d\alpha\, \nu^{\alpha-2} F(\omega; \alpha), \tag{2.18}$$

where the scaling function is expressed in terms of the conformal transform $f(u; \alpha)$ as

$$F(\omega, \alpha) = \frac{1}{(2\pi)^2}\, 2^{-\alpha} e^{-i\pi\alpha/2}\, \frac{\sin\pi\alpha}{\Gamma(\alpha)} \int_0^\infty du\, e^{i\omega u} u^{1-\alpha} f(u; \alpha). \tag{2.19}$$

To obtain the complete diagonalization of the expansions (2.16), (2.18), we define the Mellin transform with respect to the u variable of $f(u; \alpha)$;

$$f(\alpha, \tau) = \int_0^\infty du\, u^{\tau-1} f(u; \alpha), \tag{2.20}$$

so in configuration space we have

$$V(x^2, u) = \frac{1}{(2\pi i)^2} \int_C d\alpha \int_{C-i\infty}^{C+i\infty} d\tau [x^2]^{-\alpha} u^{-\tau} f(\alpha, \tau), \tag{2.21}$$

and in momentum space

$$V(q^2, \nu) = \frac{1}{(2\pi)^4} \int_C d\alpha \int_{C-i\infty}^{C+i\infty} d\tau 2^{2-2\alpha}\, e^{-i\pi\alpha/2}\, e^{-i\pi\tau/2} \times \frac{\sin\pi\alpha}{\Gamma(\alpha)}\, \Gamma(2-\alpha-\tau)(-q^2)^{\alpha-2} \omega^\tau f(\alpha, \tau), \tag{2.22}$$

where we have used the relation

$$F(\omega, \alpha) = \frac{1}{(2\pi)^3}\, 2^{-\alpha}\, e^{-i\pi\alpha/2}\, \frac{\sin\pi\alpha}{\Gamma(\alpha)}\, \omega^{\alpha-2} \times \int_{C-i\infty}^{C+i\infty} d\tau e^{-i\pi\tau/2}\, \Gamma(2-\alpha-\tau) \omega^\tau f(\alpha, \tau). \tag{2.23}$$

3. LORENTZ AND WEYL TRANSFORMS

In this section we will expand the amplitude in configuration space directly into irreducible representations of the Lorentz group. The complete diagonalization will be obtained in this case by means of the Weyl transform which is defined as the product of the Lorentz and Mellin transform in the x^2 variable. These transforms obviously commute. We start by projecting out the dependence of $\tilde{V}^F(x^2, \cosh\zeta_x) = V^F(x^2, x\cdot p)$ on $\cosh\zeta_x = (x\cdot p)/\sqrt{x^2}$ by performing the usual Lorentz transform[7]

$$\int_0^\infty \tilde{V}^F(x^2, \cosh\zeta_x)\, \mathfrak{D}_\lambda(\cosh\zeta_x)\, \sinh^2\zeta_x\, d\zeta_x = \tilde{V}^F_\lambda(x^2) \tag{3.1}$$

as defined by Eq. (A1).

The Plancherel theorem gives

$$\tilde{V}^F(x^2, \cosh\zeta_x) = \frac{i}{\pi} \int_{-i\infty}^{i\infty} d\lambda \lambda^2 \tilde{V}^F_\lambda(x^2)\, \mathfrak{D}_{-\lambda}(\cosh\zeta_x), \tag{3.2}$$

where the path has to be suitably shifted for non-L^2 functions. Possible behaviors of the Lorentz transform $\tilde{V}^F_\lambda(x^2)$ were studied in Ref. 2 in connection with simple structures suggested by ladder models. The corresponding expansion in momentum space is obtained, by computing the Lorentz transform of the Fourier kernel $e^{-iq\cdot x}$ [see Eq. (A5)], by means of the formula

$$\tilde{V}_\lambda(q^2) = \int dR\, R^3\, \frac{K_\lambda(R\sqrt{-q^2})}{R\sqrt{-q^2}}\, \tilde{V}^F_\lambda(R^2), \tag{3.3}$$

where $K_\lambda(x)$ is a modified Bessel function of third kind (Hankel function) and $R = \sqrt{x^2}$. The inversion formula reads

$$V(q^2, \nu) = 4\pi i \int_C d\lambda \lambda^2 \tilde{V}_\lambda(q^2)\, \mathfrak{C}_\lambda(\nu, q^2), \tag{3.4}$$

where $\mathfrak{C}_\lambda(\nu, q^2)$ is a second kind matrix element on the Lorentz group defined by Eq. (A7). We note that the partial wave $\tilde{V}_\lambda(q^2)$ defined by Eq. (3.3) could in principle have λ singularities originated by the Hankel function. This phenomenon is more transparent if we perform the complete diagonalization by means of the Mellin transform in the variable x^2. We define the Weyl transform as

$$\tilde{V}^F_{\lambda\rho} = \int\int \tilde{V}^F(x^2, \cosh\zeta_x)\, \mathfrak{D}_\lambda(\cosh\zeta_x) \times \sinh^2\zeta_x\, d\zeta_x (x^2)^{\rho-1} dx^2 \tag{3.5}$$

according to Eq. (A9). We call it a Weyl transform since it performs the diagonalization with respect to irreducible representations of the Weyl group. The inversion formula is given by

$$\tilde{V}^F(x^2, \cosh\zeta_x) = \frac{1}{2\pi^2} \int d\lambda \lambda^2 \int d\rho (x^2)^{-\rho}\, \mathfrak{D}_\lambda(\cosh\zeta_x) \tilde{V}^F_{\lambda\rho} \tag{3.6}$$

and in momentum space, via Eq. (A12), we have

$$V(q^2, \nu) = \frac{4}{\pi} \int d\lambda \int d\rho\, \lambda^2 2^{-2\rho}\, \Gamma(\tfrac{1}{2}\lambda - \rho + \tfrac{3}{2}) \times \Gamma(-\tfrac{1}{2} - \rho + \tfrac{3}{2})\, (-q^2)^{\rho-2}\, \mathfrak{C}_\lambda(\nu, q^2) \tilde{V}^F_{\lambda\rho}. \tag{3.7}$$

We observe that the paths of integration in the λ and ρ planes are along the imaginary axis for L^2 functions and they have to be suitably shifted for non-L^2 functions.

The Weyl transform is simply connected to the conformal transform introduced in the previous section. To see this, we start by rewriting Eq. (2.21) as

$$V(x^2, u) = \frac{2}{(2\pi i)^2} \int\int d\alpha d\rho [x^2]^{-\rho} (\cosh\zeta_x)^{2\alpha-2\rho} f(\alpha, 2\rho - 2\alpha) \tag{3.8}$$

after the change of variable $\tau = 2\rho - 2\alpha$. This connection clarifies the kinematical origin of the Lorentz-pole content of a light-cone contribution; this will be shown explicitly in the next section.

4. DECOMPOSITION OF A LIGHT-CONE SINGULARITY INTO LORENTZ POLE CONTRIBUTIONS

In this section we want to investigate the connection between the two integral representations for the causal structure functions $V^F(x^2, u)$, (2.21) and (3.6) [and their related momentum space versions (2.23) and (3.7)]. If we remember the structure of the Eq. (3.8), we see that the transformation function which relates the two expansions is nothing but the Lorentz transform of the power $(\cosh\zeta_x)^{2\rho-2\alpha}$. Its Lorentz transform is given by Eq. (A13). So we have in terms of irreducible Lorentz representations,

$$\tilde{V}^F(x^2, \cosh\zeta_x) = \frac{-1}{(2\pi i)^3} \iiint d\alpha\, d\rho d\lambda\, \lambda^2\, \frac{2^{2\rho-2\alpha-1}}{\Gamma(2\rho-2\alpha)} \times f(\alpha, 2\rho-2\alpha)\Gamma(\tfrac{1}{2}\lambda - \tfrac{1}{2} + \rho - \alpha) \times \Gamma(-\tfrac{1}{2}\lambda - \tfrac{1}{2} + \rho - \alpha)[x^2]^{-\rho}\, \mathfrak{D}_\lambda(\cosh\zeta_x) \tag{4.1}$$

and in momentum space, by means of Fourier transform,

$$V(q^2,\nu)=\frac{2}{(2\pi)^2 i}\iiint d\alpha d\rho d\lambda\lambda^2\frac{2^{-2\alpha}}{\Gamma(2\rho-2\alpha)}\times f(\alpha,2\rho-2\alpha)(-2i\sin\pi\alpha)\times\Gamma(\tfrac{1}{2}\lambda-\tfrac{1}{2}+\rho-\alpha)\,\Gamma(-\tfrac{1}{2}\lambda-\tfrac{1}{2}+\rho-\alpha)\times\Gamma(\tfrac{1}{2}\lambda-\rho+\tfrac{3}{2})\,\Gamma(-\tfrac{1}{2}\lambda-\rho+\tfrac{3}{2})(-q^2)^{\rho-2}\,\mathfrak{A}_\lambda(\nu,q^2)\tag{4.2}$$

Comparing Eqs. (3.7) and (4.2), we have the connection between the two previously introduced conformal and Weyl transforms:

$$\tilde V^F_{\lambda\rho}=\frac{1}{8\pi i}\int d\alpha\,\frac{2^{2\rho-2\alpha}}{\Gamma(2\rho-2\alpha)}\,\Gamma(\tfrac{1}{2}\lambda-\tfrac{1}{2}+\rho-\alpha)\times\Gamma(-\tfrac{1}{2}\lambda-\tfrac{1}{2}+\rho-\alpha)f(\alpha,2\rho-2\alpha)(-2i\sin\pi\alpha).\tag{4.3}$$

This result clearly means that a Weyl contribution is in general built up from an infinite sequence of light-cone singularities. In order to study the matching between light-cone singularities and Lorentz poles, we change the order of integration in the integral representation (4.2) and perform the λ-integration explicitly by means of Cauchy's theorem. We get

$$V(q^2,\nu)=\sum_{n=0}^{\infty}\frac{1}{\pi}\iint d\alpha d\rho\,\frac{(-1)^{n-1}}{(n-1)!}\,\frac{2^{-2\alpha}}{\Gamma(2\rho-2\alpha)}\times(-2i\sin\pi\alpha)f(\alpha,2\rho-2\alpha)\,\Gamma(2\rho-2\alpha-1+n)\times\Gamma(-\alpha+1+n)\Gamma(\alpha-2\rho+2-n)\times(2\rho-2\alpha-1+2n)^2(-q^2)^{\rho-2}\,\mathfrak{A}_{2\rho-2\alpha-1+2n}(\nu,q^2)+\sum_{n=0}^{\infty}\frac{1}{\pi}\iint d\alpha d\rho\,\frac{(-1)^{n-1}}{(n-1)!}\,\frac{2^{-2\alpha}}{\Gamma(2\rho-2\alpha)}(-2i\sin\pi\alpha)\times f(\alpha,2\rho-2\alpha)\,\Gamma(-\alpha+1+n)\times\Gamma(-\alpha+2\rho-2-n)\,\Gamma(-2\rho+3+n)\times(-2\rho+3+2n)^2(-q^2)^{\rho-2}\,\mathfrak{A}_{-2\rho+3+2n}(\nu,q^2),\tag{4.4}$$

where we have taken the contribution of the Lorentz poles at $\lambda=2\rho-2\alpha-1+2n$, $\lambda=-2\rho+3+2n$ and closed the integration path in the right half-plane, where the functions $\mathfrak{A}_\lambda(\nu,q^2)$ go to zero.

To see the Lorentz pole content of a light-cone contribution, we assume that the ρ integration can be performed by an appropriate deformation of the integration path in such a way that the Mellin transform is analytic in the corresponding region of the ρ plane. This corresponds, in Regge pole language, to considering the Compton amplitude with true Regge poles subtracted.[1] Note that a conventional Regge pole would correspond to a pole in the Mellin transform and would give rise to a different behavior in q^2 of the residue function.

Performing the integral, we get

$$V(q^2,\nu)=\sum_{n,m=0}^{\infty}4\int d\alpha\,\frac{(-1)^{n-1}}{(n-1)!}\,\frac{(-1)^{m-1}}{(m-1)!}\times\frac{2^{-2\alpha}\sin\pi\alpha}{\Gamma(-\alpha+2-n+m)}f(\alpha,-\alpha+2-n+m)\times\Gamma(-\alpha+1+n)\Gamma(-\alpha+1+m)\times(-\alpha+1+n+m)^2(-q^2)^{[(\alpha+2-n+m)/2]-2}\times\mathfrak{A}_{-\alpha+1+n+m}(\nu,q^2)+\sum_{n,m=0}^{\infty}4\int d\alpha\,\frac{(-1)^{n-1}}{(n-1)!}\times\frac{(-1)^{m-1}}{(m-1)!}\,\frac{2^{-2\alpha}\sin\pi\alpha}{\Gamma(-\alpha+2+n-m)}\times f(\alpha,-\alpha+2+n-m)\,\Gamma(-\alpha+1+n)\times\Gamma(-\alpha+1+m)(-\alpha+1+n+m)^2\times(-q^2)^{[(\alpha+2+n-m)/2]-2}\,\mathfrak{A}_{-\alpha+1+n+m}(\nu,q^2),\tag{4.5}$$

where we have taken the contributions of the poles at $\rho=(-\alpha+2-n+m)/2$, $\rho=(2+\alpha+n-m)/2$ in the first and second integral of (4.5), respectively. The integral has been closed in such a way that the background goes to zero by moving the integration path at infinity.

Formula (4.5) can be rewritten as

$$V(q^2,\nu)=\sum_{n,m=0}^{\infty}4\int d\alpha\left(\frac{f(\alpha,-\alpha+2-n+m)}{\Gamma(-\alpha+2-n+m)}\times(-q^2)^{[(\alpha+2-n+m)/2]-2}+n\rightleftarrows m\right)\frac{(-1)^{n-1}}{(n-1)!}\times\frac{(-1)^{m-1}}{(m-1)!}\,2^{-2\alpha}\sin\pi\alpha\,\Gamma(-\alpha+1+m)\times\Gamma(-\alpha+1+m)(-\alpha+1+n+m)^2\times\mathfrak{A}_{-\alpha+1+n+m}(\nu,q^2).\tag{4.6}$$

To see the behavior of a light-cone singularity, we pick up a contribution to the integrand in (4.6). In the Regge limit we have

$$V_\alpha(q^2,\nu)\underset{\nu\to\infty}{=}\sum_{n,m=0}^{\infty}-4\left(\frac{f(\alpha,-\alpha+2-n+m)}{\Gamma(-\alpha+2-n+m)}\times(-q^2)^m+n\rightleftarrows m\right)\frac{(-1)^{n-1}}{(n-1)!}\,\frac{(-1)^{m-1}}{(m-1)!}\times2^{-2\alpha}\sin\pi\alpha\,\Gamma(-\alpha+1+n)\times\Gamma(-\alpha+1+m)(-\alpha+1+n+m)\times e^{i\pi(2-\alpha+n+m)/2}\,\nu^{\alpha-2-n-m},\tag{4.7}$$

where we have used the asymptotic behavior of the functions $\mathfrak{A}_\lambda(\nu,q^2)$ defined in Eq. (A7). In particular we observe that the Γ functions exactly cancel each other for $\alpha=$ negative integer, corresponding to the case of a derivative of a δ singularity on the light-cone. In particular the leading pole corresponding to the α-light-cone singularity goes as

$$V_\alpha(q^2,\nu)\underset{\nu\to\infty}{\sim}\text{const}\cdot\nu^{\alpha-2}\tag{4.8}$$

with residue independent of q^2. In general the residue of the nth nonleading poles is a polynomial in q^2 of order n. These results establish that a light-cone singularity $(x_+^2)^{-\alpha}$ in the current commutator corresponds to a sequence of poles at $J_n=\alpha-n$ whose residues are polynomial of order n in the photon square mass.

APPENDIX

In this appendix we recall some formula which we need in the text. We start by recalling that the causal structure function $V^F(x^2, x\cdot p) = \tilde{V}^F(x^2, \cosh\zeta_x)$ $(\cos\zeta_x = x\cdot p/\sqrt{x^2})$ can be considered, for fixed x^2, as a bicovariant function[8] defined over $SL(2, C)$ (the universal covering group of the Lorentz group). This means that it is a function $\tilde{V}(x^2, a), a \in SL(2, C)$, which satisfies the covariance relation $\tilde{V}(x^2, a) = \tilde{V}(x^2, h_1 a h_2)$ for $h_1, h_2 \in SU(2)$. Its Lorentz transform is given by the formula

$$\int_{SL(2,C)} \tilde{V}^F(x^2, a)\mathfrak{D}^{0\lambda}_{0000}(a)d^6a = 4\pi^3\int_0^\infty \tilde{V}^F(x^2, \cosh\zeta_x) \times \mathfrak{D}_\lambda(\cosh\zeta_x)\sinh^2\zeta_x d\zeta_x = 4\pi^3\tilde{V}^F_\lambda(x^2), \quad \text{(A1)}$$

where

$$\mathfrak{D}_\lambda(\cosh\zeta_x) = d^{0\lambda}_{000}(\cosh\zeta_x) = (\sinh\lambda\zeta_x)/(\lambda\sinh\zeta_x) \quad \text{(A2)}$$

is a matrix element of an irreducible representation of the type $(0, \lambda)$. This function is called an elementary spherical harmonic of $SL(2, C)$. The Plancherel theorem gives

$$\tilde{V}^F(x^2, \cosh\zeta_x) = \frac{i}{\pi}\int_{-i\infty}^{i\infty} d\lambda\lambda^2\tilde{V}^F_\lambda(x^2)\mathfrak{D}_\lambda(\cosh\zeta_x) \quad \text{(A3)}$$

for functions L^2 over $SL(2, C)$. For non-L^2 functions the integration path must be suitably shifted.

Computation of the Lorentz transform of the inverse Fourier transform

$$V(q^2, \nu) = \int e^{-iq\cdot x}V^F(x^2, x\cdot p)d^4x \quad \text{(A4)}$$

requires the knowledge of the Lorentz transform of the Fourier kernel $e^{-iq\cdot x}$. This has been evaluated in Ref. 9, and we get

$$\int e^{-iq\cdot x}\mathfrak{D}^{0\lambda}_{0000}(a)d^6a = 2\pi^2(1/R\sqrt{-q^2})K_\lambda(R\sqrt{-q^2})\mathfrak{D}_\lambda(\nu, q^2), \quad \text{(A5)}$$

where $K_\lambda(x)$ is a modified Bessel function of third kind[10] and

$$\mathfrak{D}_\lambda(\nu, q^2) = d^{0\lambda}_{000}(\nu/\sqrt{q^2}) \quad \text{(A6)}$$

is a matrix element of a $(0, \lambda)$ representation of $SL(2, C)$ continued to imaginary values of $\cosh\zeta_q$ by means of the formulas

$$\mathfrak{D}_\lambda(\nu, q^2) = \mathfrak{C}_\lambda(\nu, q^2) - \mathfrak{C}_{-\lambda}(\nu, q^2),$$

$$\mathfrak{C}_{-\lambda}(\nu, q^2) = \frac{1}{\lambda}\frac{1}{\sqrt{\nu^2 - q^2}}(\nu + \sqrt{\nu^2 - q^2})^\lambda \times (-q^2)^{(1-\lambda)/2}e^{i\pi(1-\lambda)/2}. \quad \text{(A7)}$$

At this point we have to make a remark: In principle the projection of the Fourier kernel $e^{-iq\cdot x}$, which acts from a homogeneous space of the kind $SL(2, C)/SU(2)$ $(x^2 > 0)$ to one of the kind $SL(2, C)/SU(1, 1)$ $(q^2 < 0)$ receives contributions also from irreducible representations of the type $(M, 0)$. Nevertheless, as explained in Ref. 11, the Plancherel measure in the inversion formula in momentum space has a support $\Omega_{x,q}$ which is the intersection of the supports Ω_x, Ω_q of the Plancherel measures on the two different homogeneous spaces so

$$\Omega_{x,q} = \Omega_x \cap \Omega_q = (0, \lambda) \cap [(0, \lambda) + (M, 0)] = (0, \lambda).$$

We then obtain the Lorentz expansion in q-space in the form

$$V(q^2, \nu) = 2\pi i\int d\lambda\lambda^2\tilde{V}_\lambda(q^2)\mathfrak{D}_\lambda(\nu, q^2), \quad \text{(A8)}$$

where

$$\tilde{V}_\lambda(q^2) = \int dR\, R^3\{[K_\lambda(R\sqrt{-q^2})]/(R\sqrt{-q^2})\}\tilde{V}^F_\lambda(R^2)$$

and we have used the formula

$$\theta(x^2)\theta(x^0)d^4x = (1/\pi)R^3dRd^3X,$$

where d^3X is the invariant measure over $SL(2, C)/SU(2)$ and is defined by the formula

$$d^3ud^3X = d^6a,$$

d^3u, d^6a being the Haar measures over $SU(2), SL(2, C)$, respectively.

We observe that, to obtain complete diagonalization of the Lorentz expansion, we have to perform the Mellin transform in the variable $R = \sqrt{x^2}$. If we define the Weyl transform as

$$4\pi^3\tilde{V}^F_{\lambda\rho} = \int\tilde{V}^F(x^2, \cosh\zeta_x)\mathfrak{D}_\lambda(a)(x^2)^{\rho-1}d^6a\,dx^2, \quad \text{(A9)}$$

we get

$$V^F(x^2, x\cdot p) = \frac{1}{2\pi^2}\int_{-i\infty}^{i\infty} d\lambda\lambda^2 \times \int_{C-i\infty}^{C+i\infty} d\rho(x^2)^{-\rho}\mathfrak{D}_\lambda(\cosh\zeta_x)\tilde{V}^F_{\lambda\rho} \quad \text{(A10)}$$

and the corresponding expansion in momentum space is

$$V(q^2, \nu) = \frac{2}{\pi}\int_{-i\infty}^{i\infty} d\lambda\int_{C-i\infty}^{C+i\infty} d\rho\lambda^2\tilde{V}^F_{\lambda\rho}2^{-2\rho}(-q^2)^{\rho-2} \times \Gamma(\tfrac{1}{2}\lambda - \rho + \tfrac{3}{2})\Gamma(-\tfrac{1}{2}\lambda - \rho + \tfrac{3}{2})\mathfrak{D}_\lambda(\nu, q^2). \quad \text{(A11)}$$

To derive the last equation, we performed the Mellin transform of the function[12] $K_\lambda(x)$

$$\int\frac{dx}{x}x^{-2\rho+3}K_\lambda(x) = 2^{-2\rho+1}\Gamma(\tfrac{1}{2}\lambda - \rho + \tfrac{3}{2})\Gamma(-\tfrac{1}{2} - \rho + \tfrac{3}{2}). \quad \text{(A12)}$$

Finally we perform the Weyl transform of a light-cone singularity contribution. This is necessary to relate the expansions introduced in Secs. 2 and 3. We get[13]

$$\int(1/x^2)^\alpha f(x\cdot p, \alpha)\mathfrak{D}_\lambda(a)(x^2)^{\rho-1}d^6a\,dx^2 = 2\int dx\, x^{-2\alpha+2\rho-1}f(x, \alpha)\int(\cosh\zeta_x)^{2\alpha-2\rho}\mathfrak{D}_\lambda(a)d^6a = \pi^3 f(\alpha, -2\alpha + 2\rho)[2^{2\rho-2\alpha}/\Gamma(2\rho - 2\alpha)] \times \Gamma(\tfrac{1}{2}\lambda - \tfrac{1}{2} + \rho - \alpha)\Gamma(-\tfrac{1}{2}\lambda - \tfrac{1}{2} + \rho - \alpha). \quad \text{(A13)}$$

---

* Present address: Université de Genève (Switzerland).

1 H. Leutwyler and J. Stern, Phys. Letters **31B**, 458 (1970); J. M. Cornwall, D. Corrigan, and R. E. Norton, Phys. Rev. D **3**, 536 (1971).

2 R. Gatto and P. Menotti, Nuovo Cimento **2A**, 881 (1971).

3 S. Ferrara and G. Rossi, Nuovo Cimento Lett. **4**, 408 (1970), and Nuovo Cimento **4A**, 851 (1971).

[4] J. W. Meyer and H. Suura, Phys. Rev. **160**, 1366 (1967).
[5] I. M. Gel'fand, M. I. Graev, and N. Ya. Vilenkin, *Generalized Functions* (Academic, New York, 1966), Vol. 5, Chap. VI.
[6] S. Ferrara, R. Gatto, and A. F. Grillo, Nucl. Phys. **B34**, 349 (1971).
[7] M. Toller, Nuovo Cimento **53A**, 671 (1968).
[8] W. Rühl, *The Lorentz Group and Harmonic Analysis* (Benjamin, New York, 1970).
[9] J. S. Zmuidzinas, J. Math. Phys. **7**, 761 (1966).
[10] A. Erdélyi, *et al.*, *Higher Trascendental Functions* (McGraw-Hill, New York, 1953), Vol. 2, Chap. VII, Sec. VII. 2. 2.
[11] S. Ferrara and G. Mattioli, Ann. Phys. (N.Y.) **59**, 444 (1970).
[12] A. Erdélyi, *et al.*, *Table of Integral Transforms* (McGraw-Hill, New York, 1954), Vol. I, p. 331.
[13] W. Rühl, Commun. Math. Phys. **10**, 199 (1968).

Sliced Extensions, Irreducible Extensions, and Associated Graphs: An Analysis of Lie Algebra Extensions. I. General Theory

U. Cattaneo*
Institut de Physique Théorique, Université Catholique de Louvain, Belgium
and
Instituut voor Theoretische Fysika, Katholieke Universiteit, Nijmegen, The Netherlands
(Received 6 August 1971; Revised Manuscript Received 29 November 1971)

The Lie algebra extension problem is investigated and incorporated into a formalism which allows for a visualization of the algebraic structures involved. A finer analysis than the usual mathematical one is sometimes required for the physical applications. This brings about the consideration of sliced extensions, i.e., of extensions provided with sections. Some of these, the ω-sliced extensions, are particularly interesting. They are directly connected with natural Levi decompositions of the Lie algebras obtained from extensions. Graphs are associated with ω-sliced extensions. This is especially suitable for the study of irreducible extensions, which are the basic ones among the extensions. Their structure may be described in terms of graph theory. A subset is picked out from the set of all irreducible extensions of an arbitrary Lie algebra. Its elements, the primitive extensions, have the simplest extension structure and are characterized by the empty graph or by one-vertex graphs.

INTRODUCTION

In the past last decade several papers have considered the group extension problem in physics, following Michel's nice exposition of the usefulness of group extensions in quantum mechanics at the Istanbul Summer School, 1962.[1] Some important contributions to the problem were given by Michel himself,[2–5] especially for extensions of the Poincaré group.

But, already in his paper on continuous unitary representations of the Poincaré group,[6] Wigner had implicitly solved a group extension problem, determining all equivalence classes of analytic central extensions of the universal covering group of $P_+^\uparrow$ by $U(1)$. As is now well known, there is only one such equivalence class, that of trivial extensions. This problem was generalized by Bargmann,[7] who studied the continuous unitary projective representations of simply connected topological groups. For the exceptional case of Lie groups, Bargmann reduced the group problem to an equivalent Lie algebra extension problem. As noted by Galindo,[8] this is possible in the case studied by Bargmann, but not for every analytic group extension. There are extensions of the Lie algebra of an analytic group (i.e., a connected Lie group) which do not induce corresponding analytic group extensions. Anyway, Lie algebra extensions can be of much help in the study of analytic group extensions, since no topological difficulties are involved, and they give a good deal of information for the solution of the latter problem.[9] Galindo studied some Lie algebra extensions of the Poincaré algebra and his study was pursued in Refs. 10 and 11.

The purpose of this paper is the study of Lie algebra extensions in a constructive way. As physicists are interested in the determination of some Lie algebra extensions, we look for the tools in order to accomplish this task. We derive some conditions which must be satisfied by the considered extensions, and give a procedure for the construction. In Paper I we formulate general results, valid for any Lie algebra extension, and we analyze more extensively the case of some types of Lie algebras which frequently appear in physical problems. In the Paper II we will illustrate the results of the present paper in the case of the Lie algebras of the Euclidean, Galilean, and Poincaré groups.

Following Calabi,[12] we introduce in Sec. 1 the idea of sliced extension,[13] and examine the physical meaning of this definition. The fundamental problems of Lie algebra extension theory are then formulated and we give a procedure for the solution. Given an arbitrary extension $(\mathcal{E}, \rho)$, we can consider some particularly important related sliced extensions, the ω-sliced extensions. These give directly information about the structure of the Lie algebra $\mathcal{E}$. From the set of extensions of any Lie algebra we pick out the primitive extensions, which are the basic elements of the set. They are all extensions by Abelian Lie algebras.

Section 2 is mainly devoted to the study of irreducible extensions, which form the core of any extension. In this case, conditions are given which must be satisfied by the Lie algebras coming into the picture. We associate with an ω-sliced extension one graph which allows for a visualization of these requirements.

The paper ends with two appendices. In Appendix A we give examples of some peculiarities of the extensions by non-Abelian Lie algebras. A theorem permitting the selection of a useful complete set of 2-pseudococycles is proved in Appendix B.

Notation And Some Conventions

We consider exclusively vector spaces, modules, and Lie algebras of *finite dimension* over a fixed, but arbitrarily chosen, field **F** of characteristic 0. This is tacitly understood throughout the paper.

Capital script letters $\mathfrak{A}, \mathfrak{B}, \cdots$ denote Lie algebras, and the corresponding capital italic characters A, $B, \cdots$ their underlying vector spaces. If φ is a (linear) representation of a Lie algebra $\mathcal{G}$ on a vector space V, then V_φ is the $\mathcal{G}$-module associated with φ. The $\mathcal{G}$-module V_0 (associated with the trivial representation $\varphi = 0$) is said to be trivial, and by a basis of V_φ we understand a basis of V. If $\mathfrak{A}$ is a subset of elements of some vector space, $\mathfrak{A}$ denotes the

[4] J. W. Meyer and H. Suura, Phys. Rev. **160**, 1366 (1967).
[5] I. M. Gel'fand, M. I. Graev, and N. Ya. Vilenkin, *Generalized Functions* (Academic, New York, 1966), Vol. 5, Chap. VI.
[6] S. Ferrara, R. Gatto, and A. F. Grillo, Nucl. Phys. **B34**, 349 (1971).
[7] M. Toller, Nuovo Cimento **53A**, 671 (1968).
[8] W. Rühl, *The Lorentz Group and Harmonic Analysis* (Benjamin, New York, 1970).
[9] J. S. Zmuidzinas, J. Math. Phys. **7**, 761 (1966).
[10] A. Erdélyi, *et al.*, *Higher Trascendental Functions* (McGraw-Hill, New York, 1953), Vol. 2, Chap. VII, Sec. VII. 2. 2.
[11] S. Ferrara and G. Mattioli, Ann. Phys. (N.Y.) **59**, 444 (1970).
[12] A. Erdélyi, *et al.*, *Table of Integral Transforms* (McGraw-Hill, New York, 1954), Vol. I, p. 331.
[13] W. Rühl, Commun. Math. Phys. **10**, 199 (1968).

Sliced Extensions, Irreducible Extensions, and Associated Graphs: An Analysis of Lie Algebra Extensions. I. General Theory

U. Cattaneo*
Institut de Physique Théorique, Université Catholique de Louvain, Belgium
and
Instituut voor Theoretische Fysika, Katholieke Universiteit, Nijmegen, The Netherlands
(Received 6 August 1971; Revised Manuscript Received 29 November 1971)

The Lie algebra extension problem is investigated and incorporated into a formalism which allows for a visualization of the algebraic structures involved. A finer analysis than the usual mathematical one is sometimes required for the physical applications. This brings about the consideration of sliced extensions, i.e., of extensions provided with sections. Some of these, the ω-sliced extensions, are particularly interesting. They are directly connected with natural Levi decompositions of the Lie algebras obtained from extensions. Graphs are associated with ω-sliced extensions. This is especially suitable for the study of irreducible extensions, which are the basic ones among the extensions. Their structure may be described in terms of graph theory. A subset is picked out from the set of all irreducible extensions of an arbitrary Lie algebra. Its elements, the primitive extensions, have the simplest extension structure and are characterized by the empty graph or by one-vertex graphs.

INTRODUCTION

In the past last decade several papers have considered the group extension problem in physics, following Michel's nice exposition of the usefulness of group extensions in quantum mechanics at the Istanbul Summer School, 1962.[1] Some important contributions to the problem were given by Michel himself,[2–5] especially for extensions of the Poincaré group.

But, already in his paper on continuous unitary representations of the Poincaré group,[6] Wigner had implicitly solved a group extension problem, determining all equivalence classes of analytic central extensions of the universal covering group of $P_+^\uparrow$ by $U(1)$. As is now well known, there is only one such equivalence class, that of trivial extensions. This problem was generalized by Bargmann,[7] who studied the continuous unitary projective representations of simply connected topological groups. For the exceptional case of Lie groups, Bargmann reduced the group problem to an equivalent Lie algebra extension problem. As noted by Galindo,[8] this is possible in the case studied by Bargmann, but not for every analytic group extension. There are extensions of the Lie algebra of an analytic group (i.e., a connected Lie group) which do not induce corresponding analytic group extensions. Anyway, Lie algebra extensions can be of much help in the study of analytic group extensions, since no topological difficulties are involved, and they give a good deal of information for the solution of the latter problem.[9] Galindo studied some Lie algebra extensions of the Poincaré algebra and his study was pursued in Refs. 10 and 11.

The purpose of this paper is the study of Lie algebra extensions in a constructive way. As physicists are interested in the determination of some Lie algebra extensions, we look for the tools in order to accomplish this task. We derive some conditions which must be satisfied by the considered extensions, and give a procedure for the construction. In Paper I we formulate general results, valid for any Lie algebra extension, and we analyze more extensively the case of some types of Lie algebras which frequently appear in physical problems. In the Paper II we will illustrate the results of the present paper in the case of the Lie algebras of the Euclidean, Galilean, and Poincaré groups.

Following Calabi,[12] we introduce in Sec. 1 the idea of sliced extension,[13] and examine the physical meaning of this definition. The fundamental problems of Lie algebra extension theory are then formulated and we give a procedure for the solution. Given an arbitrary extension $(\mathcal{E}, \rho)$, we can consider some particularly important related sliced extensions, the ω-sliced extensions. These give directly information about the structure of the Lie algebra $\mathcal{E}$. From the set of extensions of any Lie algebra we pick out the primitive extensions, which are the basic elements of the set. They are all extensions by Abelian Lie algebras.

Section 2 is mainly devoted to the study of irreducible extensions, which form the core of any extension. In this case, conditions are given which must be satisfied by the Lie algebras coming into the picture. We associate with an ω-sliced extension one graph which allows for a visualization of these requirements.

The paper ends with two appendices. In Appendix A we give examples of some peculiarities of the extensions by non-Abelian Lie algebras. A theorem permitting the selection of a useful complete set of 2-pseudococycles is proved in Appendix B.

Notation And Some Conventions

We consider exclusively vector spaces, modules, and Lie algebras of *finite dimension* over a fixed, but arbitrarily chosen, field **F** of characteristic 0. This is tacitly understood throughout the paper.

Capital script letters $\mathfrak{A}, \mathfrak{B}, \cdots$ denote Lie algebras, and the corresponding capital italic characters A, $B, \cdots$ their underlying vector spaces. If φ is a (linear) representation of a Lie algebra $\mathcal{G}$ on a vector space V, then V_φ is the $\mathcal{G}$-module associated with φ. The $\mathcal{G}$-module V_0 (associated with the trivial representation $\varphi = 0$) is said to be trivial, and by a basis of V_φ we understand a basis of V. If $\mathfrak{A}$ is a subset of elements of some vector space, $\mathfrak{A}$ denotes the

vector space generated by $\mathfrak{A}$ and, if V and W are subspaces of the underlying vector space of a Lie algebra, $[V, W]$ is defined as $\overline{\{[v,w] | v \in V, w \in W\}}$.[14] Suppose that an equivalence relation R is defined in a set $\mathfrak{S}$. Then, if $x \in \mathfrak{S}$, we write $\underline{x}$ for the equivalence class of x, tacitly understanding "modulo R" if no misunderstanding is possible.

The direct sum (resp. the direct product) of vector spaces, modules, and Lie algebras is symbolized by $\oplus$ (resp. by $\times$), and $\mathfrak{A} \dotplus \mathfrak{B}$ means a semidirect sum of $\mathfrak{A}$ and $\mathfrak{B}$, with ideal $\mathfrak{B}$. If both $\mathfrak{A}$ and $\mathfrak{B}$ are sets, resp. vector spaces, Lie algebras, or $\mathcal{G}$-modules, then $\mathfrak{A} \to \mathfrak{B}$ (morphism of $\mathfrak{A}$ into $\mathfrak{B}$) denotes a map, resp. a homomorphism of vector spaces, Lie algebras, or $\mathcal{G}$-modules. $\twoheadrightarrow$, $\rightarrowtail$ and $\twoheadrightarrow$ (or $\approx$) mean epi-, resp. mono- and isomorphisms (i.e., surjective, resp. injective and bijective morphisms). If σ is a linear map of $\mathfrak{A}$ into $\mathfrak{B}$, we write σ_v for the induced morphism $A \to B$.

The following symbols are also used:

$\subset$: proper set inclusion;

$\mathbf{N} = \mathbf{N}^* \cup \{0\}$: set of positive integers;

$\mathfrak{C}(\mathfrak{A})$: center of $\mathfrak{A}$;

$\mathrm{D}(\mathfrak{A})$, $\mathrm{I}(\mathfrak{A})$, $\Delta(\mathfrak{A}) = \mathrm{D}(\mathfrak{A})/\mathrm{I}(\mathfrak{A})$: Lie algebras of, respectively, all derivations, inner derivations, and outer derivations of $\mathfrak{A}$;

$\mathrm{L}(\mathfrak{A}, \mathfrak{B})$: vector space of all linear maps of $\mathfrak{A}$ into $\mathfrak{B}$;

$\hom(\mathfrak{A}, \mathfrak{B})$: set of all Lie algebra homomorphisms of $\mathfrak{A}$ into $\mathfrak{B}$;

$\mathcal{GL}(V)$: Lie algebra of all endomorphisms of V.

We refer to Refs. 10 and 11 for the basic concepts of Lie algebra extension theory. In particular, by an extension of $\mathfrak{B}$ by $\mathfrak{A}$ we mean an ordered pair $(\mathcal{E}, \rho)$, where $\rho: \mathcal{E} \twoheadrightarrow \mathfrak{B}$ and $\mathfrak{A} = \mathrm{Ker}\, \rho$.

1. SLICED AND PRIMITIVE EXTENSIONS

A. Sliced Extensions and Their Equivalence Classes

The concept of equivalent extensions is a fundamental one in the mathematical theory of Lie algebra extensions. In fact, the principal problem of this theory is the analysis of the set $\mathrm{ext}(\mathfrak{B}, \mathfrak{A})$ of equivalence classes of extensions of $\mathfrak{B}$ by $\mathfrak{A}$, for a given pair of Lie algebras $\mathfrak{B}$ and $\mathfrak{A}$. Notice that all extensions of an equivalence class have the same character,[11] and therefore we can consider the sets $\mathrm{ext}(\mathfrak{B}, \mathfrak{A}, \Phi)$ of equivalence classes of extensions of $\mathfrak{B}$ by $\mathfrak{A}$ with character Φ for any $\Phi \in \hom(\mathfrak{B}, \Delta(\mathfrak{A}))$. The union of these sets is $\mathrm{ext}(\mathfrak{B}, \mathfrak{A})$.

Also for physical problems the equivalence classes of extensions can be the relevant objects. For example, the study of continuous unitary projective representations of connected Lie groups gives rise to the problem of central Lie algebra extensions by the Lie algebra of $U(1)$.[7,15] Two equivalent extensions are related to associated projective representations, whereas two inequivalent extensions give nonsimilar projective representations. This is the root of the well-known Bargmann's superselection rule of the nonrelativistic mass. Another example of the relevance of equivalence classes of extensions, but now in the case of (abstract) group extensions, was given by Lurçat and Michel.[16] They found that, between the equivalence classes of central extensions of $P_+^\uparrow$ by the unitary gauge group whose infinitesimal generators are the operators of the super-conserved charges (electric, baryonic, and total leptonic), only one is realized in nature. The "physical extensions" are characterized by the relation

$$(-1)^{2j} = (-1)^{b+l},$$

where j is the total angular momentum of any state of a system of particles, b its baryonic number, and l its total leptonic number ($l = l_\mu + l_e$).

But the pertinence of the mathematical partition in classes of equivalent extensions to the physical reality depends strongly on the problem considered. Occasionally, a coarser subdivision can be suitable, for instance, that given by the consideration of extension types.[10,11] We have an example of this in the case of group extensions: The crystallographic space groups can be obtained from extensions of point groups by free Abelian groups,[17] and in crystallography isomorphic space groups are usually identified. Sometimes a finer partition of the set of extensions can be needed in order to give the solution of a physical problem. In fact, even if we know that an extension $(\mathcal{E}, \rho)$ of $\mathfrak{B}$ by $\mathfrak{A}$ appears in a given problem, this is, in general, still not sufficient for the physical interpretation of the result: One must know the connection of $\mathfrak{B}$ and $\mathfrak{A}$ in $\mathcal{E}$. This was emphasized by Michel[3,4] in the case of group extensions, and the same reasons fit obviously in the algebraic case too. An example of the occurence of such a problem will be published elsewhere.[18]

The relationship of $\mathfrak{B}$ and $\mathfrak{A}$ can be established by considering, together with $(\mathcal{E}, \rho)$, also a section σ of $(\mathcal{E}, \rho)$ over $\mathfrak{B}$,[10] and constructing then a new object, a sliced extension.[12]

Definition 1: *A sliced extension of* $\mathfrak{B}$ *by* $\mathfrak{A}$ (with character Φ) related to $(\mathcal{E}, \rho)$ is an ordered triple $(\mathcal{E}, \rho, \sigma)$, where $(\mathcal{E}, \rho)$ is an extension of $\mathfrak{B}$ by $\mathfrak{A}$ (with character Φ) and σ a section of $(\mathcal{E}, \rho)$ over $\mathfrak{B}$.

Any sliced extension $(\mathcal{E}, \rho, \sigma)$ of $\mathfrak{B}$ by $\mathfrak{A}$ determines uniquely a natural decomposition

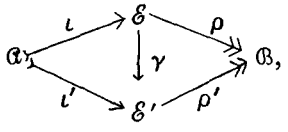
$$E = \sigma_v(B) \oplus A$$

of the vector space E underlying $\mathcal{E}$, where B is isomorphic to $\sigma_v(B)$ by σ_v. In fact, σ_v is a monomorphism by the definition of section.

For any extension $(\mathcal{E}, \rho)$ of $\mathfrak{B}$ by $\mathfrak{A}$ we have a collection of sliced extensions: one for any section of $(\mathcal{E}, \rho)$ over $\mathfrak{B}$. Consequently, there are a lot of possibilities for the connection of $\mathfrak{B}$ and $\mathfrak{A}$ in $\mathcal{E}$. In general, only few of these are significant for a physical problem. The example in Ref. 18 is an illustration of this assertion: The charge and hypercharge operators for a hadronic system are obtained from two different sliced extensions of the one-dimensional Lie algebra of the electric charge by $\mathcal{SU}(2)_t$ (the Lie algebra of isospin). Actually, in this problem we have two physically relevant equivalence classes of sliced extensions, according to the following.

Definition 2: The sliced extensions $(\mathcal{E}, \rho, \sigma)$ and $(\mathcal{E}', \rho', \sigma')$ of $\mathfrak{B}$ by $\mathfrak{A}$ are said to be *equivalent*, if there

is a Lie algebra homomorphism γ of $\mathcal{E}$ into $\mathcal{E}'$ such that $\sigma' = \gamma\circ\sigma$ and the following diagram is commutative,



ι and ι' being the canonical monomorphisms.

If $(\mathcal{E},\rho,\sigma)$ and $(\mathcal{E}',\rho',\sigma')$ are equivalent, we write $(\mathcal{E},\rho,\sigma)\overset{\gamma}{\equiv}(\mathcal{E}',\rho',\sigma')$, where γ is the homomorphism (actually an isomorphism) which establishes the equivalence according to Definition 2. This requires $(\mathcal{E},\rho)\overset{\gamma}{\equiv}(\mathcal{E}',\rho')$ too, i.e., the equivalence of $(\mathcal{E},\rho)$ and $(\mathcal{E}',\rho')$ by γ.[10] It is easy to see that Definition 2 gives rise to an equivalence relation in the set of sliced extensions.

With a sliced extension $(\mathcal{E},\rho,\sigma)$ of $\mathfrak{B}$ by $\mathfrak{A}$, with character Φ, is associated one and only one $(2;\mathfrak{B},\mathfrak{A},\Phi)$-pseudococycle[11] (φ,f_2) defined by

$$f_2(b,b') = [\sigma(b),\sigma(b')] - \sigma([b,b']),\quad \varphi(b) = \mathrm{ad}_{\mathfrak{A}}\sigma(b),$$

$\mathrm{ad}_{\mathfrak{A}}\colon \mathcal{E}\to \mathrm{D}(\mathfrak{A})$ being defined by

$$(\mathrm{ad}_{\mathfrak{A}}e)a = [e,a] \quad \text{for all } a\in\mathfrak{A}.$$

We call (φ,f_2) the $(2;\mathfrak{B},\mathfrak{A},\Phi)$-pseudococycle of $(\mathcal{E},\rho,\sigma)$. The Lie algebra structure of $\mathcal{E}$ can now be given as follows. Any element $e\in\mathcal{E}$ can be written as $e=\sigma(b)+a$, where $b\in\mathfrak{B}$ and $a\in\mathfrak{A}$ are uniquely determined by e. Then

$$[e,e'] = [\sigma(b)+a,\sigma(b')+a'] = \sigma([b,b']) + [a,a'] + \varphi(b)a' - \varphi(b')a + f_2(b,b'). \quad (1.1)$$

If the extension $(\mathcal{E},\rho)$ is kept fixed, and we consider all sections of $(\mathcal{E},\rho)$ over $\mathfrak{B}$, we obtain a set of representatives of the equivalence classes of sliced extensions related to elements of $\underline{(\mathcal{E},\rho)}$. But two elements $(\mathcal{E},\rho,\sigma)$ and $(\mathcal{E},\rho,\sigma')$ of this set still can be equivalent. This is the case if there exists $\gamma:\mathcal{E}\to\mathcal{E}$ such that $\gamma(a)=a$ and $\gamma(\sigma(b))=\sigma'(b)$, i.e., if

$$[\sigma(b)+f_1(b),\sigma(b')+f_1(b')] = \sigma([b,b']) + f_1([b,b']) + f_2(b,b') \quad (1.2)$$

and

$$[\sigma(b)+f_1(b),a] = [\sigma(b),a] \quad \text{for all } a\in\mathfrak{A},\quad b,b'\in\mathfrak{B},$$

where $f_1=\sigma'-\sigma$. (1.2) implies $\mathrm{Im}f_1\subseteq\mathfrak{C}(\mathfrak{A})$ and $\delta_1(\varphi)f_1=0$. Conversely, if these conditions are satisfied, $(\mathcal{E},\rho,\sigma)\overset{\gamma}{\equiv}(\mathcal{E},\rho,\sigma')$. Thus it is plain how to select, starting from $(\mathcal{E},\rho,\sigma)$, one representative from each equivalence class of sliced extensions related to elements of $\underline{(\mathcal{E},\rho)}$. We can summarize these remarks by saying that any extension generates a complete set of representatives of the corresponding equivalence classes of sliced extensions.

It must be noted that the relation between sliced extensions and pseudococycles gives rise to a surjection of the set of sliced extensions onto the set of pseudococycles which is, in general, not injective. We have the following.

Proposition 1: Let $(\mathcal{E},\rho,\sigma)$ and $(\mathcal{E}',\rho',\sigma')$ be sliced extensions of $\mathfrak{B}$ by $\mathfrak{A}$ with character Φ and let $(\mathcal{E},\rho)\overset{\gamma}{\equiv}(\mathcal{E}',\rho')$. Suppose further that (φ,f_2) and (φ',f_2') are, respectively, the $(2;\mathfrak{B},\mathfrak{A},\Phi)$-pseudococycles of $(\mathcal{E},\rho,\sigma)$ and $(\mathcal{E}',\rho',\sigma')$. Then $(\varphi,f_2)=(\varphi',f_2')$ if and only if $\mathrm{Im}f_1\subseteq\mathfrak{C}(\mathfrak{A})$ and $\delta_1(\varphi)f_1=0$, where $f_1=\gamma\circ\sigma-\sigma'$.

Proof: If $\mathrm{Im}f_1\subseteq\mathfrak{C}(\mathfrak{A})$, then

$$\varphi'(b) = \mathrm{ad}_{\mathfrak{A}}\sigma'(b) = \mathrm{ad}_{\mathfrak{A}}(\gamma\circ\sigma)(b) = \varphi(b) \quad \text{for all } b\in\mathfrak{B}$$

since

$$(\mathrm{ad}_{\mathfrak{A}}(\gamma\circ\sigma)(b))a = [(\gamma\circ\sigma)(b),a] = \gamma([\sigma(b),a]) = [\sigma(b),a] \quad \text{for all } a\in\mathfrak{A}.$$

If in addition $\delta_1(\varphi)f_1=0$, we have

$$f_2'(b,b') = [\sigma'(b),\sigma'(b')] - \sigma'([b,b']) = \gamma\{[\sigma(b),\sigma(b')] - \sigma([b,b'])\} - (\delta_1(\varphi)f_1)(b,b') = f_2(b,b') \quad \text{for all } b,b'\in\mathfrak{B}.$$

It is obvious that, conversely, $\varphi'=\varphi$ requires $\mathrm{Im}f_1\subseteq\mathfrak{C}(\mathfrak{A})$ and $f_2'=f_2$ requires $\delta_1(\varphi)f_1=0$. ■

Clearly, if $(\mathcal{E},\rho)\not\equiv(\mathcal{E}',\rho')$, then $(\mathcal{E},\rho,\sigma)$ and $(\mathcal{E}',\rho',\sigma')$ have never the same $(2;\mathfrak{B},\mathfrak{A},\Phi)$-pseudococycle, whatever the sections σ and σ' are. The necessary and sufficient conditions of Proposition 1 are obviously satisfied if $(\mathcal{E},\rho,\sigma)\overset{\gamma}{\equiv}(\mathcal{E}',\rho',\sigma')$: Two equivalent sliced extensions have always the same pseudococycle. Proposition 1 gives rise to a new equivalence relation in the set of all sliced extensions of $\mathfrak{B}$ by $\mathfrak{A}$ with character Φ: $(\mathcal{E},\rho,\sigma)$ and $(\mathcal{E}',\rho',\sigma')$ are equivalent if there exists $\gamma:\mathcal{E}\to\mathcal{E}'$ such that $(\mathcal{E},\rho)\overset{\gamma}{\equiv}(\mathcal{E}',\rho')$, and $f_1=\gamma\circ\sigma-\sigma'$ satisfies $\mathrm{Im}f_1\subseteq\mathfrak{C}(\mathfrak{A})$ and $\delta_1(\mathrm{ad}_{\mathfrak{A}}\circ\sigma)f_1=0$. There is then a natural bijection of the quotient of the set of sliced extensions of $\mathfrak{B}$ by $\mathfrak{A}$ with character Φ by this new equivalence relation onto $\mathfrak{Z}^2_\Phi(\mathfrak{B},\mathfrak{A})$.[11]

If we consider the sliced extension $(\mathcal{E},\rho,\sigma)$ of $\mathfrak{B}$ by $\mathfrak{A}$ with character Φ, keeping $(\mathcal{E},\rho)$ fixed we let σ run over the full set of sections of $(\mathcal{E},\rho)$ over $\mathfrak{B}$, then the $(2;\mathfrak{B},\mathfrak{A},\Phi)$-pseudococycle (φ,f_2) of $(\mathcal{E},\rho,\sigma)$ runs over the full equivalence class $\underline{(\varphi,f_2)}$. We can summarize this by saying that any extension is associated with an equivalence class of 2-pseudococycles.

B. Fundamental Problems of the Lie Algebra Extension Theory

We are now faced with the following general problems:

(I) Determine all extensions of $\mathfrak{B}$ by $\mathfrak{A}$, $\mathfrak{B}$ and $\mathfrak{A}$ being an arbitrary pair of Lie algebras. It suffices to determine a complete set of representatives of the equivalence classes of extensions of $\mathfrak{B}$ by $\mathfrak{A}$, i.e., one representative from each class.

(II) Determine all sliced extensions of $\mathfrak{B}$ by $\mathfrak{A}$. Again it is sufficient to determine a complete set of representatives of the equivalence classes of sliced extensions of $\mathfrak{B}$ by $\mathfrak{A}$.

The main theorem of Lie algebra extension theory[11] supplies us with a bijection

$$\alpha:\mathrm{ext}(\mathfrak{B},\mathfrak{A},\Phi)\rightarrowtail\mathfrak{H}^2_\Phi(\mathfrak{B},\mathfrak{A}), \quad (1.3)$$

so established that the Lie multiplication in the Lie algebra $\mathcal{E}$ of an extension $(\mathcal{E},\rho)$ of $\mathfrak{B}$ by $\mathfrak{A}$ with charac-

ter Φ is given by (1.1) if $(\varphi, f_2) \in \mathfrak{H}^2_\Phi(\mathfrak{B}, \mathfrak{A})$ is the image of $(\underline{\mathcal{E}, \rho}) \in \text{ext}(\mathfrak{B}, \mathfrak{A}, \Phi)$ by α.

This is the analog of Schreier's theorem for extensions of abstract groups.[19] $\text{ext}(\mathfrak{B}, \mathfrak{A})$ is the union of the sets $\text{ext}(\mathfrak{B}, \mathfrak{A}, \Phi)$ for all $\Phi \in \text{hom}(\mathfrak{B}, \Delta(\mathfrak{A}))$, and we have to consider all the corresponding sets $\mathfrak{H}^2_\Phi(\mathfrak{B}, \mathfrak{A})$. It must be remarked that we can have $\mathfrak{H}^2_\Phi(\mathfrak{B}, \mathfrak{A}) = \emptyset$ for some Φ [and consequently $\text{ext}(\mathfrak{B}, \mathfrak{A}, \Phi) = \emptyset$]. The Mori–Hochschild theory of Lie algebra kernels states that $\mathfrak{H}^2_\Phi(\mathfrak{B}, \mathfrak{A}) \neq \emptyset$ if and only if $\text{Obs}(\mathfrak{B}, \mathfrak{A}, \Phi) = 0$ (Refs. 11, 20, 21).

The bijection (1.3) allows the translation of problem (I) into the following:

(I′) For any $\Phi \in \text{hom}(\mathfrak{B}, \Delta(\mathfrak{A}))$ such that $\text{Obs}(\mathfrak{B}, \mathfrak{A}, \Phi) = 0$ determine a complete set of $(2; \mathfrak{B}, \mathfrak{A}, \Phi)$-pseudococycles.

The solution of problem (II) does not meet particular difficulties. Suppose that we have a representative (φ, f_2) of an equivalence class of $(2; \mathfrak{B}, \mathfrak{A}, \Phi)$-pseudococycles. We define a Lie algebra extension $(\mathcal{E}, \rho)$ of the equivalence class $\alpha^{-1}(\varphi, f_2)$ as follows. Let E' be the vector space $B \times A$. Consider a Lie algebra $\mathcal{E}'$ with the Lie multiplication

$$[(b,a),(b',a')] = ([b,b'],[a,a'] + \varphi(b)a' - \varphi(b')a + f_2(b,b')) \quad \text{for all } a,a' \in \mathfrak{A},\ b,b' \in \mathfrak{B}.$$

$\mathcal{E}'$ becomes $\mathcal{E} \supseteq \mathfrak{A}$ by the natural identification of $(\{0\}, \mathfrak{A}) \subseteq \mathcal{E}'$ with $\mathfrak{A}$. We also define ρ by $\rho((b,a)) = b$ and a section σ of $(\mathcal{E}, \rho)$ over $\mathfrak{B}$ by $\sigma(b) = (b, 0)$, and then $(\mathcal{E}, \rho, \sigma)$ is a sliced extension of $\mathfrak{B}$ by $\mathfrak{A}$ with character Φ and $(2; \mathfrak{B}, \mathfrak{A}, \Phi)$-pseudococycle (φ, f_2). Notice that $(b,a) = \sigma(b) + a$, and therefore the Lie product of (b,a) and (b',a') coincides with that given by (1.1).

A complete set of representatives of the equivalence classes of sliced extensions related to elements of $(\underline{\mathcal{E}, \rho})$ is now easily obtained starting from $(\mathcal{E}, \rho, \sigma)$ (see Sec. 1A). We have already remarked that, in a physical problem, only few equivalence classes of sliced extensions are allowed: The relation $\sigma' = \sigma + f_1$ induces a relation between physical quantities, which must be satisfied in nature. In the example of Ref. 18 we obtain the Gell-Mann–Nishijima formula.

The present paper is mainly devoted to the solution of problem (I) [or (I′)]. This is a more difficult problem, and it depends strongly on the Lie algebras involved. We will give some general results as hints for its solution, and a deeper characterization for special types of Lie algebras which often appear in physical problems.

When $\mathfrak{A}$ is an Abelian Lie algebra, problem (I′) reduces to the study of the vector spaces $H^2(\mathfrak{B}, A_\Phi)$ (Refs. 10, 11, 22). The difficulty in this case resides in the fact that one must know the representations Φ of $\mathfrak{B}$ on A which are, in general, hard to compute.[23] Once the representations are known, we have to solve a simple problem of linear algebra.

When $\mathfrak{A}$ is not Abelian, problem (I′) is, in general, very difficult. The analysis of $\mathfrak{H}^2_\Phi(\mathfrak{B}, \mathfrak{A})$ requires the knowledge of $\Phi \in \text{hom}(\mathfrak{B}, \Delta(\mathfrak{A}))$ and consequently of $\Delta(\mathfrak{A})$, the Lie algebra of outer derivations of $\mathfrak{A}$. The computation of $\Delta(\mathfrak{A})$ is not very hard, since $H^1(\mathfrak{A}, A_{\text{ad}})$ is its underlying vector space, ad being the adjoint representation of $\mathfrak{A}$ (on A). But only little is known about the relation between the Lie algebra structure of $\mathfrak{A}$ and that of $\Delta(\mathfrak{A})$.[24,25] Moreover, if we suppose to have $\Phi \in \text{hom}(\mathfrak{B}, \Delta(\mathfrak{A}))$, it must be proved that $\text{Obs}(\mathfrak{B}, \mathfrak{A}, \Phi) = 0$. A sufficient, but obviously not necessary, condition is $H^3(\mathfrak{B}, C(\mathfrak{A})_\Psi) = \{0\}$, where $C(\mathfrak{A})$ is the vector space underlying $\mathfrak{C}(\mathfrak{A})$. $\Psi : \mathfrak{B} \to \mathfrak{GL}(C(\mathfrak{A}))$ is defined by $\Psi(b)c = \varphi(b)c$ for all $c \in C(\mathfrak{A})$, φ being any prerepresentation of $\mathfrak{B}$ into $D(\mathfrak{A})$ lifted over Φ.[21] Ψ does not depend on the particular choice of φ, but only on Φ. We call it the central character determined by Φ.

Given $\Phi \in \text{hom}(\mathfrak{B}, \Delta(\mathfrak{A}))$ such that $\text{Obs}(\mathfrak{B}, \mathfrak{A}, \Phi) = 0$, we must compute $\mathfrak{H}^2_\Phi(\mathfrak{B}, \mathfrak{A})$, and for $\mathfrak{A}$ non-Abelian this is, in general, quite arduous. But the analysis of the Abelian case can be of help since we have a bijection.

$$\beta : \mathfrak{H}^2_\Phi(\mathfrak{B}, \mathfrak{A}) \rightarrowtail H^2(\mathfrak{B}, C(\mathfrak{A})_\Psi), \tag{1.4}$$

where Ψ is the central character determined by Φ. The bijection is defined as follows. Select a $(2; \mathfrak{B}, \mathfrak{A}, \Phi)$-pseudococycle (φ, g_2) which is kept fixed. Let now (φ', f_2') be any $(2; \mathfrak{B}, \mathfrak{A}, \Phi)$-pseudococycle; then, by Lemma 1 of Ref. 11, there exists $(\varphi, f_2) \in (\underline{\varphi', f_2'})$, φ being the fixed prerepresentation. $f_2 - g_2$ can be identified with an element of $Z^2(\mathfrak{B}, C(\mathfrak{A})_\Psi)$, since $\text{Im}(f_2 - g_2) \subseteq \mathfrak{C}(\mathfrak{A})$. The bijection (1.4) is such that $\beta(\underline{\varphi', f_2'}) = \underline{f_2 - g_2}$. Clearly, if we replace (φ, g_2) by another $(2; \mathfrak{B}, \mathfrak{A}, \Phi)$-pseudococycle of the same equivalence class, we have exactly the same bijection β. We can then replace (1.4) by

$$\beta_{(\underline{\varphi, g_2})} : \mathfrak{H}^2_\Phi(\mathfrak{B}, \mathfrak{A}) \rightarrowtail H^2(\mathfrak{B}, C(\mathfrak{A})_\Psi), \tag{1.5}$$

where $(\underline{\varphi, g_2})$ means that the bijection (1.4) is defined as above picking out one element of the class $(\underline{\varphi, g_2})$. Any element of $\mathfrak{H}^2_\Phi(\mathfrak{B}, \mathfrak{A})$ determines a different bijection.

$H^2(\mathfrak{B}, C(\mathfrak{A})_\Psi)$ being a vector space, we can transport its structure on $\text{ext}(\mathfrak{B}, \mathfrak{A}, \Phi)$ by means of a bijection $\alpha^{-1} \circ \beta^{-1}_{(\underline{\varphi, g_2})}$ [$(\underline{\varphi, g_2}) \in \mathfrak{H}^2_\Phi(\mathfrak{B}, \mathfrak{A})$ and α of (1.3)]. Suppose that there exists one and only one equivalence class of inessential extensions of $\mathfrak{B}$ by $\mathfrak{A}$ with character Φ. We choose the bijection $\lambda = \alpha^{-1} \circ \beta^{-1}_{(\underline{\varphi, 0})}$ of $H^2(\mathfrak{B}, C(\mathfrak{A})_\Psi)$ onto $\text{ext}(\mathfrak{B}, \mathfrak{A}, \Phi)$, and then the unique class of inessential extensions is the image of $0 \in H^2(\mathfrak{B}, C(\mathfrak{A})_\Psi)$ by λ. Through λ, $\text{ext}(\mathfrak{B}, \mathfrak{A}, \Phi)$ becomes a vector space $\text{Ext}(\mathfrak{B}, \mathfrak{A}, \Phi)$ naturally isomorphic to $H^2(\mathfrak{B}, C(\mathfrak{A})_\Psi)$. We have a generalization of the Chevalley–Eilenberg theory[10,22] for $\mathfrak{A}$ non-Abelian. Clearly, this generalization is not possible if there exist many or no equivalence classes of inessential extensions: We can always supply $\text{ext}(\mathfrak{B}, \mathfrak{A}, \Phi)$ with a vector space structure, but this is not a natural one for the extension theory. We will show in Appendix A, by some examples, that the cases of no or many equivalence classes of inessential extensions really occur. If $f_1 \in L(\mathfrak{B}, \mathfrak{A})$, let $\Delta_1(\varphi) f_1$ be the bilinear alternating map of $\mathfrak{B} \times \mathfrak{B}$ into $\mathfrak{A}$ defined by

$$(\Delta_1(\varphi) f_1)(b,b') = (\delta_1(\varphi) f_1)(b,b') + [f_1(b), f_1(b')]. \tag{1.6}$$

We can give the following uniqueness criterion.

Proposition 2: In order that there exists one and only one equivalence class of inessential extensions of $\mathfrak{B}$ by $\mathfrak{A}$ with character Φ, it is necessary and sufficient that

(1) there be a representation $\varphi \in \hom(\mathfrak{B}, D(\mathfrak{A}))$ lifted over Φ,

and

(2) for any $f_1 \in L(\mathfrak{B}, \mathfrak{A})$ with

$$\mathrm{Im}(\Delta_1(\varphi)f_1) \subseteq \mathcal{C}(\mathfrak{A}) \tag{1.7}$$

there is $f_1' \in L(\mathfrak{B}, \mathfrak{A})$ such that $\mathrm{Im}f_1' \subseteq \mathcal{C}(\mathfrak{A})$ and $\delta_1(\varphi)f_1' = \Delta_1(\varphi)f_1$.

Proof: Necessity: Condition (1) is necessary by Theorem 1 of Ref. 11. Suppose that (2) is not satisfied. Consider $f_1 \in L(\mathfrak{B}, \mathfrak{A})$ such that no $f_1' \in L(\mathfrak{B}, \mathfrak{A})$ with $\mathrm{Im}f_1' \subseteq \mathcal{C}(\mathfrak{A})$ and $\delta_1(\varphi)f_1' = \Delta_1(\varphi)f_1$ exists. If $\varphi' = \varphi + \mathrm{ad}\circ f_1$, then $(\varphi', \Delta_1(\varphi)f_1) \in \mathfrak{Z}^2_\Phi(\mathfrak{B}, \mathfrak{A})$ and $\underline{(\varphi', \Delta_1(\varphi)f_1)} = \underline{(\varphi, 0)}$. But, by (1.7), φ' is a representation, and hence $(\varphi', 0) \in \mathfrak{Z}^2_\Phi(\mathfrak{B}, \mathfrak{A})$. Moreover, $\underline{(\varphi', 0)} \neq \underline{(\varphi', \Delta_1(\varphi)f_1)}$ by our assumption, i.e., $\underline{(\varphi', 0)} \neq \underline{(\varphi, 0)}$. Then we have two equivalence classes of inessential extensions $\alpha^{-1}\underline{(\varphi', 0)}$ and $\alpha^{-1}\underline{(\varphi, 0)}$, α being the bijection (1.3).

Sufficiency: Suppose that the conditions (1) and (2) are satisfied and consider $\varphi' \in \hom(\mathfrak{B}, D(\mathfrak{A}))$ lifted over Φ. There exists $f_1 \in L(\mathfrak{B}, \mathfrak{A})$ such that $\varphi' = \varphi + \mathrm{ad}\circ f_1$, and then $\underline{(\varphi', \Delta_1(\varphi)f_1)} = \underline{(\varphi, 0)}$ by (1.7). As $\underline{(\varphi', \Delta_1(\varphi)f_1)} = \underline{(\varphi', 0)}$, since $\Delta_1(\varphi)f_1 = \delta_1(\varphi)f_1' = \delta_1(\varphi')f_1'$, we get $\underline{(\varphi, 0)} = \underline{(\varphi', 0)}$. This implies the existence of one and only one equivalence class of inessential extensions.[26] ■

Now we can prove the following result quoted by Galindo[8]:

Corollary: A necessary and sufficient condition that any inessential extension by $\mathfrak{A}$ with character 0 be trivial is that the extension $(\mathfrak{A}, \mathrm{ad})$ of $I(\mathfrak{A})$ by $\mathcal{C}(\mathfrak{A})$ be inessential.

Proof: If $\mathfrak{A}$ and $\mathfrak{B}$ are arbitrary Lie algebras, there exists always one equivalence class of inessential extensions of $\mathfrak{B}$ by $\mathfrak{A}$ with character 0: the class of trivial extensions. A representation $\varphi \in \hom(\mathfrak{B}, D(\mathfrak{A}))$ lifted over 0 is then the trivial one. Suppose that the class of trivial extensions is the unique equivalence class of inessential extensions by $\mathfrak{A}$ with character 0, and consider $\mathfrak{B} = I(\mathfrak{A})$. If σ is any section of $(\mathfrak{A}, \mathrm{ad})$ over $I(\mathfrak{A})$, we can define $f_1 \in L(I(\mathfrak{A}), \mathfrak{A})$ by $f_1(b) = \sigma(b)$ for all $b \in I(\mathfrak{A})$. Then f_1 satisfies (1.7) with $\varphi = 0$, $\Delta_1(0)f_1$ being the factor set associated with σ, and $(\mathfrak{A}, \mathrm{ad})$ is inessential by condition (2) of Proposition 2. Conversely, if $(\mathfrak{A}, \mathrm{ad})$ is an inessential (central) extension of $I(\mathfrak{A})$ by $\mathcal{C}(\mathfrak{A})$, there is a section σ of $(\mathfrak{A}, \mathrm{ad})$ over $I(\mathfrak{A})$ such that $\mathfrak{A} = \sigma(I(\mathfrak{A})) \oplus \mathcal{C}(\mathfrak{A})$. $\mathfrak{B}$ being an arbitrary Lie algebra, choose $\varphi = 0$ and consider the projection $\pi_{\mathcal{C}(\mathfrak{A})}$ of $\mathfrak{A}$ onto $\mathcal{C}(\mathfrak{A})$. If $f_1 \in (L(\mathfrak{B}, \mathfrak{A})$ satisfies (1.7), we must have $(\Delta_1(0)f_1)(b, b') = -\pi_{\mathcal{C}(\mathfrak{A})}f_1([b, b'])$, for all $b, b' \in \mathfrak{B}$. This implies $\delta_1(0)f_1' = \Delta_1(0)f_1$ if we choose $f_1' \in L(\mathfrak{B}, \mathfrak{A})$ defined by $f_1'(b) = \pi_{\mathcal{C}(\mathfrak{A})}f_1(b)$. The wanted result follows from Proposition 2.

Independently of the uniqueness or not of the equivalence class of inessential extensions, the bijections (1.5) are very useful in solving problem (I'). We can select an equivalence class of $\mathfrak{Z}^2_\Phi(\mathfrak{B}, \mathfrak{A})$ and, keeping an element (φ, g_2) of this class fixed, choose a representative $(2; \mathfrak{B}, C(\mathfrak{A})_\psi)$-cocycle h_2 from each equivalence class of $Z^2(\mathfrak{B}, C(\mathfrak{A})_\psi)$. Then we consider $(\varphi, f_2) \in \mathfrak{Z}^2_\Phi(\mathfrak{B}, \mathfrak{A})$, where $f_2 = h_2 + g_2$. If h_2 runs over all classes of $Z^2(\mathfrak{B}, C(\mathfrak{A})_\psi)$, (φ, f_2) runs over all classes of $\mathfrak{Z}^2_\Phi(\mathfrak{B}, \mathfrak{A})$. Moreover, if $\underline{h_2} \neq \underline{h_2'}$ and $f_2' = h_2' + g_2$, then $\underline{(\varphi, f_2)} \neq \underline{(\varphi, f_2')}$. Thus we have to choose suitable representatives of the equivalence classes of $Z^2(\mathfrak{B}, C(\mathfrak{A})_\psi)$. This can be done by means of the Hochschild–Serre theorem.[10,27] The reduction theorem of Ref. 11, which is a generalization, in the case of $\mathfrak{A}$ non-Abelian, of the Hochschild–Serre theorem for $n = 2$, allows an appropriate choice of (φ, g_2). We will prove in Appendix B a theorem which implies the previous remarks. Actually the reduction theorem is a stronger result, and its application to our problem already includes the choice of h_2 according to the Hochschild–Serre theorem. The meaning of the reduction theorem is the following. Let $(\mathcal{E}, \rho)$ be an extension of $\mathfrak{B}$ by $\mathfrak{A}$ with character Φ, and suppose that $\mathfrak{D}$ is an ideal of $\mathfrak{B}$ such that $\mathfrak{B}/\mathfrak{D}$ is semisimple. By Levi's theorem[14,28,29] any extension of a semisimple Lie algebra is inessential. Therefore, $\mathfrak{B}$ has a subalgebra $\mathfrak{S}$ isomorphic to $\mathfrak{B}/\mathfrak{D}$ by the canonical epimorphism $\mathfrak{B} \twoheadrightarrow \mathfrak{B}/\mathfrak{D}$, and $\mathfrak{B} = \mathfrak{S} \dotplus \mathfrak{D}$. Then there exists a sliced extension $(\mathcal{E}, \rho, \sigma)$ with $(2; \mathfrak{B}, \mathfrak{A}, \Phi)$-pseudococycle (φ, f_2) which satisfies

$$f_2(s, b) = 0 \quad \text{for all } s \in \mathfrak{S},\ b, b' \in \mathfrak{B}$$
$$\varphi(s)f_2(b, b') = f_2([s, b], b') + f_2(b, [s, b']), \tag{1.8}$$

i.e., f_2 is ω-orthogonal to $\mathfrak{S}$, with $\omega = \varphi \mid \mathfrak{S}$.[11] ω is a representation of $\mathfrak{S}$ into $D(\mathfrak{A})$ lifted over $\Phi \mid \mathfrak{S}$, and

$$[\varphi(s), \varphi(b)] = \varphi([s, b]) \quad \text{for all } s \in \mathfrak{S},\ b \in \mathfrak{B}.$$

It follows also, from the proof of the theorem, that ω can be chosen in such a way that, for any $s \in \mathfrak{S}$, $\omega(s) = 0$ if and only if $\Phi(s) = 0$. We say that such an ω is isomorphically lifted over $\Phi \mid \mathfrak{S}$. If $(\mathcal{E}', \rho')$ is any other extension of $\mathfrak{B}$ by $\mathfrak{A}$ with character Φ, there exists a sliced extension $(\mathcal{E}', \rho', \sigma')$ with $(2; \mathfrak{B}, \mathfrak{A}, \Phi)$-pseudococycle (φ, f_2') (the same φ as above), where f_2' is ω-orthogonal to $\mathfrak{S}$. Our remarks fit in particular if $\mathfrak{D} = \mathfrak{R}$, the radical of $\mathfrak{B}$, i.e., if $\mathfrak{B} = \mathfrak{S} \dotplus \mathfrak{R}$ is a Levi decomposition, $\mathfrak{S}$ being a maximal semisimple subalgebra of $\mathfrak{B}$.

Obviously, the reduction theorem gives a trivial result whenever the maximal semisimple subalgebra of $\mathfrak{B}$ is $\{0\}$, i.e., if $\mathfrak{B}$ is solvable. But many Lie algebras of interest in physics are not solvable (for example, the Lie algebras of the Euclidean, Galilean, and Poincaré groups).

C. ω-Sliced and Primitive Extensions

The Lie algebra $\mathcal{E}$ obtained from an extension $(\mathcal{E}, \rho)$ of $\mathfrak{B}$ by $\mathfrak{A}$ has some well determined Levi decompositions induced by those of $\mathfrak{B}$ and $\mathfrak{A}$, as appears from the following result which generalizes Theorem 5 of Ref. 11.

Theorem 1 (structure theorem): Let $\mathfrak{A} = \mathfrak{S}_1 \dotplus \mathfrak{R}_1$ and $\mathfrak{B} = \mathfrak{S}_2 \dotplus \mathfrak{R}_2$ be arbitrary Levi decompositions of

$\mathfrak{A}$ and $\mathfrak{B}$, and suppose that $D^n\mathfrak{R}_1 = \{0\}, D^m\mathfrak{R}_2 = \{0\}$. If $(\mathcal{E},\rho)$ is an extension of $\mathfrak{B}$ by $\mathfrak{A}$, there exist a sliced extension $(\mathcal{E},\rho,\sigma)$ and a Levi decomposition

$$\mathcal{E} = (\sigma(\mathfrak{S}_2) \oplus \mathfrak{S}_1) \dotplus \mathfrak{R},$$

where $\sigma(\mathfrak{S}_2)$ is isomorphic to $\mathfrak{S}_2$ by ρ and $[\sigma_v(R_2), S_1] = \{0\}$. The radical $\mathfrak{R}$ of $\mathcal{E}$ is such that $R = \sigma_v(R_2) \oplus R_1$ and $D^{m+n}\mathfrak{R} = \{0\}$.

Proof: $\mathfrak{R}_1$ is an ideal of $\mathcal{E}$, hence $(\mathcal{E}/\mathfrak{R}_1, \rho_q)$ is an extension of $\mathfrak{B}$ by $\mathfrak{A}/\mathfrak{R}_1$, where ρ_q: $\mathcal{E}/\mathfrak{R}_1 \twoheadrightarrow \mathfrak{B}$ is the epimorphism obtained from ρ by passing to the quotient. $\mathfrak{A}/\mathfrak{R}_1$ is semisimple, and there is an isomorphism η: $\mathfrak{S}_1 \rightarrowtail \mathfrak{A}/\mathfrak{R}_1$, the restriction to $\mathfrak{S}_1$ of the canonical epimorphism of $\mathfrak{A}$ onto $\mathfrak{A}/\mathfrak{R}_1$. Then there is a sliced extension $(\mathcal{E}/\mathfrak{R}_1, \rho_q, \sigma_q)$ such that

$$\mathcal{E}/\mathfrak{R}_1 = \sigma_q(\mathfrak{B}) \oplus \eta(\mathfrak{S}_1), \tag{1.9}$$

σ_q being a monomorphism. Put $\mathfrak{S} = \sigma_q(\mathfrak{S}_2) \oplus \eta(\mathfrak{S}_1)$ and consider the canonical epimorphism ρ^*: $\mathcal{E} \twoheadrightarrow \mathcal{E}/\mathfrak{R}_1$. $(\mathcal{E},\rho^*)$ is an extension of $\mathcal{E}/\mathfrak{R}_1$ by $\mathfrak{R}_1$ (with character Φ^*), and by the reduction theorem we may consider a sliced extension $(\mathcal{E},\rho^*,\sigma^*)$ whose $(2;\mathcal{E}/\mathfrak{R}_1, \mathfrak{R}_1, \Phi^*)$-pseudococycle (φ^*, f_2^*) is such that f_2^* is ω^*-orthogonal to $\mathfrak{S}$, with $\omega^* = \varphi^* \mid \mathfrak{S}$ isomorphically lifted over $\Phi^* \mid \mathfrak{S}$. Moreover, σ^* can be chosen to satisfy[26]

$$\sigma^*(\eta(s_1)) = s_1 \quad \text{for all } s_1 \in \mathfrak{S}_1.$$

Put $\sigma^* \circ \sigma_q = \sigma$ and consider the vector space $M = \sigma_v(R_2) \oplus R_1 \cdot \mathcal{E}$ induces on M the structure of an ideal $\mathfrak{M}$: If $m = \sigma(r_2) + r_1$ is any element of M and $e = \sigma^*(\underline{e}) + r_1'$ any element of $\mathcal{E}$, then

$$[e,m] = \{\sigma^*([\underline{e}, \sigma_q(r_2)]) + [r_1', r_1] + \varphi^*(\underline{e})r_1 - \varphi^*(\sigma_q(r_2))r_1' + f_2^*(\underline{e}, \sigma_q(r_2))\} \in M,$$

since $\sigma^*([\underline{e}, \sigma_q(r_2)]) \in \sigma_v(R_2)$. $\mathfrak{M}$ is a solvable ideal too:

$$[\sigma(r_2), \sigma(r_2')] = \{\sigma([r_2, r_2']) + f_2^*(\sigma_q(r_2), \sigma_q(r_2'))\} \in \mathfrak{M}$$

and

$$[\sigma(r_2), r_1] = \varphi^*(\sigma_q(r_2))r_1 \in \mathfrak{R}_1 \quad \text{for all } r_2, r_2' \in \mathfrak{R}_2, r_1 \in \mathfrak{R}_1,$$

therefore,

$$D^1M \subseteq \sigma_v(D^1R_2) \oplus R_1$$

and analogously

$$D^iM \subseteq \sigma_v(D^iR_2) \oplus R_1,$$

where $D^iM\,(D^iR_2)$ means the vector space underlying $D^i\mathfrak{M}\,(D^i\mathfrak{R}_2)$. For $i = m$ we obtain $D^mM \subset R_1$ and infer $D^{m+n}\mathfrak{M} = \{0\}$. Besides, $\sigma^* \mid \mathfrak{S}$ is a monomorphism, and thus $\sigma^*(\mathfrak{S}) = \mathfrak{S}_1 \oplus \sigma(\mathfrak{S}_2)$ is a semisimple Lie algebra supplementary to $\mathfrak{M}$ in $\mathcal{E}$ and $\sigma(\mathfrak{S}_2)$ is isomorphic to $\mathfrak{S}_2$ by ρ. Hence $\mathfrak{M} = \mathfrak{R}$, the biggest solvable ideal, and

$$\mathcal{E} = (\sigma(\mathfrak{S}_2) \oplus \mathfrak{S}_1) \dotplus \mathfrak{R}.$$

Moreover,

$$[\sigma(r_2), s_1] = \sigma^*([\sigma_q(r_2), \eta(s_1)]) = 0 \quad \text{for all } r_2 \in \mathfrak{R}_2,\ s_1 \in \mathfrak{S}_1,$$

by (1.9). We end the proof observing that σ is really a section of $(\mathcal{E},\rho)$ over $\mathfrak{B}$. Indeed $\rho = \rho_q \circ \rho^*$ and $\rho(\sigma(b)) = b$ for all $b \in \mathfrak{B}$. ■

Notice that, choosing $(\mathcal{E},\rho,\sigma)$ as in the proof, we have

$$f_2(b,b') = f_2^*(\sigma_q(b), \sigma_q(b')) \in \mathfrak{R}_1 \quad \text{for all } b, b' \in \mathfrak{B}.$$

An immediate consequence is then the following.

Corollary: Let $\mathfrak{A}$ and $\mathfrak{B}$ be as in Theorem 1. If $(\mathcal{E},\rho)$ is an extension of $\mathfrak{B}$ by $\mathfrak{A}$ with character Φ, there is a sliced extension $(\mathcal{E},\rho,\sigma)$ whose $(2;\mathfrak{B},\mathfrak{A},\Phi)$-pseudococycle (φ,f_2) has f_2 ω-orthogonal to $\mathfrak{S}_2$, $\omega = \varphi \mid \mathfrak{S}_2$ being isomorphically lifted over $\Phi \mid \mathfrak{S}_2$, and satisfies $\varphi(\mathfrak{B})\mathfrak{S}_1 = \{0\}$ and $\mathrm{Im}f_2 \subseteq \mathfrak{R}_1$.

If $(\mathcal{E}',\rho')$ is any other extension of $\mathfrak{B}$ by $\mathfrak{A}$ with character Φ, there is a sliced extension $(\mathcal{E}',\rho',\sigma')$ with $(2;\mathfrak{B},\mathfrak{A},\Phi)$-pseudococycle (φ,f_2'), where f_2' is ω-orthogonal to $\mathfrak{S}_2$, φ being the prerepresentation of the corollary and $f_2' = f_2 + h_2$, $h_2 \in Z^2(\mathfrak{B}, C(\mathfrak{A})_\psi)$ (see the theorem of Appendix B). It follows that $\mathrm{Im}f_2' \subseteq \mathfrak{R}_1$. This suggests the following.

Definition 3: We say that a sliced extension of $\mathfrak{B}$ by $\mathfrak{A}$ with character Φ is an *ω-sliced extension* for the Levi decompositions $\mathfrak{A} = \mathfrak{S}_1 \dotplus \mathfrak{R}_1$ and $\mathfrak{B} = \mathfrak{S}_2 \dotplus \mathfrak{R}_2$, if its $(2;\mathfrak{B},\mathfrak{A},\Phi)$-pseudococycle (φ,f_2) has f_2 ω-orthogonal to $\mathfrak{S}_2$, with $\omega = \varphi \mid \mathfrak{S}_2$ isomorphically lifted over $\Phi \mid \mathfrak{S}_2$, and $\varphi(\mathfrak{B})\mathfrak{S}_1 = \{0\}$, $\mathrm{Im}f_2 \subseteq \mathfrak{R}_1$.

The corollary can now be stated as follows: With $(\mathcal{E},\rho)$ as in Theorem 1, there exists an ω-sliced extension $(\mathcal{E},\rho,\sigma)$ for the Levi decompositions $\mathfrak{A} = \mathfrak{S}_1 \dotplus \mathfrak{R}_1$ and $\mathfrak{B} = \mathfrak{S}_2 \dotplus \mathfrak{R}_2$.

ω-sliced extensions are clearly dependent on the chosen Levi decompositions. But the result of the preceding Corollary is valid for any fixed Levi decomposition of $\mathfrak{A}$ and $\mathfrak{B}$. Whenever we speak of ω-sliced extensions throughout this paper, it must be understood as "for the Levi decompositions $\mathfrak{A} = \mathfrak{S}_1 \dotplus \mathfrak{R}_1$ and $\mathfrak{B} = \mathfrak{S}_2 \dotplus \mathfrak{R}_2$."

Note that, even if $\mathfrak{B}$ is solvable and hence any $(2;\mathfrak{B},\mathfrak{A},\Phi)$-pseudococycle (φ,f_2) has f_2 0-orthogonal to $\mathfrak{S}_2 = \{0\}$, the result of the corollary is not trivial: 0-sliced extensions can be selected according to it with $(2;\mathfrak{B},\mathfrak{A},\Phi)$-pseudococycles (φ,f_2) satisfying $\varphi(\mathfrak{B})\mathfrak{S}_1 = \{0\}$ and $\mathrm{Im}f_2 \subseteq \mathfrak{R}_1$.

Let $\Delta_1(\varphi)f_1$ be the bilinear alternating map of $\mathfrak{B} \times \mathfrak{B}$ into $\mathfrak{A}$ defined by (1.6). The following proposition answers the question about the relation between ω- and ω'-sliced extensions.

Proposition 3: Let $(\mathcal{E},\rho,\sigma)$ and $(\mathcal{E},\rho,\sigma')$ be sliced extensions of $\mathfrak{B}$ by $\mathfrak{A}$ with character Φ. Suppose, moreover, that $(\mathcal{E},\rho,\sigma)$ is an ω-sliced extension with $(2;\mathfrak{B},\mathfrak{A},\Phi)$-pseudococycle (φ,f_2). A necessary and sufficient condition that $(\mathcal{E},\rho,\sigma')$ be an ω'-sliced extension is that $f_1 = \sigma' - \sigma$ satisfy

(1) $\mathrm{Im}f_1 \subseteq \mathcal{C}_{\mathfrak{A}}(\mathfrak{S}_1)$, the centralizer of $\mathfrak{S}_1$ in $\mathfrak{A}$,

(2) $(\Delta_1(\varphi)f_1)(s_2, b) = 0$ for all $s_2 \in \mathfrak{S}_2$, $b \in \mathfrak{B}$.

Proof: If (φ',f_2') is the $(2;\mathfrak{B},\mathfrak{A},\Phi)$-pseudococycle of $(\mathcal{E},\rho,\sigma')$, then $f_2' = f_2 + \Delta_1(\varphi)f_1$. Condition (2) means $(\Delta_1(\varphi)f_1)_{s_2} = 0$ for all $s_2 \in \mathfrak{S}_2$ (in the notation

of Ref. 11), and obviously it is necessary in order to have $(\varphi', f_2') \in \mathfrak{Z}^2_{\Phi,\omega}(\mathfrak{B}, \mathfrak{S}_2, \mathfrak{A})$, $\omega' = \varphi' | \mathfrak{S}_2$. Moreover, $\varphi'(\mathfrak{B})\mathfrak{S}_1 = \{0\}$ requires $\mathrm{Im} f_1 \subseteq \mathfrak{C}_{\mathfrak{A}}(\mathfrak{S}_1)$, since

$$\varphi'(b) = \varphi(b) + \mathrm{ad} f_1(b) \quad \text{for all } b \in \mathfrak{B},$$

and $\varphi(\mathfrak{B})\mathfrak{S}_1 = \{0\}$. Suppose now that, conversely, conditions (1) and (2) are satisfied. We first prove $\mathfrak{C}_{\mathfrak{A}}(\mathfrak{S}_1) \subseteq \mathfrak{R}_1$. Let $s_1 + r_1$ be any element of $\mathfrak{C}_{\mathfrak{A}}(\mathfrak{S}_1)$, with $s_1 \in \mathfrak{S}_1$, $r_1 \in \mathfrak{R}_1$. This requires

$[s_1 + r_1, s_1'] = 0$ for all $s_1' \in \mathfrak{S}_1$, and thus we get $s_1 = 0$,

i.e., $\mathfrak{C}_{\mathfrak{A}}(\mathfrak{S}_1) \subseteq \mathfrak{R}_1$, from the semisimplicity of $\mathfrak{S}_1$. For all $b, b' \in \mathfrak{B}$, $\varphi(b) f_1(b') \in \mathfrak{R}_1$, since $f_1(b') \in \mathfrak{C}_{\mathfrak{A}}(\mathfrak{S}_1) \subseteq \mathfrak{R}_1$ and $\mathfrak{R}_1$ is a characteristic ideal of $\mathfrak{A}$. It follows that $\mathrm{Im}(\Delta_1(\varphi) f_1) \subseteq \mathfrak{R}_1$, and then $\mathrm{Im} f_2' \subseteq \mathfrak{R}_1$, $\varphi'(\mathfrak{B})\mathfrak{S}_1 = \{0\}$. Moreover, as $(f_2')_{s_2} = 0$,

$$s_2(\omega') \cdot f_2' = \delta_1(\varphi')(f_2')_{s_2} = 0 \quad \text{for all } s_2 \in \mathfrak{S}_2$$

by (III. 4) of Ref. 11, i.e., $(\varphi', f_2') \in \mathfrak{Z}^2_{\Phi,\omega'}(\mathfrak{B}, \mathfrak{S}_2, \mathfrak{A})$ and $(\mathcal{E}, \rho, \sigma')$ is an ω'-sliced extension. ■

If ω is any representation of $\mathfrak{S}_2$ into $\mathrm{D}(\mathfrak{A})$, we can equip A with the $\mathfrak{S}_2$-module structure associated with ω. By Weyl's theorem[28,29] we obtain a semisimple $\mathfrak{S}_2$-module A_ω. In particular, every ω-sliced extension of $\mathfrak{B}$ by $\mathfrak{A}$ with character Φ determines one $\mathfrak{S}_2$-module A_ω, which can be kept fixed for all extensions of $\mathfrak{B}$ by $\mathfrak{A}$ with character Φ (remark following Corollary to Theorem 1).

Definition 4: We call *fundamental* $\mathfrak{S}_2$*-module* of an ω-sliced extension of $\mathfrak{B}$ by $\mathfrak{A}$ with character Φ and $(2; \mathfrak{B}, \mathfrak{A}, \Phi)$-pseudococycle (φ, f_2) the semisimple $\mathfrak{S}_2$-module $A(\varphi, f_2)$ induced by A_ω on $\overline{\mathrm{Im} f_2}$.

The vector space $\overline{\mathrm{Im} f_2} \subseteq A$, generated by $\mathrm{Im} f_2$, becomes in fact an $\mathfrak{S}_2$-submodule of A_ω by (1.8). Two cases are particularly important:

(i) $A(\varphi, f_2) = \{0\}$: $(\mathcal{E}, \rho)$ is then an inessential extension;

(ii) $A(\varphi, f_2)$ simple $\mathfrak{S}_2$-module.

Consider the set $\mathfrak{E}(\mathfrak{S}_2, \mathfrak{R}_2)$ of isomorphism classes of simple fundamental $\mathfrak{S}_2$-modules of ω-sliced extensions of $\mathfrak{B}$ (ω variable). If $\mathfrak{S}_2'$ is another Levi factor of $\mathfrak{B}$, there is a corresponding set $\mathfrak{E}(\mathfrak{S}_2', \mathfrak{R}_2)$ and a natural bijection of $\mathfrak{E}(\mathfrak{S}_2', \mathfrak{R}_2)$ onto $\mathfrak{E}(\mathfrak{S}_2, \mathfrak{R}_2)$ which allows the identification of these two sets. We will now show that $\mathfrak{E}(\mathfrak{S}_2, \mathfrak{R}_2)$ is a finite set. In order to do this, we define a representation $\Omega : \mathfrak{S}_2 \to \mathfrak{gl}(R_2)$ by

$$\Omega(s)r = [s, r] \quad \text{for any } r \in \mathfrak{R}_2, \tag{1.10}$$

and we obtain an associated $\mathfrak{S}_2$-module R_{2_Ω}. If $A(\varphi, f_2)$ is the fundamental $\mathfrak{S}_2$-module of an ω-sliced extension of $\mathfrak{B}$, there is an epimorphism (of $\mathfrak{S}_2$-modules)

$$\overset{2}{\wedge} R_{2_\Omega} \twoheadrightarrow A(\varphi, f_2)$$

given by

$$r \wedge r' \mapsto f_2(r, r') \quad \text{for all } r, r' \in R_{2_\Omega},$$

where the $\mathfrak{S}_2$-module $\overset{2}{\wedge} R_{2_\Omega}$ is the second exterior power of R_{2_Ω} (associated with the representation given by the diagonal action of $\mathfrak{S}_2$ on $\overset{2}{\wedge} R_2$).[28] From the semisimplicity of $\overset{2}{\wedge} R_{2_\Omega}$ it follows then easily that $\mathfrak{E}(\mathfrak{S}_2, \mathfrak{R}_2)$ is a finite set.

The following result is now plain if we define the isotypical components and the length of semisimple modules as Bourbaki (Ref. 30, §3, Secs. 4 and 5) does.

Proposition 4: Let A_ω be the $\mathfrak{S}_2$-module determined by an arbitrary ω-sliced extension of $\mathfrak{B}$ by $\mathfrak{A}$ with character Φ. If the lengths of the isotypical components of A_ω of types belonging to $\mathfrak{E}(\mathfrak{S}_2, \mathfrak{R}_2)$ are all 0, then every extension of $\mathfrak{B}$ by $\mathfrak{A}$ with character Φ is inessential.

We now pick out some important extensions which are the cornerstones of the set of all extensions of $\mathfrak{B}$.

Definition 5: An essential extension of $\mathfrak{B}$ by $\mathfrak{A}$ with character Φ is a *primitive extension* if A_ω is a simple $\mathfrak{S}_2$-module for any representation ω of $\mathfrak{S}_2$ into $\mathrm{D}(\mathfrak{A})$ isomorphically lifted over $\Phi | \mathfrak{S}_2$. We call also primitive the truly trivial extensions of $\mathfrak{B}$, i.e., the extensions of $\mathfrak{B}$ by $\{0\}$.

If $(\mathcal{E}, \rho)$ is a primitive extension, then any extension of the equivalence class $(\mathcal{E}, \rho)$ is primitive too. We may speak of classes of primitive extensions.

Theorem 2: Every primitive extension is an extension with Abelian kernel.

Proof: Let $(\mathcal{E}, \rho)$ be a primitive extension of $\mathfrak{B}$ by $\mathfrak{A}$ and $(\mathcal{E}, \rho, \sigma)$ an ω-sliced extension. As

$$A_\omega = S_{1_{\omega'}} \oplus R_{1_{\omega''}},$$

where ω' and ω'' are, respectively, the subrepresentations of ω on S_1 and R_1, the primitivity of $(\mathcal{E}, \rho)$ implies $\mathfrak{S}_1 = \{0\}$. Then $\mathfrak{A}$ is a solvable Lie algebra and $D^1\mathfrak{A} \subset \mathfrak{A}$ or $\mathfrak{A} = \{0\}$. But $D^1\mathfrak{A}$ is a characteristic ideal of $\mathfrak{A}$; hence it carries the $\mathfrak{S}_2$-module structure induced by A_ω. Again from the simplicity of A_ω we conclude that $\mathfrak{A} = \{0\}$ or $D^1\mathfrak{A} = \{0\}$, i.e., $\mathfrak{A}$ is Abelian. ■

It follows in particular from Theorem 2 that we have actually a unique representation of $\mathfrak{S}_2$ into $\mathrm{D}(\mathfrak{A})$ lifted over $\Phi | \mathfrak{S}_2$. Moreover, since for an ω-sliced extension $(\mathcal{E}, \rho, \sigma)$ the fundamental $\mathfrak{S}_2$-module $A(\varphi, f_2) \subseteq R_{1_{\omega''}}$, the preceding considerations give the following.

Corollary: Let $(\mathcal{E}, \rho, \sigma)$ be an ω-sliced extension of $\mathfrak{B}$ by $\mathfrak{A}$ with simple fundamental $\mathfrak{S}_2$-module $A(\varphi, f_2)$. Suppose further that $\mathfrak{A}$ induces on $A(\varphi, f_2)$ the structure of a Lie algebra $\mathfrak{A}(\varphi, f_2)$. Then $\mathfrak{A}(\varphi, f_2)$ is Abelian.

A few comments are now necessary.

(1) Consider a primitive extension $(\mathcal{E}, \rho)$ of $\mathfrak{B}$ by $\mathfrak{A}$ and an ω-sliced extension $(\mathcal{E}, \rho, \sigma)$ with fundamental $\mathfrak{S}_2$-module $A(\varphi, f_2)$. $\mathfrak{A}$ is Abelian by Theorem 2 and $A(\varphi, f_2) = A_\omega$. Besides, φ is an irreducible representation of $\mathfrak{B}$ into $\mathrm{D}(\mathfrak{A})$ and A becomes a simple $\mathfrak{B}$-module A_φ or $A = \{0\}$. If $\mathfrak{N}$ is the nilradical of $\mathfrak{B}$, i.e., the intersection of the kernels of all finite-dimensional irreducible representations of $\mathfrak{B}$ (the "radical nilpotent" of Bourbaki,[29] which is contained in the "nil radical" of Jacobson[14]), then by definition $\varphi(\mathfrak{N}) = \{0\}$.

(2) Let $(\mathcal{E},\rho)$ and $(\mathcal{E}',\rho')$ be essential primitive extensions of $\mathfrak{B}=\mathfrak{S}_2\dotplus\mathfrak{N}$ by $\mathfrak{A}$ with character Φ, and suppose that $\wedge^2 N_\Omega$ is a simple $\mathfrak{S}_2$-module, where Ω is given by (1.10). Then $\mathcal{E}\approx\mathcal{E}'$, i.e., $(\mathcal{E},\rho)$ and $(\mathcal{E}',\rho')$ are of the same type.[11]

In order to see this, let $(\mathcal{E},\rho,\sigma)$ and $(\mathcal{E}',\rho',\sigma')$ be ω-sliced extensions with, respectively, $(2;\mathfrak{B},\mathfrak{A},\Phi)$-pseudococylces (φ,f_2) and (φ',f_2'). Notice that $\varphi=\varphi'$, since $\mathfrak{A}$ is Abelian. It is now possible to define an isomorphism $\theta:\mathcal{E}\rightarrowtail\mathcal{E}'$ by

$$\theta(\sigma(b))=\sigma'(b),\qquad \theta(f_2(b,b'))=f_2'(b,b')$$
$$\text{for all } b,b'\in\mathfrak{B}.$$

In fact, two arbitrary elements $e,e'\in\mathcal{E}$ can be written as $e=\sigma(b)+\sum_{i,j}\alpha_{ij}f_2(n_i,n_j)$ and $e'=\sigma(b')+\sum_{i,j}\alpha'_{ij}f_2(n_i,n_j)$, $\{n_i\}$ being a basis of $\mathfrak{N}$ and $\{\alpha_{ij}^{(\prime)}\}\subset\mathbf{F}$. Let $b=s+n$ and $b'=s'+n'$, where $s,s'\in\mathfrak{S}_2$, $n,n'\in\mathfrak{N}$. Then

$$\begin{aligned}[\theta(e),\theta(e')]&=\sigma'([b,b'])+\sum_{i,j}\alpha'_{ij}\\&\quad\times\{f_2'([s,n_i],n_j)+f_2'(n_i,[s,n_j])\}\\&\quad-\sum_{i,j}\alpha_{ij}\{f_2'([s',n_i],n_j)+f_2'(n_i,[s',n_j])\}+f_2'(n,n')\\&=\theta([e,e']),\end{aligned}$$

and obviously $\operatorname{Ker}\theta=\{0\}$.

2. IRREDUCIBLE EXTENSIONS AND GRAPH THEORY

A. Irreducible Extensions

Consider a $(2;\mathfrak{B},\mathfrak{A},\Phi)$-pseudococycle (φ,f_2). Starting from the vector space $\operatorname{Im}f_2$, define some other subspaces of A as follows:

$$\varphi^i(\mathfrak{B})\operatorname{Im}f_2=\{(\varphi(b_1))^{i_1}(\varphi(b_2))^{i_2}\cdots(\varphi(b_j))^{i_j}a\,|\,b_1,b_2,\cdots,b_j\in\mathfrak{B};a\in\operatorname{Im}f_2;\sum_{k=1}^{j}i_k=i\text{ with }i_k,j\in\mathbf{N}^*\},\ i\in\mathbf{N}^*,$$

and

$$\varphi^0(\mathfrak{B})\operatorname{Im}f_2=\overline{\operatorname{Im}f_2}.$$

Then

$$A'=\sum_{i\in\mathbf{N}}\varphi^i(\mathfrak{B})\operatorname{Im}f_2$$

is a subspace of A, but, in general, $\varphi^j(\mathfrak{B})\operatorname{Im}f_2\cap\varphi^k(\mathfrak{B})\operatorname{Im}f_2\neq\{0\}$. $\mathfrak{A}$ induces on A' a Lie algebra structure, because of the following.

Theorem 3: Let $(\mathcal{E},\rho,\sigma)$ be a sliced extension of $\mathfrak{B}$ by $\mathfrak{A}$ with character Φ and $(2;\mathfrak{B},\mathfrak{A},\Phi)$-pseudococycle (φ,f_2). The vector space $A'=\sum_{i\in\mathbf{N}}\varphi^i(\mathfrak{B})\operatorname{Im}f_2$ becomes a Lie algebra $\mathfrak{A}'$ with

$$[\varphi^i(\mathfrak{B})\operatorname{Im}f_2,\varphi^j(\mathfrak{B})\operatorname{Im}f_2]\subseteq\varphi^{i+j+2}(\mathfrak{B})\operatorname{Im}f_2+\varphi^{i+j+1}(\mathfrak{B})\operatorname{Im}f_2.$$

Proof: φ is a prerepresentation of $\mathfrak{B}$ into $D(\mathfrak{A})$ associated with f_2,[11] then

$$\begin{aligned}&\varphi(b_1)\varphi(b_2)f_2(b_3,b_4)-\varphi(b_2)\varphi(b_1)f_2(b_3,b_4)\\&\quad=\varphi([b_1,b_2])f_2(b_3,b_4)+[f_2(b_1,b_2),f_2(b_3,b_4)]\\&\quad\text{for all } b_1,b_2,b_3,b_4\in\mathfrak{B}.\end{aligned}$$

Therefore

$$[\overline{\operatorname{Im}f_2},\overline{\operatorname{Im}f_2}]\subseteq\varphi^2(\mathfrak{B})\operatorname{Im}f_2+\varphi(\mathfrak{B})\operatorname{Im}f_2,$$

and analogously

$$[\varphi^i(\mathfrak{B})\operatorname{Im}f_2,\overline{\operatorname{Im}f_2}]\subseteq\varphi^{i+2}(\mathfrak{B})\operatorname{Im}f_2+\varphi^{i+1}(\mathfrak{B})\operatorname{Im}f_2\qquad\text{for all } i\in\mathbf{N}^*.$$

With $i\in\mathbf{N}^*$ and the induction assumption

$$[\varphi^{i-1}(\mathfrak{B})\operatorname{Im}f_2,\varphi^j(\mathfrak{B})\operatorname{Im}f_2]\subseteq\varphi^{i+j+1}(\mathfrak{B})\operatorname{Im}f_2+\varphi^{i+j}(\mathfrak{B})\operatorname{Im}f_2\qquad\text{for all } j\in\mathbf{N},$$

we find

$$\begin{aligned}&[\varphi^i(\mathfrak{B})\operatorname{Im}f_2,\varphi^j(\mathfrak{B})\operatorname{Im}f_2]\\&\quad=[[\sigma_v(B),\varphi^{i-1}(\mathfrak{B})\operatorname{Im}f_2],\varphi^j(\mathfrak{B})\operatorname{Im}f_2]\\&\quad\subseteq[[\varphi^{i-1}(\mathfrak{B})\operatorname{Im}f_2,\varphi^j(\mathfrak{B})\operatorname{Im}f_2],\sigma_v(B)]\\&\qquad+[[\varphi^j(\mathfrak{B})\operatorname{Im}f_2,\sigma_v(B)],\varphi^{i-1}(\mathfrak{B})\operatorname{Im}f_2]\\&\quad\subseteq[\varphi^{i+j+1}(\mathfrak{B})\operatorname{Im}f_2,\sigma_v(B)]+[\varphi^{i+j}(\mathfrak{B})\operatorname{Im}f_2,\sigma_v(B)]\\&\qquad+[\varphi^{j+1}(\mathfrak{B})\operatorname{Im}f_2,\varphi^{i-1}(\mathfrak{B})\operatorname{Im}f_2]\\&\quad\subseteq\varphi^{i+j+2}(\mathfrak{B})\operatorname{Im}f_2+\varphi^{i+j+1}(\mathfrak{B})\operatorname{Im}f_2.\end{aligned}$$

The conclusion is obvious. ■

Define

$$E'=\overline{\{\sigma(b)+a\,|\,b\in\mathfrak{B};a\in\mathfrak{A}'\}}.$$

E' becomes a Lie algebra $\mathcal{E}'\subseteq\mathcal{E}$ with the Lie algebra structure induced by $\mathcal{E}$. Hence $(\mathcal{E}',\rho')$, where $\rho'=\rho|\mathcal{E}'$, is an extension of $\mathfrak{B}$ by $\mathfrak{A}'$. Choosing another sliced extension $(\mathcal{E},\rho,\sigma')$ we obtain, in general, a different extension $(\mathcal{E}'',\rho'')$. In particular, $(\mathcal{E}',\rho')$ and $(\mathcal{E}'',\rho'')$ can be irreducible and different even if $\mathfrak{A}$ is Abelian, contrarily to a claim of uniqueness in Theorem 4 of Ref. 10. Now we will study $(\mathcal{E}',\rho')$, generalizing some ideas of Ref. 11.

Definition 6: Let $(\mathcal{E},\rho,\sigma)$ be a sliced extension of $\mathfrak{B}$ by $\mathfrak{A}$ with character Φ and $(2;\mathfrak{B},\mathfrak{A},\Phi)$-pseudococycle (φ,f_2). We say that $\overline{\operatorname{Im}f_2}$ is *maximal in* A if $\overline{\operatorname{Im}f_2}\neq A$ and there exists no ideal $\mathfrak{A}'$ of $\mathcal{E}$ such that $\overline{\operatorname{Im}f_2}\subseteq A'\subset A$.

Theorem 4: With the notation of Definition 6, a necessary condition in order to have the irreducibility of $(\mathcal{E},\rho)$ is that $\overline{\operatorname{Im}f_2}=A$ or $\overline{\operatorname{Im}f_2}$ be maximal in A.

Proof: Suppose $(\mathcal{E},\rho)$ irreducible. If $\overline{\operatorname{Im}f_2}$ is not maximal in A, then $\overline{\operatorname{Im}f_2}=A$ or there exists an ideal $\mathfrak{A}'$ of $\mathcal{E}$ such that $\overline{\operatorname{Im}f_2}\subseteq A'\subset A$. In the latter case we have an induced sliced extension $(\mathcal{E}/\mathfrak{A}',\rho_q,\sigma_q)$ of $\mathfrak{B}$, where $\sigma_q(b)=\overline{\sigma(b)}$ (mod $\mathfrak{A}'$), with $(2;\mathfrak{B},\mathfrak{A}/\mathfrak{A}',\Phi_q)$-pseudococycle (φ_q,f_{2q}). Actually $f_{2q}=0$, and therefore $(\mathcal{E}/\mathfrak{A}',\rho_q)$ is inessential. By the Irreducibility Criterion (Theorem 7 of Ref. 11) $(\mathcal{E},\rho)$ is reducible, and this contradicts our assumption. ■

Theorem 5: In order that an extension $(\mathcal{E}, \rho)$ of $\mathfrak{B}$ by $\mathfrak{A}$ with character Φ be irreducible, a necessary condition is that

$$A = \sum_{i \in \mathbf{N}} \varphi^i(\mathfrak{B})\, \mathrm{Im} f_2,$$

where (φ, f_2) is the $(2; \mathfrak{B}, \mathfrak{A}, \Phi)$-pseudococycle of an arbitrary sliced extension related to $(\mathcal{E}, \rho)$.

Proof: Let $(\mathcal{E}, \rho)$ be irreducible and consider $A' = \sum_{i \in \mathbf{N}} \varphi^i(\mathfrak{B})\, \mathrm{Im} f_2$. Suppose $A' \subset A$. We have $\overline{\mathrm{Im} f_2} \subseteq A' \subset A$ and, by Theorem 3, A' becomes a Lie algebra $\mathfrak{A}'$ which, moreover, is an ideal of $\mathcal{E}$. This means that $\overline{\mathrm{Im} f_2}$ is not maximal in A and, by Theorem 4, the reducibility of $(\mathcal{E}, \rho)$ contrary to our assumption. ■

The importance of irreducible extensions is easily recognized: They form the core of any extension. The determination of the extensions of $\mathfrak{B}$ by $\mathfrak{A}$ is simplified once the irreducible extensions of $\mathfrak{B}$ by some subalgebras of $\mathfrak{A}$ are known. Now we will restrict ourselves mainly to the study of irreducible extensions.

From the structure theorem we see immediately that all extensions by a nonsolvable Lie algebra are reducible, and, therefore, we always have $\mathcal{S}_1 = \{0\}$ and $\mathfrak{A} = \mathcal{R}_1$ whenever irreducible extensions are considered.

Let $(\mathcal{E}, \rho)$ be an irreducible extension of $\mathfrak{B}$ by $\mathfrak{A}$ with character Φ, and let again (φ, f_2) be the $(2; \mathfrak{B}, \mathfrak{A}, \Phi)$-pseudococycle of a sliced extension $(\mathcal{E}, \rho, \sigma)$. By Theorem 5, A can be expressed as a sum of subspaces obtained from $\overline{\mathrm{Im} f_2}$ by iterated application of the linear operators $\varphi(b)$ for all $b \in \mathfrak{B}$. If we choose an ω-sliced extension, we can expect to obtain a better characterization of the conditions which must be satisfied by $\mathfrak{A}$. In fact we know that, in this case, A becomes a semisimple $\mathcal{S}_2$-module A_ω. Moreover, $\overline{\mathrm{Im} f_2}$ becomes the fundamental $\mathcal{S}_2$-module $A(\varphi, f_2)$ and, for any $i \in \mathbf{N}$, we can consider the $\mathcal{S}_2$-module $\varphi^i(\mathcal{R}_2)A(\varphi, f_2)$ instead of $\varphi^i(\mathfrak{B})\, \mathrm{Im} f_2$, since

$$\begin{aligned}
&\omega(s)(\varphi(r_1))^{i_1}(\varphi(r_2))^{i_2} \cdots (\varphi(r_j))^{i_j} a \\
&= \Bigg(\sum_{k=1}^{j} \sum_{l=1}^{i_k} (\varphi(r_1))^{i_1} (\varphi(r_2))^{i_2} \cdots (\varphi(r_k))^{i_k - l} \varphi([s, r_k]) \\
&\quad \times (\varphi(r_k))^{l-1} \cdots (\varphi(r_j))^{i_j} a + (\varphi(r_1))^{i_1} (\varphi(r_2))^{i_2} \\
&\quad \times \cdots (\varphi(r_j))^{i_j} \omega(s) a \Bigg) \in \varphi^i(\mathcal{R}_2) A(\varphi, f_2)
\end{aligned}$$

for all $s \in \mathcal{S}_1$ and all elements $(\varphi(r_1))^{i_1} \cdots (\varphi(r_j))^{i_j} a$ which generate $\varphi^i(\mathcal{R}_2)A(\varphi, f_2)$ $[\sum_{k=1}^{j} i_k = i,\ a \in A(\varphi, f_2)]$. Note that if $A_{1\omega_1}$ and $A_{2\omega_2}$ are $\mathcal{S}_2$-submodules of A_ω, then $[A_1, A_2]$ becomes an $\mathcal{S}_2$-submodule $[A_{1\omega_1}, A_{2\omega_2}]$ too. Then we may replace "sliced extension" by "ω-sliced extension," "vector space" by "$\mathcal{S}_2$-module", $\overline{\mathrm{Im} f_2}$ by $A(\varphi, f_2)$, $\varphi^i(\mathfrak{B})\mathrm{Im} f_2$ by $\varphi^i(\mathcal{R}_2)A(\varphi, f_2)$, A' by $A'_{\omega'}$, and A by A_ω in Theorems 3, 4, 5 and Definition 6 without changing the outcomes. Besides, some previous results can now be improved:

(i) In the particular case of extensions of a Lie algebra with Abelian radical $\mathcal{R}_2$, it is easy to prove (see proof of Theorem 3) that

$$[\varphi^i(\mathcal{R}_2)A(\varphi, f_2), \varphi^j(\mathcal{R}_2)A(\varphi, f_2)] \subseteq \varphi^{i+j+2}(\mathcal{R}_2)A(\varphi, f_2),$$

since

$$[A(\varphi, f_2), A(\varphi, f_2)] \subseteq \varphi^2(\mathcal{R}_2)A(\varphi, f_2).$$

(ii) Consider an irreducible extension $(\mathcal{E}, \rho)$ of $\mathfrak{B}$ by $\mathfrak{A}$ with character 0. There is a 0-sliced extension $(\mathcal{E}, \rho, \sigma)$ with fundamental $\mathcal{S}_2$-module $A(0, f_2)$. From Theorem 3 and Theorem 5 it is plain that $\mathfrak{A} = \mathfrak{A}(0, f_2)$ is Abelian, i.e., all irreducible extensions with character 0 are central.

(iii) An obvious consequence of Theorem 5 is the following.

Corollary: Let $(\mathcal{E}, \rho, \sigma)$ be an ω-sliced extension of $\mathfrak{B}$ by $\mathfrak{A}$ with character Φ and fundamental $\mathcal{S}_2$-module $A(\varphi, f_2) \neq A_\omega$. Suppose further that $\varphi(r)a = 0$ for all $r \in \mathcal{R}_2$, $a \in A(\varphi, f_2)$. Then the extension $(\mathcal{E}, \rho)$ is reducible.

$(\mathcal{E}, \rho)$ being the same irreducible extension as above, suppose that there exists $n \in \mathbf{N}$ such that $\varphi^n(\mathcal{R}_2)$ $A(\varphi, f_2) = \{0\}$. Then $\mathfrak{A}$ is a nilpotent Lie algebra. In particular this is always true if the radical $\mathcal{R}$ of $\mathcal{E}$ is nilpotent. Now we give some sufficient conditions in order that $\mathcal{R}$ be nilpotent. Notice that the radical $\mathcal{R}_2$ of $\mathfrak{B} = \mathcal{S}_2 \dotplus \mathcal{R}_2$ becomes an $\mathcal{S}_2$-module R_{2_Ω} associated with the representation Ω given by (1.10).

Theorem 6: Let $(\mathcal{E}, \rho)$ be an irreducible extension of $\mathfrak{B}$ by $\mathfrak{A}$. If $\mathfrak{B} = \mathcal{S}_2 \dotplus \mathcal{R}_2$ with $\mathcal{R}_2$ nilpotent, and if the $\mathcal{S}_2$-modules R_{2_Ω} and $\overset{2}{\wedge} R_{2_\Omega}$ have no trivial $\mathcal{S}_2$-submodules $\neq \{0\}$, then the radical $\mathcal{R}$ of $\mathcal{E}$ is nilpotent.

Proof: Let Φ be the character of $(\mathcal{E}, \rho)$ and consider an ω-sliced extension $(\mathcal{E}, \rho, \sigma)$ with $(2; \mathfrak{B}, \mathfrak{A}, \Phi)$ pseudococycle (φ, f_2). If $f_2 = 0$, $(\mathcal{E}, \rho)$ is inessential and the irreducibility requires $\mathfrak{A} = \{0\}$. In this case, $\mathcal{R}$ is nilpotent since $\sigma(\mathcal{R}_2) = \mathcal{R}$, σ being an isomorphism. Suppose now $f_2 \neq 0$ and let $\varphi' \in \hom(\mathcal{S}_2, D(\mathcal{R}))$ be defined by $\varphi'(s)r = [\sigma(s), r]$ for all $s \in \mathcal{S}_2$, $r \in \mathcal{R}$. In particular, we have $\varphi'(s)a = \omega(s)a$ for all $a \in A$. If $\mathcal{R}$ is not nilpotent, the biggest nilpotent ideal $\mathfrak{N}$ of $\mathcal{R}$ (the "nil radical" of Jacobson[14]) is a proper subalgebra of $\mathcal{R}$. Since $\mathfrak{N}$ is a characteristic ideal of $\mathcal{E}$, we have a proper submodule $N_{\varphi''}$ of the $\mathcal{S}_2$-module $R_{\varphi'}$, φ'' being the subrepresentation of φ' on N. By Weyl's theorem there is a supplementary $\mathcal{S}_2$-module of $N_{\varphi''}$ in $R_{\varphi'}$, and this is a trivial module R'_0 since any derivation of $\mathcal{R}$ sends elements of $\mathcal{R}$ into $\mathfrak{N}$.[14,29] Therefore,

$$R_{\varphi'} = N_{\varphi''} \oplus R'_0.$$

$\sigma_v(R_2) \subset R$ becomes also an $\mathcal{S}_2$-module $(\sigma_v(R_2))_{\varphi'''}$ of $R_{\varphi'}$, φ''' being the subrepresentation of φ' on $\sigma_v(R_2)$. Moreover, by our assumptions,

$$(\sigma_v(R_2))_{\varphi'''} \cap R'_0 = \{0\},$$

since $(\sigma_v(R_2))_{\varphi'''}$ is isomorphic to the $\mathcal{S}_2$-module R_{2_Ω}, and

$$A(\varphi, f_2) \cap R'_0 = \{0\},$$

since $A(\varphi, f_2)$ is the direct sum of nontrivial simple $\mathcal{S}_2$-modules. Then there is an extension $(\mathcal{E}', \rho')$ of $\mathfrak{B}$ with

$$\mathcal{E}' = [(\sigma(\mathcal{S}_2) \oplus \mathcal{S}_1) \oplus \mathfrak{N}] \subset \mathcal{E}$$

and $\rho' = \rho|\mathcal{E}'$, and this contradicts the irreducibility of $(\mathcal{E}, \rho)$. It follows that $\mathcal{R} = \mathfrak{N}$, i.e., $\mathcal{R}$ is nilpotent. ■

Notice that if we assume $\mathcal{R}_2$ Abelian, the fundamental $\mathcal{S}_2$-module $A(\varphi, f_2)$ of $(\mathcal{E}, \rho, \sigma)$ is contained in $N_{\varphi''}$, since

$$f_2(r, r') = [\sigma(r), \sigma(r')] \quad \text{for all } r, r' \in \mathcal{R}_2$$

and $D^1\mathcal{R} \subseteq \mathfrak{N}$.

Corollary 1: Let $(\mathcal{E}, \rho)$ be an irreducible extension of $\mathcal{B}$ by $\mathcal{A}$ and let $\mathcal{B} = \mathcal{S}_2 \dotplus \mathcal{R}_2$ have Abelian radical. In order that the radical $\mathcal{R}$ of $\mathcal{E}$ be nilpotent, a sufficient condition is that the $\mathcal{S}_2$-module R_{2_Ω} have no trivial $\mathcal{S}_2$-submodules $\neq \{0\}$.

The biggest nilpotent ideal of a Lie algebra contains all nilpotent ideals. This implies the following.

Corollary 2: Let $(\mathcal{E}, \rho)$ be an irreducible extension of $\mathcal{B}$ by a nilpotent Lie algebra $\mathcal{A}$. Suppose further that $\mathcal{B} = \mathcal{S}_2 \dotplus \mathcal{R}_2$ with nilpotent $\mathcal{R}_2$. In order that the radical $\mathcal{R}$ of $\mathcal{E}$ be nilpotent, it is sufficient that the $\mathcal{S}_2$-module R_{2_Ω} have no trivial $\mathcal{S}_2$-submodules $\neq \{0\}$.

Any subalgebra of a nilpotent Lie algebra is nilpotent. It follows:

Corollary 3: With the assumptions of Theorem 6 (or Corollary 1) the Lie algebra $\mathcal{A}$ is nilpotent.

Let $(\mathcal{E}, \rho)$ be an irreducible extension of $\mathcal{B}$ by $\mathcal{A}$ with character Φ, and let $\mathcal{R}$ be nilpotent. Consider an ω-sliced extension $(\mathcal{E}, \rho, \sigma)$ with fundamental $\mathcal{S}_2$-module $A(\varphi, f_2)$. By Theorem 5 there exists $n \in \mathbf{N}$ such that

$$\varphi^n(\mathcal{R}_2)A(\varphi, f_2) = \{0\}, \qquad \varphi^j(\mathcal{R}_2)A(\varphi, f_2) \neq \{0\} \quad \text{for all } j < n \tag{2.1}$$

and

$$A_\omega = \sum_{i=0}^{n-1} \varphi^i(\mathcal{R}_2)A(\varphi, f_2) \quad \text{for } n \in \mathbf{N}^*, \qquad A_\omega = \{0\} \quad \text{for } n = 0. \tag{2.2}$$

Consider $(\mathcal{E}, \rho, \sigma')$ with $\sigma' \neq \sigma$ and fundamental $\mathcal{S}_2$-module $A(\varphi', f_2')$. There exists $n' \in \mathbf{N}$ such that the relations (2.1) and (2.2), with n replaced by n' and $A(\varphi, f_2)$ by $A(\varphi', f_2')$, are satisfied. Now we will show that $n' = n$. Since (φ, f_2) and (φ', f_2') are equivalent elements of $\mathfrak{Z}^2_\Phi(\mathcal{B}, \mathcal{A})$, there is $f_1 \in \mathrm{L}(\mathcal{B}, \mathcal{A})$ such that

$$\varphi'(r) = \varphi(r) + \mathrm{ad} f_1(r)$$

and

$$f_2'(r, r') = f_2(r, r') + (\delta_1(\varphi)f_1)(r, r') + [f_1(r), f_1(r')] \quad \text{for all } r, r' \in \mathcal{R}_2.$$

But

$$\mathrm{Im} f_1 \subseteq \sum_{i=0}^{n} \varphi^i(\mathcal{R}_2)A(\varphi, f_2),$$

and hence

$$(\varphi'(r_1))^{i_1}(\varphi'(r_2))^{i_2} \cdots (\varphi'(r_j))^{i_j} f_2'(r, r') = (\varphi(r_1))^{i_1}(\varphi(r_2))^{i_2} \cdots (\varphi(r_j))^{i_j} f_2(r, r') + g^{i_1 \cdots i_j}_{r_1 \cdots r_j}(r, r'),$$

where

$$\sum_{k=1}^{j} i_k = i, \qquad r_1, \ldots, r_j, r, r' \in \mathcal{R}_2$$

$$\text{and } g^{i_1 \cdots i_j}_{r_1 \cdots r_j}(r, r') \in \sum_{k \geq i} \varphi^k(\mathcal{R}_2)A(\varphi, f_2).$$

This implies the existence of $m \leq n$ such that $g^{i_1 \cdots i_{j'}}_{r_1 \cdots r_{j'}}(r, r') = 0$ for all $r_1, \ldots, r_{j'}, r, r' \in \mathcal{R}_2$ and all $i_1, \ldots, i_{j'}$ with $\sum_{k=1}^{j'} i_k = m$. Consequently $n' \leq n$. Conversely, if we start with $(\mathcal{E}, \rho, \sigma')$, we obtain $n \leq n'$, and therefore $n = n'$. This allows the partition of the set of irreducible extensions of $\mathcal{B}$ into two disjoint parts.

(1) We say that an extension $(\mathcal{E}, \rho)$ of $\mathcal{B}$ belongs to $\mathfrak{E}_n(\mathcal{B})$ $(n \in \mathbf{N})$ if it is irreducible and if there exists an ω-sliced extension $(\mathcal{E}, \rho, \sigma)$ with fundamental $\mathcal{S}_2$-module $A(\varphi, f_2)$ such that (2.1) and (2.2) are satisfied.

(2) We say that $(\mathcal{E}, \rho)$ belongs to $\mathfrak{E}_\infty(\mathcal{B})$ if it is irreducible and if no $n \in \mathbf{N}$ exists such that the fundamental $\mathcal{S}_2$-module $A(\varphi, f_2)$ of an arbitrary ω-sliced extension $(\mathcal{E}, \rho, \sigma)$ satisfies (2.1) and (2.2).

The extensions belonging to $\mathfrak{E}_n(\mathcal{B})$ are irreducible extensions by nilpotent Lie algebras. However, extensions of $\mathcal{B}$ by Abelian Lie algebras which belong to $\mathfrak{E}_\infty(\mathcal{B})$ are not ruled out.

The following result follows immediately from the remarks above.

Theorem 7: In order that $(\mathcal{E}, \rho)$ be an element of $\mathfrak{E}_n(\mathcal{B})$ with $n \in \mathbf{N}^*$, it is necessary that (2.1) and (2.2) be satisfied for an arbitrary ω-sliced extension of $\mathcal{B}$ by $\mathcal{A}$ related to $(\mathcal{E}, \rho)$ and with fundamental $\mathcal{S}_2$-module $A(\varphi, f_2)$. If $n = 0$, $\mathcal{A} = \{0\}$ is a necessary and sufficient condition.

B. Graphs Associated with an ω-Sliced Extension

Consider the fundamental $\mathcal{S}_2$-module $A(\varphi, f_2)$ of an ω-sliced extension $(\mathcal{E}, \rho, \sigma)$ of $\mathcal{B}$ by $\mathcal{A}$ and the $\mathcal{S}_2$-submodules $\varphi^i(\mathcal{R}_2)A(\varphi, f_2)$ $(i \in \mathbf{N})$ of A_ω. Weyl's theorem allows a deeper analysis of these $\mathcal{S}_2$-modules, since they are all semisimple and then expressible as sum of simple $\mathcal{S}_2$-submodules.

Let

$$\sum_{i \in \mathbf{N}} \varphi^i(\mathcal{R}_2)A(\varphi, f_2) = A'_\omega \subseteq A_\omega.$$

A'_ω can be given as sum of a family of simple $\mathcal{S}_2$-submodules. There are, in general, many families for which the sum is A'_ω, and some give A'_ω as a direct sum. We will now construct a family which is important for our purposes. Let

$$A(\varphi, f_2) = \bigoplus_{i_0 \in I^0} A^{i_0}_{\omega_{i_0}}$$

be a direct sum decomposition of $A(\varphi, f_2)$ in simple $\mathcal{S}_2$-submodules, where ω_{i_0} is the subrepresentation of ω on A^{i_0} and I^0 the index set of the family $\left(A^{i_0}_{\omega_{i_0}}\right)$.

For any $i_0 \in I^0$ take a direct sum decomposition of $\varphi(\mathcal{R}_2)A^{i_0}_{\omega_{i_0}}$ indexed by $I^1_{i_0}$ and let $I^1 = \cup_{i_0 \in I^0} I^1_{i_0}$.

Take afterwards, for any $i_1 \in I^1$, a direct sum decomposition of $\varphi(\mathcal{R}_2)A^{i_1}_{\omega_{i_1}}$ indexed by $I^2_{i_1}$, let $I^2 =$

$\cup_{i_1 \in I^1} I^2_{i_1}$, and so on. We have ultimately a family of simple $\mathcal{S}_2$-submodules indexed by $I' = I/R$ such that $A'_{\omega'} = \sum_{i \in I'} A^i_{\omega_i}$, where $I = \cup_{j \in \mathbf{N}} I^j$ and R is the equivalence relation in I obtained by the identification of i and j whenever $A^i_{\omega_i} = A^j_{\omega_j}$. Notice that $I^{k+1}_{i^k} = \emptyset$, as well as $I^j = \emptyset$ and $I = \emptyset$, are not excluded ($k, j \in \mathbf{N}$). We call a *principal family* of $A'_{\omega'}$ any family of simple $\mathcal{S}_2$-submodules with sum $A'_{\omega'}$ which is defined as above. If, moreover, $A'_{\omega'} = A_\omega$, we say that $(A^i_{\omega_i})_{i \in I'}$ is a *maximal principal family*.

Now it is possible to associate a graph[31,32] with the ω-sliced extension $(\mathcal{E}, \rho, \sigma)$, i.e., an ordered pair $(\mathfrak{H}, \Gamma)$, where

$$\mathfrak{H} = \{A^j_{\omega_j} \mid A^j_{\omega_j} \in (A^i_{\omega_i})_{i \in I'}\}$$

is the set of vertices and Γ is a multivalued mapping of $\mathfrak{H}$ into $\mathfrak{H}$ defined as follows:

$$\Gamma A^{i_j}_{\omega_{i_j}} = \{A^k_{\omega_k} \mid k \in I^{j+1}_{i_j}\} \quad \text{for all } i_j \in I^j,\ j \in \mathbf{N}.$$

We say that $(\mathfrak{H}, \Gamma)$ is the graph associated with $(\mathcal{E}, \rho, \sigma)$ by the principal family $(A^i_{\omega_i})_{i \in I'}$, or simply a graph associated with $(\mathcal{E}, \rho, \sigma)$. We will also speak of a "graph of $(\mathcal{E}, \rho)$," meaning a graph associated with an ω-sliced extension related to $(\mathcal{E}, \rho)$. If $(A^i_{\omega_i})_{i \in I'}$ is a maximal principal family, we call the graph associated by it a *maximal graph*. Clearly, if there is a maximal graph associated with $(\mathcal{E}, \rho, \sigma)$, all other graphs associated with $(\mathcal{E}, \rho, \sigma)$ are maximal. Note that, for the same reasons as in Sec. 1A, a given graph can be associated with many ω-sliced extensions.

In the graphs associated with ω-sliced extensions, a pair of vertices can be joined either by zero, one, or by two arcs (directed edges). In this latter case the arcs have opposite directions. But, in general, we do not have a digraph (directed graph), in the sense of Ref. 32, since loops are allowed. Moreover, we consider also infinite graphs and a graph with $\mathfrak{H} = \emptyset$ which will be called the empty graph.

The graphs of extensions belonging to $\mathfrak{E}_n(\mathcal{B})$ ($n \in \mathbf{N}$) are finite and without circuits (directed cycles and loops in Ref. 32). The primitive essential extensions have only one-vertex graphs, and the inessential primitive extensions (the truly trivial extensions) are characterized by the empty graph.

Define a subgraph of $(\mathfrak{H}, \Gamma)$ as in Ref. 31 (induced subgraph in Ref. 32), i.e., as a graph $(\mathfrak{K}, \Gamma_{\mathfrak{K}})$, where $\mathfrak{K} \subseteq \mathfrak{H}$ and $\Gamma_{\mathfrak{K}}$ is given by

$$\Gamma_{\mathfrak{K}} x = \Gamma x \cap \mathfrak{K} \quad \text{for all } x \in \mathfrak{K}.$$

Definition 7: Let $(\mathfrak{H}, \Gamma)$ be the graph associated with an ω-sliced extension of $\mathcal{B}$ by $\mathcal{A}$ by a principal family $(A^i_{\omega_i})_{i \in I'}$. We call a *fundamental subgraph* of $(\mathfrak{H}, \Gamma)$ the graph $(\mathfrak{H}', \Gamma_{\mathfrak{H}'})$ where

$$\mathfrak{H}' = \{A^{i_0}_{\omega_{i_0}} \mid i_0 \in I^0\}.$$

In any graph $(\mathfrak{H}, \Gamma)$ the transitive closure $\hat{\Gamma}$ of Γ (Ref. 31) is a multivalued map of $\mathfrak{H}$ into $\mathfrak{H}$ defined by

$$\hat{\Gamma} x = \{x\} \cup \Gamma x \cup \Gamma^2 x \cup \cdots \quad \text{for all } x \in \mathfrak{H}.$$

$\hat{\Gamma} x$ is the set of vertices reachable from x. With the notation of Definition 7 we can define the set

$$\hat{\mathfrak{H}}' = \bigcup_{i_0 \in I^0} \hat{\Gamma} A^{i_0}_{\omega_{i_0}}$$

and the subgraph $(\hat{\mathfrak{H}}', \Gamma_{\hat{\mathfrak{H}}'})$. Then by definition $(\mathfrak{H}, \Gamma) = (\hat{\mathfrak{H}}', \Gamma_{\hat{\mathfrak{H}}'})$.

Define a vertex basis of a nonempty graph $(\mathfrak{H}, \Gamma)$ as a minimal collection of vertices from which all vertices are reachable, i.e., as a set $\mathfrak{B} \subseteq \mathfrak{H}$ such that

(i) $b, b' \in \mathfrak{B}$ and $b \neq b'$ imply $b \notin \hat{\Gamma} b'$ and $b' \notin \hat{\Gamma} b$,

(ii) for any $x \in \mathfrak{H}$ there exists $b \in \mathfrak{B}$ such that $x \in \hat{\Gamma} b$.

It is clear that a graph $(\mathfrak{H}, \Gamma)$ associated with an ω-sliced extension $(\mathcal{E}, \rho, \sigma)$ has vertex bases which are subsets of $\mathfrak{H}'$. If $(\mathfrak{H}, \Gamma)$ is a finite graph without circuits, then the vertex basis is unique, and all vertices of this basis are of indegree 0 (Ref. 32), i.e., they are not terminal points of any arc. This is the case for the graphs of extensions belonging to $\mathfrak{E}_n(\mathcal{B})$ ($n \in \mathbf{N}^*$).

The set of ω-sliced extensions can be enlarged by removing the conditions $\varphi(\mathcal{B})\mathcal{S}_1 = \{0\}$ and $\mathrm{Im} f_2 \subseteq \mathcal{R}_1$ of Definition 3. We say that $(\mathcal{E}, \rho, \sigma)$ is a *quasi-ω-sliced extension* of $\mathcal{B}$ by $\mathcal{A}$ with character Φ if its $(2; \mathcal{B}, \mathcal{A}, \Phi)$-pseudococycle $(\varphi, f_2) \in \mathfrak{Z}^2_{\Phi,\omega}(\mathcal{B}, \mathcal{S}_2, \mathcal{A})$. We have omitted, as usual, "for the Levi decomposition $\mathcal{B} = \mathcal{S}_2 \dotplus \mathcal{R}_2$." It is clear that, also in the case of quasi-ω-sliced extensions, we can still speak of the fundamental $\mathcal{S}_2$-module $A(\varphi, f_2)$ of $(\mathcal{E}, \rho, \sigma)$, of A_ω, of the graph $(\mathfrak{H}, \Gamma)$ associated with $(\mathcal{E}, \rho, \sigma)$ by a principal family $(A^i_{\omega_i})_{i \in I'}$, and of the fundamental subgraph of $(\mathfrak{H}, \Gamma)$. Theorem 5 may be improved as follows.

Theorem 8: An extension $(\mathcal{E}, \rho)$ of $\mathcal{B}$ by $\mathcal{A}$ is irreducible if and only if there is a maximal graph associated with $(\mathcal{E}, \rho, \sigma)$ for all ω-sliced extensions $(\mathcal{E}, \rho, \sigma)$ related to $(\mathcal{E}, \rho)$.

Proof: The condition is a necessary one by Theorem 5. In order to prove the converse, let $(\mathcal{E}, \rho)$ be reducible. Then there exists an extension $(\mathcal{E}', \rho')$ of $\mathcal{B}$ by $\mathcal{A}' \subset \mathcal{A}$, where $\mathcal{E}' \subset \mathcal{E}$ and $\rho' = \rho \mid \mathcal{E}'$. If $(\mathcal{E}', \rho', \sigma')$ is a quasi-ω-sliced extension with fundamental $\mathcal{S}_2$-module $A(\varphi', f'_2)$, consider the section $\sigma = \sigma'$ (up to a canonical monomorphism) of $(\mathcal{E}, \rho)$ over $\mathcal{B}$. $(\mathcal{E}, \rho, \sigma)$ is also a quasi-ω-sliced extension and, if $A(\varphi, f_2)$ is its fundamental $\mathcal{S}_2$-module, $A(\varphi, f_2) = A(\varphi', f'_2)$, since $f_2 = f'_2$ and $\varphi(b) \mid a' = \varphi'(b)\, a'$ for all $b \in \mathcal{B}$, $a' \in \mathcal{A}'$. Let $(A^j_{\omega_j})_{j \in J'}$ be a principal family of $A'_{\omega'}$. There exists a principal family $(A^i_{\omega_i})_{i \in I'}$ of A_ω with $J \subset I$ and $J^{j+1}_{i_j} = I^{j+1}_{i_j}$ for all $J^{j+1}_{i_j} \subseteq J$, and we consider the graph $(\mathfrak{H}, \Gamma)$ associated with $(\mathcal{E}, \rho, \sigma)$ by this family. Let $(\mathfrak{K}, \Theta)$ be the graph associated with $(\mathcal{E}', \rho', \sigma')$ by $(A^j_{\omega_j})_{j \in J'}$. We have $(\mathfrak{K}, \Theta) = (\mathfrak{K}, \Gamma)$, since $(\hat{\mathfrak{K}}', \Theta_{\hat{\mathfrak{K}}'}) = (\hat{\mathfrak{H}}', \Gamma_{\hat{\mathfrak{H}}'})$. Suppose that for any ω-sliced extension related to $(\mathcal{E}, \rho)$ there is an associated maximal graph. Then $\mathcal{A}$ is solvable, and every quasi-ω-sliced extension is ω-sliced. It follows that $(\mathfrak{H}, \Gamma)$ is maximal and $\mathcal{E}' = \mathcal{E}$, which is absurd. ■

In particular, we get from Theorem 8 that all primitive extensions are irreducible. The outdegree of a vertex is the cardinal number of the set of arcs with initial points in the given vertex. The following result is now obvious.

Corollary: Let $(\mathcal{E},\rho,\sigma)$ be an ω-sliced extension of $\mathfrak{B}$ by $\mathfrak{A}$ with fundamental $\mathfrak{S}_2$-module $A(\varphi,f_2)$ and let $(\mathfrak{H},\Gamma)$ be a graph associated with $(\mathcal{E},\rho,\sigma)$. Suppose $A(\varphi,f_2)\subset A_\omega$. If all vertices of the fundamental subgraph of $(\mathfrak{H},\Gamma)$ have outdegrees 0, the extension $(\mathcal{E},\rho)$ is reducible.

Let $(\mathcal{E},\rho)$ be an extension of a Lie algebra $\mathfrak{B}$ with Abelian radical $\mathfrak{R}_2$. The descending central series $(C^i\mathfrak{R})_{i\in\mathbf{N}}$ of ideals of the radical $\mathfrak{R}$ of $\mathcal{E}$ can be of much help to decide about the irreducibility of $(\mathcal{E},\rho)$. The subalgebras $C^i\mathfrak{R}$ are ideals of $\mathcal{E}$. Moreover, for $i\in\mathbf{N}^*$, they are nilpotent and contained in $\mathfrak{A}$. In fact, let $(\mathcal{E},\rho,\sigma)$ be an ω-sliced extension with fundamental $\mathfrak{S}_2$-module $A(\varphi,f_2)$. Any element of $\mathfrak{R}$ can be written uniquely as $\sigma(r)+a$, where $r\in\mathfrak{R}_2$, $a\in\mathfrak{A}$, and otherwise

$$[\sigma(r)+a,\sigma(r')+a']=\{[a,a']+\varphi(r)a'-\varphi(r')a+f_2(r,r')\}\in\mathfrak{A}\quad\text{for all } r,r'\in\mathfrak{R}_2, a,a'\in\mathfrak{A}.$$

Hence $C^i\mathfrak{R}\subseteq\mathfrak{A}$ for all $i\in\mathbf{N}^*$.

We have the following possibilities:

(1) $C^1\mathfrak{R}=\{0\}$, i.e., $(\mathcal{E},\rho)$ is inessential. If $(\mathcal{E},\rho)$ is irreducible, it must be truly trivial, i.e., $\mathfrak{A}=\{0\}$.

(2) $C^1\mathfrak{R}\neq\{0\}$, $C^1\mathfrak{R}\subset\mathfrak{A}$. We can consider the induced extension $(\mathcal{E}/C^1\mathfrak{R},\rho_q)$ of $\mathfrak{B}$ by $\mathfrak{A}/C^1\mathfrak{R}$, which is inessential since $\mathfrak{R}/C^1\mathfrak{R}$ is Abelian. By the irreducibility criterion (Theorem 7 of Ref. 11) $(\mathcal{E},\rho)$ is reducible.

The previous analysis gives that $C^1\mathfrak{R}=\mathfrak{A}$ is a necessary condition for the irreducibility of an extension of $\mathfrak{B}$ (with Abelian radical) by $\mathfrak{A}$.

Theorem 9: Let $(\mathcal{E},\rho)$ be an extension of $\mathfrak{B}$ by $\mathfrak{A}$ and let $\mathfrak{B}=\mathfrak{S}_2\dotplus\mathfrak{R}_2$ with Abelian $\mathfrak{R}_2$. Suppose that the radical $\mathfrak{R}$ of $\mathcal{E}$ is nilpotent. In order that $(\mathcal{E},\rho)$ be irreducible, it is necessary and sufficient that $C^1\mathfrak{R}=\mathfrak{A}$.

Proof: We must prove the sufficiency of the condition. If $\mathfrak{A}=\{0\}$, the statement is obviously true. Suppose $\mathfrak{A}\neq\{0\}$, $C^1\mathfrak{R}=\mathfrak{A}$, and let $(\mathcal{E},\rho,\sigma)$ be an ω-sliced extension with fundamental $\mathfrak{S}_2$-module $A(\varphi,f_2)$. By the structure theorem

$$R=\sigma_v(R_2)\oplus A,$$

since $C^1\mathfrak{R}=\mathfrak{A}$ implies the nilpotency of $\mathfrak{A}$. Therefore,

$$A=C^1R=[\sigma_v(R_2),\sigma_v(R_2)]+[\sigma_v(R_2),A]+C^1A=[\sigma_v(R_2),\sigma_v(R_2)]+C^2R,$$

where C^iR, C^1A are the vector spaces underlying $C^i\mathfrak{R}, C^1\mathfrak{A}$.

In general,

$$C^jR=(\mathrm{ad}\sigma_v(R_2))^j\sigma_v(R_2)+C^{j+1}R\quad\text{for all } j\in\mathbf{N},$$

and

$$A=\sum_{j=1}^{m}(\mathrm{ad}\sigma_v(R_2))^j\sigma_v(R_2)+C^{m+1}R\quad\text{for all } m\in\mathbf{N}^*,\tag{2.3}$$

where

$$(\mathrm{ad}\sigma_v(R_2))^j\sigma_v(R_2)=[\sigma_v(R_2),[\sigma_v(R_2),\cdots[\sigma_v(R_2),\sigma_v(R_2)]\cdots]]$$

(with j Lie products). $\mathfrak{R}$ being nilpotent, there exists an $n\in\mathbf{N}^*$ such that $C^{n+1}\mathfrak{R}=\{0\}$. It follows from (2.3) that

$$A_\omega=\sum_{i=0}^{n-1}\varphi^i(\mathfrak{R}_2)A(\varphi,f_2).$$

As this is true for any ω-sliced extension related to $(\mathcal{E},\rho)$, by Theorem 8 $(\mathcal{E},\rho)$ is irreducible. ■

Notice that if the radical $\mathfrak{R}_2$ of $\mathfrak{B}$ is an Abelian Lie algebra, the necessary conditions of Theorem 7 become sufficient since they imply $C^1\mathfrak{R}=\mathfrak{A}$ with $\mathfrak{R}$ nilpotent, i.e.,

Theorem 10: Let $(\mathcal{E},\rho,\sigma)$ be an ω-sliced extension of $\mathfrak{B}$ by $\mathfrak{A}$ with fundamental $\mathfrak{S}_2$-module $A(\varphi,f_2)$. Suppose further that the radical $\mathfrak{R}_2$ of $\mathfrak{B}$ is Abelian. Then $(\mathcal{E},\rho)$ belongs to $\mathfrak{E}_n(\mathfrak{B})$ $(n\in\mathbf{N}^*)$ if and only if

$$\varphi^n(\mathfrak{R}_2)A(\varphi,f_2)=\{0\},\qquad \varphi^j(\mathfrak{R}_2)A(\varphi,f_2)\neq\{0\}\quad\text{for all } j<n$$

and

$$A_\omega=\sum_{i=0}^{n-1}\varphi^i(\mathfrak{R}_2)A(\varphi,f_2).$$

This theorem is particularly significant for the case of the Lie algebras of the Poinçaré and Euclidean groups which have Abelian radical. The extensions of these Lie algebras will be considered in Paper II.

ACKNOWLEDGMENTS

This paper is based partly on a doctoral dissertation presented to the University of Louvain. The author expresses his gratitude to Professor D. Speiser for his guidance in this research and for the many stimulating conversations. He also has the pleasure to thank Professor F. Cerulus, Professor A. Janner, Professor G. Emch, and Dr. T. Janssen for enlightening discussions and critical remarks. The author acknowledges gratefully the Instituut voor Theoretische Fysika, Katholieke Universiteit Nijmegen, where the paper was completed, for the warm hospitality extended to him.

APPENDIX A

We consider the four-dimensional nilpotent Lie algebra $\mathfrak{A}$ generated by the basis $\{a_1,a_2,a_3,a_4\}$ with the Lie products

$$[a_1,a_2]=a_4,\quad [a_1,a_3]=[a_1,a_4]=[a_2,a_3]=[a_2,a_4]=[a_3,a_4]=0.$$

We see that $\mathfrak{C}(\mathfrak{A})$ is generated by a_3 and a_4. In order to compute $\mathrm{D}(\mathfrak{A})$, $\mathrm{I}(\mathfrak{A})$, and $\Delta(\mathfrak{A})$, we first determine $Z^1(\mathfrak{A},A_{\mathrm{ad}})$, $B^1(\mathfrak{A},A_{\mathrm{ad}})$, $H^1(\mathfrak{A},A_{\mathrm{ad}})$ and then the commutators of the obtained derivations. $\mathrm{D}(\mathfrak{A})$ is the ten-dimensional Lie algebra generated by the basis $\{D_i\}_{1\leq i\leq 10}$ defined in Table I. The Lie products are given in Table II.

D_9 and D_{10} generate the Abelian Lie algebra $\mathrm{I}(\mathfrak{A})$, $\{D_j\}_{4\leq j\leq 10}$ generates the seven-dimensional radical $\mathfrak{R}$ of $\mathrm{D}(\mathfrak{A})$, and $\{D_1,D_2,D_3\}$ a (simple) Levi subalgebra. Note that $D^1\mathfrak{R}$ is generated by $\{D_4,D_5,D_6,D_9,D_{10}\}$ and $D^2\mathfrak{R}=\mathrm{I}(\mathfrak{A})$, but $\mathfrak{R}$ is not nilpotent.

TABLE I. Action of the basis $\{D_i\}_{1\leqslant i\leqslant 10}$ of $D(\mathfrak{A})$ on the basis $\{a_j\}_{1\leqslant j\leqslant 4}$ of $\mathfrak{A}$

D_i \ a_j	a_1	a_2	a_3	a_4
D_1	$\frac{1}{2}a_2$	$\frac{1}{2}a_1$	0	0
D_2	$\frac{1}{2}a_2$	$-\frac{1}{2}a_1$	0	0
D_3	$\frac{1}{2}a_1$	$-\frac{1}{2}a_2$	0	0
D_4	a_3	0	0	0
D_5	0	a_3	0	0
D_6	0	0	a_4	0
D_7	0	0	a_3	0
D_8	$\frac{1}{2}a_1$	$\frac{1}{2}a_2$	0	a_4
D_9	0	a_4	0	0
D_{10}	$-a_4$	0	0	0

$\Delta(\mathfrak{A})$ is the eight-dimensional Lie algebra generated by a basis $\{\Delta_i\}_{1\leqslant i\leqslant 8}$ with the Lie products obtained from those of $D(\mathfrak{A})$ by changing D_k into Δ_k for $1 \leqslant k \leqslant 8$ and putting $D_9 = D_{10} = 0$.

Now we will study three particularly interesting examples of extensions of $\mathfrak{B}$ by $\mathfrak{A}$, where $\mathfrak{B}$ is the Abelian Lie algebra generated by b and b'. Notice that $\mathrm{ext}(\mathfrak{B}, \mathfrak{A}, \Phi) \neq \emptyset$ for all $\Phi \in \hom(\mathfrak{B}, \Delta(\mathfrak{A}))$, since $H^3(\mathfrak{B}, C(\mathfrak{A})_\psi) = \{0\}$. Moreover, any $(2; \mathfrak{B}, \mathfrak{A}, \Phi)$-pseudocochain is a $(2; \mathfrak{B}, \mathfrak{A}, \Phi)$-pseudococycle.

(1) Define $\Phi \in \hom(\mathfrak{B}, \Delta(\mathfrak{A}))$ by $\Phi(b) = \Delta_4$ and $\Phi(b') = \Delta_6$. Then any prerepresentation φ lifted over Φ is given by

$$\varphi(b) = D_4 + \alpha D_9 + \beta D_{10}$$

and

$$\varphi(b') = D_6 + \alpha' D_9 + \beta' D_{10} \quad (\alpha, \alpha', \beta, \beta' \in \mathbf{F}).$$

But

$$[\varphi(b), \varphi(b')] = D_{10},$$

i.e., we have no $\varphi \in \hom(\mathfrak{B}, D(\mathfrak{A}))$ lifted over Φ, and all extensions of $\mathfrak{B}$ by $\mathfrak{A}$ with character Φ are essential. φ is a proper prerepresentation associated with a bilinear alternating map f_2 given by

$$f_2(b, b') = a_2 + \gamma a_3 + \delta a_4 \quad (\gamma, \delta \in \mathbf{F}),$$

and $(\varphi, f_2) \in \mathfrak{Z}^2_\Phi(\mathfrak{B}, \mathfrak{A})$. According to the discussion in Sec. 1B, a complete set of $(2; \mathfrak{B}, \mathfrak{A}, \Phi)$-pseudococycles can be constructed if we compute a complete set of $(2; \mathfrak{B}, C(\mathfrak{A})_\psi)$-cocycles. The central character determined by Φ satisfies

$$\Psi(b) = 0, \quad \Psi(b')a_3 = a_4, \quad \Psi(b')a_4 = 0,$$

and

$$(\delta_1 f_1)(b, b') = \epsilon a_4 \quad (\epsilon \in \mathbf{F})$$

gives an arbitrary element of $B^2(\mathfrak{B}, C(\mathfrak{A})_\psi)$. Therefore,

$$\mathfrak{H}^2_\Phi(\mathfrak{B}, \mathfrak{A}) = \left\{ \underline{(\varphi, f_2)} \;\middle|\; \begin{array}{l} \varphi(b) = D_4,\ \varphi(b') = D_6, \\ f_2(b, b') = a_2 + \gamma a_3\ (\gamma \in \mathbf{F}) \end{array} \right\},$$

i.e., we have a set equipollent to $\mathbf{F}$ of equivalence classes of $(2; \mathfrak{B}, \mathfrak{A}, \Phi)$-pseudococycles. The corresponding equivalence classes of extensions are constructed by means of the bijection (1.3).

(2) Choose $\Phi = 0$. Then any prerepresentation φ lifted over Φ is given by $\varphi(b) = \alpha D_9 + \beta D_{10}$ and $\varphi(b') = \alpha' D_9 + \beta' D_{10}$ $(\alpha, \alpha', \beta, \beta' \in \mathbf{F})$, i.e., φ is a representation. If $(\varphi, f_2) \in \mathfrak{Z}^2_0(\mathfrak{B}, \mathfrak{A})$, then $\mathrm{Im} f_2 \subseteq \mathfrak{C}(\mathfrak{A})$, and therefore

$$f_2(b, b') = \gamma a_3 + \delta a_4 \quad (\gamma, \delta \in \mathbf{F}).$$

Consider a particular φ given by $\varphi(b) = D_9$ and $\varphi(b') = D_{10}$. Any $f_1 \in L(\mathfrak{B}, \mathfrak{A})$ such that $\varphi = \mathrm{ad} \circ f_1$ satisfies

$$f_1(b) = a_1 + \gamma'' a_3 + \delta'' a_4, \quad f_1(b') = a_2 + \gamma' a_3 + \delta' a_4 \quad (\gamma'', \gamma', \delta'', \delta' \in \mathbf{F}),$$

and thus

$$(\delta(0) f_1)(b, b') + [f_1(b), f_1(b')] = a_4.$$

Hence $\underline{(0, 0)} \neq \underline{(\varphi, 0)}$, and we have two different equivalence classes of inessential extensions. This is in agreement with the Corollary to Proposition 2, since $(\mathfrak{A}, \mathrm{ad})$ is an essential extension of $1(\mathfrak{A})$ by $\mathfrak{C}(\mathfrak{A})$. Actually, if we proceed as in (1), we obtain a set equipollent to $\mathbf{F}$ of equivalence classes of inessential extensions which is a proper subset of $\mathrm{ext}(\mathfrak{B}, \mathfrak{A}, 0)$.

(3) Let Φ be given by $\Phi(b) = \Delta_4$ and $\Phi(b') = \Delta_5$, and proceed as in (1). Any φ lifted over Φ is a representation of $\mathfrak{B}$ into $D(\mathfrak{A})$. Consider $\varphi, \varphi' \in \hom(\mathfrak{B}, D(\mathfrak{A}))$ given by

TABLE II. Lie products of the basis elements D_i $(1 \leqslant i \leqslant 10)$ defined in Table I.

D_i \ D_j	D_1	D_2	D_3	D_4	D_5	D_6	D_7	D_8	D_9	D_{10}
D_1	0	D_3	D_2	$-\frac{1}{2}D_5$	$-\frac{1}{2}D_4$	0	0	0	$\frac{1}{2}D_{10}$	$\frac{1}{2}D_9$
D_2	$-D_3$	0	D_1	$\frac{1}{2}D_5$	$-\frac{1}{2}D_4$	0	0	0	$\frac{1}{2}D_{10}$	$-\frac{1}{2}D_9$
D_3	$-D_2$	$-D_1$	0	$-\frac{1}{2}D_4$	$\frac{1}{2}D_5$	0	0	0	$\frac{1}{2}D_9$	0
D_4	$\frac{1}{2}D_5$	$-\frac{1}{2}D_5$	$\frac{1}{2}D_4$	0	0	D_{10}	$-D_4$	$\frac{1}{2}D_4$	0	0
D_5	$\frac{1}{2}D_4$	$\frac{1}{2}D_4$	$-\frac{1}{2}D_5$	0	0	$-D_9$	$-D_5$	$\frac{1}{2}D_5$	0	0
D_6	0	0	0	$-D_{10}$	D_9	0	D_6	$-D_6$	0	0
D_7	0	0	0	D_4	D_5	$-D_6$	0	0	0	0
D_8	0	0	0	$-\frac{1}{2}D_4$	$-\frac{1}{2}D_5$	D_6	0	0	$\frac{1}{2}D_9$	D_{10}
D_9	$-\frac{1}{2}D_{10}$	$-\frac{1}{2}D_{10}$	$-\frac{1}{2}D_9$	0	0	0	0	$-\frac{1}{2}D_9$	0	0
D_{10}	$-\frac{1}{2}D_9$	$\frac{1}{2}D_9$	0	0	0	0	0	$-D_{10}$	0	0

$$\varphi(b) = D_4, \quad \varphi(b') = D_5 \quad \text{and}$$
$$\varphi'(b) = D_4, \quad \varphi'(b') = D_5 + D_9.$$

$(\varphi, 0), (\varphi', 0) \in \mathfrak{Z}^2_\Phi(\mathfrak{B}, \mathfrak{A})$ and $(\varphi, 0) \neq (\varphi', 0)$; hence we have two different equivalence classes of inessential extensions. All elements of $\mathrm{ext}(\mathfrak{B}, \mathfrak{A}, \Phi)$ are actually equivalence classes of inessential extensions.

APPENDIX B

Let $\Pi_\mathfrak{A}$ be the canonical epimorphism $D(\mathfrak{A}) \twoheadrightarrow \Delta(\mathfrak{A})$ and let φ be a given prerepresentation of $\mathfrak{B}$ into $D(\mathfrak{A})$. Consider the following sets

$$\mathfrak{C}^2_\varphi(\mathfrak{B}, \mathfrak{A}) = \left\{(\varphi, f_2) \,\middle|\, \begin{array}{l} f_2 \in A_2(\mathfrak{B}, \mathfrak{A}); [\varphi(b), \varphi(b')] = \varphi([b, b']) \\ + \mathrm{ad} f_2(b, b') \text{ for all } b, b' \in \mathfrak{B} \end{array}\right\},$$

$$\mathfrak{Z}^2_\varphi(\mathfrak{B}, \mathfrak{A}) = \mathfrak{C}^2_\varphi(\mathfrak{B}, \mathfrak{A}) \cap \mathfrak{Z}^2_{\Pi_\mathfrak{A} \circ \varphi}(\mathfrak{B}, \mathfrak{A}),$$

where $A_2(\mathfrak{B}, \mathfrak{A})$ is the vector space of the bilinear alternating maps of $\mathfrak{B} \times \mathfrak{B}$ into $\mathfrak{A}$.

If, besides, $\varphi | \mathfrak{S} = \omega$ is a representation of $\mathfrak{S} \subseteq \mathfrak{B}$ into $D(\mathfrak{A})$, we can also consider

$$\mathfrak{C}^2_\varphi(\mathfrak{B}, \mathfrak{S}, \mathfrak{A}) = \mathfrak{C}^2_\varphi(\mathfrak{B}, \mathfrak{A}) \cap \mathfrak{C}^2_{\Pi_\mathfrak{A} \circ \varphi, \omega}(\mathfrak{B}, \mathfrak{S}, \mathfrak{A}),$$
$$\mathfrak{Z}^2_\varphi(\mathfrak{B}, \mathfrak{S}, \mathfrak{A}) = \mathfrak{Z}^2_\varphi(\mathfrak{B}, \mathfrak{A}) \cap \mathfrak{C}^2_{\Pi_\mathfrak{A} \circ \varphi, \omega}(\mathfrak{B}, \mathfrak{S}, \mathfrak{A}).$$

The equivalence relations R in $\mathfrak{Z}^2_\Phi(\mathfrak{B}, \mathfrak{A})$ and $R(\omega)$ in $\mathfrak{Z}^2_{\Phi,\omega}(\mathfrak{B}, \mathfrak{S}, \mathfrak{A})$ give rise,[11] respectively, to the equivalence relations R′ in $\mathfrak{Z}^2_\varphi(\mathfrak{B}, \mathfrak{A})$ and $R'(\varphi | \mathfrak{S})$ in $\mathfrak{Z}^2_\varphi(\mathfrak{B}, \mathfrak{S}, \mathfrak{A})$ as follows:

R′: We say that $(\varphi, f_2), (\varphi, f_2') \in \mathfrak{Z}^2_\varphi(\mathfrak{B}, \mathfrak{A})$ are equivalent if there exists $f_1 \in L(\mathfrak{B}, \mathfrak{A})$ such that $\mathrm{Im} f_1 \subseteq \mathfrak{C}(\mathfrak{A})$ and

$$f_2'(b, b') = f_2(b, b') + (\delta_1(\varphi) f_1)(b, b') \quad \text{for all } b, b' \in \mathfrak{B}. \tag{B1}$$

$R'(\varphi | \mathfrak{S})$: We say that $(\varphi, f_2), (\varphi, f_2') \in \mathfrak{Z}^2_\varphi(\mathfrak{B}, \mathfrak{S}, \mathfrak{A})$ are equivalent if there exists $f_1 \in L(\mathfrak{B}, \mathfrak{A})(\varphi | \mathfrak{S})$-orthogonal to $\mathfrak{S}$ such that $\mathrm{Im} f_1 \subseteq \mathfrak{C}(\mathfrak{A})$ and (B1) is satisfied.

Now it is possible to define

$$\mathfrak{H}^2_\varphi(\mathfrak{B}, \mathfrak{A}) = \mathfrak{Z}^2_\varphi(\mathfrak{B}, \mathfrak{A})/R'$$

and

$$\mathfrak{H}^2_\varphi(\mathfrak{B}, \mathfrak{S}, \mathfrak{A}) = \mathfrak{Z}^2_\varphi(\mathfrak{B}, \mathfrak{S}, \mathfrak{A})/R'(\varphi | \mathfrak{S}),$$

and to state the following.

Theorem: Let $\mathfrak{A}$ and $\mathfrak{B}$ be Lie algebras and let $\Phi \in \hom(\mathfrak{B}, \Delta(\mathfrak{A}))$. Suppose that $\mathfrak{D}$ is an ideal of $\mathfrak{B}$ such that $\mathfrak{B}/\mathfrak{D}$ is semisimple, and let $\mathfrak{S}$ be a subalgebra of $\mathfrak{B}$ isomorphic to $\mathfrak{B}/\mathfrak{D}$ by the canonical epimorphism $\mathfrak{B} \twoheadrightarrow \mathfrak{B}/\mathfrak{D}$. Then there are a prerepresentation φ of $\mathfrak{B}$ into $D(\mathfrak{A})$ lifted over Φ, such that $\varphi | \mathfrak{S}$ is a representation isomorphically lifted over $\Phi | \mathfrak{S}$, and a bijection

$$\mathfrak{H}^2_\Phi(\mathfrak{B}, \mathfrak{A}) \approx \mathfrak{H}^2_\varphi(\mathfrak{B}, \mathfrak{S}, \mathfrak{A}).$$

Proof: Let $(\varphi, f_2) \in \mathfrak{Z}^2_{\Phi,\omega}(\mathfrak{B}, \mathfrak{S}, \mathfrak{A})$. If for any other element (φ', f_2') of $\mathfrak{Z}^2_{\Phi,\omega}(\mathfrak{B}, \mathfrak{S}, \mathfrak{A})$ such that $(\varphi, f_2) \neq (\varphi', f_2') \pmod{R(\omega)}$ there exists $(\varphi, f_2'') \in (\varphi', f_2') \pmod{R(\omega)}$ we have a bijection

$$\mathfrak{H}^2_{\Phi,\omega}(\mathfrak{B}, \mathfrak{S}, \mathfrak{A}) \rightarrowtail \mathfrak{H}^2_\varphi(\mathfrak{B}, \mathfrak{S}, \mathfrak{A}) \tag{B2}$$

given by $(\varphi, f_2) \pmod{R(\omega)} \mapsto (\varphi, f_2) \pmod{R'(\omega)}$. This requires the existence of $f_1 \in L(\mathfrak{B}, \mathfrak{A})$ ω-orthogonal to $\mathfrak{S}$ such that

$$\varphi'(b) = \varphi(b) + \mathrm{ad} f_1(b), \quad f_2'(o, b') = f_2''(b, b') + (\delta_1(\varphi) f_1)(b, b') + [f_1(b), f_1(b')] \text{ for all } b, b' \in \mathfrak{B}. \tag{B3}$$

By Lemma 1 and the techniques of the proof of Theorem 2 of Ref. 11, there exists $f_1' \in L(\mathfrak{B}, \mathfrak{A})$ which satisfies (B3), and this implies $f_1'(s) \in \mathfrak{C}(\mathfrak{A})$ and $(\delta_1(\varphi) f_1')_s = 0$ for all $s \in \mathfrak{S}$.

We need the following result.

Lemma: Let $\mathfrak{B}, \mathfrak{A}, \mathfrak{D}, \mathfrak{S}$, and Φ be as in the theorem. Suppose, moreover, that φ is a prerepresentation of $\mathfrak{B}$ into $D(\mathfrak{A})$ lifted over Φ which satisfies

$$[\varphi(s), \varphi(b)] = \varphi([s, b]) \quad \text{for all } s \in \mathfrak{S}, \ b \in \mathfrak{B},$$

and let $f_1' \in L(\mathfrak{B}, \mathfrak{A})$ be such that $\mathrm{Im}(f_1' | \mathfrak{S}) \subseteq \mathfrak{C}(\mathfrak{A})$, $(\delta_1(\varphi) f_1')_s = 0$ for all $s \in \mathfrak{S}$.

There exists $g_1 \in Z^1(\mathfrak{B}, C(\mathfrak{A})_\Psi)$, Ψ being the central character determined by Φ, such that $(f_1')_s = (g_1)_s$ and $s(\varphi) \cdot f_1' = s(\varphi) \cdot g_1$ for all $s \in \mathfrak{S}$.

Proof: The proof is analogous to that of Lemma 3 of Ref. 11. We equip $L(\mathfrak{B}, \mathfrak{A})$ with the $\mathfrak{S}$-module structure associated with the representation $\Sigma: \mathfrak{S} \to \mathfrak{GL}(L(\mathfrak{B}, \mathfrak{A}))$ given by $\Sigma(s) h_1 = s(\varphi) \cdot h_1$ for all $h_1 \in L(\mathfrak{B}, \mathfrak{A})$. Then, $\delta_0(\varphi) C(\mathfrak{A})$ becomes an $\mathfrak{S}$-submodule of $(L(\mathfrak{B}, \mathfrak{A}))_\Sigma$. Proceeding as in the proof of Lemma 3 of Ref. 11, we find $g_1' \in Z^1(\mathfrak{B}, C(\mathfrak{A})_\Psi)$ such that $g_1 = g_1' + \delta_0(\varphi) a, a \in C(\mathfrak{A})$, and $(f_1')_s = (g_1)_s$ for all $s \in \mathfrak{S}$. But $(\delta_1(\varphi) f_1')_s = (\delta_1(\varphi) g_1)_s = 0$, and hence $s(\varphi) \cdot f_1' = s(\varphi) \cdot g_1$ by (III. 4) of Ref. 11. ■

We have an $f_1 \in L(\mathfrak{B}, \mathfrak{A})$, ω-orthogonal to $\mathfrak{S}$ and satisfying (B3), if we put $f_1 = f_1' - g_1$. This proves the existence of the bijection (B2), which is obviously true also if $\mathfrak{H}^2_{\Phi,\omega}(\mathfrak{B}, \mathfrak{S}, \mathfrak{A}) = \emptyset$. Besides, the reduction theorem supplies us with a representation ω of $\mathfrak{S}$ into $D(\mathfrak{A})$ lifted over $\Phi | \mathfrak{S}$ and with a bijection

$$\mathfrak{H}^2_\Phi(\mathfrak{B}, \mathfrak{A}) \rightarrowtail \mathfrak{H}^2_{\Phi,\omega}(\mathfrak{B}, \mathfrak{S}, \mathfrak{A}).$$

ω can be chosen isomorphically lifted over $\Phi | \mathfrak{S}$. This follows from the proof of the reduction theorem.[11] Combining these results we obtain the statement. ■

The meaning of the theorem is the following: A complete set of $(2; \mathfrak{B}, \mathfrak{A}, \Phi)$-pseudococycles is given by a complete set of elements of $\mathfrak{Z}^2_\varphi(\mathfrak{B}, \mathfrak{S}, \mathfrak{A})$ [for some prerepresentation φ of $\mathfrak{B}$ into $D(\mathfrak{A})$]. Notice that if we pick out one element of this latter set, say (φ, g_2), and we consider any other element of it, say (φ, f_2), then $h_2 = f_2 - g_2$ satisfies $\delta_2(\varphi) h_2 = 0$ and $\mathrm{Im} h_2 \subseteq \mathfrak{C}(\mathfrak{A})$. This gives the connection with the analysis of Sec. 1B.

---

* Postal address: Instituut voor Theoretische Fysika, Katholieke Universiteit, Driehuizerweg 200, Nijmegen (Netherlands).

1 L. Michel, "Invariance in Quantum Mechanics and Group Extension," in *Group-Theoretical Concepts and Methods in Elementary Particle Physics*, edited by F. Gürsey (Gordon and Breach, New York, 1965).

[2] L. Michel, Nucl. Phys. **57**, 356 (1964).
[3] L. Michel, Phys. Rev. **137**, B405 (1965).
[4] L. Michel, "Relativistic Invariance and Internal Symmetries" in *Axiomatic Field Theory*, edited by M. Chretien and S. Deser (Gordon and Breach, New York, 1966).
[5] L. Michel, "Relations entre symétries internes et invariance relativiste," in *Application of Mathematics to Problems in Theoretical Physics*, edited by F. Lurçat (Gordon and Breach, New York, 1967).
[6] E. Wigner, Ann. Math. **40**, 149 (1939).
[7] V. Bargmann, Ann. Math. **59**, 1 (1954).
[8] A. Galindo, J. Math. Phys. **8**, 768 (1967).
[9] R. A. Macaulay, Trans. Amer. Math. Soc. **95**, 530 (1960).
[10] U. Cattaneo, Commun. Math. Phys. **13**, 226 (1969).
[11] U. Cattaneo, Commun. Math. Phys. **20**, 220 (1971).
[12] L. Calabi, Ann. Mat. Pura Appl. **32**, 295 (1951).
[13] "Sliced extension" is used as translation of Calabi's "extension tranchée". It is not to be confused with the term "slice" used in the literature on actions of compact Lie groups on manifolds [cf., e.g., (1) R. S. Palais, "Slices and Equivariant Imbeddings," in *Seminar on Transformation Groups*, Annals of Mathematical Studies Vol. 46, edited by A. Borel (Princeton U.P., Princeton, N.J., 1960); (2) L. Michel, "Applications of Group Theory to Quantum Physics. Algebraic Aspects," in *Group Representations in Mathematics and Physics*, Lecture Notes in Physics, Vol. 6, edited by V. Bargmann (Springer-Verlag, Heidelberg, 1970)].
[14] N. Jacobson, *Lie Algebras* (Interscience, New York, 1962).
[15] D. J. Simms, *Lie Groups and Quantum Mechanics*, Lecture Notes in Mathematics, Vol. 52, (Springer-Verlag, Heidelberg, 1968).
[16] F. Lurçat and L. Michel, Nuovo Cimento **21**, 574 (1961).
[17] E. Ascher and A. Janner, Helv. Phys. Acta **38**, 551 (1965).
[18] U. Cattaneo, Nuovo Cimento **7A**, 839 (1972).
[19] O. Schreier, Monatsh. Math. Phys. **34**, 164 (1926); Abhandl. Math. Sem. Hamburg **4**, 321 (1926).
[20] M. Mori, J. Math. Soc. Japan **5**, 171 (1953).
[21] G. Hochschild, Amer. J. Math. **76**, 698, 763 (1954).
[22] C. Chevalley and S. Eilenberg, Trans. Amer. Math. Soc. **63**, 85 (1948).
[23] C. George and M. Lévy-Nahas, J. Math. Phys. **7**, 980 (1966).
[24] G. F. Leger, Jr., Proc. Amer. Math. Soc. **4**, 511 (1953); **10**, 10 (1959); Duke Math. J. **30**, 637 (1963).
[25] S. Tôgô, Can. J. Math. **13**, 201 (1961).
[26] U. Cattaneo, thesis (Louvain, 1970).
[27] G. Hochschild and J. P. Serre, Ann. Math. **57**, 591 (1953).
[28] J. P. Serre, *Lie Algebras and Lie Groups* (Benjamin, New York, 1965).
[29] N. Bourbaki, *Groupes et algèbres de Lie* (Hermann, Paris, 1960), Chap. 1, ASI 1285.
[30] N. Bourbaki, *Algébre* (Hermann, Paris, 1958), Chap. 8, ASI 1261.
[31] C. Berge, *Théorie des graphes et ses applications* (Dunod, Paris, 1967).
[32] F. Harary, *Graph Theory* (Addison-Wesley, Reading, Mass., 1969).

Sliced Extensions, Irreducible Extensions, and Associated Graphs: An Analysis of Lie Algebra Extensions. II. Application to Euclidean, Poincaré, and Galilean Algebras

U. Cattaneo*
Instituut voor Theoretische Fysika, Katholieke Universiteit, Nijmegen, The Netherlands
(Received 6 August 1971; Revised Manuscript Received 29 November 1971)

The results of a preceding paper on Lie algebra extensions and sliced extensions are applied to the Lie algebras $\mathcal{E}(3)$, $\mathcal{P}$, and $\mathcal{G}$ of the Euclidean, resp. Poincaré and Galilean groups. The primitive extensions are analyzed in detail. A procedure for the construction of irreducible extensions is illustrated by some examples, using diagrams which picture the graphs of the extensions. It is proved that all extensions by $\mathcal{E}(3)$, $\mathcal{P}$, and $\mathcal{G}$ are inessential.

INTRODUCTION

In the previous paper,[1] hereafter referred to as I, we have explored the fundamental problems of Lie algebra extension theory. Our principal aim was to get some hints for the construction of extensions. The results so obtained are especially useful for extensions of Lie algebras with nontrivial Levi subalgebras, a property typical of many Lie algebras used in physics. This is the case, for example, of the Euclidean, Poincaré, and Galilean algebras, i.e., of the (real) Lie algebras $\mathcal{E}(3)$, $\mathcal{P}$, and $\mathcal{G}$ of the Euclidean group of a three-dimensional Euclidean space E_3, resp. of the (ten-dimensional) Poincaré and Galilean groups. In this paper we apply the results of I to $\mathcal{E}(3)$, $\mathcal{P}$, and $\mathcal{G}$.

The paper is organized as follows. Section 1 is devoted to the extensions of $\mathcal{E}(3)$ with special emphasis on the irreducible ones. We study the primitive extensions and prove that the graphs of irreducible extensions are all finite and without circuits.

In Sec. 2 we consider extensions of $\mathcal{P}$. A close analogy with the results of Sec. 1 appears in the most cases. One important discrepancy is pointed out.

Section 3 concerns extensions of $\mathcal{G}$. Here, the problem is more involved. The primitive extensions are analyzed in detail. In particular, we show explicitly how the essential primitive extensions of $\mathcal{G}$ by a one-dimensional Lie algebra arise from trivial extensions of $\mathcal{P}$ when the Inönü–Wigner contraction of $\mathcal{P}$ in $\mathcal{G}$ is performed. The graphs of irreducible extensions of $\mathcal{G}$ may allow circuits. We give the example of one finite graph associated with an ω-sliced extension of $\mathcal{G}$ and having a loop.

In Sec. 4 we prove that all extensions by $\mathcal{E}(3)$, $\mathcal{P}$, and $\mathcal{G}$ are inessential. The proof requires some results, collected in the Appendix, about the Lie algebras $D(\mathcal{E}(3))$, $D(\mathcal{P})$, and $D(\mathcal{G})$.

The conventions and the notation of I are used throughout the paper. However, if the contrary is not explicitly stated, the base field is always **R**, the field of real numbers. The metric and the totally antisymmetric tensors are denoted by the usual symbols g, resp. ϵ.

1. EXTENSIONS OF $\mathcal{E}(3)$

The Lie algebra $\mathcal{E}(3)$ has a well-known Levi decomposition

$$\mathcal{E}(3) = \mathcal{SO}(3) \dot{+} \mathcal{T}(3),$$

where $\mathcal{SO}(3)$ and $\mathcal{T}(3)$ are, respectively, the Lie algebras of the (proper) rotation and translation groups of E_3. We choose a basis $\{r_i\}_{1\le i<3}$ of $\mathcal{SO}(3)$ and a basis $\{t_j\}_{1<j<3}$ of $\mathcal{T}(3)$, in such a way that $\{r_i, t_j\}_{1\le i,j<3}$ is a basis of $\mathcal{E}(3)$ with the Lie products

$$[r_i, r_j] = \sum_k \epsilon_{ijk} r_k, \qquad [r_i, t_j] = \sum_k \epsilon_{ijk} t_k, \qquad [t_i, t_j] = 0.$$

In order to apply the results of I, we consider standard representatives of the isomorphism classes of simple $\mathcal{SO}(3)$-modules constructed as follows. Let $\mathfrak{D}^{(j)}$ ($2j \in \mathbf{N}$) be an irreducible continuous complex representation of $SU(2)$ on a (complex) vector space $D^{(j)}$ of dimension $(2j + 1)$. For example, we can take

[2] L. Michel, Nucl. Phys. **57**, 356 (1964).
[3] L. Michel, Phys. Rev. **137**, B405 (1965).
[4] L. Michel, "Relativistic Invariance and Internal Symmetries" in *Axiomatic Field Theory*, edited by M. Chretien and S. Deser (Gordon and Breach, New York, 1966).
[5] L. Michel, "Relations entre symétries internes et invariance relativiste," in *Application of Mathematics to Problems in Theoretical Physics*, edited by F. Lurçat (Gordon and Breach, New York, 1967).
[6] E. Wigner, Ann. Math. **40**, 149 (1939).
[7] V. Bargmann, Ann. Math. **59**, 1 (1954).
[8] A. Galindo, J. Math. Phys. **8**, 768 (1967).
[9] R. A. Macaulay, Trans. Amer. Math. Soc. **95**, 530 (1960).
[10] U. Cattaneo, Commun. Math. Phys. **13**, 226 (1969).
[11] U. Cattaneo, Commun. Math. Phys. **20**, 220 (1971).
[12] L. Calabi, Ann. Mat. Pura Appl. **32**, 295 (1951).
[13] "Sliced extension" is used as translation of Calabi's "extension tranchée". It is not to be confused with the term "slice" used in the literature on actions of compact Lie groups on manifolds [cf., e.g., (1) R. S. Palais, "Slices and Equivariant Imbeddings," in *Seminar on Transformation Groups*, Annals of Mathematical Studies Vol. 46, edited by A. Borel (Princeton U.P., Princeton, N.J., 1960); (2) L. Michel, "Applications of Group Theory to Quantum Physics. Algebraic Aspects," in *Group Representations in Mathematics and Physics*, Lecture Notes in Physics, Vol. 6, edited by V. Bargmann (Springer-Verlag, Heidelberg, 1970)].
[14] N. Jacobson, *Lie Algebras* (Interscience, New York, 1962).
[15] D. J. Simms, *Lie Groups and Quantum Mechanics*, Lecture Notes in Mathematics, Vol. 52, (Springer-Verlag, Heidelberg, 1968).
[16] F. Lurçat and L. Michel, Nuovo Cimento **21**, 574 (1961).
[17] E. Ascher and A. Janner, Helv. Phys. Acta **38**, 551 (1965).
[18] U. Cattaneo, Nuovo Cimento **7A**, 839 (1972).
[19] O. Schreier, Monatsh. Math. Phys. **34**, 164 (1926); Abhandl. Math. Sem. Hamburg **4**, 321 (1926).
[20] M. Mori, J. Math. Soc. Japan **5**, 171 (1953).
[21] G. Hochschild, Amer. J. Math. **76**, 698, 763 (1954).
[22] C. Chevalley and S. Eilenberg, Trans. Amer. Math. Soc. **63**, 85 (1948).
[23] C. George and M. Lévy-Nahas, J. Math. Phys. **7**, 980 (1966).
[24] G. F. Leger, Jr., Proc. Amer. Math. Soc. **4**, 511 (1953); **10**, 10 (1959); Duke Math. J. **30**, 637 (1963).
[25] S. Tôgô, Can. J. Math. **13**, 201 (1961).
[26] U. Cattaneo, thesis (Louvain, 1970).
[27] G. Hochschild and J. P. Serre, Ann. Math. **57**, 591 (1953).
[28] J. P. Serre, *Lie Algebras and Lie Groups* (Benjamin, New York, 1965).
[29] N. Bourbaki, *Groupes et algèbres de Lie* (Hermann, Paris, 1960), Chap. 1, ASI 1285.
[30] N. Bourbaki, *Algébre* (Hermann, Paris, 1958), Chap. 8, ASI 1261.
[31] C. Berge, *Théorie des graphes et ses applications* (Dunod, Paris, 1967).
[32] F. Harary, *Graph Theory* (Addison-Wesley, Reading, Mass., 1969).

Sliced Extensions, Irreducible Extensions, and Associated Graphs: An Analysis of Lie Algebra Extensions. II. Application to Euclidean, Poincaré, and Galilean Algebras

U. Cattaneo*
Instituut voor Theoretische Fysika, Katholieke Universiteit, Nijmegen, The Netherlands
(Received 6 August 1971; Revised Manuscript Received 29 November 1971)

The results of a preceding paper on Lie algebra extensions and sliced extensions are applied to the Lie algebras $\mathcal{E}(3)$, $\mathcal{P}$, and $\mathcal{G}$ of the Euclidean, resp. Poincaré and Galilean groups. The primitive extensions are analyzed in detail. A procedure for the construction of irreducible extensions is illustrated by some examples, using diagrams which picture the graphs of the extensions. It is proved that all extensions by $\mathcal{E}(3)$, $\mathcal{P}$, and $\mathcal{G}$ are inessential.

INTRODUCTION

In the previous paper,[1] hereafter referred to as I, we have explored the fundamental problems of Lie algebra extension theory. Our principal aim was to get some hints for the construction of extensions. The results so obtained are especially useful for extensions of Lie algebras with nontrivial Levi subalgebras, a property typical of many Lie algebras used in physics. This is the case, for example, of the Euclidean, Poincaré, and Galilean algebras, i.e., of the (real) Lie algebras $\mathcal{E}(3)$, $\mathcal{P}$, and $\mathcal{G}$ of the Euclidean group of a three-dimensional Euclidean space E_3, resp. of the (ten-dimensional) Poincaré and Galilean groups. In this paper we apply the results of I to $\mathcal{E}(3)$, $\mathcal{P}$, and $\mathcal{G}$.

The paper is organized as follows. Section 1 is devoted to the extensions of $\mathcal{E}(3)$ with special emphasis on the irreducible ones. We study the primitive extensions and prove that the graphs of irreducible extensions are all finite and without circuits.

In Sec. 2 we consider extensions of $\mathcal{P}$. A close analogy with the results of Sec. 1 appears in the most cases. One important discrepancy is pointed out.

Section 3 concerns extensions of $\mathcal{G}$. Here, the problem is more involved. The primitive extensions are analyzed in detail. In particular, we show explicitly how the essential primitive extensions of $\mathcal{G}$ by a one-dimensional Lie algebra arise from trivial extensions of $\mathcal{P}$ when the Inönü–Wigner contraction of $\mathcal{P}$ in $\mathcal{G}$ is performed. The graphs of irreducible extensions of $\mathcal{G}$ may allow circuits. We give the example of one finite graph associated with an ω-sliced extension of $\mathcal{G}$ and having a loop.

In Sec. 4 we prove that all extensions by $\mathcal{E}(3)$, $\mathcal{P}$, and $\mathcal{G}$ are inessential. The proof requires some results, collected in the Appendix, about the Lie algebras $D(\mathcal{E}(3))$, $D(\mathcal{P})$, and $D(\mathcal{G})$.

The conventions and the notation of I are used throughout the paper. However, if the contrary is not explicitly stated, the base field is always **R**, the field of real numbers. The metric and the totally antisymmetric tensors are denoted by the usual symbols g, resp. ϵ.

1. EXTENSIONS OF $\mathcal{E}(3)$

The Lie algebra $\mathcal{E}(3)$ has a well-known Levi decomposition

$$\mathcal{E}(3) = \mathcal{SO}(3) \dot{+} \mathcal{T}(3),$$

where $\mathcal{SO}(3)$ and $\mathcal{T}(3)$ are, respectively, the Lie algebras of the (proper) rotation and translation groups of E_3. We choose a basis $\{r_i\}_{1\le i<3}$ of $\mathcal{SO}(3)$ and a basis $\{t_j\}_{1<j<3}$ of $\mathcal{T}(3)$, in such a way that $\{r_i, t_j\}_{1\le i,j<3}$ is a basis of $\mathcal{E}(3)$ with the Lie products

$$[r_i, r_j] = \sum_k \epsilon_{ijk} r_k, \qquad [r_i, t_j] = \sum_k \epsilon_{ijk} t_k, \qquad [t_i, t_j] = 0.$$

In order to apply the results of I, we consider standard representatives of the isomorphism classes of simple $\mathcal{SO}(3)$-modules constructed as follows. Let $\mathfrak{D}^{(j)}$ ($2j \in \mathbf{N}$) be an irreducible continuous complex representation of $SU(2)$ on a (complex) vector space $D^{(j)}$ of dimension $(2j + 1)$. For example, we can take

as $\mathfrak{D}^{(j)}$ the spinor representation $\mathfrak{S}_{2j}$ (on S_{2j}) defined in Ref. 2. As is well known, $\mathfrak{so}(3)$ is isomorphic to the Lie algebra $\mathfrak{su}(2)$ of $SU(2)$. Let ι be an isomorphism of $\mathfrak{so}(3)$ onto $\mathfrak{su}(2)$, and let $d\mathfrak{D}^{(j)}$ be the differential representation of $\mathfrak{su}(2)$ associated with $\mathfrak{D}^{(j)}$, canonically obtained from the differential of $\mathfrak{D}^{(j)}$.[3] Then $d\mathfrak{D}^{(j)}\circ\iota$ is an irreducible complex representation of $\mathfrak{so}(3)$ on $D^{(j)}$. For the sake of simplicity, we still denote this representation by $\mathfrak{D}^{(j)}$. If $j \in \mathbf{N}$, we get a Lie algebra representation of real type, and, if $j \in \{\frac{1}{2}, \frac{3}{2}, \frac{5}{2}, \dots\}$, one of quaternionic type[4] (potentially real, resp. pseudoreal in the group terminology of Wigner[5]). In the former case, there are isomorphic real forms of $\mathfrak{D}^{(j)}$ on the $(2j+1)$-dimensional real vector spaces of the fixed points of invariant anti-involutions of the first kind[4] (cf., however, Ref. 6), and we can choose one such real form $\mathfrak{D}^{\{j\}}$ (on a vector space $D^{\{j\}}$) for each $j \in \mathbf{N}$. In the second case, $\mathfrak{D}^{(j)}$ induces canonically one irreducible real representation $\mathfrak{D}^{\{j\}}$ of $\mathfrak{so}(3)$ on the $2(2j+1)$-dimensional vector space $D^{\{j\}}$ obtained from $D^{(j)}$ by considering it as a real vector space.[4] $\{D^{\{j\}}_{\mathfrak{D}^{\{j\}}}\}_{2j\in\mathbf{N}}$ is a complete set of representatives of the isomorphism classes of simple $\mathfrak{so}(3)$-modules.

An extension $(\mathcal{E}, \rho)$ of $\mathcal{E}(3)$ by an arbitrary Lie algebra $\mathfrak{A}$ being given, we consider an ω-sliced extension $(\mathcal{E}, \rho, \sigma)$ with fundamental $\mathfrak{so}(3)$-module $A(\varphi, f_2)$. If Ω is the representation of $\mathfrak{so}(3)$ given by formula (1.10) of I, $\overset{2}{\wedge} T(3)_\Omega$ is isomorphic to $D^{\{1\}}_{\mathfrak{D}^{\{1\}}}$. It follows that $A(\varphi, f_2)$ is a simple $\mathfrak{so}(3)$-module isomorphic to $D^{\{1\}}_{\mathfrak{D}^{\{1\}}}$ too, or $A(\varphi, f_2) = \{0\}$. Hence, by Proposition 4 of I, if the isotypical component of type $D^{\{1\}}_{\mathfrak{D}^{\{1\}}}$ of the $\mathfrak{so}(3)$-module A_ω has length 0, then all extensions of $\mathcal{E}(3)$ by $\mathfrak{A}$ with character $\Phi = \Pi_{\mathfrak{A}}\circ\varphi$ are inessential. Pursuing the analysis of $(\mathcal{E}, \rho)$, we find that $(\varphi, f_2) \in \mathfrak{Z}^2_{\Phi,\omega}(\mathcal{E}(3), \mathfrak{so}(3), \mathfrak{A})$ requires

$$\sum_{i,j,k} \epsilon_{ijk}\varphi(t_i)f_2(t_j, t_k) = 0. \tag{1.1}$$

If $f_2 = 0$, (1.1) is always satisfied. If $A(\varphi, f_2) \neq \{0\}$, $\varphi(\mathcal{T}(3))A(\varphi, f_2)$ is isomorphic to an $\mathfrak{so}(3)$-submodule of $\overset{2}{\otimes} D^{\{1\}}_{\mathfrak{D}^{\{1\}}}$. Take a basis $\{a_i\}_{1\le i\le 3}$ of $A(\varphi, f_2) \neq \{0\}$, where

$$a_i = \tfrac{1}{2}\sum_{j,k} \epsilon_{ijk} f_2(t_j, t_k). \tag{1.2}$$

This basis satisfies

$$\varphi(r_i)a_j = \sum_k \epsilon_{ijk}a_k \quad \text{for all } i, j \in \{1, 2, 3\},$$

and (1.1) becomes

$$\sum_i \varphi(t_i)a_i = 0. \tag{1.1'}$$

Tables I and II, whose meaning is self-evident, show immediately the consequence of condition (1.1'): $\varphi(\mathcal{T}(3))A(\varphi, f_2)$ does not have any $\mathfrak{so}(3)$-submodule isomorphic to $D^{\{0\}}_{\mathfrak{D}^{\{0\}}}$. Notice that, in the tables, $\mathfrak{D}^{(j)}$ ($j \in \{0, 1, 2, 3\}$) stands also for every representation isomorphic to $\mathfrak{D}^{(j)}$.

Now suppose that, in the above extension $(\mathcal{E}, \rho)$, $\mathfrak{A}$ is an Abelian Lie algebra. We can write $f_2 \in Z^2(\mathcal{E}(3), \mathfrak{so}(3), A_\Phi)$, after identification of φ with Φ.[7,8] Let $f_2' \in Z^2(\mathcal{E}(3), \mathfrak{so}(3), A_\Phi)$ with $A(\Phi, f_2') = A(\Phi, f_2)$. If $f_2 \neq 0$, consider the $\mathbf{R}$-linear map κ of $A(\Phi, f_2)$ onto itself given by

$$\kappa f_2(t_i, t_j) = f_2'(t_i, t_j) \quad \text{for all } i, j \in \{1, 2, 3\}.$$

Actually κ is an $\mathfrak{so}(3)$-endomorphism of $A(\Phi, f_2)$, since it satisfies

$$\kappa\Phi(r)a = \Phi(r)\kappa a \quad \text{for all } r \in \mathfrak{so}(3),\ a \in A(\Phi, f_2).$$

By Schur's lemma[9] κ is then an $\mathfrak{so}(3)$-automorphism. Take the complexifications $\mathfrak{so}(3)_{(\mathbf{C})}$ and $A(\Phi, f_2)_{(\mathbf{C})}$ of $\mathfrak{so}(3)$, resp. $A(\Phi, f_2)$. There is a unique $\mathfrak{so}(3)_{(\mathbf{C})}$-automorphism $\kappa_{(\mathbf{C})}$ such that the following diagram is commutative:

$$\begin{array}{ccc} A(\Phi, f_2) & \xrightarrow{\xi} & A(\Phi, f_2)_{(\mathbf{C})} \\ \downarrow \kappa & & \downarrow \kappa_{(\mathbf{C})} \\ A(\Phi, f_2) & \xrightarrow{\xi} & A(\Phi, f_2)_{(\mathbf{C})} \end{array}$$

with ξ the canonical complexification map. $A(\Phi, f_2)_{(\mathbf{C})}$ is a simple $\mathfrak{so}(3)_{(\mathbf{C})}$-module and, by a well-known corollary to Schur's lemma, $\kappa_{(\mathbf{C})} = \lambda_{(\mathbf{C})} I_{A(\Phi, f_2)_{(\mathbf{C})}}$ ($\lambda_{(\mathbf{C})} \in \mathbf{C}$, $\lambda_{(\mathbf{C})} \neq 0$). Then $\kappa = \lambda I_{A(\Phi, f_2)}$ ($\lambda \in \mathbf{R}$, $\lambda \neq 0$), i.e., $f_2' = \lambda f_2$. This result can also be easily found by applying the same method which will be used in Sec. 2 for the case of $\mathfrak{L}$-modules. Notice that $\underline{\lambda} f_2 = \underline{\lambda'} f_2$ and $f_2 \neq 0$ imply $\lambda = \lambda'$, because $\underline{\lambda} f_2 = \lambda \underline{f_2}$. If we consider the vector space $A' = \overline{\mathrm{Im} f_2} \neq \{0\}$ and if we define Φ': $\mathcal{E}(3) \to \mathfrak{gl}(A')$ by

$$\Phi'(r)a' = \Phi(r)a' \quad \text{for all } r \in \mathfrak{so}(3), a' \in A',$$
$$\Phi'(t) = 0 \quad \text{for all } t \in \mathcal{T}(3),$$

then we get

$$H^2(\mathcal{E}(3), \mathfrak{so}(3), A'_{\Phi'}) = \left\{ \lambda \underline{h_2} \;\middle|\; \begin{array}{l} \lambda \in \mathbf{R};\ h_2 \in A_2(\mathcal{E}(3), A'_{\Phi'}) \text{ with} \\ h_2(e, e') = f_2(e, e') \text{ for all } e, e' \in \mathcal{E}(3) \end{array} \right\}, \tag{1.3}$$

where $A_2(\mathcal{E}(3), A'_{\Phi'})$ is the vector space of bilinear alternating maps of $\mathcal{E}(3) \times \mathcal{E}(3)$ into $A'_{\Phi'}$. (1.3) is obviously true for $f_2 = 0$ too. Furthermore, if $f_2 \neq 0$, the existence of $f_1 \in C^1(\mathcal{E}(3), \mathfrak{so}(3), A_\Phi)$ and of a pair of elements $t, t' \in \mathcal{T}(3)$ such that $(\delta_1 f_1)(t, t') \in A(\Phi, f_2)$ and $(\delta_1 f_1)(t, t') \neq 0$ is a necessary and sufficient condition in order that $(\underline{\mathcal{E}, \rho})$ be an equivalence class of inessential extensions.

Keeping the same notations as above, let $(\mathcal{E}, \rho)$ be a primitive extension. By Theorem 2 of I it is an extension with Abelian kernel, and by definition $A(\varphi, f_2) = A_\omega$, i.e., $\mathfrak{A}$ is three-dimensional or $\{0\}$. On account of the fact that $\mathcal{T}(3)$ is the nilradical of $\mathcal{E}(3)$, $\Phi(\mathcal{T}(3)) = \{0\}$. If $(\mathcal{E}, \rho)$ is essential, $\{\sigma(r_i), \sigma(t_j), a_k\}_{1\le i,j,k\le 3}$ is a basis of $\mathcal{E}$ with a_k defined by (1.2) and with the Lie products

$$[\sigma(r_i), \sigma(r_j)] = \sum_k \epsilon_{ijk}\sigma(r_k), \qquad [\sigma(r_i), \sigma(t_j)] = \sum_k \epsilon_{ijk}\sigma(t_k),$$
$$[\sigma(r_i), a_j] = \sum_k \epsilon_{ijk}a_k, \qquad [\sigma(t_i), a_j] = 0,$$

TABLE I. Basis elements of some $S\mathcal{O}(3)$-submodules (up to canonical identifications) of the pth tensorial power $\overset{p}{\otimes} D^{\{1\}}_{\mathfrak{D}\{1\}}$ $(p = 1, 2, 3)$.

$D^{\{1\}}_{\mathfrak{D}\{1\}}$	$\overset{2}{\otimes} D^{\{1\}}_{\mathfrak{D}\{1\}} = D^{\{1\,0\}}_{\mathfrak{D}\{0\}} \oplus D^{\{1\,1\}}_{\mathfrak{D}\{1\}} \oplus D^{\{1\,2\}}_{\mathfrak{D}\{2\}}$			$\overset{3}{\otimes} D^{\{1\}}_{\mathfrak{D}\{1\}} = (D^{\{1\}}_{\mathfrak{D}\{1\}} \otimes D^{\{1\,0\}}_{\mathfrak{D}\{0\}}) \oplus (D^{\{1\}}_{\mathfrak{D}\{1\}} \otimes D^{\{1\,1\}}_{\mathfrak{D}\{1\}}) \oplus (D^{\{1\}}_{\mathfrak{D}\{1\}} \otimes D^{\{1\,2\}}_{\mathfrak{D}\{2\}})$		
	$D^{\{1\,0\}}_{\mathfrak{D}\{0\}}$	$D^{\{1\,1\}}_{\mathfrak{D}\{1\}}$	$D^{\{1\,2\}}_{\mathfrak{D}\{2\}}$	$D^{\{1\}}_{\mathfrak{D}\{1\}} \otimes D^{\{1\,0\}}_{\mathfrak{D}\{0\}} = D^{\{1\,0\,1\}}_{\mathfrak{D}\{1\}}$	$D^{\{1\}}_{\mathfrak{D}\{1\}} \otimes D^{\{1\,1\}}_{\mathfrak{D}\{1\}} = D^{\{1\,1\,0\}}_{\mathfrak{D}\{0\}} \oplus D^{\{1\,1\,1\}}_{\mathfrak{D}\{1\}} \oplus D^{\{1\,1\,2\}}_{\mathfrak{D}\{2\}}$	
					$D^{\{1\,1\,0\}}_{\mathfrak{D}\{0\}}$	$D^{\{1\,1\,1\}}_{\mathfrak{D}\{1\}}$
			$e^{\{1\,2\}}_1 = e^{(1)}_1 \otimes e^{(1)}_2 + e^{(1)}_2 \otimes e^{(1)}_1$			
			$e^{\{1\,2\}}_2 = e^{(1)}_2 \otimes e^{(1)}_3 + e^{(1)}_3 \otimes e^{(1)}_2$			
			$e^{\{1\,2\}}_3 = e^{(1)}_3 \otimes e^{(1)}_1 + e^{(1)}_1 \otimes e^{(1)}_3$			
		$\{e^{\{1\,1\}}_i\}_{1\le i\le 3}$ where $e^{\{1\,1\}}_i = \sum_{j,k} \epsilon_{ijk} (e^{(1)}_j \otimes e^{(1)}_k)$		$\{e^{\{1\,0\,1\}}_i\}_{1\le i\le 3}$ where $e^{\{1\,0\,1\}}_i = e^{(1)}_i \otimes e^{(1\,0)}$		$\{e^{\{1\,1\,1\}}_i\}_{1\le i\le 3}$ where $e^{\{1\,1\,1\}}_i = \sum_{j,k} \epsilon_{ijk} (e^{(1)}_j \otimes e^{(1\,1)}_k)$
$\{e^{(1)}_i\}_{1\le i\le 3}$	$e^{(1\,0)} = \sum_i (e^{(1)}_i \otimes e^{(1)}_i)$		$e^{\{1\,2\}}_4 = e^{(1)}_1 \otimes e^{(1)}_1 - e^{(1)}_2 \otimes e^{(1)}_2$		$e^{(1\,1\,0)} = \sum_i (e^{(1)}_i \otimes e^{(1\,1)}_i)$	
			$e^{\{1\,2\}}_5 = e^{(1)}_1 \otimes e^{(1)}_1 - e^{(1)}_3 \otimes e^{(1)}_3$			

$$[\sigma(t_i), \sigma(t_j)] = \sum_k \epsilon_{ijk} a_k, \qquad [a_i, a_j] = 0.$$

The (maximal) graph associated with $(\mathcal{E}, \rho, \sigma)$ can be represented by means of the one-vertex diagram

$$\overset{\{1\}}{\circledcirc}, \tag{1.4}$$

whose meaning is self-evident. The elements of $H^2(\mathcal{E}(3), S\mathcal{O}(3), A_\Phi)$ are given by (1.3), and may be bijectively associated with those of $\mathrm{Ext}(\mathcal{E}(3), A_\Phi)$ $[= \mathrm{Ext}(\mathcal{E}(3), \mathfrak{A}, \Phi)]$ by means of the Hochschild–Serre theorem and of the bijection (1.3) of I. If $f_2 = 0$, we get in this manner the equivalence class of inessential primitive extensions (truly trivial extensions). If $f_2 \neq 0$, only for $\lambda \neq 0$ we have classes of essential primitive extensions. From remark (2) at the end of Sec. 1 of I we infer that the essential primitive extensions of $\mathcal{E}(3)$ are all of the same type. We emphasize that the same diagram (1.4) actually represents the graphs of all essential primitive extensions, i.e., of all extensions belonging to $\mathfrak{E}_1(\mathcal{E}(3))$.

The graphs of the extensions belonging to $\mathfrak{E}_2(\mathcal{E}(3))$ are pictured in Table III. With any irreducible ω-sliced extension $(\mathcal{E}, \rho, \sigma)$ is associated one diagram, and we will hereafter refer to it as to a "diagram of $(\mathcal{E}, \rho)$." It is a diagram which represents graphs of $(\mathcal{E}, \rho)$. But, naturally, there are a lot of extensions (even nonequivalent) with the same diagram. The meaning of Table III will be cleared by the following example. Take the diagram

$$\overset{\{1\}}{\circledcirc} \longrightarrow \overset{\{1\}}{\circ} \tag{1.5}$$

Suppose that it represents the graph associated with an irreducible ω-sliced extension $(\mathcal{E}, \rho, \sigma)$ of $\mathcal{E}(3)$ by $\mathfrak{A}$ with fundamental $S\mathcal{O}(3)$-module $A(\varphi, f_2)$. Then we read from (1.5) that

$$A_\omega = A(\varphi, f_2) \oplus A'_{\omega'},$$

where ω' is the subrepresentation of ω on A'. $A(\varphi, f_2)$ and $A'_{\omega'}$ are both isomorphic to $D^{\{1\}}_{\mathfrak{D}\{1\}}$, and the base vertex $A(\varphi, f_2)$ is symbolized by $\overset{\{1\}}{\circledcirc}$. The arrow gives the action of $\varphi(\mathcal{T}(3))$, i.e., $\varphi(\mathcal{T}(3))A(\varphi, f_2) = A'_{\omega'}$, $\varphi^2(\mathcal{T}(3))A(\varphi, f_2) = \{0\}$. Tables I and II show that if $\{a_i\}_{1\le i\le 3}$ is the basis of $A(\varphi, f_2)$ given by (1.2),

The basis $\{e_i^{(1)}\}_{1\le i\le 3}$ of $D^{\{1\}}_{\mathfrak{D}\{1\}}$ satisfies $\mathfrak{D}^{\{1\}}(r_i)e_j^{(1)} = \sum_k \epsilon_{ijk} e_k^{(1)}$ $(i,j,k \in \{1,2,3\})$.

	$D^{\{1\}}_{\mathfrak{D}\{1\}} \otimes D^{\{12\}}_{\mathfrak{D}\{12\}} = D^{\{121\}}_{\mathfrak{D}\{1\}} \oplus D^{\{122\}}_{\mathfrak{D}\{2\}} \oplus D^{\{123\}}_{\mathfrak{D}\{3\}}$		
$D^{\{112\}}_{\mathfrak{D}\{12\}}$	$D^{\{121\}}_{\mathfrak{D}\{1\}}$	$D^{\{122\}}_{\mathfrak{D}\{2\}}$	$D^{\{123\}}_{\mathfrak{D}\{3\}}$
$e_1^{\{112\}} = e_1^{(1)} \otimes e_2^{(11)}$	$e_1^{\{121\}} = 3e_2^{(1)} \otimes e_1^{(12)}$	$e_1^{\{122\}} = e_1^{(1)} \otimes e_3^{(12)}$	$e_1^{\{123\}} = e_1^{(1)} \otimes e_3^{(12)}$
$+ e_2^{(1)} \otimes e_1^{(11)}$	$+ 3e_3^{(1)} \otimes e_3^{(12)}$	$- e_2^{(1)} \otimes e_2^{(12)}$	$+ e_3^{(1)} \otimes e_5^{(12)}$
	$+ 2e_1^{(1)} \otimes (e_4^{(12)} + e_5^{(12)})$	$- 2e_3^{(1)} \otimes e_4^{(12)}$	
$e_2^{\{112\}} = e_2^{(1)} \otimes e_3^{(11)}$	$e_2^{\{121\}} = 3e_3^{(1)} \otimes e_2^{(12)}$	$e_2^{\{122\}} = e_2^{(1)} \otimes e_1^{(12)}$	$e_2^{\{123\}} = e_2^{(1)} \otimes e_1^{(12)}$
$+ e_3^{(1)} \otimes e_2^{(11)}$	$+ 3e_1^{(1)} \otimes e_1^{(12)}$	$- e_3^{(1)} \otimes e_3^{(12)}$	$- e_1^{(1)} \otimes e_4^{(12)}$
	$- 2e_2^{(1)} \otimes (2e_4^{(12)} - e_5^{(12)})$	$+ 2e_1^{(1)} \otimes (e_4^{(12)} - e_5^{(12)})$	
$e_3^{\{112\}} = e_3^{(1)} \otimes e_1^{(11)}$	$e_3^{\{121\}} = 3e_1^{(1)} \otimes e_3^{(12)}$	$e_3^{\{122\}} = e_3^{(1)} \otimes e_2^{(12)}$	$e_3^{\{123\}} = e_3^{(1)} \otimes e_3^{(12)}$
$+ e_1^{(1)} \otimes e_3^{(11)}$	$+ 3e_2^{(1)} \otimes e_2^{(12)}$	$- e_1^{(1)} \otimes e_1^{(12)}$	$- e_1^{(1)} \otimes e_5^{(12)}$
	$+ 2e_3^{(1)} \otimes (e_4^{(12)} - 2e_5^{(12)})$	$+ 2e_2^{(1)} \otimes e_5^{(12)}$	
		$e_4^{\{122\}} = - e_1^{(1)} \otimes e_2^{(12)}$	$e_4^{\{123\}} = e_1^{(1)} \otimes e_1^{(12)}$
$e_4^{\{112\}} = e_1^{(1)} \otimes e_1^{(11)}$		$- e_2^{(1)} \otimes e_3^{(12)}$	$+ e_2^{(1)} \otimes e_4^{(12)}$
$- e_2^{(1)} \otimes e_2^{(11)}$		$+ 2e_3^{(1)} \otimes e_1^{(12)}$	
		$e_5^{\{122\}} = e_3^{(1)} \otimes e_1^{(12)}$	$e_5^{\{123\}} = e_1^{(1)} \otimes e_2^{(12)}$
$e_5^{\{112\}} = e_1^{(1)} \otimes e_1^{(11)}$		$+ e_1^{(1)} \otimes e_2^{(12)}$	$+ e_2^{(1)} \otimes e_3^{(12)}$
$- e_3^{(1)} \otimes e_3^{(11)}$		$- 2e_2^{(1)} \otimes e_3^{(12)}$	$+ e_3^{(1)} \otimes e_1^{(12)}$
			$e_6^{\{123\}} = e_1^{(1)} \otimes e_1^{(12)}$
			$- e_3^{(1)} \otimes e_2^{(12)}$
			$+ e_2^{(1)} \otimes e_5^{(12)}$
			$e_7^{\{123\}} = e_2^{(1)} \otimes e_2^{(12)}$
			$- e_1^{(1)} \otimes e_3^{(12)}$
			$- e_3^{(1)} \otimes e_4^{(12)}$

$\{a_i'\}_{1\le i\le 3}$ with $a_i' = \sum_{j,k} \epsilon_{ijk}\varphi(t_j)a_k$ is a basis of $A'_{\omega'}$. The Lie algebra $\mathfrak{A}$ is Abelian, as follows from Theorems 3 and 5 of I. Actually, by the same theorems, all extensions belonging to $\mathfrak{E}_2(\mathcal{E}(3))$ have Abelian kernels. However, the kernel of an extension belonging to $\mathfrak{E}_3(\mathcal{E}(3))$ can be non-Abelian. Take, for example, the diagram

$$\{1\} \qquad \{1\} \qquad \{1\} \tag{1.6}$$
$$\circledcirc \longrightarrow \circ \longrightarrow \circ$$

Keeping the same notation as before, we get

$$A_\omega = A(\varphi, f_2) \oplus A'_{\omega'} \oplus A''_{\omega''},$$

where $A''_{\omega''}$ too is isomorphic to $D^{\{1\}}_{\mathfrak{D}\{1\}}$. Besides, $\varphi^2(\mathcal{T}(3))A(\varphi, f_2) = A''_{\omega''}$, $\varphi^3(\mathcal{T}(3))A(\varphi, f_2) = \{0\}$, and there is a basis $\{a_i''\}_{1\le i\le 3}$ of $A''_{\omega''}$, where $a_i'' = \sum_j \{\varphi(t_j) \times \varphi(t_i)a_j - \varphi(t_j)\varphi(t_j)a_i\}$. The diagram (1.6) does not give the Lie algebra structure of $\mathfrak{A}$: It determines only the $\mathfrak{SO}(3)$-module structure of A_ω. Some diagrams represent graphs of extensions with Abelian kernels as well as graphs of extensions with non-Abelian kernels. However, this is not the case of diagram (1.6). Notice that if $\mathfrak{A}$ is an Abelian, resp. non-Abelian Lie algebra, the Lie algebra structure of $\mathcal{E}$ requires, for all $t, t' \in \mathcal{T}(3)$,

$$[\varphi(t), \varphi(t')] = 0,$$

resp. $$[\varphi(t), \varphi(t')] = \mathrm{ad} f_2(t, t').$$

Furthermore, if $A_{1\omega_1}$ and $A_{2\omega_2}$ are arbitrary $\mathfrak{SO}(3)$-submodules of A_ω, then $[A_{1\omega_1}, A_{2\omega_2}]$ must be an $\mathfrak{SO}(3)$-submodule of A_ω too. In the case of diagram (1.6), by Theorems 3 and 5 of I and on account of Schur's lemma, we must have

$$[a_i, a_j] = \lambda \sum_k \epsilon_{ijk} a_k'', \quad [a_i, a_j'] = 0, \quad [a_i, a_j''] = 0,$$
$$[a_i', a_j'] = 0, \quad [a_i', a_j''] = 0, \quad [a_i'', a_j''] = 0,$$
$$\text{for all } i, j \in \{1, 2, 3\},$$

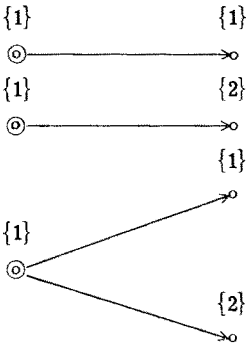
where $\lambda \in \mathbf{R}$ and is different from 0.

The assumptions of Theorems 6, 9, and 10 of I are satisfied by all irreducible extensions of $\mathcal{E}(3)$ which

TABLE II. Diagonal action of $S\mathcal{O}(3)$ on the basis elements of ${}^{\gamma}D^{\{1\}}_{\mathfrak{D}\{1\}}$ ($p = 1, 2, 3$) given in Table I.

	${}^{\{j\}(r_k)}b_i^{\{j\}}$																
	$b_i^{\{j\}}$	$b_1^{(0)} = \{e^{(110)}$		$b_i^{(1)} = \{e_i^{(1)}, e_i^{(111)}, e_i^{(101)}$			$b_i^{(2)} = e_i^{(112)}, e_i^{(112)}, e_i^{(122)}$					$b_i^{(3)} = e_i^{(123)}$						
${}^{\{j\}(r_k)}\mathfrak{D}^{\{j\}}$		$b_1^{(0)}$		$b_1^{(1)}$	$b_2^{(1)}$	$b_3^{(1)}$	$b_1^{(2)}$	$b_2^{(2)}$	$b_3^{(2)}$	$b_4^{(2)}$	$b_5^{(2)}$	$b_1^{(3)}$	$b_2^{(3)}$	$b_3^{(3)}$	$b_4^{(3)}$	$b_5^{(3)}$	$b_6^{(3)}$	$b_7^{(3)}$
${}^{\{j\}(r_1)}\mathfrak{D}^{\{j\}}$		0		0	$b_3^{(1)}$	$-b_2^{(1)}$	$b_3^{(2)}$	$2(b_4^{(2)} - b_5^{(2)})$	$-b_1^{(2)}$	$-b_2^{(2)}$	$b_2^{(2)}$	$-b_6^{(3)}$	$b_5^{(3)}$	$-b_5^{(3)}$	$-b_7^{(3)}$	$2(b_3^{(3)} - b_2^{(3)})$	$3b_1^{(3)} + 2b_7^{(3)}$	$-2b_6^{(3)} + 3b_4^{(3)}$
${}^{\{j\}(r_2)}\mathfrak{D}^{\{j\}}$		0		$-b_3^{(1)}$	0	$b_1^{(1)}$	$-b_2^{(2)}$	$b_1^{(2)}$	$2b_5^{(2)}$	$-b_3^{(2)}$	$-2b_3^{(2)}$	$-3b_3^{(3)}$	$-b_7^{(3)}$	$3b_1^{(3)}$	$-b_5^{(3)}$	$2b_6^{(3)}$	$-2b_5^{(3)}$	$2b_3^{(3)} + b_2^{(3)}$
${}^{\{j\}(r_3)}\mathfrak{D}^{\{j\}}$		0		$b_2^{(1)}$	$-b_1^{(1)}$	0	$-2b_4^{(2)} - b_3^{(2)}$		$b_2^{(2)}$	$2b_1^{(2)}$	$b_1^{(2)}$	$b_5^{(3)}$	$-3b_4^{(3)}$	$-b_6^{(3)}$	$3b_2^{(3)}$	$2b_7^{(3)}$	$b_3^{(3)} + 2b_2^{(3)}$	$-2b_5^{(3)}$

TABLE III. Diagrams of the extensions belonging to $\mathfrak{E}_2(\mathcal{E}(3))$.



then belong to $\mathfrak{E}_n(\mathcal{E}(3))$ ($n \in \mathbf{N}$) and are extensions by nilpotent Lie algebras. Their graphs are finite, without circuits and, for $n \neq 0$, with vertex bases consisting of the unique element represented by $\overset{\{1\}}{\odot}$. Therefore, all these graphs are weakly connected.[10,11] They can be pictured by diagrams as in the case of $\mathfrak{E}_2(\mathcal{E}(3))$. We emphasize that, on account of Theorem 7 of I, the $S\mathcal{O}(3)$-module A_ω determined by an arbitrary ω-sliced irreducible extension is isomorphic to the direct sum of a family of simple $S\mathcal{O}(3)$-modules belonging to $\{D^{\{j\}}_{\mathfrak{D}\{j\}}\}_{j\in\mathbf{N}}$.

2. EXTENSIONS OF $\mathcal{P}$

Lie algebra extensions of $\mathcal{P}$ have been dealt with in other papers.[7,8,12] Here, we will merely recall some results and give a few details.

We consider the Levi decomposition

$$\mathcal{P} = \mathcal{L} \dotplus \mathcal{T},$$

where $\mathcal{L}$ and $\mathcal{T}$ are, respectively, the Lie algebras of the Lorentz and of the translation groups of the four-dimensional Minkowskian space of relativistic space-time events. Let $\{l_{\mu\nu}, t_\rho\}_{0\leqslant\mu,\nu,\rho\leqslant 3}$ be the usual basis of $\mathcal{P}$ with the Lie products

$$[l_{\mu\nu}, l_{\rho\sigma}] = g_{\mu\rho}l_{\nu\sigma} + g_{\nu\sigma}l_{\mu\rho} - g_{\mu\sigma}l_{\nu\rho} - g_{\nu\rho}l_{\mu\sigma},$$
$$[l_{\mu\nu}, t_\rho] = g_{\mu\rho}t_\nu - g_{\nu\rho}t_\mu, \qquad [t_\rho, t_\sigma] = 0.$$

$\{l_{\mu\nu}\}_{0\leqslant\mu,\nu\leqslant 3}$ and $\{t_\rho\}_{0\leqslant\rho\leqslant 3}$ are, respectively, bases of $\mathcal{L}$ and $\mathcal{T}$.

Take an irreducible continuous complex representation $\mathfrak{D}^{(j_1,j_2)}(2j_1, 2j_2 \in \mathbf{N})$ of $SL(2, \mathbf{C})$ on a vector space $D^{(j_1,j_2)}$ of dimension $(2j_1 + 1)(2j_2 + 1)$, for example, the spinor representation $S_{2j_1,2j_2}$ (on $\tilde{R}_{2j_1,2j_2}$) of Ref. 2. $\mathcal{L}$ being isomorphic to $S\mathcal{L}(2, \mathbf{C})$ [the Lie algebra of $SL(2, \mathbf{C})$], and proceeding as we did in Sec. 1 for the representations of $S\mathcal{O}(3)$, we get a representation of $\mathcal{L}$ on $D^{(j_1,j_2)}$ still denoted $\mathfrak{D}^{(j_1,j_2)}$. Suppose first $j_1 = j_2 = j$. $\mathfrak{D}^{(j,j)}$ is of real type and we can choose one real form $\mathfrak{D}^{\{j,j\}}$ of $\mathfrak{D}^{(j,j)}$ on a vector space $D^{\{j,j\}}$ defined as referred in Sec. 1. If $j_1 \neq j_2$, $\mathfrak{D}^{(j_1,j_2)}$ is neither of real nor of quaternionic type, and it induces a real representation $\mathfrak{D}^{\{j_1,j_2\}}$ on the $[2(2j_1 + 1)(2j_2 + 1)]$-dimensional real vector space $D^{\{j_1,j_2\}}$ obtained from $D^{(j_1,j_2)}$ by restriction of the field $\mathbf{C}$ to $\mathbf{R}$. A complete set of representatives of the isomorphism classes of simple $\mathcal{L}$-modules is given by

$$\left\{D^{\{j_1,j_2\}}_{\mathfrak{D}\{j_1,j_2\}}\right\}_{\substack{2j_1,2j_2\in\mathbf{N}\cdot 4\\ j_1>j_2}}$$

Using the same notation as in Sec. 1, let now $(\mathcal{E}, \rho, \sigma)$ be an ω-sliced extension of $\mathcal{P}$ by an arbitrary Lie algebra $\mathfrak{A}$ and let $A(\varphi, f_2)$ be its fundamental $\mathcal{L}$-module. $\overset{2}{\wedge} T_\Omega$ is isomorphic to $D^{\{1,0\}}_{\mathfrak{D}\{1,0\}}$, and this implies that $A(\varphi, f_2)$ is isomorphic to $D^{\{1,0\}}_{\mathfrak{D}\{1,0\}}$ or $A(\varphi, f_2) = \{0\}$. Because of $(\varphi, f_2) \in \mathfrak{Z}^{\,2}_{\Phi,\omega}(\mathcal{P}, \mathcal{L}, \mathfrak{A})$, we get

$$\sum_{\nu,\rho,\sigma} \epsilon_{\mu\nu\rho\sigma}\,\varphi(t^\nu)f_2(t^\rho, t^\sigma) = 0 \quad \text{for all } \mu \in \{0, 1, 2, 3\},$$

where

$$t^\tau = \sum_\upsilon g^{\tau\upsilon}t_\upsilon.$$

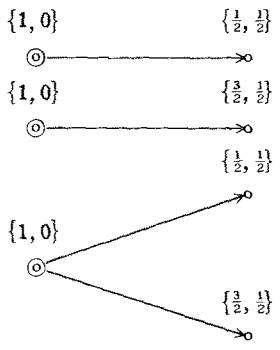
It follows that there is no simple $\mathcal{L}$-submodule of $\varphi(\mathcal{T})A(\varphi, f_2)$ isomorphic to $D^{\{1/2,1/2\}}_{\mathfrak{D}\{1/2,1/2\}}$ and generated by $\{\sum_{\nu,\rho,\sigma}\epsilon_{\mu\nu\rho\sigma}\varphi(t^\nu)f_2(t^\rho, t^\sigma)\}_{0\leqslant\mu\leqslant 3}$. However, $\varphi(\mathcal{T})A(\varphi, f_2)$, which is isomorphic to an $\mathcal{L}$-submodule of $D^{\{1/2,1/2\}}_{\mathfrak{D}\{1/2,1/2\}} \otimes D^{\{1,0\}}_{\mathfrak{D}\{1,0\}}$, can have one simple $\mathcal{L}$-submodule isomorphic to $D^{\{1/2,1/2\}}_{\mathfrak{D}\{1/2,1/2\}}$ and generated by $\{\sum_\nu \varphi(t^\nu)f_2(t_\nu, t_\mu)\}_{0\leqslant\mu\leqslant 3}$. Notice that, on account of Proposition 4 of I, if the length of the isotypical component of type $D^{\{1,0\}}_{\mathfrak{D}\{1,0\}}$ of A_ω is 0, then all extensions of $\mathcal{P}$ by $\mathfrak{A}$ with character Φ are inessential (Proposition 3 of Ref. 8).

As usual, if $\mathfrak{A}$ is an Abelian Lie algebra we identify φ and Φ, and we have $f_2 \in Z^2(\mathcal{P}, \mathcal{L}, A_\Phi)$. Let $f_2 \neq 0$. Taking $f_2' \in Z^2(\mathcal{P}, \mathcal{L}, A_\Phi)$ with $A(\Phi, f_2') = A(\Phi, f_2)$, and letting κ be the $\mathbf{R}$-linear map of $A(\Phi, f_2)$ onto itself given by

$$\kappa f_2(t_\mu, t_\nu) = f_2'(t_\mu, t_\nu) \quad \text{for all } \mu, \nu \in \{0, 1, 2, 3\},$$

we conclude that κ is an $\mathcal{L}$-automorphism of $A(\Phi, f_2)$ by Schur's lemma. But now we can no more follow the method of Sec. 1, since $A(\Phi, f_2)_{(\mathbf{C})}$ is not a simple

TABLE IV. Diagrams of the extensions belonging to $\mathfrak{E}_2(\mathcal{P})$



$\mathfrak{L}_{(\mathbb{C})}$-module. However, the number of linearly independent $\mathfrak{L}$-automorphisms of $A(\Phi, f_2)$ is given by the length of the isotypical component of type $D^{\{0,0\}}_{\mathfrak{D}^{\{0,0\}}}$ in the second tensorial power $\overset{2}{\otimes} D^{\{1,0\}}_{\mathfrak{D}^{\{1,0\}}}$.[9] As this length is 2, and as we know that f_2 and f_2^* are linearly independent bilinear alternating maps of $\mathcal{P} \times \mathcal{P}$ into $\mathfrak{A}$ if f_2^* is defined by

$$f_2^*(l_{\mu\nu}, l_{\rho\sigma}) = 0, \quad f_2^*(l_{\mu\nu}, t_\rho) = 0, \quad f_2^*(t_\mu, t_\nu) = \tfrac{1}{2}\sum_{\tau,\upsilon} \epsilon_{\mu\nu\tau\upsilon} f_2(t^\tau, t^\upsilon)$$

for all $\mu, \nu, \rho, \sigma \in \{0, 1, 2, 3\}$,

we get

$$f_2' = \lambda f_2 + \lambda' f_2^* \quad (\lambda, \lambda' \in \mathbf{R}).$$

Hence, with $A' = \overline{\mathrm{Im} f_2} \neq \{0\}$ and $\Phi'\colon \mathcal{P} \to \mathfrak{GL}(A')$ given by

$$\Phi'(l)a' = \Phi(l)a' \quad \text{for all } l \in \mathfrak{L}, a' \in A',$$
$$\Phi'(t) = 0 \quad \text{for all } t \in \mathcal{T},$$

we obtain

$$H^2(\mathcal{P}, \mathfrak{L}, A'_{\Phi'}) = \left\{ \lambda \underline{h}_2 + \lambda' \underline{h}_2^* \;\middle|\; \begin{array}{l} \lambda, \lambda' \in \mathbf{R};\ h_2, h_2^* \in A_2(\mathcal{P}, A'_{\Phi'}) \text{ with} \\ h_2(p, p') = f_2(p, p'),\ h_2^*(p, p') \\ = f_2^*(p, p') \quad \text{for all } p, p' \in \mathcal{P} \end{array} \right\}, \tag{2.1}$$

and dim $H^2(\mathcal{P}, \mathfrak{L}, A'_{\Phi'}) = 2$, and not 1 as stated in Ref. 7. It follows also from (2.1) that $(\underline{\mathcal{E}}, \underline{\rho})$ is an equivalence class of inessential extensions if and only if there are a linear map $f_1 \in C^1(\mathcal{P}, \mathfrak{L}, A_\Phi)$ and a pair of elements $t, t' \in \mathcal{T}$ satisfying

$$(\delta_1 f_1)(t, t') = f_2(t, t').$$

The extensions belonging to $\mathfrak{E}_1(\mathcal{P})$, i.e., the essential primitive extensions of $\mathcal{P}$, have Abelian kernels of dimension six. They can be constructed, using standard methods, starting from the elements of $H^2(\mathcal{P}, \mathfrak{L}, A_\Phi)$, which are given by (2.1) since $\Phi(\mathcal{T}) = \{0\}$.[7,8,12] $\overset{\{1,0\}}{\circledcirc}$ is the diagram of the extensions belonging to $\mathfrak{E}_1(\mathcal{P})$, which are all of the same type.

In Table IV we list the diagrams of all extensions belonging to $\mathfrak{E}_2(\mathcal{P})$. On account of Theorems 3 and 5 of I, $\mathfrak{E}_2(\mathcal{P})$ is again a set of extensions with Abelian kernels. Examples of extensions belonging to $\mathfrak{E}_3(\mathcal{P})$ and with non-Abelian kernels were given in Ref. 8. The irreducible extensions of $\mathcal{P}$ satisfy the assumptions of Theorems 6, 9, and 10 of I (cf. Proposition 5 and Theorems 11, 12 of Ref. 8). Therefore, they belong all to $\mathfrak{E}_n(\mathcal{P})$ ($n \in \mathbf{N}$); their graphs are finite and without circuits, weakly connected and, for $n \neq 0$, $\overset{\{1,0\}}{\circledcirc}$ represents the unique element of their vertex bases.

3. EXTENSIONS OF $\mathcal{G}$

We choose a basis $\{r_i, n_j, t_k, h_0\}_{1 \leq i,j,k \leq 3}$ of $\mathcal{G}$ with the Lie products

$$[r_i, r_j] = \sum_k \epsilon_{ijk} r_k, \quad [n_i, n_j] = 0, \quad [t_i, t_j] = 0,$$
$$[r_i, n_j] = \sum_k \epsilon_{ijk} n_k, \quad [r_i, t_j] = \sum_k \epsilon_{ijk} t_k, \quad [r_i, h_0] = 0,$$
$$[n_i, t_j] = 0, \quad [n_i, h_0] = t_i, \quad [t_i, h_0] = 0.$$

$\{r_i\}_{1 \leq i \leq 3}$, $\{n_j\}_{1 \leq j \leq 3}$, $\{t_k\}_{1 \leq k \leq 3}$, and $\{h_0\}$ generate $\mathcal{SO}(3)$, $\mathcal{N}(3)$, $\mathcal{T}(3)$, and $\mathcal{K}$ which are, respectively, the Lie algebras of the groups of space rotations, of Galilean boosts, of space translations, and of time translations of the four-dimensional space of nonrelativistic space-time events. The radical of $\mathcal{G}$ is

$$\mathcal{R}_2 = \mathcal{N}(3) \dotplus (\mathcal{T}(3) \oplus \mathcal{K}),$$

and we have a Levi decomposition

$$\mathcal{G} = \mathcal{SO}(3) \dotplus \mathcal{R}_2.$$

Let now $(\mathcal{E}, \rho, \sigma)$ be an ω-sliced extension of $\mathcal{G}$ by an arbitrary Lie algebra $\mathfrak{A}$, and let $A(\varphi, f_2)$ be its fundamental $\mathcal{SO}(3)$-module. As $\overset{2}{\wedge} R_{2_\Omega}$ is isomorphic to an $\mathcal{SO}(3)$-submodule of $D^{\{2\}}_{\mathfrak{D}^{\{2\}}} \times 5D^{\{1\}}_{\mathfrak{D}^{\{1\}}} \times D^{\{0\}}_{\mathfrak{D}^{\{0\}}}$, $A(\varphi, f_2)$ is, in general, not simple. Proposition 4 of I means in this case that if the isotypical components of type $D^{\{0\}}_{\mathfrak{D}^{\{0\}}}$, $D^{\{1\}}_{\mathfrak{D}^{\{1\}}}$, and $D^{\{2\}}_{\mathfrak{D}^{\{2\}}}$ of the $\mathcal{SO}(3)$-module A_ω have all lengths 0, then every extension of $\mathcal{G}$ by $\mathfrak{A}$ with character $\Phi = \Pi_{\mathfrak{a}} \circ \varphi$ is inessential.

Let $f_1 \in \mathrm{L}(\mathcal{G}, \mathfrak{A})$ be given by

$$f_1(r_i) = 0, \quad f_1(n_j) = 0, \quad f_1(t_k) = -f_2(n_k, h_0), \quad f_1(h_0) = 0$$

for all $i, j, k \in \{1, 2, 3\}$.

f_1 is actually ω-orthogonal to $\mathcal{SO}(3)$ and, moreover, it satisfies

$$(\delta_1(\varphi) f_1)(n_k, h_0) = f_2(n_k, h_0) \quad \text{for all } k \in \{1, 2, 3\}.$$

For that reason, we will hereafter suppose that $(\mathcal{E}, \rho, \sigma)$ has been chosen in such a way that

$$f_2(n, h) = 0 \quad \text{for all } n \in \mathcal{N}(3), h \in \mathcal{K}. \tag{3.1}$$

The conditions which must be satisfied by φ and f_2 in order that $(\varphi, f_2) \in \mathfrak{Z}^{\,2}_{\Phi,\omega}(\mathcal{G}, \mathcal{SO}(3), \mathfrak{A})$ are now more involved as in the cases of Secs. 1 and 2. For all $i, j, k \in \{1, 2, 3\}$ we get

$$\varphi(n_i) f_2(n_j, n_k) + \varphi(n_j) f_2(n_k, n_i) + \varphi(n_k) f_2(n_i, n_j) = 0, \tag{3.2}$$

$$\varphi(t_i) f_2(t_j, t_k) + \varphi(t_j) f_2(t_k, t_i) + \varphi(t_k) f_2(t_i, t_j) = 0, \tag{3.3}$$

$$\varphi(h_0)f_2(n_i,n_j)+f_2(n_i,t_j)-f_2(n_j,t_i)=0, \tag{3.4}$$

$$\varphi(t_i)f_2(t_j,h_0)+\varphi(t_j)f_2(h_0,t_i)+\varphi(h_0)f_2(t_i,t_j)=0, \tag{3.5}$$

$$\varphi(n_i)f_2(t_j,h_0)+\varphi(h_0)f_2(n_i,t_j)-f_2(t_j,t_i)=0, \tag{3.6}$$

$$\varphi(t_i)f_2(t_j,n_k)+\varphi(t_j)f_2(n_k,t_i)+\varphi(n_k)f_2(t_i,t_j)=0, \tag{3.7}$$

$$\varphi(n_i)f_2(n_j,t_k)+\varphi(n_j)f_2(t_k,n_i)+\varphi(t_k)f_2(n_i,n_j)=0. \tag{3.8}$$

Notice that, as

$$[\varphi(r_i),\varphi(h_0)]=0 \quad \text{for all } i\in\{1,2,3\},$$

$\varphi(h_0)|A'_{\omega'}$ is an $\mathfrak{SO}(3)$-monomorphism or 0 for each simple $\mathfrak{SO}(3)$-submodule $A'_{\omega'}$ of A_ω, on account of Schur's lemma. Formulas (3.2)–(3.8) give rise to the following system of relations between elements of $A(\varphi,f_2)$ and $\varphi(\mathfrak{R}_2)A(\varphi,f_2)$ which, if different from 0, are basis elements constructed according to Tables I and II:

$$\sum_{i,j,k}\epsilon_{ijk}\varphi(n_i)f_2(n_j,n_k)=0; \tag{3.2'}$$

$$\sum_{i,j,k}\epsilon_{ijk}\varphi(t_i)f_2(t_j,t_k)=0; \tag{3.3'}$$

$$\sum_{j,k}\epsilon_{ijk}\varphi(h_0)f_2(n_j,n_k)=-2\sum_{j,k}\epsilon_{ijk}f_2(n_j,t_k); \tag{3.4'}$$

$$2\sum_{j,k}\epsilon_{ijk}\varphi(t_j)f_2(t_k,h_0)+\sum_{j,k}\epsilon_{ijk}\varphi(h_0)f_2(t_j,t_k)=0; \tag{3.5'}$$

$$\sum_i\varphi(n_i)f_2(t_i,h_0)+\sum_i\varphi(h_0)f_2(n_i,t_i)=0; \tag{3.6'}$$

$$\sum_{j,k}\epsilon_{ijk}\varphi(n_j)f_2(t_k,h_0)+\sum_{j,k}\epsilon_{ijk}\varphi(h_0)f_2(n_j,t_k)=-\sum_{j,k}\epsilon_{ijk}f_2(t_j,t_k); \tag{3.6''}$$

$$\begin{aligned}
&\{\varphi(n_1)f_2(t_2,h_0)+\varphi(n_2)f_2(t_1,h_0)\}+\varphi(h_0)\{f_2(n_1,t_2)+f_2(n_2,t_1)\}=0,\\
&\{\varphi(n_2)f_2(t_3,h_0)+\varphi(n_3)f_2(t_2,h_0)\}+\varphi(h_0)\{f_2(n_2,t_3)+f_2(n_3,t_2)\}=0,\\
&\{\varphi(n_3)f_2(t_1,h_0)+\varphi(n_1)f_2(t_3,h_0)\}+\varphi(h_0)\{f_2(n_3,t_1)+f_2(n_1,t_3)\}=0,\\
&\{\varphi(n_1)f_2(t_1,h_0)-\varphi(n_2)f_2(t_2,h_0)\}+\varphi(h_0)\{f_2(n_1,t_1)-f_2(n_2,t_2)\}=0,\\
&\{\varphi(n_1)f_2(t_1,h_0)-\varphi(n_3)f_2(t_3,h_0)\}+\varphi(h_0)\{f_2(n_1,t_1)-f_2(n_3,t_3)\}=0;
\end{aligned} \tag{3.6'''}$$

$$2\sum_{i,j,k}\epsilon_{ijk}\varphi(t_i)f_2(n_j,t_k)+\sum_{i,j,k}\epsilon_{ijk}\varphi(n_i)f_2(t_j,t_k)=0; \tag{3.7'}$$

$$\begin{aligned}
&3\sum_{i,j,k,l}\epsilon_{1ij}\epsilon_{jkl}\varphi(t_i)f_2(n_k,t_l)-\{3\varphi(t_2)[f_2(n_1,t_2)+f_2(n_2,t_1)]+3\varphi(t_3)[f_2(n_3,t_1)+f_2(n_1,t_3)]+2\varphi(t_1)[f_2(n_1,t_1)-f_2(n_2,t_2)]\\
&\quad+2\varphi(t_1)[f_2(n_1,t_1)-f_2(n_3,t_3)]\}+4\sum_i\varphi(t_1)f_2(n_i,t_i)-3\sum_{i,j,k,l}\epsilon_{1ij}\epsilon_{jkl}\varphi(n_i)f_2(t_k,t_l)=0,
\end{aligned}$$

$$\begin{aligned}
&3\sum_{i,j,k,l}\epsilon_{2ij}\epsilon_{jkl}\varphi(t_i)f_2(n_k,t_l)-\{3\varphi(t_3)[f_2(n_2,t_3)+f_2(n_3,t_2)]+3\varphi(t_1)[f_2(n_1,t_2)+f_2(n_2,t_1)]\\
&\quad-4\varphi(t_2)[f_2(n_1,t_1)-f_2(n_2,t_2)]+2\varphi(t_2)[f_2(n_1,t_1)-f_2(n_3,t_3)]\}+4\sum_i\varphi(t_2)f_2(n_i,t_i)\\
&\quad-3\sum_{i,j,k,l}\epsilon_{2ij}\epsilon_{jkl}\varphi(n_i)f_2(t_k,t_l)=0,
\end{aligned} \tag{3.7''}$$

$$\begin{aligned}
&3\sum_{i,j,k,l}\epsilon_{3ij}\epsilon_{jkl}\varphi(t_i)f_2(n_k,t_l)-\{3\varphi(t_1)[f_2(n_3,t_1)+f_2(n_1,t_3)]+3\varphi(t_2)[f_2(n_2,t_3)+f_2(n_3,t_2)]\\
&\quad+2\varphi(t_3)[f_2(n_1,t_1)-f_2(n_2,t_2)]-4\varphi(t_3)[f_2(n_1,t_1)-f_2(n_3,t_3)]\}+4\sum_i\varphi(t_3)f_2(n_i,t_i)-3\sum_{i,j,k,l}\epsilon_{3ij}\epsilon_{jkl}\varphi(n_i)f_2(t_k,t_l)=0;\\
&\{\varphi(t_1)[f_2(n_3,t_1)+f_2(n_1,t_3)]-\varphi(t_2)[f_2(n_2,t_3)+f_2(n_3,t_2)]-2\varphi(t_3)[f_2(n_1,t_1)-f_2(n_2,t_2)]\}\\
&\quad-\sum_{i,j}\{\epsilon_{2ij}\varphi(t_1)f_2(n_i,t_j)+\epsilon_{1ij}\varphi(t_2)f_2(n_i,t_j)\}+\sum_{i,j}\{\epsilon_{2ij}\varphi(n_1)f_2(t_i,t_j)+\epsilon_{1ij}\varphi(n_2)f_2(t_i,t_j)\}=0,\\
&\{\varphi(t_2)[f_2(n_1,t_2)+f_2(n_2,t_1)]-\varphi(t_3)[f_2(n_3,t_1)+f_2(n_1,t_3)]+2\varphi(t_1)[f_2(n_1,t_1)-f_2(n_2,t_2)]\\
&\quad-2\varphi(t_1)[f_2(n_1,t_1)-f_2(n_3,t_3)]\}-\sum_{i,j}\{\epsilon_{3ij}\varphi(t_2)f_2(n_i,t_j)+\epsilon_{2ij}\varphi(t_3)f_2(n_i,t_j)\}\\
&\quad+\sum_{i,j}\{\epsilon_{3ij}\varphi(n_2)f_2(t_i,t_j)+\epsilon_{2ij}\varphi(n_3)f_2(t_i,t_j)\}=0,
\end{aligned}$$

$$\{\varphi(t_3)[f_2(n_2,t_3)+f_2(n_3,t_2)]-\varphi(t_1)[f_2(n_1,t_2)+f_2(n_2,t_1)]+2\varphi(t_2)[f_2(n_1,t_1)-f_2(n_3,t_3)]\}$$
$$-\sum_{i,j}\{\epsilon_{1ij}\varphi(t_3)f_2(n_i,t_j)+\epsilon_{3ij}\varphi(t_1)f_2(n_i,t_j)\}+\sum_{i,j}\{\epsilon_{1ij}\varphi(n_3)f_2(t_i,t_j)+\epsilon_{3ij}\varphi(n_1)f_2(t_i,t_j)\}=0, \qquad (3.7''')$$

$$\{-\varphi(t_1)[f_2(n_2,t_3)+f_2(n_3,t_2)]-\varphi(t_2)[f_2(n_3,t_1)+f_2(n_1,t_3)]+2\varphi(t_3)[f_2(n_1,t_2)+f_2(n_2,t_1)]\}$$
$$-\sum_{i,j}\{\epsilon_{1ij}\varphi(t_1)f_2(n_i,t_j)-\epsilon_{2ij}\varphi(t_2)f_2(n_i,t_j)\}+\sum_{i,j}\{\epsilon_{1ij}\varphi(n_1)f_2(t_i,t_j)-\epsilon_{2ij}\varphi(n_2)f_2(t_i,t_j)\}=0,$$

$$\{\varphi(t_3)[f_2(n_1,t_2)+f_2(n_2,t_1)]+\varphi(t_1)[f_2(n_2,t_3)+f_2(n_3,t_2)]-2\varphi(t_2)[f_2(n_3,t_1)+f_2(n_1,t_3)]\}$$
$$-\sum_{i,j}\{\epsilon_{1ij}\varphi(t_1)f_2(n_i,t_j)-\epsilon_{3ij}\varphi(t_3)f_2(n_i,t_j)\}+\sum_{i,j}\{\epsilon_{1ij}\varphi(n_1)f_2(t_i,t_j)-\epsilon_{3ij}\varphi(n_3)f_2(t_i,t_j)\}=0.$$

Formulas (3.8′), (3.8″), and (3.8‴) are then obtained, respectively, from (3.7′), (3.7″), and (3.7‴) making the substitutions $n_i \longleftrightarrow t_i$ for all $i \in \{1,2,3\}$. Notice that the system of relations (3.6‴), for example, is obtained from whatever one of its elements by application of the $\mathfrak{so}(3)$-module operations. Obviously, the same statement is valid for each of the systems (3.2′)–(3.8‴).

Using the same notation as before, we suppose that $(\mathcal{E},\rho)$ is an essential primitive extension, and we make the usual identifications of Φ and φ. As $\mathfrak{N}(3)\oplus\mathfrak{T}(3)$ is the nilradical of $\mathcal{G}$, $\Phi(\mathfrak{N}(3))=\Phi(\mathfrak{T}(3))=\{0\}$. Besides, by Schur's lemma,

$$\Phi(h_0)=\gamma I_{A_\Phi} \quad (\gamma\in\mathbf{R}) \qquad (3.9)$$

(cf. Sec. 1). We will prove that $\gamma=0$. In fact:

(a) $\Phi(h_0)=0$ by (3.6‴) if A_ω is isomorphic to $D^{\{2\}}_{\mathfrak{D}\{2\}}$, i.e., if it is generated by $\{b_i\}_{1\le i\le 5}$ with $b_1=f_2(n_1,t_2)+f_2(n_2,t_1)$, $b_2=f_2(n_2,t_3)+f_2(n_3,t_2)$, $b_3=f_2(n_3,t_1)+f_2(n_1,t_3)$, $b_4=f_2(n_1,t_1)-f_2(n_2,t_2)$, and $b_5=f_2(n_1,t_1)-f_2(n_3,t_3)$. Analogously, if A_ω is isomorphic to $D^{\{0\}}_{\mathfrak{D}\{0\}}$ and it is generated by $\sum_i f_2(n_i,t_i)$, (3.6′) requires $\Phi(h_0)=0$.

(b) Let A_ω be isomorphic to $D^{\{1\}}_{\mathfrak{D}\{1\}}$. Suppose $\gamma\neq 0$ and take $f_1\in L(\mathcal{G},\mathfrak{A})$ ω-orthogonal to $\mathfrak{so}(3)$, given by $f_1(r_i)=0$, $f_1(n_j)=0$, $f_1(t_k)=-(1/\gamma)f_2(t_k,h_0)$, $f_1(h_0)=0$ for all $i,j,k\in\{1,2,3\}$. As

$$(\delta_1 f_1)(t_k,h_0)=f_2(t_k,h_0),$$

we can then choose f_2 in such a way that

$$f_2(t,h)=0 \quad \text{for all } t\in\mathfrak{T}(3),\ h\in\mathfrak{K}. \qquad (3.10)$$

Besides, on account of (3.4′), (3.5′), (3.6″), and (3.9), we get

$$\gamma\sum_{j,k}\epsilon_{ijk}f_2(n_j,n_k)=-2\sum_{j,k}\epsilon_{ijk}f_2(n_j,t_k), \qquad (3.11)$$

$$\gamma\sum_{j,k}\epsilon_{ijk}f_2(t_j,t_k)=0, \qquad (3.12)$$

$$\gamma\sum_{j,k}\epsilon_{ijk}f_2(n_j,t_k)=-\sum_{j,k}\epsilon_{ijk}f_2(t_j,t_k). \qquad (3.13)$$

If (3.10) is satisfied, (3.11)–(3.13) require $f_2=0$. As $f_2\neq 0$, $\gamma=0$ and $\sum_{j,k}\epsilon_{ijk}f_2(n_j,t_k)=\sum_{j,k}\epsilon_{ijk}\times f_2(t_j,t_k)=0$ for all $i\in\{1,2,3\}$.

From the previous considerations we can infer that $(\mathcal{E},\rho)$ is an essential primitive extension if and only if one of the following possibilities is satisfied.

(1) A_ω is isomorphic to $D^{\{2\}}_{\mathfrak{D}\{2\}}$ and it is generated by $\{b_i\}_{1\le i\le 5}$;

(2) A_ω is isomorphic to $D^{\{0\}}_{\mathfrak{D}\{0\}}$ and it is generated by $a_0=\sum_i f_2(n_i,t_i)$;

(3′) A_ω, isomorphic to $D^{\{1\}}_{\mathfrak{D}\{1\}}$, is generated by $\{a_i'\}_{1\le i\le 3}$ with

$$a_i'=f_2(t_i,h_0) \qquad (3.14)$$

and

$$\sum_{j,k}\epsilon_{ijk}f_2(n_j,n_k)\neq\lambda a_i' \text{ for all } i\in\{1,2,3\},\lambda\in\mathbf{R}-\{0\};$$

(3″) A_ω is isomorphic to $D^{\{1\}}_{\mathfrak{D}\{1\}}$, it is generated by $\{a_i''\}_{1\le i\le 3}$ with

$$a_i''=\tfrac{1}{2}\sum_{j,k}\epsilon_{ijk}f_2(n_j,n_k) \qquad (3.15)$$

and

$$f_2(t_i,h_0)\neq\lambda a_i'' \text{ for all } i\in\{1,2,3\},\lambda\in\mathbf{R}-\{0\};$$

(3‴) A_ω is isomorphic to $D^{\{1\}}_{\mathfrak{D}\{1\}}$ too, and it is generated by $\{a_i'\}_{1\le i\le 3}$ with $a_i'=\lambda a_i''$ ($\lambda\in\mathbf{R},\lambda\neq 0$), where a_i' and a_i'' are given by (3.14), resp. (3.15).

In the cases (1) and (2)

$$H^2(\mathcal{G},\mathfrak{so}(3),A_\omega)=\{\lambda\underline{f}_2\,|\,\lambda\in\mathbf{R}\}$$

(with the corresponding A_ω and f_2), by Schur's lemma. $\overset{\{2\}}{\circledcirc}$, resp. $\overset{\{0\}}{\circledcirc}$, are the diagrams of any essential extension of $\mathcal{G}$ by such an A_ω. Letting now A_ω be as in (3‴), we define $f_2',f_2''\in Z^2(\mathcal{G},\mathfrak{so}(3),A_\omega)$ such that $f_2'(t_i,h_0)=a_i'$, $\frac{1}{2}\sum_{j,k}\epsilon_{ijk}f_2''(n_j,n_k)=a_i''$, and vanishing for all other basis elements of $\mathcal{G}\times\mathcal{G}$. Then

$$H^2(\mathcal{G},\mathfrak{so}(3),A_\omega)=\{\lambda'\underline{f}_2'+\lambda''\underline{f}_2''\,|\,\lambda',\lambda''\in\mathbf{R}\}.$$

The diagrams of the essential extensions of $\mathcal{G}$ by A_ω are in this case $\overset{\{1\}'}{\circledcirc}$, $\overset{\{1\}''}{\circledcirc}$, and $\overset{\{1\}'''}{\circledcirc}$, which, respectively, represent the possibilities (3′), (3″) and (3‴). On the other hand, it must be noted that, for diagrams of nonprimitive extensions, the vertices $\overset{\{1\}\text{iv}}{\circledcirc}$ and $\overset{\{1\}\text{v}}{\circledcirc}$, with bases $\{\sum_{j,k}\epsilon_{ijk}f_2(n_j,t_k)\}_{1\le i\le 3}$, resp. $\{\frac{1}{2}\sum_{j,k}\epsilon_{ijk}f_2(t_j,t_k)\}_{1\le i\le 3}$, are not, in general, ruled out.

Bargmann's superselection rule of the nonrelativistic mass[13] is a well-known consequence of the existence of central essential extensions of $\mathcal{G}$ by a one-dimensional Lie algebra $\mathfrak{A}$. On the other hand, we know that all central extensions of $\mathcal{P}$ by such an $\mathfrak{A}$ are trivial (cf. Sec. 2) and that the transition from a relativistic to a nonrelativistic theory involves the

Inönü–Wigner contraction of the Poincaré group in the Galilean one.[14] As this group contraction induces a Lie algebra contraction of $\mathcal{P}$ in $\mathcal{G}$, it is natural to ask the following question: How can it be that a trivial extension of $\mathcal{P}$ by $\mathfrak{A}$ goes in an essential extension of $\mathcal{G}$ if we contract $\mathcal{P}$ in $\mathcal{G}$? Saletan[15] has solved this problem in the general case of central extensions of an arbitrary Lie algebra $\mathfrak{B}$ by $\mathfrak{A}$ showing that if $\mathfrak{B}'$ is a contraction of $\mathfrak{B}$, $f_2 \in B^2(\mathfrak{B}, A_0)$ may go in a $(2;\mathfrak{B}', A_0^{(0)})$-cocycle $f_2^{(0)}$ which is not a coboundary. In the case of the Inönü–Wigner contraction of $\mathcal{P}$ in $\mathcal{G}$, this goes as follows. Let λ be a real parameter with $0 < \lambda \leq 1$, and consider a one-parameter family of isomorphisms $\tau(\lambda): P \rightarrowtail\!\!\!\!\twoheadrightarrow P$ given, in the basis of Sec. 2, by

$$\tau(\lambda)l_{ij} = l_{ij}, \quad \tau(\lambda)l_{i0} = \lambda l_{i0} \quad \tau(\lambda)t_i = t_i,$$
$$\tau(\lambda)t_0 = (1/\lambda)t_0 \quad (i,j \in \{1,2,3\}).$$

Then there is a corresponding one-parameter family of isomorphic Lie algebras $\mathcal{P}^{(\lambda)}$ with $P^{(\lambda)} = P$ and $\mathcal{P}^{(1)} = \mathcal{P}$, which is defined as follows. Let $p^{(\lambda)}$ denote $p \in P$ when it is considered as an element of $\mathcal{P}^{(\lambda)}$. The Lie multiplication $[\ ,\ ]^{(\lambda)}$ of $\mathcal{P}^{(\lambda)}$ is given by

$$[p^{(\lambda)}, p'^{(\lambda)}]^{(\lambda)} = \tau(\lambda)^{-1}[\tau(\lambda)p, \tau(\lambda)p'].$$

If we define a Lie multiplication $[\ ,\ ]^{(0)}$ by

$$[p^{(0)}, p'^{(0)}]^{(0)} = \lim_{\lambda \to 0}[p^{(\lambda)}, p'^{(\lambda)}]^{(\lambda)},$$

we get the Lie algebra $\mathcal{P}^{(0)} = \mathcal{G}$. From the physical point of view there are only two different cases:

(a) $0 < \lambda \leq 1$: relativistic theory, where $1/\lambda = c$ gives the velocity of light in vacuum in different units;

(b) $\lambda = 0$: nonrelativistic theory with $c \to \infty$.

Now, take $f_1 \in C^1(\mathcal{P}, A_0)$ such that $\delta_1 f_1 = f_2 \neq 0$. We choose $f_2(l_{ij}, l_{km}) = 0$, $f_2(l_{ij}, l_{k0}) = 0$, $f_2(l_{ij}, t_k) = 0$, $f_2(l_{i0}, t_i) \neq 0$ for all $i,j,k,m \in \{1,2,3\}$. Let $A_0^{(\lambda)}$ be the trivial $\mathcal{P}^{(\lambda)}$-module with underlying vector space A, and let $f_1^{(\lambda)} \in C^1(\mathcal{P}^{(\lambda)}, A_0^{(\lambda)})$, $f_2^{(\lambda)} \in C^2(\mathcal{P}^{(\lambda)}, A_0^{(\lambda)})$ be given by

$$f_1^{(\lambda)}(p^{(\lambda)}) = (1/\lambda^2)f_1(p) \quad (0 < \lambda \leq 1),$$
$$f_2^{(\lambda)}(p^{(\lambda)}, p'^{(\lambda)}) = (\delta_1(f_1^{(\lambda)} \circ \tau(\lambda)^{-1}))(\tau(\lambda)p, \tau(\lambda)p') \quad (0 < \lambda \leq 1),$$
$$f_2^{(0)} = \lim_{\lambda \to 0} f_2^{(\lambda)}.$$

Obviously, $f_2^{(\lambda)} \in Z^2(\mathcal{P}^{(\lambda)}, A_0^{(\lambda)})$ for all $\lambda \neq 0$, and Saletan[15] proved that this is also true if $\lambda = 0$. We get then

$$f_2^{(0)}(l_{\mu\nu}^{(0)}, l_{\rho\sigma}^{(0)}) = 0 \quad \text{for all } \mu,\nu,\rho,\sigma \in \{0,1,2,3\},$$
$$f_2^{(0)}(l_{ij}^{(0)}, t_k^{(0)}) = 0, \quad f_2^{(0)}(l_{ij}^{(0)}, t_0^{(0)}) = 0,$$
$$f_2^{(0)}(l_{i0}^{(0)}, t_0^{(0)}) = 0,$$

and

$$f_2^{(0)}(l_{i0}^{(0)}, t_j^{(0)}) = -g_{ij} f_1(t_0) \quad \text{for all } i,j,k \in \{1,2,3\}.$$

Furthermore, $f_2^{(0)} \notin B^2(\mathcal{G}, A_0^{(0)})$ because $(\delta_1 h_1^{(0)})(l_{i0}^{(0)}, t_i^{(0)}) = 0$ for all $h_1^{(0)} \in C^1(\mathcal{G}, A_0^{(0)})$.

The five one-vertex diagrams of the essential primitive extensions of $\mathcal{G}$ considered above do not fill the set of diagrams of extensions belonging to $\mathfrak{E}_1(\mathcal{G})$. However, they are all the connected ones (those which picture connected graphs). Obviously, the disconnected diagrams of extensions belonging to $\mathfrak{E}_1(\mathcal{G})$ are obtained by combination of the connected ones, and their maximal number of vertices is four, since $\{1\}'''$ ◎ may never be combined with $\{1\}'$ ◎ or (and) $\{1\}''$ ◎.

All extensions belonging to $\mathfrak{E}_2(\mathcal{G})$ have Abelian kernel, notwithstanding the fact that $\mathcal{R}_2$ is not Abelian. In fact, the fundamental $\mathcal{SO}(3)$-module $A(\varphi, f_2)$ of an ω-sliced extension of $\mathcal{G}$ by an arbitrary Lie algebra $\mathfrak{A}$ satisfies

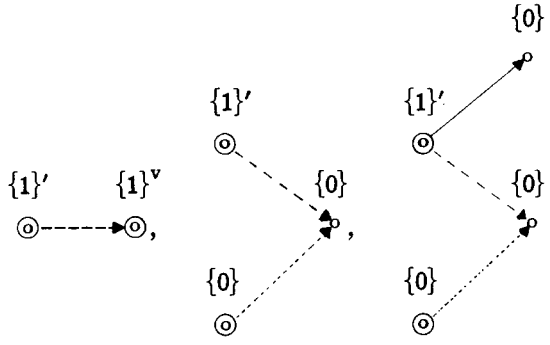
$$[A(\varphi,f_2), A(\varphi,f_2)] \subseteq \varphi^2(\mathcal{R}_2)A(\varphi,f_2)$$

because of (3.1). Following the proof of Theorem 3 of I, we get

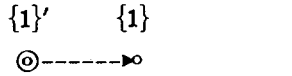
$$[\varphi^i(\mathcal{R}_2)A(\varphi,f_2), \varphi^j(\mathcal{R}_2)A(\varphi,f_2)] \subseteq \varphi^{i+j+2}(\mathcal{R}_2)A(\varphi,f_2).$$

Hence, in order that $(\mathcal{E},\rho) \in \mathfrak{E}_2(\mathcal{G})$, $\mathfrak{A}$ must be Abelian on account of Theorem 5 of I. Diagrams of irreducible extensions can be drawn as in the case of extensions of $\mathcal{E}(3)$ and $\mathcal{P}$, but we have now to use three kinds of arrows, ⟶, - - ›, and ⋯›, picturing, respectively, the action of $\varphi(\mathcal{T}(3))$, $\varphi(\mathfrak{N}(3))$, and $\varphi(\mathcal{K})$.

For example,



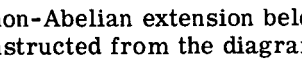
are diagrams of extensions belonging to $\mathfrak{E}_2(\mathcal{G})$. They are supposed to satisfy the relations (3.2′)–(3.8‴). However, there is no extension of $\mathcal{G}$ with diagram



(3.16)

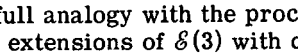
since (3.6″) is never satisfied by (3.16). This shows how much care is needed when one draws diagrams of extensions of $\mathcal{G}$, in order to satisfy (3.2′)–(3.8‴).

A non-Abelian extension belonging to $\mathfrak{E}_3(\mathcal{G})$ can be constructed from the diagram



in full analogy with the procedure shown in Sec. 1 for the extensions of $\mathcal{E}(3)$ with diagram (1.6).

Now consider an extension $(\mathcal{E},\rho)$ of $\mathcal{G}$ by $\mathfrak{A}$ with diagram



(3.17)

Taking again an ω-sliced extension related to $(\mathcal{E},\rho)$ with fundamental $\mathcal{SO}(3)$-module $A(\varphi,f_2)$, we see from (3.17) that

$$A_\omega = A(\varphi,f_2) \oplus A'_{\omega'},$$

where $A(\varphi,f_2)$ and $A'_{\omega'}$ are isomorphic to $D^{\{1\}}_{\mathfrak{D}\{1\}}$. $A(\varphi,f_2)$ is generated by $\{a_i\}_{1\leqslant i\leqslant 3}$ with $a_i = f_2(t_i,h_0)$, and $\{a'_j\}_{1\leqslant j\leqslant 3}$ with $a'_j = \varphi(h_0)f_2(t_j,h_0)$ is a basis of $A'_{\omega'}$. Moreover, $\varphi(\mathfrak{N}(3))A(\varphi,f_2) = \varphi(\mathcal{T}(3))A(\varphi,f_2) = \varphi(\mathcal{T}(3))A'_{\omega'} = \varphi(\mathfrak{N}(3))A'_{\omega'} = \{0\}$. As $\varphi(h_0)\neq 0$, we conclude that, for any $i\in\{1,2,3\}$,

$$\varphi(h_0)a'_i = \lambda a'_i \quad (\lambda\in\mathbf{R},\ \lambda\neq 0)$$

by Schur's lemma.

Hence the radical of $\mathcal{E}$ is not nilpotent, and $(\mathcal{E},\rho)$ belongs to $\mathfrak{E}_\infty(\mathcal{G})$. Unlike $\mathfrak{E}_\infty(\mathcal{E}(3))$ and $\mathfrak{E}_\infty(\mathcal{P})$, $\mathfrak{E}_\infty(\mathcal{G})$ is not empty. Actually $\mathfrak{A}$ is an Abelian Lie algebra, because of Theorems 3 and 5 of I and since the Lie multiplication is a bilinear alternating map.

4. EXTENSIONS BY $\mathcal{E}(3)$, $\mathcal{P}$, AND $\mathcal{G}$

Let $\mathfrak{A}$ mean $\mathcal{E}(3)$, $\mathcal{P}$, or $\mathcal{G}$. Suppose $\mathfrak{B}$ is an arbitrary Lie algebra, and let $\Phi\in\hom(\mathfrak{B},\Delta(\mathfrak{A}))$. If Ψ is the central character determined by Φ, we get $H^2(\mathfrak{B},C(\mathfrak{A})_\Psi) = \{0\}$ since $\mathcal{C}(\mathfrak{A}) = \{0\}$. Hence, on account of the bijection (1.5) of I, there is only one equivalence class of extensions of $\mathfrak{B}$ by $\mathfrak{A}$ with character Φ, or $\mathrm{ext}(\mathfrak{B},\mathfrak{A},\Phi) = \emptyset$. In the Appendix we will show that $(\mathrm{D}(\mathfrak{A}),\Pi_{\mathfrak{A}})$ is an inessential extension of $\Delta(\mathfrak{A})$ by $\mathrm{I}(\mathfrak{A})$. Therefore, for arbitrary $\mathfrak{B}$ and arbitrary $\Phi\in\hom(\mathfrak{B},\Delta(\mathfrak{A}))$, $\mathrm{Obs}(\mathfrak{B},\mathfrak{A},\Phi) = 0$, there is one and only one equivalence class of extensions of $\mathfrak{B}$ by $\mathfrak{A}$ with character Φ, and all these extensions are inessential (cf. Sec. 2 of Ref. 8).

ACKNOWLEDGMENTS

It is a pleasure to acknowledge the benefit of a number of helpful conversations with Professor A. Janner, Professor G. Emch, and Dr. T. Janssen.

APPENDIX: THE LIE ALGEBRAS $\mathrm{D}(\mathcal{E}(3))$, $\mathrm{D}(\mathcal{P})$, AND $\mathrm{D}(\mathcal{G})$

A Lie algebra $\mathfrak{A}$ being given, define $D_a\in\mathrm{I}(\mathfrak{A})$ $(a\in\mathfrak{A})$ by

$$D_a a' = [a,a'] \quad \text{for all } a'\in\mathfrak{A}.$$

If $\mathfrak{A}\neq\{0\}$ and $\{a_i\}$ is a basis of $\mathfrak{A}$, then $\{D_{a_i}\}$ is a basis of $\mathrm{I}(\mathfrak{A})$ because

$$[D_{a_1},D_{a_2}] = D_{[a_1,a_2]} \quad \text{for all } a_1,a_2\in\mathfrak{A}. \tag{A1}$$

Letting $\mathfrak{A}$ be $\mathcal{E}(3)$, $\mathcal{P}$, or $\mathcal{G}$ and considering the bases given in Sec. 1, resp. 2 and 3, we look for a basis of $\mathrm{D}(\mathfrak{A})$ which embeds $\{D_{a_i}\}$.

1. $\mathrm{D}(\mathcal{E}(3))$

If $f_1\in Z^1(\mathcal{E}(3),\mathcal{SO}(3),E(3)_{\mathrm{ad}})$ is given by $f_1(r_i)=0$, $f_1(t_j)=t_j$ for all $i,j\in\{1,2,3\}$, we get

$$H^1(\mathcal{E}(3),E(3)_{\mathrm{ad}}) = \{\lambda\underline{f}_1 \mid \lambda\in\mathbf{R}\}. \tag{A2}$$

Hence $\Delta(\mathcal{E}(3))$ is a one-dimensional Lie algebra.

According to (A2), we can define a derivation D of $\mathcal{E}(3)$ by $Dr_i = 0$, $Dt_j = t_j$, getting a basis $\{D_{r_i},D_{t_j},D\}_{1\leqslant i,j\leqslant 3}$ of $\mathrm{D}(\mathcal{E}(3))$. The basis elements have Lie products given by (A1) and by

$$[D_{r_i},D] = 0, \qquad [D_{t_j},D] = -D_{t_j}.$$

$(\mathrm{D}(\mathcal{E}(3)),\Pi_{\mathcal{E}(3)})$ is an inessential extension of $\Delta(\mathcal{E}(3))$ by $\mathrm{I}(\mathcal{E}(3))$.

2. $\mathrm{D}(\mathcal{P})$

After the appropriate change in notation, in particular the substitution of the bases $\{r_i\}_{1\leqslant i\leqslant 3}$ and $\{t_j\}_{1\leqslant j\leqslant 3}$, respectively, by $\{l_{\mu\nu}\}_{0\leqslant\mu,\nu\leqslant 3}$ and $\{t_\rho\}_{0\leqslant\rho\leqslant 3}$, all results obtained in the case of $\mathrm{D}(\mathcal{E}(3))$ are valid for $\mathrm{D}(\mathcal{P})$ too.[12,16,17]

3. $\mathrm{D}(\mathcal{G})$

Let $f_1,f'_1\in Z^1(\mathcal{G},\mathcal{SO}(3),G_{\mathrm{ad}})$ be such that

$$f_1(r_i)=0,\quad f_1(n_j)=n_j,\quad f_1(t_k)=0,\quad f_1(h_0)=-h_0,$$
$$f'_1(r_i)=0,\quad f'_1(n_j)=0,\quad f'_1(t_k)=t_k,\quad f'_1(h_0)=h_0,$$
$$\text{for all } i,j,k\in\{1,2,3\}.$$

Then

$$H^1(\mathcal{G},G_{\mathrm{ad}}) = \{\lambda\underline{f}_1 + \lambda'\underline{f}'_1 \mid \lambda,\lambda'\in\mathbf{R}\}. \tag{A3}$$

On account of (A3) we can define $D,D'\in\mathrm{D}(\mathcal{G})$ by

$$Dr_i = 0,\quad Dn_j = n_j,\quad Dt_k = 0,\quad Dh_0 = -h_0,$$
$$D'r_i = 0,\quad D'n_j = 0,\quad D't_k = t_k,\quad D'h_0 = h_0.$$

$\{D_{r_i},D_{n_j},D_{t_k},D_{h_0},D,D'\}_{1\leqslant i,j,k\leqslant 3}$ is a basis of $\mathrm{D}(\mathcal{G})$ with the Lie products given by (A1) and by

$$[D_{r_i},D] = 0,\quad [D_{n_j},D] = -D_{n_j},\quad [D_{t_k},D] = 0,$$
$$[D_{h_0},D] = D_{h_0},\quad [D_{r_i},D'] = 0,\quad [D_{n_j},D'] = 0,$$
$$[D_{t_k},D'] = -D_{t_k},\quad [D_{h_0},D'] = -D_{h_0},$$
$$[D,D'] = 0.$$

$\Delta(\mathcal{G})$ is a two-dimensional Abelian Lie algebra, and $(\mathrm{D}(\mathcal{G}),\Pi_{\mathcal{G}})$ is an inessential extension of $\Delta(\mathcal{G})$ by $\mathrm{I}(\mathcal{G})$.

---

* On leave of absence from Institut de Physique Théorique, Université Catholique, Louvain (Belgium). Postal address: Instituut voor Theoretische Fysika, Katholieke Universiteit, Driehuizerweg 200, Nijmegen (Netherlands).

1 U. Cattaneo, J. Math. Phys. **13**, 504 (1972).
2 M. A. Naïmark, *Les représentations linéaires du groupe de Lorentz* (Dunod, Paris, 1962).
3 G. Hochschild, *The Structure of Lie Groups* (Holden-Day, San Francisco, 1965).
4 J. Tits, *Tabellen zu den einfachen Lie Gruppen und ihren Darstellungen*, Lecture Notes in Mathematics, Vol. 40 (Springer-Verlag, Heidelberg, 1967).
5 E. P. Wigner, *Group Theory and Its Application to the Quantum Mechanics of Atomic Spectra* (Academic, New York, 1959).
6 H. Freudenthal and H. De Vries, *Linear Lie Groups* (Academic, New York, 1969).
7 U. Cattaneo, Commun. Math. Phys. **13**, 226 (1969).
8 U. Cattaneo, Commun. Math. Phys. **20**, 220 (1971).
9 C. Chevalley, *Théorie des groupes de Lie (Groupes algébriques—Théorèmes généraux sur les algèbres de Lie)* (Hermann, Paris, 1968).
10 C. Berge, *Théorie des graphes et ses applications* (Dunod, Paris, 1967).
11 F. Harary, *Graph Theory* (Addison-Wesley, Reading, Mass., 1969).

[12] A. Galindo, J. Math. Phys. **8**, 768 (1967).
[13] V. Bargmann, Ann. Math. **59**, 1 (1954).
[14] E. Inönü and E. P. Wigner, Proc. Natl. Acad. Sci. (U.S.) **39**, 510 (1953).
[15] E. J. Saletan, J. Math. Phys. **2**, 1 (1961).
[16] L. Michel, "The Group of Automorphisms of the Poincaré Group," in *Lectures in Theoretical Physics*, edited by W. E. Brittin and A. O. Barut (University of Colorado Press, Boulder, 1965).
[17] L. Michel, "Relations entre symétries internes et invariance relativiste," in *Application of Mathematics to Problems in Theoretical Physics*, edited by F. Lurçat (Gordon and Breach, New York, 1967).

Generalization of Euler Angles to N-Dimensional Orthogonal Matrices*

David K. Hoffman, Richard C. Raffenetti, and Klaus Ruedenberg
Department of Chemistry and Institute for Atomic Research, Iowa State University, Ames, Iowa 50010
(Received 22 October 1971)

An algorithm is presented whereby an N-dimensional orthogonal matrix can be represented in terms of $\frac{1}{2}N(N-1)$ *independent* parameters $\theta_k^{(\nu)}[\nu = 2, 3, \ldots, N; k = 1, 2, \ldots, (\nu - 1)]$. The parameters have the character of angles, whose compact domains are defined in a manner such that there exists a one-to-one correspondence between the points in the parameter space and the group of orthogonal matrices. Explicit formulas are given which express all matrix elements in terms of the angles, and formulas are given which express the angles in terms of the matrix elements. Special choices of angles give block-diagonal matrices. For three-dimensional matrices, the parametrization is equivalent to that of Euler.

1. INTRODUCTION

Orthogonal transformations occur frequently in theoretical physics and theoretical chemistry. An orthogonal matrix has N^2 different elements which satisfy $N(N+1)/2$ constraints, viz. the orthonormality conditions. Thus, the number of independent variables is only $N(N-1)/2$, and sometimes it is desirable to have a convenient representation of the elements in terms of a set of *independent parameters*. In a previous communication on this subject,[1] a parametric representation was given, which could be considered as a generalization of the Eulerian angles known for the three-dimensional case.

Three questions had been left unanswered in that investigation: (1) The geometrical significance of the angular variables introduced; (2) the boundaries of the domains of these variables in the parameter-space; and (3) the algebraic inversion formulas of the angular variables in terms of the elements of the orthogonal matrix. The second of these is essential whenever an integration over the parameter space is required, as might be the case when the orthogonal matrices are considered as forming a Lie group. The third is essential in many practical situations, when initial values are given or interpolation is desired.

The treatment of the problem, developed in the present investigation, contains the answers to these outstanding questions. Thereby, the problem of the parametrization of a general orthogonal matrix is completely solved.

2. INDEPENDENT ANGULAR PARAMETERS FOR AN N-DIMENSIONAL ORTHOGONAL BASIS-TRANSFORMATION

A. Parametrization of a Unit Vector

Consider an arbitrary unit vector in an N-dimensional vector space spanned by the orthonormal basis vectors $\mathbf{e}_1, \mathbf{e}_2, \ldots, \mathbf{e}_N$. Such a unit vector can be expressed by the following parametric representation:

$$\mathbf{a} = \mathbf{e}_1 \sin\theta_1 + \cos\theta_1(\mathbf{e}_2 \sin\theta_2 + \cos\theta_2\{\mathbf{e}_3 \sin\theta_3 + \cos\theta_3[\cdots\cos\theta_{N-2}(\mathbf{e}_{N-1}\sin\theta_{N-1} + \mathbf{e}_N\cos\theta_{N-1})\cdots]\}) \tag{1}$$

in terms of $(N-1)$ arbitrary angles. This decomposition can be obtained by the following sequence of progressive projections:

$$\begin{aligned} \mathbf{f}_1 &= \mathbf{a}, \\ \mathbf{f}_k &= s_k\mathbf{e}_k + c_k\mathbf{f}_{k+1}, \quad k = 1, 2, \ldots, (N-1), \\ \mathbf{f}_N &= \mathbf{e}_N, \end{aligned} \tag{2}$$

where, for all k,

$$\langle \mathbf{f}_k | \mathbf{f}_k \rangle = 1, \quad \langle \mathbf{f}_{k+1} | \mathbf{e}_k \rangle = 0 \tag{3}$$

and

$$s_k = \sin\theta_k, \quad c_k = \cos\theta_k \tag{4}$$

with

$$\begin{aligned} \theta_k &= \text{angle between } \mathbf{f}_k \text{ and } \mathbf{f}_{k+1} \\ &= (\tfrac{1}{2}\pi) - (\text{angle between } \mathbf{f}_k \text{ and } \mathbf{e}_k). \end{aligned} \tag{5}$$

In particular, therefore,

$$\theta_N = \tfrac{1}{2}\pi, \quad s_N = 1, \quad c_N = 0. \tag{6}$$

It is apparent that $\sin\theta_k$ can be positive or negative. But by definition, $\mathbf{f}_{k+1}$ is taken in the direction of the projection of $\mathbf{f}_k$ onto the space spanned by $\mathbf{e}_{k+1}, \mathbf{e}_{k+2}, \ldots, \mathbf{e}_N$. Hence, $\cos\theta_k$ can be restricted to being positive. The only exception is θ_{N-1} because $\mathbf{f}_N = \mathbf{e}_N$ is fixed, and so $\cos\theta_{N-1}$ can be positive as well as negative. Consequently, if the domains of the angles are chosen as follows,

$$\left.\begin{aligned} &-\tfrac{1}{2}\pi \le \theta_k \le \tfrac{1}{2}\pi \quad \text{for } k = 1, 2, \ldots, (N-2), \\ &-\pi \le \theta_{N-1} < \pi, \\ &\theta_N = \tfrac{1}{2}\pi, \end{aligned}\right\} \tag{7}$$

then all possible directions of $\mathbf{a}$ are uniquely related to the angles $\theta_1 \cdots \theta_{N-1}$.

From Eq. (2) follows that the projections $\mathbf{f}_j$ can be expressed as

$$\mathbf{f}_j = \sum_{k=j}^{N} (c_j c_{j+1} \cdots c_{k-1} c_k)\left(\frac{s_k}{c_k}\right)\mathbf{e}_k. \tag{8}$$

B. Orthogonal Transformation to a Basis Containing One Arbitrary Unit Vector

With the help of this representation of $\mathbf{a}$, it is possible to construct an orthogonal basis which contains $\mathbf{a}$ as

[12] A. Galindo, J. Math. Phys. **8**, 768 (1967).
[13] V. Bargmann, Ann. Math. **59**, 1 (1954).
[14] E. Inönü and E. P. Wigner, Proc. Natl. Acad. Sci. (U.S.) **39**, 510 (1953).
[15] E. J. Saletan, J. Math. Phys. **2**, 1 (1961).
[16] L. Michel, "The Group of Automorphisms of the Poincaré Group," in *Lectures in Theoretical Physics*, edited by W. E. Brittin and A. O. Barut (University of Colorado Press, Boulder, 1965).
[17] L. Michel, "Relations entre symétries internes et invariance relativiste," in *Application of Mathematics to Problems in Theoretical Physics*, edited by F. Lurçat (Gordon and Breach, New York, 1967).

Generalization of Euler Angles to N-Dimensional Orthogonal Matrices*

David K. Hoffman, Richard C. Raffenetti, and Klaus Ruedenberg
Department of Chemistry and Institute for Atomic Research, Iowa State University, Ames, Iowa 50010
(Received 22 October 1971)

An algorithm is presented whereby an N-dimensional orthogonal matrix can be represented in terms of $\frac{1}{2}N(N-1)$ *independent* parameters $\theta_k^{(\nu)}[\nu = 2, 3, \ldots, N; k = 1, 2, \ldots, (\nu - 1)]$. The parameters have the character of angles, whose compact domains are defined in a manner such that there exists a one-to-one correspondence between the points in the parameter space and the group of orthogonal matrices. Explicit formulas are given which express all matrix elements in terms of the angles, and formulas are given which express the angles in terms of the matrix elements. Special choices of angles give block-diagonal matrices. For three-dimensional matrices, the parametrization is equivalent to that of Euler.

1. INTRODUCTION

Orthogonal transformations occur frequently in theoretical physics and theoretical chemistry. An orthogonal matrix has N^2 different elements which satisfy $N(N+1)/2$ constraints, viz. the orthonormality conditions. Thus, the number of independent variables is only $N(N-1)/2$, and sometimes it is desirable to have a convenient representation of the elements in terms of a set of *independent parameters*. In a previous communication on this subject,[1] a parametric representation was given, which could be considered as a generalization of the Eulerian angles known for the three-dimensional case.

Three questions had been left unanswered in that investigation: (1) The geometrical significance of the angular variables introduced; (2) the boundaries of the domains of these variables in the parameter-space; and (3) the algebraic inversion formulas of the angular variables in terms of the elements of the orthogonal matrix. The second of these is essential whenever an integration over the parameter space is required, as might be the case when the orthogonal matrices are considered as forming a Lie group. The third is essential in many practical situations, when initial values are given or interpolation is desired.

The treatment of the problem, developed in the present investigation, contains the answers to these outstanding questions. Thereby, the problem of the parametrization of a general orthogonal matrix is completely solved.

2. INDEPENDENT ANGULAR PARAMETERS FOR AN N-DIMENSIONAL ORTHOGONAL BASIS-TRANSFORMATION

A. Parametrization of a Unit Vector

Consider an arbitrary unit vector in an N-dimensional vector space spanned by the orthonormal basis vectors $\mathbf{e}_1, \mathbf{e}_2, \ldots, \mathbf{e}_N$. Such a unit vector can be expressed by the following parametric representation:

$$\mathbf{a} = \mathbf{e}_1 \sin\theta_1 + \cos\theta_1(\mathbf{e}_2 \sin\theta_2 + \cos\theta_2\{\mathbf{e}_3 \sin\theta_3 + \cos\theta_3[\cdots\cos\theta_{N-2}(\mathbf{e}_{N-1}\sin\theta_{N-1} + \mathbf{e}_N\cos\theta_{N-1})\cdots]\}) \tag{1}$$

in terms of $(N-1)$ arbitrary angles. This decomposition can be obtained by the following sequence of progressive projections:

$$\begin{aligned} \mathbf{f}_1 &= \mathbf{a}, \\ \mathbf{f}_k &= s_k\mathbf{e}_k + c_k\mathbf{f}_{k+1}, \quad k = 1, 2, \ldots, (N-1), \\ \mathbf{f}_N &= \mathbf{e}_N, \end{aligned} \tag{2}$$

where, for all k,

$$\langle \mathbf{f}_k | \mathbf{f}_k \rangle = 1, \quad \langle \mathbf{f}_{k+1} | \mathbf{e}_k \rangle = 0 \tag{3}$$

and

$$s_k = \sin\theta_k, \quad c_k = \cos\theta_k \tag{4}$$

with

$$\begin{aligned} \theta_k &= \text{angle between } \mathbf{f}_k \text{ and } \mathbf{f}_{k+1} \\ &= (\tfrac{1}{2}\pi) - (\text{angle between } \mathbf{f}_k \text{ and } \mathbf{e}_k). \end{aligned} \tag{5}$$

In particular, therefore,

$$\theta_N = \tfrac{1}{2}\pi, \quad s_N = 1, \quad c_N = 0. \tag{6}$$

It is apparent that $\sin\theta_k$ can be positive or negative. But by definition, $\mathbf{f}_{k+1}$ is taken in the direction of the projection of $\mathbf{f}_k$ onto the space spanned by $\mathbf{e}_{k+1}, \mathbf{e}_{k+2}, \ldots, \mathbf{e}_N$. Hence, $\cos\theta_k$ can be restricted to being positive. The only exception is θ_{N-1} because $\mathbf{f}_N = \mathbf{e}_N$ is fixed, and so $\cos\theta_{N-1}$ can be positive as well as negative. Consequently, if the domains of the angles are chosen as follows,

$$\left.\begin{aligned} &-\tfrac{1}{2}\pi \le \theta_k \le \tfrac{1}{2}\pi \quad \text{for } k = 1, 2, \ldots, (N-2), \\ &-\pi \le \theta_{N-1} < \pi, \\ &\theta_N = \tfrac{1}{2}\pi, \end{aligned}\right\} \tag{7}$$

then all possible directions of $\mathbf{a}$ are uniquely related to the angles $\theta_1 \cdots \theta_{N-1}$.

From Eq. (2) follows that the projections $\mathbf{f}_j$ can be expressed as

$$\mathbf{f}_j = \sum_{k=j}^{N} (c_j c_{j+1} \cdots c_{k-1} c_k)\left(\frac{s_k}{c_k}\right)\mathbf{e}_k. \tag{8}$$

B. Orthogonal Transformation to a Basis Containing One Arbitrary Unit Vector

With the help of this representation of $\mathbf{a}$, it is possible to construct an orthogonal basis which contains $\mathbf{a}$ as

one of its basis vectors. It is the new orthonormal basis set $\mathbf{b}_1, \mathbf{b}_2, \ldots, \mathbf{b}_N$, whose vectors $\mathbf{b}_\nu$ are defined by

$$\mathbf{b}_\nu = (c_1 c_2 \cdots c_{\nu-1})^{-1} \partial_\nu \mathbf{a}, \qquad \nu = 1, 2, \ldots, (N-1), \tag{9}$$

$$\mathbf{b}_N = \mathbf{a}, \tag{10}$$

where ∂_ν denotes $\partial/\partial\theta_\nu$. Before proving this contention, we note that, according to Eq. (2),

$$\partial_\nu \mathbf{a} = (c_1 c_2 \ldots c_{\nu-1}) \partial_\nu (s_\nu \mathbf{e}_\nu + c_\nu \mathbf{f}_{\nu+1}),$$

whence

$$\mathbf{b}_\nu = c_\nu \mathbf{e}_\nu - s_\nu \mathbf{f}_{\nu+1}, \qquad \nu = 1, 2, \ldots, (N-1). \tag{11}$$

As for the proof, we observe that, for $\mu < \nu$,

$$\begin{aligned} \partial_\mu \partial_\nu \mathbf{a} &= [\partial_\mu (c_1 c_2 \cdots c_{\nu-1})][\partial_\nu (s_\nu \mathbf{e}_\nu + e_\nu \mathbf{f}_{\nu+1})] \\ &= (-s_\mu / c_\mu) \partial_\nu \mathbf{a}. \end{aligned} \tag{12}$$

Now since $\mathbf{a}$ is normalized, one has

$$(\mathbf{a} \cdot \partial_\nu \mathbf{a}) = \tfrac{1}{2} \partial_\nu (\mathbf{a}^2) = 0, \tag{13}$$

which shows that $(\mathbf{b}_N \cdot \mathbf{b}_\nu) = 0$ for $\nu \neq N$. By virtue of Eqs. (12) and (13), one obtains also

$$(\mathbf{a} \cdot \partial_\mu \partial_\nu \mathbf{a}) = (-s_\mu / c_\mu)(\mathbf{a} \cdot \partial_\nu \mathbf{a}) = 0,$$

whence

$$(\partial_\mu \mathbf{a}) \cdot (\partial_\nu \mathbf{a}) = \partial_\mu [(\mathbf{a} \cdot \partial_\nu \mathbf{a})] - (\mathbf{a} \cdot \partial_\mu \partial_\nu \mathbf{a}) = 0 \tag{14}$$

showing that $(\mathbf{b}_\mu \cdot \mathbf{b}_\nu) = 0$ for $\mu \neq \nu$, $\mu \neq N$, $\nu \neq N$, which demonstrates the contention.

From Eqs. (10) and (11) the transformation to the new basis can be written

$$\mathbf{b}_k = \sum_{i=1}^{N} \mathbf{e}_i A_{ik}. \tag{15}$$

The matrix $\mathbf{A}$ has the form

$$\mathbf{A} = \left(\begin{array}{c|c} \text{I} & \text{II} \\ \hline \text{III} & \text{IV} \end{array} \right)$$

and, in the four indicated regions, the elements are

Region I: $A_{ii} = c_i, i = 1, 2, \ldots, (N-1),$ (16)

Region II: $A_{iN} = (c_1 c_2 \cdots c_i)(s_i / c_i), \quad i = 1, 2, \ldots, N,$ (17)

Region III: $A_{ik} = -(c_k c_{k+1} \cdots c_{i-1} c_i)(s_i s_k / c_i c_k), \quad i > k,$ (18)

Region IV: $A_{ik} = 0, \quad i < k < N.$ (19)

According to Eqs. (10), (11) and Eq. (1), the particular parameter choice $\theta_1 = \theta_2 = \cdots = \theta_{N-1} = 0$ yields

$$\mathbf{b}_\nu = \mathbf{e}_\nu \quad \text{for } \nu = 1, 2, \ldots, (N-1)$$

$$\mathbf{b}_N = \mathbf{a} = \mathbf{f}_1 = \mathbf{f}_2 = \cdots = \mathbf{f}_N = \mathbf{e}_N,$$

whence

$$A_{ik}(\text{all } \theta_k = 0) = \delta_{ik}. \tag{20}$$

Since all transformations considered here can be obtained by continuously varying the θ_k from zero to their respective finite values, it follows that $\det(\mathbf{A}) = 1$ (and not -1), i.e., $\mathbf{A}$ always represents a proper rotation.

C. Transformation to a Basis Containing N Arbitrary Orthogonal Unit Vectors

In order to construct the transformation from the original basis $\mathbf{e}_1 \mathbf{e}_2 \cdots \mathbf{e}_N$ to an arbitrary new orthonormal basis $\mathbf{a}_1 \mathbf{a}_2, \ldots, \mathbf{a}_N$ we proceed as follows.

Consider first the vector $\mathbf{a}_N$; identify it with the vector $\mathbf{a}$ of the previous section; and form a new basis $\mathbf{b}_1^{(N)}, \mathbf{b}_2^{(N)}, \ldots, \mathbf{b}_{N-1}^{(N)}, \mathbf{b}_N^{(N)} = \mathbf{a}_N$, with the help of a transformation $\mathbf{A}^{(N)}$ of the form defined in Eqs. (15)–(19). Denote the angular variables by $\theta_i^{(N)}$.

Next consider the vector $\mathbf{a}_{N-1}$. It lies in the space spanned by $\mathbf{b}_1^{(N)}, \mathbf{b}_2^{(N)}, \ldots, \mathbf{b}_{N-1}^{(N)}$, and the method of the previous sections can now be used to construct a new basis $\mathbf{b}_1^{(N-1)}, \mathbf{b}_2^{(N-1)}, \ldots, \mathbf{b}_{N-2}^{(N-1)}, \mathbf{b}_{N-1}^{(N-1)} = \mathbf{a}_{N-1}, \mathbf{b}_N^{(N-1)} = \mathbf{a}_N$ by an $(N-1)$ dimensional transformation $\mathbf{A}_{N-1}$ from $\mathbf{b}_1^{(N)} \cdots \mathbf{b}_{N-1}^{(N)}$ to $\mathbf{b}_1^{(N-1)} \cdots \mathbf{b}_{N-1}^{(N-1)}$ according to Eqs. (15)–(19) with new angular variables $\theta_k^{(N-1)}$.

In an analogous fashion, we treat successively the vectors $\mathbf{a}_{N-2}, \mathbf{a}_{N-3}, \ldots, \mathbf{a}_3, \mathbf{a}_2$. The general step in this sequence of transformations leads from a basis

$$\mathbf{b}_1^{(\nu+1)}, \mathbf{b}_2^{(\nu+1)}, \ldots, \mathbf{b}_\nu^{(\nu+1)}, \mathbf{b}_{(\nu+1)}^{(\nu+1)} = \mathbf{a}_{\nu+1},$$
$$\mathbf{b}_{\nu+2}^{(\nu+1)} = \mathbf{a}_{\nu+2}, \ldots, \mathbf{b}_N^{(\nu+1)} = \mathbf{a}_N,$$

to the basis

$$\mathbf{b}_1^{(\nu)}, \mathbf{b}_2^{(\nu)}, \ldots, \mathbf{b}_{\nu-1}^{(\nu)}, \mathbf{b}_\nu^{(\nu)} = \mathbf{a}_\nu, \mathbf{b}_{\nu+1}^{(\nu)} = \mathbf{a}_{\nu+1}, \ldots, \mathbf{b}_N^{(\nu)} = \mathbf{a}_N$$

by means of a transformation

$$\mathbf{b}_k^{(\nu)} = \sum_i \mathbf{b}_i^{(\nu+1)} B_{ik}^{(\nu)}, \tag{21}$$

where $\mathbf{B}^{(\nu)}$ is the matrix

$$\mathbf{B}^{(\nu)} = \left(\begin{array}{c|c} \mathbf{A}^{(\nu)} & 0 \\ \hline 0 & \mathbf{I}^{(N-\nu)} \end{array} \right). \tag{22}$$

Here $\mathbf{I}^{(N-\nu)}$ is the $(N-\nu)$-dimensional unit matrix, and $\mathbf{A}^{(\nu)}$ is a ν-dimensional matrix of the general form discussed in Eqs. (15)–(19), with the angular variables

$$\theta_i^{(\nu)}, \; s_i^{(\nu)} = \sin\theta_i^{(\nu)}, \quad c_i^{(\nu)} = \cos\theta_i^{(\nu)}, \qquad i = 1, 2, \ldots, (\nu-1). \tag{23}$$

Note that the $\theta_i^{(\nu)}$ have the domains

$$\begin{aligned} &-\tfrac{1}{2}\pi \leq \theta_i^{(\nu)} \leq \tfrac{1}{2}\pi \quad \text{for } i = 1, 2, \ldots, (\nu-2), \\ &-\pi \leq \theta_{\nu-1}^{(\nu)} < \pi, \\ &\theta_\nu^{(\nu)} = \tfrac{1}{2}\pi. \end{aligned} \tag{24}$$

The composition of the consecutive transformations eventually yields the basis $\mathbf{a}_1 \mathbf{a}_2 \cdots \mathbf{a}_N$ in terms of the basis $\mathbf{e}_1 \mathbf{e}_2 \cdots \mathbf{e}_N$ by means of the resulting transformation

$$\mathbf{a}_n = \sum_{j=1}^{N} \mathbf{e}_j T_{jn}, \qquad n = 1, \ldots, N, \tag{25}$$

where $\mathbf{T}$ is the matrix product

$$\mathbf{T} = \mathbf{B}^{(N)}\mathbf{B}^{(N-1)} \cdots \mathbf{B}^{(3)}\mathbf{B}^{(2)}. \tag{26}$$

Since $\mathbf{T}$ is orthogonal, the inverse of Eq. (25) is

$$\mathbf{e}_i = \sum_n T_{jn}\mathbf{a}_n. \tag{27}$$

Furthermore, it is readily seen from Eqs. (21), (25), and (26) that the intermediate bases $\mathbf{b}_k^{(\nu)}$ can be expressed as

$$\mathbf{b}_k^{(\nu)} = \sum_n T_{kn}^{(\nu-1)}\mathbf{a}_n \tag{28}$$

with the definitions

$$\mathbf{T}^{(\nu)} = \mathbf{B}^{(\nu)}\mathbf{B}^{(\nu-1)}\mathbf{B}^{(\nu-2)} \cdots \mathbf{B}^{(3)}\mathbf{B}^{(2)}. \tag{29}$$

3. CONSTRUCTION OF THE TRANSFORMATION MATRIX FROM THE ANGULAR PARAMETERS

From Eq. (29) follows that $\mathbf{T}$ can be constructed by the sequence of recurrence steps

$$\mathbf{T}^{(1)} = \mathbf{I}, \tag{30}$$

$$\mathbf{T}^{(\nu)} = \mathbf{B}^{(\nu)}\mathbf{T}^{(\nu-1)}, \tag{31}$$

$$\mathbf{T} = \mathbf{T}^{(N)}. \tag{32}$$

Now the recurrence step of Eq. (31) is equivalent to the basis transformation from the $\mathbf{b}_k^{(\nu)}$ to the $\mathbf{b}_k^{(\nu+1)}$. Indeed, the inverse of Eq. (21) is given by

$$\mathbf{b}_k^{(\nu+1)} = \sum_l B_{kl}^{(\nu)}\mathbf{b}_l^{(\nu)} \tag{33}$$

and, by virtue of the representations given in Eq. (28), this is equivalent to

$$T_{kn}^{(\nu)} = \sum_l B_{kl}^{(\nu)}T_{ln}^{(\nu-1)}, \tag{34}$$

which is identical with Eq. (31).

In order to obtain simple formulas for Eq. (33), we observe that the original transformation of Eq. (21) can be expressed with the help of Eqs. (2), (10), (11), which yield

$$\mathbf{b}_k^{(\nu)} = c_k^{(\nu)}\mathbf{b}_k^{(\nu+1)} - s_k^{(\nu)}\mathbf{f}_{k+1}^{(\nu+1)}, \tag{35}$$

$$\mathbf{f}_k^{(\nu+1)} = s_k^{(\nu)}\mathbf{b}_k^{(\nu+1)} + c_k^{(\nu)}\mathbf{f}_{k+1}^{(\nu+1)}, \tag{36}$$

for $k = 1, 2, \ldots, (\nu - 1)$, and

$$\mathbf{b}_\nu^{(\nu)} = \mathbf{f}_1^{(\nu+1)} = \mathbf{a}^\nu, \tag{37}$$

$$\mathbf{f}_\nu^{(\nu+1)} = \mathbf{b}_\nu^{(\nu+1)}, \tag{38}$$

$$\mathbf{b}_j^{(\nu)} = \mathbf{b}_j^{(\nu+1)} = \mathbf{a}_j, \quad j = (\nu + 1), (\nu + 2), \ldots, N. \tag{39}$$

This set of equations can be inverted to give

$$\mathbf{b}_k^{(\nu+1)} = c_k^{(\nu)}\mathbf{b}_k^{(\nu)} + s_k^{(\nu)}\mathbf{f}_k^{(\nu+1)}, \tag{40}$$

$$\mathbf{f}_{k+1}^{(\nu+1)} = -s_k^{(\nu)}\mathbf{b}_k^{(\nu)} + c_k^{(\nu)}\mathbf{f}_k^{(\nu+1)}, \tag{41}$$

for $k = 1, 2, \ldots, (\nu - 1)$, and

$$\mathbf{f}_1^{(\nu+1)} = \mathbf{b}_\nu^{(\nu)} = \mathbf{a}_\nu, \tag{42}$$

$$\mathbf{b}_\nu^{(\nu+1)} = \mathbf{f}_\nu^{(\nu+1)}, \tag{43}$$

$$\mathbf{b}_j^{(\nu+1)} = \mathbf{b}_j^{(\nu)} = \mathbf{a}_j, \quad j = (\nu + 1), \ldots, N. \tag{44}$$

Since the set of Eqs. (40)–(44) is equivalent to the recurrence relation (33), it follows that the desired recurrence relation (31), (34) can now be expressed in the form

$$T_{kn}^{(\nu)} = c_k^{(\nu)}T_{kn}^{(\nu-1)} + s_k^{(\nu)}F_{kn}^{(\nu)}, \tag{45}$$

$$F_{k+1}^{(\nu)} = -s_k^{(\nu)}T_{kn}^{(\nu-1)} + c_k^{(\nu)}F_{kn}^{(\nu)}, \tag{46}$$

for $k = 1, 2, \ldots, (\nu - 1)$, and

$$F_{1n}^{(\nu)} = T_{\nu n}^{(\nu-1)} = \delta_{\nu n}, \tag{47}$$

$$T_{\nu n}^{(\nu)} = F_{\nu n}^{(\nu)}, \tag{48}$$

$$T_{jn}^{(\nu)} = \delta_{jn}, \quad j = (\nu + 1), (\nu + 2), \ldots, N. \tag{49}$$

Here the quantities $F_{kn}^{(\nu)}$ are defined by the expansions

$$\mathbf{f}_k^{(\nu+1)} = \sum_n F_{kn}^{(\nu)}\mathbf{a}_n. \tag{50}$$

Equations (45)–(49) represent a self-contained recurrence procedure, which starts with

$$T_{kn}^{(1)} = \delta_{kn}. \tag{51}$$

Given the matrix $\mathbf{T}^{(\nu-1)}$, Eqs. (46) and (47) yield all quantities $F_{kn}^{(\nu)}$ and, with these, the elements of the matrix $\mathbf{T}^{(\nu)}$ are obtained from Eqs. (45), (48), and (49).

Thus, for a set of $N(N-1)/2$ angles $\theta_k^{(\nu)}[\nu = 2, \ldots, N; k = 1, \ldots, (\nu - 1)]$ in the domain stipulated by Eq. (24), an orthogonal matrix $\mathbf{T}$ with determinant + 1 is obtained by applying the recursive scheme (45)–(51). If, instead of Eq. (51), one uses

$$\begin{aligned} T_{11}^{(1)} &= -1, \\ T_{kn}^{(1)} &= \delta_{kn} \quad \text{for all other elements} \end{aligned} \tag{52}$$

as a starting point, then an orthogonal matrix with determinant -1 is obtained.

4. DETERMINATION OF THE ANGULAR PARAMETERS FROM THE TRANSFORMATION MATRIX

From the definition (22) of $\mathbf{B}^{(\nu)}$ and the basic recurrence relation (31), it is apparent that the matrices $\mathbf{T}^{(\nu)}$ have the structure

$$\mathbf{T}^{(\nu)} = \left(\begin{array}{c|c} \hat{\mathbf{T}}^{(\nu)} & 0 \\ \hline 0 & \mathbf{I}^{(N-\nu)} \end{array}\right), \tag{53}$$

so that the recurrence step of Eq. (31) reduces to

$$\hat{\mathbf{T}}^{(\nu)} = \mathbf{A}^{(\nu)}\begin{pmatrix} \hat{\mathbf{T}}^{(\nu-1)} & 0 \\ 0 & 1 \end{pmatrix},$$

that is,

$$T_{k\nu}^{(\nu)} = A_{k\nu}^{(\nu)}, \quad k = 1, 2, \ldots, \nu.$$

By virtue of the form of $\mathbf{A}^{(\nu)}$ given in Eqs. (16)–(19), this becomes

$$T_{k\nu}^{(\nu)} = c_1^{(\nu)}c_2^{(\nu)} \cdots c_{k-1}^{(\nu)}s_k^{(\nu)}, \quad k = 1, 2, \ldots, \nu. \tag{54}$$

These identities furnish the following recurrence

scheme for determining the angles $\theta_k^{(\nu)}$ from the matrix elements $T_{k\nu}^{(\nu)}$: Find $\theta_1^{(\nu)}\theta_2^{(\nu)}\ldots\theta_{\nu-2}^{(\nu)}$ from

$$\begin{aligned}\sin\theta_1^{(\nu)} &= T_{1\nu}^{(\nu)},\\ \sin\theta_k^{(\nu)} &= T_{k\nu}^{(\nu)}/(\cos\theta_1^{(\nu)}\cos\theta_2^{(\nu)}\cdots\cos\theta_{k-1}^{(\nu)}),\end{aligned}\tag{55}$$

and the conditions $-\frac{1}{2}\pi \le \theta_k^{(\nu)} \le \frac{1}{2}\pi$. Finally, find $\theta_{\nu-1}^{(\nu)}$ from

$$\begin{aligned}\sin\theta_{\nu-1}^{(\nu)} &= T_{\nu-1,\nu}^{(\nu)}/(\cos\theta_1^{(\nu)}\cos\theta_2^{(\nu)}\cdots\cos\theta_{\nu-2}^{(\nu)}),\\ \cos\theta_{\nu-1}^{(\nu)} &= T_{\nu\nu}^{(\nu)}/(\cos\theta_1^{(\nu)}\cos\theta_2^{(\nu)}\cdots\cos\theta_{\nu-2}^{(\nu)}),\end{aligned}\tag{56}$$

and the condition $-\pi \le \theta_{\nu-1}^{(\nu)} < \pi$. If any one of the angles $\theta_k^{(\nu)}$ should turn out to be equal $\pm(\pi/2)$ (so that $\cos\theta_k^{(\nu)} = 0$), then all subsequent angles $\theta_{k+1}^{(\nu)}\cdots\theta_{\nu-1}^{(\nu)}$ are arbitrary and can be set equal to zero.

Instead of Eqs. (55), (56), the following alternative recurrence scheme can be employed. First $\theta_{\nu-1}^{(\nu)}$ is determined from

$$\begin{aligned}\tan\theta_{\nu-1}^{(\nu)} &= T_{\nu-1,\nu}^{(\nu)}/T_{\nu\nu}^{(\nu)},\\ \text{sign }(\sin\theta_{\nu-1}^{(\nu)}) &= \text{sign}(T_{\nu-1,\nu}^{(\nu)}),\\ \text{sign }(\cos\theta_{\nu-1}^{(\nu)}) &= \text{sign}(T_{\nu\nu}^{(\nu)}),\\ -\pi \le \theta_{\nu-1}^{(\nu)} &< \pi.\end{aligned}\tag{57}$$

Then $\theta_{\nu-2}^{(\nu)}, \theta_{\nu-1}^{(\nu)}, \ldots, \theta_1^{(\nu)}$ are found from

$$\begin{aligned}\tan\theta_k^{(\nu)} &= T_{k\nu}^{(\nu)}/(T_{k+1,\nu}^{(\nu)}/\sin\theta_{k+1}^{(\nu)}),\\ \text{sign }(\sin\theta_k^{(\nu)}) &= \text{sign}(T_{k\nu}^{(\nu)}),\\ -\pi/2 \le \theta_k^{(\nu)} &\le +\pi/2.\end{aligned}\tag{58}$$

This recurrence procedure is numerically preferable when a two-argument arctangent is available which automatically finds an angle α such that

$$\begin{aligned}&\tan\alpha = N/D, \quad -\pi \le \alpha < \pi,\\ &\text{sign }(\sin\alpha) = \text{sign}(N), \quad \text{sign }(\cos\alpha) = \text{sign}(D).\end{aligned}$$

If, for a certain value k, one finds $\theta_k^{(\nu)}$ such that $T_{k+1,\nu}^{(\nu)}/s_{k+1}^{(\nu)}) = 1$, then all subsequent $T_{k-1,\nu}^{(\nu)}, T_{k-2,\nu}^{(\nu)}, \ldots, T_{1\nu}^{(\nu)}$ will vanish and one has to put $\theta_{k-1}^{(\nu)} = \theta_{k-2}^{(\nu)} = \cdots\theta_1^{(\nu)} = 0$.

Consequently, all $\frac{1}{2}N(N-1)$ angles $\theta_k^{(\nu)}$, $\nu = 2, 3, \ldots, N$, $k = 1, 2, \ldots, (\nu-1)$, representing a given orthogonal matrix $\mathbf{T}$ can be found, if all intermediate matrices $\mathbf{T}^{(2)}, \mathbf{T}^{(3)}, \ldots, \mathbf{T}^{(N)}$ are known. Now $\mathbf{T}^{(N)}$ is identical with $\mathbf{T}$, and thus known. The remaining $\mathbf{T}^{(\nu)}$, for $\nu = (N-1), (N-2), \ldots, 2$, can therefore be calculated, if a recurrence step leading from $\mathbf{T}^{(\nu)}$ to $\mathbf{T}^{(\nu-1)}$ is available. Such a recurrence relation is obtained by inserting the representations of Eq. (28) and (50) into the set of Eqs. (35)–(38), which yields

$$T_{kn}^{(\nu-1)} = c_k^{(\nu)}T_{kn}^{(\nu)} - s_k^{(\nu)}F_{k+1,n}^{(\nu)},\tag{59}$$

$$F_{kn}^{(\nu)} = s_k^{(\nu)}T_{kn}^{(\nu)} + c_k^{(\nu)}F_{k+1,n}^{(\nu)},\tag{60}$$

for $k = 1, 2, \ldots, (\nu-1)$, and

$$T_{\nu n}^{(\nu-1)} = F_{1n}^{(\nu)} = \delta_{\nu n},\tag{61}$$

$$F_{\nu n}^{(\nu)} = T_{\nu n}^{(\nu)},\tag{62}$$

$$T_{jn}^{(\nu-1)} = \delta_{jn}, \quad j = (\nu+1), (\nu+2), \ldots, N.\tag{63}$$

Given the matrix $\mathbf{T}^{(\nu)}$ and the angles $\theta_k^{(\nu)}$, Eqs. (60) and (62) yield all quantities $F_{kn}^{(\nu)}$ and, with these, the elements of $\mathbf{T}^{(\nu-1)}$ are obtained from Eqs. (59), (61), and (63).

Equations (55)–(63) provide thus the means to obtain a parametric representation in terms of $\frac{1}{2}N(N-1)$ angles for an arbitrary orthogonal matrix $\mathbf{T}$.

5. DISCUSSION

A. Relation to Previous Formulation

The construction procedure of Eqs. (45)–(52) is identical to that given by Raffenetti and Ruedenberg,[1] except for some changes in notation. In the previous paper the quantities $\gamma_{kn}, \mathbf{T}^{(n)}, \mathbf{t}^{(n)}, s_{kn}^{(n)}$ were used. The relation to the symbols used here is as follows.

Ref. 1	this work
γ_{kn}	$\theta_k^{(n)}$
$\mathbf{T}^{(n)}$	$\hat{\mathbf{T}}^{(n)}$
$s_{kn}^{(n)}$	$-F_{kn}^{(n)}$
$\mathbf{t}^{(n)}$	$\begin{pmatrix}\hat{\mathbf{T}}^{(n-1)} & 0\\ 0 & 1\end{pmatrix}$

In Ref. 1, there was also given a recurrence procedure for calculating the derivatives of $\mathbf{T}$ with respect to the angular parameters.

B. Block Diagonal Form

It was furthermore shown in Ref. (1) that the construction by Eqs. (45)–(52) is equivalent to expressing T as the product,

$$\begin{aligned}\mathbf{T} = \mathbf{a}_{n-1,n}\cdot\mathbf{a}_{n-2,n}\cdot\mathbf{a}_{n-3,n}\cdot\mathbf{a}_{n-4,n}\cdot\cdots\cdot\mathbf{a}_{2,n}\cdot\mathbf{a}_{1,n}\\ \times\mathbf{a}_{n-2,n-1}\cdot\mathbf{a}_{n-3,n-1}\cdot\mathbf{a}_{n-4,n-1}\cdot\cdots\cdot\mathbf{a}_{2,n-1}\cdot\mathbf{a}_{1,n-1}\\ \times\mathbf{a}_{n-3,n-2}\cdot\mathbf{a}_{n-4,n-2}\cdot\cdots\cdot\mathbf{a}_{2,n-2}\cdot\mathbf{a}_{1,n-2}\\ \times\cdots\cdots\cdots\cdots\\ \times\cdots\cdots\cdots\\ \times\mathbf{a}_{2,3}\cdot\mathbf{a}_{1,3}\\ \times\mathbf{a}_{1,2}\end{aligned}$$

where each $\mathbf{a}_{pq}$ is the "2 × 2-type" orthogonal matrix, which is identical with the unit matrix except for the diagonal elements in the pth and qth column which are $\cos\theta_p^{(q)}$, the element on the intersection of the pth row and the qth column which is $\sin\theta_p^{(q)}$, and the element on the intersection of the qth row and the pth column which is $-\sin\theta_p^{(q)}$.

Since each of these factor matrices has the nonzero off-diagonal elements in a different position, the present formulation is very convenient for the construction of a matrix which is block-diagonal or even equivalent to being block-diagonal (i.e., which can be made block-diagonal by appropriate permutations between the rows and between the columns). Such a matrix is obtained by simply choosing $\theta_p^{(q)} = 0$ for all index pairs pq for which the block-diagonal form requires $T_{pq} = 0$.

C. Accuracy

Computer programs were written for both processes:

the construction of T from $N(N-1)/2$ angles $\theta_k^{(\nu)}$ and the determination of $N(N-1)/2$ angles $\theta_k^{(\nu)}$ from a given orthogonal matrix $\mathbf{T}$. Test cases were run where, first, $\mathbf{T}$ was evaluated from arbitrarily chosen angles and, then, the angles were recalculated by the inverse process. It was observed, somewhat surprisingly, that no significant figures were lost in carrying out the two processes in succession. A documented copy of the FORTRAN program has been submitted to the Quantum Chemistry Program Exchange at Indiana University.

6. ALTERNATIVE PARAMETRIZATION

There are many different parametrizations which can be formulated that are similar to the one outlined in the previous sections. For example, instead of Eq. (2), one could write

$$\bar{\mathbf{f}}_\nu = \bar{c}_\nu \mathbf{e}_\nu + \bar{s}_\nu \bar{\mathbf{f}}_{\nu+1},$$

where the barred quantities are the counterparts of the unbarred functions previously defined. However, in order that the transformation $\bar{\mathbf{A}}$ [corresponding to $\mathbf{A}$ of Eq. (15)] have determinant + 1, it is necessary to replace the definitions (9), (10) by

$$\bar{\mathbf{b}}_1 = \mathbf{a},$$
$$\bar{\mathbf{b}}_{\nu+1} = (\bar{s}_1 \cdots \bar{s}_{\nu-1})^{-1} \partial_\nu \mathbf{a}.$$

The matrix $\bar{\mathbf{A}}$ has the structure

$$\bar{A}_{ij} = 0, \quad j - i \geq 2,$$
$$\bar{A}_{ij} = -\bar{s}_i, \quad j - i = 1,$$
$$\bar{A}_{i1} = \bar{s}_1 \cdots \bar{s}_{i-1} \bar{c}_i,$$
$$\bar{A}_{ij} = \bar{c}_{j-1} \bar{s}_j \bar{s}_{j+1} \cdots \bar{s}_{i-1} \bar{c}_i, \quad i \geq j \geq 2$$

and equals the unit matrix when all angles are zero.

We introduce the succession of basis sets

$$\bar{\mathbf{b}}_1^{(1)} = \mathbf{a}_1, \bar{\mathbf{b}}_2^{(1)}, \ldots, \bar{\mathbf{b}}_N^{(1)}$$
$$\bar{\mathbf{b}}_1^{(2)} = \mathbf{a}_1, \bar{\mathbf{b}}_2^{(2)} = \mathbf{a}_2, \ldots, \bar{\mathbf{b}}_N^{(2)}$$
$$\vdots \qquad\qquad \vdots$$
$$\bar{\mathbf{b}}_1^{(N)} = \mathbf{a}_1, \bar{\mathbf{b}}_2^{(N)} = \mathbf{a}_2, \ldots, \bar{\mathbf{b}}_N^{(N)} = \mathbf{a}_N,$$

which are related by the transformations

$$\bar{\mathbf{b}}_k^{(\nu+1)} = \sum_i \bar{\mathbf{b}}_i^{(\nu)} \bar{B}_{ik}^{(N-\nu)}$$

where

$$\bar{\mathbf{B}}^{(\mu)} = \left(\begin{array}{c|c} \mathbf{I}^{(N-\mu)} & 0 \\ \hline 0 & \bar{\mathbf{A}}^{(\mu)} \end{array}\right).$$

The matrix $\bar{\mathbf{B}}^{(\mu)}$ is parametrized by the $\mu - 1$ angles,

$$0 \leq \bar{\vartheta}_i^{(\mu)} \leq \pi, \quad i = (N - \mu + 1) \text{ to } (N-2),$$
$$0 \leq \bar{\vartheta}_{N-1}^{(\mu)} < 2\pi.$$

With this choice of angles the equations equivalent to (35) and (36) are

$$\bar{\mathbf{b}}_{k+1}^{(\nu)} = -s_k^{(N-\nu+1)} \bar{\mathbf{b}}_k^{(\nu-1)} + \bar{c}_k^{(N-\nu+1)} \bar{\mathbf{f}}_{k+1}^{(\nu)},$$
$$\bar{\mathbf{f}}_k^{(\nu)} = \bar{c}_k^{(N-\nu+1)} \bar{\mathbf{b}}_k^{(\nu-1)} + \bar{s}_k^{(N-\nu+1)} \bar{\mathbf{f}}_{k+1}^{(\nu)},$$

which can be inverted to yield

$$\bar{\mathbf{b}}_k^{(\nu-1)} = \bar{c}_k^{(N-\nu+1)} \bar{\mathbf{f}}_k^{(\nu)} - \bar{s}_k^{(N-\nu+1)} \bar{\mathbf{b}}_{k+1}^{(\nu)},$$
$$\bar{\mathbf{f}}_{k+1}^{(\nu)} = \bar{s}_k^{(N-\nu+1)} \bar{\mathbf{f}}_k^{(\nu)} + \bar{c}_k^{(N-\nu+1)} \bar{\mathbf{b}}_{k+1}^{(\nu)}.$$

These equations together with the conditions

$$\mathbf{f}_N^{(j)} = \mathbf{b}_N^{(j-1)},$$
$$\mathbf{f}_\nu^{(\nu)} = \mathbf{a}_\nu$$

yield the recurrence steps by which $\bar{\mathbf{T}}$ can be decomposed or constructed.

The matrices $\bar{\mathbf{T}}^{(\nu)}$ are defined by

$$\bar{\mathbf{T}}^{(\nu)} = \bar{\mathbf{B}}^{(\nu)} \bar{\mathbf{B}}^{(\nu-1)} \cdots \bar{\mathbf{B}}^{(3)} \bar{\mathbf{B}}^{(2)}$$

and $\bar{\mathbf{T}}$ by

$$\bar{\mathbf{T}} = \bar{\mathbf{T}}^{(N)}.$$

They can be constructed recursively by the relations

$$\bar{T}_{ik}^{(\nu+1)} = \bar{c}_i^{(\nu+1)} \bar{F}_{ik}^{(\nu)} - \bar{s}_i^{(\nu+1)} \bar{T}_{i+1,k}^{(\nu)},$$
$$\bar{F}_{i+1,k}^{(\nu)} = \bar{s}_i^{(\nu+1)} \bar{F}_{ik}^{(\nu)} + \bar{c}_i^{(\nu+1)} \bar{T}_{i+1,k}^{(\nu)},$$

with the aid of the initial condition

$$\bar{F}_{N-\nu,k}^{(\nu)} = \delta_{N-\nu,k}.$$

Here the matrix $\bar{\mathbf{F}}^{(\nu)}$ is defined by the relations

$$\sum_k \bar{F}_{ik}^{(\nu)} \mathbf{a}_k = \bar{\mathbf{f}}_i^{(N-\nu)}.$$

The matrices $\bar{\mathbf{T}}^{(\nu)}$ can be decomposed by the relations

$$\bar{T}_{i+1,k}^{(\nu)} = -\bar{s}_i^{(\nu+1)} T_{ik}^{(\nu+1)} + \bar{c}_i^{(\nu+1)} \bar{F}_{i+1,k}^{(\nu)},$$
$$\bar{F}_{ik}^{(\nu)} = \bar{c}_i^{(\nu+1)} \bar{T}_{ik}^{(\nu+1)} + \bar{s}_i^{(\nu+1)} \bar{F}_{i+1,k}^{(\nu)},$$

and the condition

$$\bar{F}_{Nk}^{(\nu)} = \bar{T}_{Nk}^{(\nu+1)}.$$

The angles of $\bar{\mathbf{T}}$ can be determined from the relations

$$\cos\bar{\vartheta}_{N-\nu+1}^{(\nu)} = \bar{T}_{N-\nu+1,N-\nu+1}^{(\nu)}$$
$$\cos\bar{\vartheta}_i^{(\nu)} = \bar{T}_{i,N-\nu+1}^{(\nu)} / (\sin\bar{\vartheta}_{N-\nu+1}^{(\nu)} \cdots \sin\vartheta_{i-1}^{(\nu)}),$$
$$i = N - \nu + 2 \text{ to } N - 1,$$
$$\sin\bar{\vartheta}_{N-1}^{(\nu)} = \bar{T}_{N,N-\nu+1} / (\sin\bar{\vartheta}_{N-\nu+1}^{(\nu)} \cdots \sin\bar{\vartheta}_{N-2}^{(\nu)}).$$

As was the case with the transformation $\mathbf{T}$, the transformation $\bar{\mathbf{T}}$ can be written as a product of $\frac{1}{2}N(N-1)$ "(2×2)-type" orthogonal matrices. This is implied by the existence of the recurrence relations given above. However, in contrast with the former case when there was *one* 2×2 matrix with off-diagonal elements in the position of *any* given off-diagonal element of $\mathbf{T}$, *all* 2×2 matrices occurring *here* have the general structure

$$\begin{pmatrix} d_{nn} & d_{n,n+1} \\ d_{n+1,n} & d_{n+1,n+1} \end{pmatrix} = \begin{pmatrix} c & -s \\ s & c \end{pmatrix},$$
$$d_{ij} = \delta_{ij}, \quad i,j \neq n, n+1,$$

where d_{ij} is an element of the 2×2 matrix and c and s are the cosine and sine, respectively, of the associated angle.

* Work was performed in the Ames Laboratory of the U.S. Atomic Energy Commission, Contribution No. 3101.

[1] R. C. Raffenetti and K. Ruedenberg, Intern. J. Quantum Chem. 3S, 625 (1970).

Dynamical Quantization

E. Onofri and M. Pauri
Istituto di Fisica dell'Università, Parma, Italy
and
Istituto Nazionale di Fisica Nucleare, Sezione di Milano Milano, Italy.

(Received 29 October 1971)

An intrinsic quantization procedure based on higher symmetries of classical dynamical systems and utilizing the techniques of van Hove and Souriau is proposed. The procedure is intrinsically Hamiltonian but not explicitly canonical in that the Heisenberg algebra plays no fundamental role. The proposed method is applied to the n-dimensional harmonic oscillator and to the n-dimensional hydrogen atom. This approach seems to provide the first intrinsic justification of the success of ordinary correspondential quantization for this last system.

1. INTRODUCTION

Purpose of the present paper is to illustrate a proposal for a new point of view in quantization theory. The presentation given here has a rather provisional character from a mathematical standpoint, leaving rigorous proofs and technical refinements to a subsequent discussion.

As is well known, the usual correspondential procedure of quantization (Schrödinger method) is a rather artificial operation which does not reflect, in general, intrinsic geometric or perhaps dynamical properties of the physical system to be quantized. As a consequence, it depends strictly on such arbitrary and local objects as the canonical coordinates used for the Hamiltonian description: Precisely, the Schrödinger quantization must be carried out in Cartesian coordinates while the quantization in generalized coordinates is always to be deduced by means of a suitable transformation (see, for instance, Refs. 1, 2, 3). This unsatisfactory situation becomes a serious drawback as soon as the phase space Ω cannot be constructed in the usual way over a configuration space M (cotangent bundle $\Omega = T^*M$); in fact, in this case, the procedure completely fails.

Another unappealing feature of this method is that it does not give a general prescription to obtain self-adjoint operators from a suitable class of classical dynamical variables. As a consequence, the mathematical structure of the operators constructed by the correspondential procedure must be analyzed case by case.

A final point about the Schrödinger correspondential quantization is the lack of uniqueness.[4,5] This is usually taken as a further shortcoming; however, the fact that different quantum observables (and thus different processes of measure) may "correspond" to the same classical variable appears to be inherent to the epistemological status of quantum mechanics and, in our opinion, should not be considered a weakness of the theory.

The need for a more clear foundation of quantization has given rise to many interesting investigations. These works can be divided into two classes. The first one is concerned with algebraic rules for assigning operators to classical variables is some coordinate-invariant way[4-9]; the subsequent discussion should make clear that this approach is inadequate. The second class is concerned with a quantization based on the intrinsic geometrical structure of the canonical transformations and follows a group-theoretical rather than a simply algebraic approach.[10-16] The present paper belongs to this last class of works.

Our approach originates from the belief that if some typical attributes are to be maintained in the abstract process of quantization from classical dynamics, these should be global symmetry properties which embody configurational and dynamical features. Thus we propose an intrinsic quantization scheme for the Hamiltonian systems which admit a global maximal symmetry group satisfying certain reasonable conditions. It is clear that the consequent range of application of the theory is very narrow.[17] However, it contains in particular the dynamical systems which provide the basic models for all the nonrelativistic particle interactions, namely the oscillator and the Coulomb problems. In this connection we claim that our approach "explains" the up to now rather mysterious quantization of the hydrogen atom.[13,14] In a sense we can say that our quantization is system-dependent. At the same time we are convinced that a general intrinsic quantization procedure is a problem which cannot be settled within the framework of ordinary quantum mechanics.

Our procedure can be roughly divided into the following steps. A first stage corresponds to the identification of the global symmetry at the classical level together with an enlarged "noninvariance" transformation group of the phase space called the "quantization group". An essential property of this last group is that it always contains the Hamiltonian of the given dynamical system expressed as a function of a single infinitesimal generator. The second stage is the construction of a "contact" manifold $\tilde{\Omega}_{2n+1}$ over the phase space ("espace fibré quantifiant" of Souriau, see Ref. 12), in which a global action of the quantization group is defined. The third step is the consequent construction of a unitary reducible representation $\mathcal{R}^{(1/\hbar)}$ of the quantization group within a suitable subspace of the Hilbert space of Lebesgue square-integrable functions on $\tilde{\Omega}_{2n+1}$. The final step is the reduction of $\mathcal{R}^{(1/\hbar)}$ and the selection of the irreducible "quantum representation" $\tilde{\mathcal{R}}$. The second and third steps realize a process which can be called a prequantization.[18]

The true quantization is accomplished by the fourth step which defines uniquely the quantum description in terms of a complete set of observables which include the Hamiltonian. In this sense, the procedure

* Work was performed in the Ames Laboratory of the U.S. Atomic Energy Commission, Contribution No. 3101.

[1] R. C. Raffenetti and K. Ruedenberg, Intern. J. Quantum Chem. 3S, 625 (1970).

Dynamical Quantization

E. Onofri and M. Pauri
Istituto di Fisica dell'Università, Parma, Italy
and
Istituto Nazionale di Fisica Nucleare, Sezione di Milano Milano, Italy.

(Received 29 October 1971)

An intrinsic quantization procedure based on higher symmetries of classical dynamical systems and utilizing the techniques of van Hove and Souriau is proposed. The procedure is intrinsically Hamiltonian but not explicitly canonical in that the Heisenberg algebra plays no fundamental role. The proposed method is applied to the n-dimensional harmonic oscillator and to the n-dimensional hydrogen atom. This approach seems to provide the first intrinsic justification of the success of ordinary correspondential quantization for this last system.

1. INTRODUCTION

Purpose of the present paper is to illustrate a proposal for a new point of view in quantization theory. The presentation given here has a rather provisional character from a mathematical standpoint, leaving rigorous proofs and technical refinements to a subsequent discussion.

As is well known, the usual correspondential procedure of quantization (Schrödinger method) is a rather artificial operation which does not reflect, in general, intrinsic geometric or perhaps dynamical properties of the physical system to be quantized. As a consequence, it depends strictly on such arbitrary and local objects as the canonical coordinates used for the Hamiltonian description: Precisely, the Schrödinger quantization must be carried out in Cartesian coordinates while the quantization in generalized coordinates is always to be deduced by means of a suitable transformation (see, for instance, Refs. 1, 2, 3). This unsatisfactory situation becomes a serious drawback as soon as the phase space Ω cannot be constructed in the usual way over a configuration space M (cotangent bundle $\Omega = T^*M$); in fact, in this case, the procedure completely fails.

Another unappealing feature of this method is that it does not give a general prescription to obtain self-adjoint operators from a suitable class of classical dynamical variables. As a consequence, the mathematical structure of the operators constructed by the correspondential procedure must be analyzed case by case.

A final point about the Schrödinger correspondential quantization is the lack of uniqueness.[4,5] This is usually taken as a further shortcoming; however, the fact that different quantum observables (and thus different processes of measure) may "correspond" to the same classical variable appears to be inherent to the epistemological status of quantum mechanics and, in our opinion, should not be considered a weakness of the theory.

The need for a more clear foundation of quantization has given rise to many interesting investigations. These works can be divided into two classes. The first one is concerned with algebraic rules for assigning operators to classical variables is some coordinate-invariant way[4-9]; the subsequent discussion should make clear that this approach is inadequate. The second class is concerned with a quantization based on the intrinsic geometrical structure of the canonical transformations and follows a group-theoretical rather than a simply algebraic approach.[10-16] The present paper belongs to this last class of works.

Our approach originates from the belief that if some typical attributes are to be maintained in the abstract process of quantization from classical dynamics, these should be global symmetry properties which embody configurational and dynamical features. Thus we propose an intrinsic quantization scheme for the Hamiltonian systems which admit a global maximal symmetry group satisfying certain reasonable conditions. It is clear that the consequent range of application of the theory is very narrow.[17] However, it contains in particular the dynamical systems which provide the basic models for all the nonrelativistic particle interactions, namely the oscillator and the Coulomb problems. In this connection we claim that our approach "explains" the up to now rather mysterious quantization of the hydrogen atom.[13,14] In a sense we can say that our quantization is system-dependent. At the same time we are convinced that a general intrinsic quantization procedure is a problem which cannot be settled within the framework of ordinary quantum mechanics.

Our procedure can be roughly divided into the following steps. A first stage corresponds to the identification of the global symmetry at the classical level together with an enlarged "noninvariance" transformation group of the phase space called the "quantization group". An essential property of this last group is that it always contains the Hamiltonian of the given dynamical system expressed as a function of a single infinitesimal generator. The second stage is the construction of a "contact" manifold $\tilde{\Omega}_{2n+1}$ over the phase space ("espace fibré quantifiant" of Souriau, see Ref. 12), in which a global action of the quantization group is defined. The third step is the consequent construction of a unitary reducible representation $\mathcal{R}^{(1/\hbar)}$ of the quantization group within a suitable subspace of the Hilbert space of Lebesgue square-integrable functions on $\tilde{\Omega}_{2n+1}$. The final step is the reduction of $\mathcal{R}^{(1/\hbar)}$ and the selection of the irreducible "quantum representation" $\tilde{\mathcal{R}}$. The second and third steps realize a process which can be called a prequantization.[18]

The true quantization is accomplished by the fourth step which defines uniquely the quantum description in terms of a complete set of observables which include the Hamiltonian. In this sense, the procedure

provides an intrinsic Hamiltonian quantization. The usual canonical description is possibly recovered at the end, in the sense that canonical operators Q, P, irreducibly represented, may be definable within the quantal representation in such a way that the Hamiltonian, as a function of them, takes the usual correspondential expression.

The mathematical technique we use is due to the investigations of van Hove[10] and especially of Souriau,[12] with some adaptations and developments indicated by the works of Auslander and Kostant[18,19] and Hurt.[16]

The essential points of van Hove's and Souriau's results are briefly reviewed and discussed in Secs. 2A and 2C. In Section 2B the problem of quantization is discussed in its generality starting from the standard historical formulation (Dirac problem). Section 3 is devoted to the statement of the proposed dynamical quantization. Finally, in Sec. 4, the theory is applied to two basic classes of dynamical systems with higher symmetries: the n-dimensional isotropic harmonic oscillator and the n-dimensional hydrogen atom.

We adopt the notations of Ref. 20.

2. GENERAL PROBLEMS OF INTRINSIC QUANTIZATION

A. Euclidean "Prequantization"

Van Hove's investigation[10] is concerned with a Euclidean phase space $\Omega_{2n} = R^{2n}$. Consider the infinite Lie pseudogroup Γ of "contact" transformations, consisting of those $C^{(\infty)}$ global diffeomorphisms of $R^{2n+1}(s, q_1, \ldots, q_n, p_1, \ldots, p_n)$ onto itself which leave the 1-form

$$\theta = -ds + \sum_1^n p_i dq_i \tag{1}$$

invariant. The general form of an element $\gamma \in \Gamma$ is

$$\gamma: \begin{cases} q_i' = q_i'(q,p), & i = 1, \ldots, n, \\ p_i' = p_i'(q,p), \\ s' = s + \pi_\gamma(q,p), \end{cases} \tag{2}$$

where

$$d\pi_\gamma(q,p) = \sum_1^n (p_i' dq_i' - p_i dq_i) \tag{3}$$

and $\pi_\gamma(q,p)$ is the "generating function"[21] of the transformation $(q,p) \to (q',p')$, which is a canonical transformation since

$$\omega' - \omega \equiv \sum_1^n (dp_i' \wedge dq_i' - dp_i \wedge dq_i) = \mathbf{d}\left(\sum_1^n [p_i' dq_i' - p_i dq_i]\right) = 0. \tag{4}$$

A one-parameter subgroup $\gamma(\tau) \subset \Gamma$ is characterized by its "infinitesimal generating function" $f(q,p)$:

$$\frac{ds(\tau)}{d\tau} = f(q,p) - \sum_1^n p_i \frac{\partial f}{\partial p_i}, \qquad \frac{dq_i(\tau)}{d\tau} = \frac{\partial f(q,p)}{\partial p_i}, \quad \frac{dp_i(\tau)}{d\tau} = -\frac{\partial f(q,p)}{\partial q_i}. \tag{5}$$

Not every $C^{(\infty)}$ function $f(q,p)$ generates one-parameter subgroups of Γ, i.e., not every $C^{(\infty)}$ vector field of the form

$$X[f] = \left(f - \sum_1^n p_i \frac{\partial f}{\partial p_i}\right)\frac{\partial}{\partial s} + \{f, \cdots\} \tag{6}$$

is "complete."[22] Let us call $\tilde{\mathfrak{F}}_\Gamma \subset \tilde{\mathfrak{F}}(R^{2n})$ the family of $C^{(\infty)}$ functions $f(q,p)$ such that $X[f]$ is "complete."

Van Hove first defines the following unitary representation $\mathfrak{R}$ of Γ within the Hilbert space of the Lebesgue square-integrable functions over R^{2n+1}:

$$\mathfrak{R}: \forall \gamma \in \Gamma, \ \phi \in L_2(R^{2n+1}): \ [U_\gamma \phi](q,p,s) = \phi(\gamma^{-1}(q,p,s)). \tag{7}$$

This is a completely reducible representation of Γ which can be reduced into a direct integral over the generalized subspaces of functions $\phi(q,p,s)$ satisfying the condition

$$\phi(q,p,s+s') = e^{i\alpha s'}\phi(q,p,s), \alpha \text{ any real number.} \tag{8}$$

From Eq. (8) it follows

$$\phi(q,p,s) = e^{i\alpha s}\psi(q,p). \tag{9}$$

Therefore, the representations $\mathfrak{R}^{(\alpha)}$ defined by

$$\mathfrak{R}^{(\alpha)}: \forall \gamma \in \Gamma, \ \psi \in L_2(R^{2n}) \qquad [U_\gamma \psi](q,p) = e^{i\alpha\pi_\gamma(\gamma^{-1}(q,p))}\psi(\gamma^{-1}(q,p)), \tag{10}$$

result faithful (unless $\alpha = 0$) irreducible unitary representations of Γ.

Each multiplier representation $\mathfrak{R}^{(\alpha)}$ associates a self-adjoint operator $H^{(\alpha)}[f]$ in $L_2(R^{2n})$ to every $f(q,p) \in \tilde{\mathfrak{F}}_\Gamma$: In fact, since $f(q,p)$ generates a one-parameter subgroup $\gamma_f(\tau) \subset \Gamma$, $\mathfrak{R}^{(\alpha)}(\gamma_f(\tau))$ is a continuous one-parameter group of unitary transformations

$$U^{(\alpha)}_{\gamma_f(\tau)} = \exp(i\tau H^{(\alpha)}[f]); \tag{11}$$

then, by the Stone–von Neumann theorem, $H^{(\alpha)}[f]$ is self-adjoint. The explicit representation of $H^{(\alpha)}[f]$, according to Eq. (5), is

$$H^{(\alpha)}(f) = \alpha\left(f - \sum_1^n p_i \frac{\partial f}{\partial p_i}\right) + i\{f, \cdots\}. \tag{12}$$

As shown by van Hove, the operators

$$K[f] = (1/\alpha)H^{(\alpha)}[f], \quad \text{for } \alpha = \hbar^{-1}, \tag{13}$$

solve the restricted Dirac problem, i.e., establish a map from classical dynamical variables f to self-adjoint operators $K[f]$ with suitable domains in a Hilbert space, such that

$$\begin{cases} K[f_1]K[f_2] - K[f_2]K[f_1] = i\hbar K[\{f_1, f_2\}], \\ K[1] = \mathbf{1}.[23] \end{cases} \tag{14}$$

B. The Dirac Problem

The full Dirac problem, which can be considered as the standard historical formulation of the quantum conditions,[24] requires, in addition to Eqs. (13)–(14), the fundamental operators $K[q], K[p]$ to be irreducibly represented, a condition which is necessary in order to define the quantum kinematics uniquely (see also Weyl[25]).

It is easy to see that the correspondence given by Eq. (13) does not provide a solution of the full Dirac problem. Actually the operators

$$K[q_i] \equiv Q_i = q_i + i\hbar \frac{\partial}{\partial p_i}, \qquad K[p_i] \equiv P_i = -i\hbar \frac{\partial}{\partial q_i} \tag{15}$$

commute with

$$Q_i' = i\hbar \frac{\partial}{\partial p_i}, \qquad P_i' = p_i + i\hbar \frac{\partial}{\partial q_i}. \tag{16}$$

It is possible, of course, to restrict to a "subspace" corresponding to

$$Q_i'\psi(q, p) = \lambda_i \psi(q, p) \Rightarrow \psi(q, p) = e^{-(i/\hbar)\lambda_i p_i} \chi(q), \tag{17}$$

Then the restrictions of Q, P

$$\begin{cases} \hat{Q}_i \chi(q) = (q_i + \lambda_i)\chi(q) \\ \hat{P}_i \chi(q) = -i\hbar \dfrac{\partial}{\partial q_i} \chi(q) \end{cases} \tag{18}$$

are irreducible and still self-adjoint. However, such a subspace is not invariant in general under the unitary transformation group generated by the Hamiltonian. Therefore, in this way, we would have an intrinsic quantization of the kinematics alone, i.e., nothing more than the usual scheme.

This latter, on the other hand, besides its being an artificial and nonintrinsic procedure, does not represent a solution of the Dirac problem either. As a matter of fact, it satisfies the irreducibility condition, but it contains no general prescription to assign self-adjoint operators to a suitable family of classical observables. More precisely, an assignment is given, once and for all, only for the so-called Heisenberg algebra $\mathfrak{L}_H$: $q, p, 1$, while all the other dynamical variables are left to be defined ad hoc as operator functions of $q, p, 1$; in this way, however, one is not assured *a priori* that such functions define self-adjoint operators. One can be sure at most to obtain symmetric operators by following certain "ordering rules" in the construction,[4-6] but the whole procedure may not be consistent with the restricted Dirac problem as well.

We stress that this failure must be common to all the purely algebraic approaches such as those given in Refs. 4–9. The point is that the basic feature of both the classical and quantum systems is not only a Lie structure of observables but also a mapping of observables to one-parameter subgroups of canonical or unitary transformations, respectively. Thus we cannot expect a function $f(q, p) \notin \mathfrak{F}_\Gamma$ to correspond to a self-adjoint operator whatever the mapping may be, owing to the Stone–von Neumann theorem.

On the other hand, the Dirac problem as stated in its historical form is not a well-defined mathematical statement, so that the question of its general solvability cannot be asked directly. A mathematically consistent formulation of the problem has been given by van Hove, who was able to prove a nonexistence theorem. We do not know whether the extra assumptions introduced by him are really necessary from a general point of view, but we will not discuss this point here. Van Hove's theorem gives a negative answer to the question asked by the full Dirac problem for the whole family $\mathfrak{F}_\Gamma$ of classical observables.

At the same time van Hove has given a positive theorem according to which the representation $R^{(1/\hbar)}$ provides the unique solution of the full Dirac problem for the subfamily $\mathfrak{F}_L \subset \mathfrak{F}_\Gamma$ of polynomial functions, quadratic at most in the canonical variables, i.e., the generating functions of the subgroup $\Gamma_L \subset \Gamma$ of linear symplectic transformations. In physical terms this means that a quantization is given for any harmonic oscillator in the sense that what is quantized is the Heisenberg algebra plus the Hamiltonian and some other constants of the motion. In the case of the family $\mathfrak{F}_L$ the results turn out to be just the ordinary correspondential ones; however, the construction given by van Hove contains ad hoc prescriptions in the final stages which weaken somewhat its intrinsic nature.

According to van Hove's results, it is not possible to quantize in an intrinsic way sufficiently large systems of classical observables to be considered as possible observables of the classical mechanics as a whole, independently of any specifical dynamical system. As Hermann[13] remarks, this does not preclude the possibility of quantizing special classes of dynamical systems, and it seems to be an open topic of research to analyse the structure of classical systems for which this is possible. In this connection Segal[14] has stressed the fact that the success of the ordinary quantization for such a fundamental system as the hydrogen atom is not yet well understood. In order to clarify these relevant points, we propose to change the main point of view on general quantization in a sense we are going to outline in the remainder of the present paper.

First of all, let us remark that in the case of the subfamily $\mathfrak{F}_L$, which provides the only positive example we know of a true intrinsic quantization, the elements $f(q, p)$ close a finite Lie algebra and, what is more, do generate a global Lie group. In the second place, let us observe that in all the relevant theorems it must be always assumed that the Poisson bracket of two given elements $f_1, f_2 \in \mathfrak{F}_\Gamma$ also belongs to $\mathfrak{F}_\Gamma$, a fact which is not assured *a priori* since $\mathfrak{F}_\Gamma$ is not closed neither under linear combination nor under Poisson bracket. (An example is given by: $f_1 = \frac{1}{2}p^2 \in \mathfrak{F}_\Gamma$, $f_2 = \frac{1}{6}q^3 \in \mathfrak{F}_\Gamma$, $f_1 + f_2 \notin \mathfrak{F}_\Gamma$, $\{f_1, f_2\} = -\frac{1}{2}pq^2 \notin \mathfrak{F}_\Gamma$ see Footnote 22.) We recall that the neighborhood of the identity of an infinite Lie pseudogroup is not expected to share the structure of a Lie algebra.[26] Now, the important point is that for elements $f(q, p)$ there is an essential link between their property of closure under Lie bracket and their belonging to $\mathfrak{F}_\Gamma$. As a matter of fact it can be proved the following important result (Palais)[27]:

Theorem: If the functions $f_1 \cdots f_k$ are such that

(a) $f_i \in \mathfrak{F}_\Gamma$, $i = 1, \ldots, k$,

(b) they generate a finite-dimensional Lie algebra $\mathfrak{L}$ under the Poisson bracket (i.e., $f_{ij} = \{f_i, f_j\}$, $f_{ijq} = \{f_{ij}, f_q\}$, etc. are linear combinations of a finite number $r \geq k$ of functions),

then

(a′) every $X[f]$ is "complete" ($f \in \mathfrak{L}$),

(b′) there exists a unique global action on R^{2n+1} of the simply connected Lie group L whose Lie algebra is $\mathfrak{L}$.

All the above considerations strongly indicate that to attack the problem of intrinsic quantization, we must take into account a finite-dimensional subalgebra of $\mathfrak{F}_\Gamma$ (or better a Lie subgroup of Γ). On the other hand, the negative result contained in van Hove's investigation indicate that we must abandon the Heisenberg algebra or, in other words, the canonical way *"strictu sensu"* of quantization. What we propose to do now is to work with a finite Lie subgroup of Γ which is specific for a given dynamical system in that it reflects its intrinsic dynamical properties. Having identified this group, which we shall call the "quantization group" $\tilde{\mathcal{G}}$, and its particular global canonical realization $\mathcal{K}_{\tilde{\mathcal{G}}}$ in the phase space Ω_{2n} of the system, we are led to consider the corresponding unitary representation of the form $\mathcal{R}^{(1/\hbar)}$. This representation is now a reducible one[18] (unlike for the full group Γ). Then the irreducibility of the Heisenberg algebra representation is to be substituted in a natural way by an irreducibility condition on the above representation. This will finally provide the "quantization" of the dynamical system in terms of a unitary representation $\tilde{\mathcal{R}}$ of $\tilde{\mathcal{G}}$ which we shall call the "quantal representation." This irreducibility condition means that all the observables of the quantized dynamical system are to be found within the enveloping algebra $\mathfrak{E}_{\tilde{\mathcal{G}}}$ of $\tilde{\mathcal{G}}$. It must be clear, however, that only the observables belonging to the Lie algebra $\mathfrak{G}$ are intrinsically quantized, i.e., are unambiguously determined by a canonical vector field, while the other ones do not possess an intrinsic classical counterpart. Since the quantization procedure does not preserve the functional relations, the quantization of elements of the enveloping algebra $\mathfrak{E}_{\tilde{\mathcal{G}}}$ is not a well-defined intrinsic process and becomes ambiguous in the sense of the "ordering rules." We have stressed in the introduction that from our point of view this is not a drastic shortcoming and can be admitted on general grounds. In particular, the quantization of the canonical variables themselves is not defined in general also in the case of Euclidean phase spaces.

It may happen, however, (and in significant cases actually does) that the canonical operators are recovered at the end within the quantal representation $\tilde{\mathcal{R}}$. This would "explain" the effectiveness of the Schrödinger quantization in such cases. For these reasons, our procedure, which is entitled to be called a Hamiltonian quantization, is not a "canonical" quantization.

As the reader will realize, there are important questions in the final stage of our quantization method which we have not yet fully explored. The most important one is the choice of the quantal representation $\tilde{\mathcal{R}}$. The results we obtain are quite satisfactory, but we have not yet been able to exploit a deep geometrical reason to be found at the classical level of the procedure which forces the selection of $\tilde{\mathcal{R}}$. Without this geometrical foundation, the whole process would finally appear as a construction which reduces to "explain" *a posteriori* the effects of quantization and such that it could be plainly bypassed by dealing with the unitary representations of $\tilde{\mathcal{G}}$ since the beginning. That this is not the right interpretation of the facts, however, is corroborated by the existence of a deep structural similarity between the representation $\tilde{\mathcal{R}}$ and the global canonical realization $\mathcal{K}_{\tilde{\mathcal{G}}}$

Another relevant question is the uniqueness of the procedure, though, in our view, this does not represent a crucial point. Anyway, it is a remarkable fact that, in the simple (and singular) cases in which different ways of quantization appear to be possible, they bring all to the same result.

Finally, relevant technical problems may appear from a practical point of view in connection with the reduction of the unitary representation $\mathcal{R}^{(1/\hbar)}$. A great deal of work in this field has been done by Gel'fand and co-workers.[28] Investigations also more close to our problems have recently been performed by Kostant and Auslander.[18,19] Implications of all these results to our program have still to be pursued. On the other hand, the mathematical problems at the classical level should be considered completely clarified. The local theory of canonical realizations has been given in Refs. 29. Results on the global aspects of this theory can be found in several recent works.[12,18,30]

C. Non-Euclidean Prequantization

According to the proposed point of view, we are led to "quantize", in a sense to be still completely specified, global canonical realizations of Lie groups. Now, the underlying manifolds of these realizations (phase spaces in our context) are not in general Euclidean manifolds. Therefore, we are forced to go beyond van Hove's approach.

It must be remarked that the need for a generalization of the theory to arbitrary differentiable manifolds is apparent also independently of our group-theoretical considerations. For later convenience it is worthwhile to clear this point by means of a basic example. Consider the n-dimensional Kepler problem; the phase space is "naturally" Euclidean:

$$\Omega_{2n} = T^*(R^n - \{0\}), \tag{19}$$

but the Hamiltonian vector field $(dH)^\#$, where

$$H = \mathbf{p}^2/2m - k/|\mathbf{x}| \tag{20}$$

is *not* "complete" since orbits with zero angular momentum reach the point $\mathbf{x} = 0$ within a finite lapse of time. This means that $H \notin \mathfrak{F}_\Gamma$, i.e., H cannot even be prequantized; altogether, it has recently been shown by Moser[31] that the phase space can be enlarged in such a way that the Hamiltonian vector field becomes complete: this "globalization" (Ω^*_{2n}, X^*_H) (in the sense of Palais[27]) is characterized by a phase space Ω^*_{2n} which is no more Euclidean (unless $n = 2$). Note that this "globalization" or "regularization" is just the one implied by the dynamical symmetry $SO(n+1)$ of the dynamical system. In this connection we shall see in the following the actual topological structure of Ω^*_{2n}.

A general investigation of the intrinsic quantization problem for non-Euclidean symplectic manifolds has been given by Souriau.[12] The essential point in geometric quantization, or better "prequantization," is always to start from the phase space Ω_{2n} and to construct a contact manifold $\tilde{\Omega}_{2n+1}$ of one dimension higher which carries the prequantization procedure. In giving a generalization of van Hove technique to arbitrary symplectic manifolds, Souriau finds that,

unlike the Euclidean case, suitable conditions must be verified in order that the phase space Ω_{2n} be prequantizable. Souriau himself has already applied his method to quantize elementary systems, such as the Galilean and Lorentz free particles with spin, which cannot be dealt with using the ordinary correspondential procedure. However, to the authors' knowledge, a consistent treatment of dynamical systems with interaction has never been made.

In what follows we shall limit ourselves to illustrate in a rather free way the Souriau theory as applied to a simple significant example, emphasizing the role which can play a group structure according to our point of view. For the sake of clarity, we shall follow step by step the exposition given of van Hove's approach.

Consider, as phase space Ω_2, the two-dimensional sphere S^2, equipped with the symplectic structure defined by the global 2-form

$$\omega(\mathbf{X}, \mathbf{Y}) = l \cdot \mathbf{X} \wedge \mathbf{Y}, \quad |l| \text{ fixed}, \tag{21}$$

where l is the radius vector and $\mathbf{X}, \mathbf{Y}$ any two tangent vectors to S^2. Since

$$\int_{S^2} \omega = 4\pi l, \tag{22}$$

a global 1-form θ such that

$$\omega = \mathrm{d}\theta \tag{23}$$

does not exist [compare instead Eq. (4)]. However, we can define two open subsets U_+, U_- of S^2 such that

$$U_+ \cup U_- = S^2, \tag{24}$$

and two 1-forms $\theta_\pm^*$ such that their restrictions to U_+, U_- satisfy

$$(\omega - \mathrm{d}\theta_\pm^*) \restriction U_\pm = 0. \tag{25}$$

A simple choice is

$$\begin{aligned} U_+ &= S^2 - \{(0, 0, -l)\}, \\ U_- &= S^2 - \{(0, 0, l)\}. \end{aligned} \tag{26}$$

Since

$$\omega = l d(\cos\vartheta) \wedge d\varphi \equiv dp \wedge dq, \tag{27}$$

the most general 1-forms restricted to $U_\pm$ are

$$\theta_\pm^* = l(\cos\vartheta \mp 1)d\varphi + \alpha_\pm \tag{28}$$

where $\mathrm{d}\alpha_\pm = 0$. Without any loss of generality, we choose $\alpha_\pm = 0$.

Then, having constructed the (trivial) local bundles

$$U_+ \times S^1, U_- \times S^1, \tag{29}$$

the existence of a global bundle over S^2 follows from a general theorem[32] under the existence of a mapping s

$$s: U_+ \cap U_- \xrightarrow{\text{onto}} S^1 \tag{30}$$

(obvious in our case). As Souriau has shown, this global bundle can be given a contact structure adapted to the underlying symplectic structure of S^2 if the mapping (30) is such that

$$\theta_-^*(x) - \theta_+^*(x) = ds(x) \quad \text{for every } x \in U_+ \cap U_-. \tag{31}$$

In the present case, from

$$-l(\cos\vartheta - 1)d\varphi + l(\cos\vartheta + 1)d\varphi = 2l d\varphi, \tag{32}$$

it must be

$$s(\vartheta, \varphi) = 2l\varphi. \tag{33}$$

Setting $\oint_{S^1} ds = 2\pi\hbar$ by convention, the mapping s exists if and only if

$$2l = N\hbar. \tag{34}$$

In conclusion, the contact manifold $\tilde{\Omega}_3$ exists, and then S^2 is prequantizable, only *if the radius l is an integer or half-integer multiple of $\hbar$*. Since in this case the contact manifold is not a trivial fibre bundle, i.e., a direct product $\Omega_2 \times S^1$, it is not possible to give explicitly the global invariant one-form [in analogy with Eq. (1)] and the global contact transformations γ which leave this 1-form invariant [in analogy with Eq. (2)]. However, we can give expressions analogous to Eqs. (1), (2), (3) in local charts. The global 1-form looks like

$$\theta_\pm = -ds + \theta_\pm^*. \tag{35}$$

Correspondingly, the expression of the generic $C^{(\infty)}$ vector field [in analogy with Eq. (5)] is now

$$X_\pm[f] = \left[f - (\cos\vartheta \mp 1)\frac{\partial f}{\partial \cos\vartheta}\right]\frac{\partial}{\partial s} + \{f, \cdots\}, \tag{36}$$

Then, just in view of our preceding remarks, we consider a particular finite Lie subgroup of the full group Γ of contact transformations of $\tilde{\Omega}_3$; precisely the group $SO(3)$ which acts transitively on S^2 leaving ω invariant. This global canonical realization will be denoted $\tilde{\mathcal{K}}_{SO(3)}$; its local structure has been studied in Ref. 29; the infinitesimal generators are globally Hamiltonian with generating functions

$$\begin{cases} J_1 = l \sin\vartheta \cos\varphi = (l^2 - p^2)^{1/2} \cos q, \\ J_2 = l \sin\vartheta \sin\varphi = (l^2 - p^2)^{1/2} \sin q, \\ J_3 = l \cos\vartheta = p, \end{cases} \tag{37}$$

where (q, p) define a canonical local chart on S^2 according to Eq. (27). Then the corresponding vector fields on $\tilde{\Omega}_3$ are

$$\begin{aligned} X_\pm[J_1] &= l\sqrt{\frac{l \mp p}{l \pm p}} \cos q \frac{\partial}{\partial s} + \{J_1, \cdots\}, \\ X_\pm[J_2] &= l\sqrt{\frac{l \mp p}{l \pm p}} \sin q \frac{\partial}{\partial s} + \{J_2, \cdots\}, \\ X_\pm[J_3] &= l\frac{\partial}{\partial s} + \{J_3, \cdots\}. \end{aligned} \tag{38}$$

Assuming the existence of a global action $\tilde{\mathcal{K}}_{SO(3)}$ in the contact manifold $\tilde{\Omega}_3$, it is possible to define a representation $\mathfrak{R}$ within the Hilbert space $L_2(\tilde{\Omega}_3)$ analogous to Eq. (7). Then, we restrict to the subspace of functions ϕ such that

$$\phi \circ e^{\tau Z} = e^{in\tau/\hbar}\phi, \tag{39}$$

where Z is the *fundamental vector field*, parallel to the fibers of $\tilde{\Omega}_3$, which in any local chart is simply $\partial/\partial s$.
Then, in any local chart, Eq. (39) reads

$$\phi(q, p, s) = e^{ins/\hbar}\Psi(q, p), \tag{40}$$

analogous to Eq. (9).

Within the subspace of functions $\psi(qp)$ defined by Eq. (40), we obtain representations $\mathcal{R}^{(n/\hbar)}$ (analogous to $\mathcal{R}^{(\alpha)}$), which are now unitary reducible representations of the rotation group. The skew-adjoint infinitesimal operators of $SO(3)$ are obtained from Eq. (38) by means of the substitution $\partial/\partial s \to in/\hbar$. On the other hand, the operators corresponding to the operators K used by van Hove become

$$K[J_i] = (\hbar/n)H^{(n/\hbar)}[J_i] \tag{41}$$

and solve the Dirac problem for the angular momentum if $n = 1$. In order to reduce the representations $\mathcal{R}^{(n/\hbar)}$, it is profitable to use the connection with the theory of "classical" representations.[33] In the present case it suffices to remark that the operators $H_{\pm}^{(1/\hbar)}$ together with the multiplication operators $\frac{1}{\hbar}J_i$ provide a Hermitian irreducible representation of the "classical" algebra of the rotation group $SO(3)$, corresponding to fixed values of the invariants, given by

$$I_1 \equiv \sum_i \frac{J_i^2}{\hbar^2} = \left(\frac{l}{\hbar}\right)^2 = \left(\frac{N}{2}\right)^2, \qquad I_2 = \frac{\sum_i J_i H_i^{(1/h)}}{\sqrt{I_1}} = \frac{N}{2}. \tag{42}$$

The "classical" algebra of $SO(3)$ is isomorphic to the Lie algebra of the three-dimensional Euclidean group $E(3)$ through the correspondences $J_i \leftrightarrow P_i$, $H_i \leftrightarrow M_i$, being M_i and P_i the generators of the space rotations and translations, respectively.

Then the restriction with respect to the subgroup generated by the H_i [i.e., $SO(3)$] is obtained (see Pauli[34]) in the following form

$$\mathcal{R}^{(1/\hbar)} = \sum_{j=N/2}^{\infty \oplus} \mathfrak{D}^{(j)}, \quad j = \frac{N}{2}, \frac{N}{2}+1, \frac{N}{2}+2, \ldots. \tag{43}$$

Note that the representation $\mathcal{R}^{(1/\hbar)}$ contains all the values of the "angular momentum" higher than the classical value; also projective representations of $SO(3)$ can occur.

In the above considerations, we were not guaranteed that a global action of $SO(3)$ on the contact manifold actually existed. Thus all the results are formal, up to now. On the other hand, under certain conditions, the contact manifold can be directly constructed starting from the transitive realization $\mathcal{K}_{\mathfrak{G}}$ itself.[35] This is just the case for $SO(3)$ on S^2. In fact, $SO(3)$, together with its Maurer–Cartan invariant 1-form[36] $\theta = l\omega^3$, constitutes a contact manifold $\tilde{\Omega}_3 \equiv [SO(3), \theta]$. The fibers of this bundle are the left cosets of the subgroup of the rotations around the third axis. The base space is obviously S^2 and the normalized fundamental vector field Z is given by the left-invariant vector field corresponding to M_3/l, where M_3 generates the rotations around the third axis. Then, a local coordinate system (Euler angles) exists in the group manifold in which θ looks like

$$\theta = l(d\psi + \cos\vartheta d\varphi), \tag{44}$$

so that we recover the symplectic 2-form on S^2 induced by θ

$$\omega = ld(\cos\vartheta) \wedge d\varphi. \tag{45}$$

The global action of $SO(3)$ on itself defined by left translations leaves the 1-form θ(and the vector field Z!) invariant so that it is a global contact action. Then the unitary representation $\mathcal{R}$ of $SO(3)$ is nothing else than the regular representation, and the subspace condition (39) given above becomes

$$\phi \circ \exp(\tau M_3/l) = e^{in\tau/\hbar}\phi, \qquad 0 \leq \tau < 2\pi l \tag{46}$$

and in particular for $n = 1$

$$\phi \circ R_3(\psi) = e^{iN\psi/2}\phi, \qquad \text{where } l\psi = \tau. \tag{47}$$

This is just the condition defining the representation *induced* by $\psi \to e^{iN\psi/2}$ in Mackey theory. Then the reduction is the one given above (see Vilenkin[37]):

$$\begin{aligned}\phi(\varphi, \vartheta, \psi) &= e^{iN\psi/2}\phi(\varphi, \vartheta, 0) \\ &= e^{iN\psi/2}\sum_{j=|N/2|}^{\infty}\sum_{|m|\leq j}\alpha_m^j e^{im\varphi}P^j_{mN/2}(\cos\vartheta) \\ &= e^{iN\psi/2}\sum_{j\geq|N/2|}^{\infty}\sum_{|m|\leq j}\alpha_m^j e^{imq}P^j_{mN/2}(2p/N),\end{aligned} \tag{48}$$

where

$$\alpha_m^j = \frac{(-1)^{m-N/2}(2j+1)}{4\pi}\int_0^{2\pi}d\varphi\int_{-1}^{1}d(\cos\vartheta)\phi(\varphi, \theta, 0) \times e^{-im\varphi}P^j_{mN/2}(\cos\vartheta). \tag{49}$$

Thus the functions

$$t^j_{mN/2}(qp\psi) = e^{i(N\psi/2+mq)}P^j_{mN/2}(2p/N) \tag{50}$$

define an orthonormal basis in the subspace L_2^N defined by Eq. (47).

If we think of the point over S^2 as of a classical model of spin,[29] we can say to have quantized the system only if we have singled out an irreducible representation of $SO(3)$ corresponding to a fixed value j of the spin. In the present example, the most natural choice for the value j is clearly the lowest one, the classical value $j = N/2$. It is also clear that within the irreducible subspaces, canonical operators Q, P do not exist. This is to be expected on general grounds: Whenever the classical phase space is compact, the canonical coordinates q, p are not *global* coordinates; correspondingly, the *quantization group* is compact and the *quantal representation* is necessarily finite-dimensional.

Let us recall that, among the semisimple groups, only the three-dimensional ones, locally isomorphic to $SL(2, R)$, admit an invariant contact structure defined in a natural way within the group manifold.

3. DYNAMICAL QUANTIZATION

We expound now our procedure with a greater detail on technical aspects. Let A be a dynamical system with n degrees of freedom, Ω_{2n} its underlying symplectic manifold (phase space), and $H(q, p)$ the Hamiltonian function, in local coordinates. We make now the following assumptions[38]:

(α) The system A admits a maximal dynamical symmetry group $\mathcal{G}_0$ acting transitively on each energy surface Σ_E. Then $\Sigma_E \approx \mathcal{G}_0/K_0$, where K_0 is the stability subgroup of some point of Σ_E. As to the structure of $\mathcal{G}_0$, we require only that the derived algebra of its Lie algebra $\mathfrak{G}_0$ be $\mathfrak{G}_0$ itself.

Condition (α) implies that the Hamiltonian is a certain function $\mathcal{F}$ of the canonical invariants of $\mathcal{G}_0$ in the given canonical realization $\mathcal{K}_{\mathcal{G}_0}$, while the remaining independent functions of the canonical invariants are identically equal to constants. In general, the realization $\mathcal{K}_{\mathcal{G}_0}$ will be a *singular* one.[29] The existence of a transitive action of $\mathcal{G}_0$ on Σ_E implies that all the orbits of A on Σ_E are diffeomorphic one to another; the second part of the condition (α) guarantees, among the other things, that they are closed, as submanifolds, and consequently that A is completely degenerate, in the sense of Ref. 39. In particular, if the orbits are compact, the usual definition of complete degeneration is recovered.

(β) The canonical realization $\mathcal{K}_{\mathcal{G}_0}$ is homogeneous.

This means that the transformations of $\mathcal{K}_{\mathcal{G}_0}$ are homogeneous canonical transformations, which implies in particular that the canonical generators are certain homogeneous functions of the canonical variables.[40] Condition (β) is a sufficient condition (most likely it is also necessary) in order that the vector field corresponding to the function $\mathcal{F}^{-1}(H)$ [defined under (α)] generates a global action in Ω_{2n} of $SO(2)$ or R in the case of compact and noncompact orbits, respectively. This will allow to give a definition of a suitable "dynamical group" in which the Hamiltonian appears as a function of a single element of the Lie algebra. Let us call $\mathcal{D}$ this "Hamiltonian group" [$SO(2)$ or R] and $\mathcal{K}_{\mathcal{D}}$ its global action in Ω_{2n}.

(γ) A dynamical group $\tilde{\mathcal{G}}$ exists having the following properties: (i) It contains $\mathcal{G}_0 \otimes \mathcal{D}$ in such a way that $\mathcal{G}_0$ is the commutant of $\mathcal{D}$; (ii) it possess a global transitive canonical action $\mathcal{K}_{\tilde{\mathcal{G}}}$ in the phase space Ω_{2n}, which coincides with the realization defined in (α) when restricted to $\mathcal{G}_0 \otimes \mathcal{D}$.

Condition (γ) implies that the realization $\mathcal{K}_{\tilde{\mathcal{G}}}$ is always a *singular (degenerate)* realization, in which the generating function of the one-parameter subgroup $\mathcal{D}$ coincides with the free canonical invariant of $\mathcal{G}_0$.

It may happen that condition (α) is not strictly verified, in the sense that the dynamical system A admits a decomposition in different subsystems which meet conditions (α) and (β) separately. For instance, both the situations of compact and noncompact orbits may be realized corresponding to different open intervals of energy values (see, for example, the hydrogen atom case, in the following). In this case, condition (γ) must be formulated for these open submanifolds separately, since a realization $\mathcal{K}_{\tilde{\mathcal{G}}}$ transitive in the whole Ω_{2n} cannot exist on obvious topological grounds. What we can expect, for analytic Hamiltonians, is that, within the above submanifolds, there shall be defined canonical realizations of $\tilde{\mathcal{G}}$ which are analytic continuations of one another. It is worth noticing the strict analogy between this situation and the quantum one (see the comments about a theorem by Ingraham, in Sec. 5).

If now all the above conditions are satisfied, we obtain a dynamical quantization by means of the following final steps:

(δ) The Souriau prequantization is applied to the singular realization $\mathcal{K}_{\tilde{\mathcal{G}}}$. This provides a unitary reducible representation $\mathcal{R}^{(1/\hbar)}$ of $\tilde{\mathcal{G}}$ within the Hilbert space $L_2(\tilde{\Omega}, Z)$ of Lebesgue square-integrable functions on the contact manifold $\tilde{\Omega}_{2n+1}$ subject to the condition $\phi \circ \exp(\tau Z) = e^{i\tau/\hbar}\phi$.

We will not face here the technical problems of the construction of the contact manifold and of the prequantization of $\mathcal{K}_{\tilde{\mathcal{G}}}$. For the mathematical details the reader is referred to the work of Souriau, who has already given necessary conditions for the extension of the action of $\tilde{\mathcal{G}}$ on $\tilde{\Omega}_{2n+1}$. Sufficient conditions general enough for our purposes must yet be found; we shall discuss this point elsewhere.

Once the reduction of $\mathcal{R}^{(1/\hbar)}$ has been accomplished, the selection of the irreducible "quantal representation" $\tilde{\mathcal{R}}$ has to be made according to the following prescription:

(ϵ) The quantal irreducible representation of $\tilde{\mathcal{G}}$ must be such that every eigensubspace for the infinitesimal generator of the Hamiltonian group $\mathcal{D}$ carries an irreducible representation of its commutant $\mathcal{G}_0$ (the "degeneracy group").

The irreducibility condition accomplishes the transition from prequantization to quantization. Souriau himself has proposed some kind of restrictions for simple specifical cases but he does not formulate the general problem in terms suitable for our scheme. As said above, a definite answer to this question would be to find a procedure based on geometrical properties of $\tilde{\Omega}_{2n+1}$. Also, questions of analyticity are most likely to play a fundamental role in the selection of the quantum representation $\tilde{\mathcal{R}}$, as strongly suggested by some recent results by Auslander and Konstant,[19] Dunne[41] and Streater.[42]

We would add a final general remark about our quantization procedure. The definition of a Hamiltonian group $\mathcal{D}$, globally realized in the phase space Ω_{2n}, and its identification with a single element of the Lie algebra of $\tilde{\mathcal{G}}$, allows an intrinsic unambiguous Hamiltonian quantization. It is a remarkable fact that the usual intuitive correspondence between compact (noncompact) classical orbits and discrete (continuous) spectrum of the quantized Hamiltonian, finds a rigorous justification within our formulation. As a matter of fact, according to a theorem by Ingraham,[43] the spectrum of the quantized Hamiltonian within the irreducible representation $\tilde{\mathcal{R}}$ must be either completely discrete or completely continuous corresponding to $\mathcal{D}$ being compact or noncompact, respectively. It is important to realize that this theorem enables

us to anticipate the nature of the spectrum for nearly all the generators of $\tilde{\mathcal{G}}$ in $\tilde{\mathcal{R}}$ which are the relevant observables of the quantum system.

In the remainder of the present paper we will illustrate our general proposal in terms of the two outstanding physical examples of the n-dimensional harmonic isotropic oscillator and the n-dimensional hydrogen atom. The general group-theoretic properties of these systems at the quantum level are well known and have been discussed in many different contexts. For our purposes the most relevant references are 42, 44–53 (see also the bibliography of Ref. 54). Another system which could be easily treated is the rotator; however, we will not discuss it in the present paper.

4. TWO CLASSES OF DYNAMICALLY QUANTIZABLE SYSTEMS

A. n-Dimensional Isotropic Harmonic Oscillator

The maximal symmetry group $\mathcal{G}_0$ of the Hamiltonian is $SU(n)$. Its global canonical realization transitive on the generic energy surface Σ_E corresponds to a homogeneous space with stability subgroup $SU(n-1)$ (except for $E = 0$). Indeed

$$\Sigma_E \approx \begin{cases} \dfrac{SU(n)}{SU(n-1)} \approx S^{2n-1}, & E \neq 0, \\ \dfrac{SU(n)}{SU(n)} \approx \{0\}, & E = 0. \end{cases} \tag{51}$$

The whole phase space Ω_{2n} is filled by energy surfaces according to

$$\Omega_{2n} \equiv R^{2n} = \{0\} \cup [S^{2n-1} \times R]. \tag{52}$$

The Hamiltonian flow defines in R^{2n} a global action of $U(1)$ which, together with $SU(n)$, gives a global realization of $U(n)$. The explicit local realization can be found for instance in Ref. 54; the global one follows directly from the relation[55]

$$SO(2n) \cap Sp(n, R) = U(n). \tag{53}$$

The most simple noncompact extension of $U(n)$, namely $SU(n, 1)$, already meets all the requirements of condition (γ). The homogeneous space is

$$\Omega_{2n} = \frac{SU(n,1)}{U(n)} \approx \frac{SU(n) \times U(1) \times R^{2n}}{SU(n) \times U(1)} \approx R^{2n}. \tag{54}$$

The global realization of $SU(n, 1)$ in R^{2n} is most simply obtained starting from the linear symplectic action of $SU(n, 1)$ in $\mathbb{C}^{n+1}$ which leaves the hypersurfaces

$$\tilde{\Sigma}_k: \tilde{H} = \sum_1^n |z_i|^2 - |z_{n+1}|^2 = -k^2, \qquad k \text{ a real constant}, \tag{55}$$

invariant. $\tilde{\Sigma}_k$ is fibered by the Hamiltonian flow generated by $\tilde{H}$:

$$\mathbb{C}^{n+1} \times R \to \mathbb{C}^{n+1}: \begin{cases} z_i(\tau) = z_i(0) e^{i\tau} \\ z_{n+1}(\tau) = z_{n+1}(0) e^{-i\tau}, \end{cases} \tag{56}$$

and the action of $SU(n, 1)$ is compatible with the projection $\Pi: \tilde{\Sigma}_k \to \tilde{\Sigma}_k / X_{\tilde{H}}$; we have chosen k real so that $\tilde{\Sigma}_k \approx R^{2n} \times U(1)$, $\tilde{\Sigma}_k / X_{\tilde{H}} \approx R^{2n}$ and $\Pi(z_i, z_{n+1}) = z_k \exp(i \arg z_{n+1}) \equiv a_i \in \mathbb{C}^n$.

The action of $SU(n, 1)$ thus defined in R^{2n} is symplectic with respect to the fundamental 2-form $\omega = \frac{1}{2} i \sum_1^n da_i \wedge da_i^*$ induced by Π. The infinitesimal generators are given by $(dX_i^j)^{\#}$ with

$$X_i^j = \left[\begin{array}{c|c} a_i^* a_j + \dfrac{k^2}{n+1} \delta_{ij} & i a_i^* \sqrt{k^2 + \sum_1^n |a_i|^2} \\ \hline i a_i \sqrt{k^2 + \sum_1^n |a_i|^2} & -\sum_1^n |a_i|^2 - \dfrac{n}{n+1} k^2 \end{array}\right], \tag{57}$$

where

$$a_i = (1/\sqrt{2})(p_i + i q_i) \tag{58}$$

and the submatrix $a_i^* a_j + [k^2/(n+1)] \delta_{ij}$ generates $U(n)$.

Let us note that the construction of the realization of $SU(n, 1)$ directly provides a Souriau prequantization, the contact manifold being $\tilde{\Sigma}_k$ itself. In this case the phase space admits a global 1-form (*potential*) so that the contact manifold is a trivial bundle. The prequantization defines a reducible unitary representation of $SU(n, 1)$ in $L_2(R^{2n})$. The condition (ϵ) then selects the so-called "ladder" representation of $SU(n, 1)$, which is already known to provide the space of states for the isotropic n-dimensional harmonic oscillator. $SU(n, 1)$ is called, in this context, the "noninvariance" dynamical group of the system.[51,53]

The effectiveness of the usual correspondential quantization in terms of the Euclidean canonical variables can be recovered here owing to the fact that within the representation $\mathcal{R}$ raising and lowering operators $A_i, A_i^\dagger$ or self-adjoint canonical operators Q_i, P_i can be defined, in terms of which the generator of $U(1) \equiv \mathcal{D}$ takes the usual form.

For $n = 1$ the procedure is quite simple. A unitary reducible representation of $SU(1, 1)$ is obtained following Souriau prescriptions: The normalization conditions $\exp(2\pi\hbar Z) = 1, \theta(Z) = 1$, with Z proportional to the Hamiltonian vectorfield $X_H = (dH)^{\#}$, give $k^2 = 2\hbar$. The Hilbert space is defined as follows:

$$\mathcal{H}: \{\phi \in L_2(\mathbb{C}^2) \mid \phi(a_1 e^{i\tau}, a_2 e^{-i\tau}) = e^{i\tau} \phi(a_1, a_2)\}. \tag{59}$$

Now, $\tilde{\Sigma}_k$ is diffeomorphic to $SU(1, 1)$ through the mapping

$$a_1/\sqrt{2\hbar} \to \beta, \quad a_2/\sqrt{2\hbar} \to \alpha^*, \quad \alpha\alpha^* - \beta\beta^* = 1, \tag{60}$$

and $e^{\tau/Z}$ induces the rotations

$$\begin{pmatrix} e^{i\tau/2} & 0 \\ 0 & e^{-i\tau/2} \end{pmatrix};$$

this shows that our representation is simply the representation *induced* by $\tau \to e^{i\tau/2}$. The reduction is well known to yield

$$\mathcal{R}^{(1/\hbar)} = \mathcal{D}_+^{(1/2)} \oplus \mathcal{D}_-^{(1/2)} \oplus \int^{\oplus} \mathcal{D}^{(-1/2+i\rho)} d\mu(\rho). \tag{61}$$

The choice of $\mathcal{D}_+^{(1/2)}$ is forced by requiring that the energy spectrum be bounded from below. In this case the condition (ϵ) is not effective. Of course, the one-

dimensional systems are quite singular in our scheme, since the very concept of degeneration becomes trivial.

It happens here that also the group $E(2)$ satisfies all the requirements to be a quantization group for the one-dimensional oscillator. The global canonical realization in R^2 is a projective one in which q, p and the Hamiltonian are the infinitesimal canonical generators of the translations and of the rotations, respectively. In this way we recover in a straightforward manner the results of Streater.[15] It is remarkable that the results are quite the same, compared with the quantization by means of $SU(1,1)$, at least for what concerns the energy spectrum.

Streater discusses, in addition, the case of the potential λq^4 concluding that an intrinsic quantization with q and p irreducibly represented is not possible in this case. However this conclusion can be bypassed by our formulation. In fact action and angle variables define a diffeomorphism between this problem and the harmonic oscillator. Precisely if $H' = \frac{1}{2}(p^2 + \lambda q^4)$, we have

$$H_{\text{oscill}} = (\lambda^{-1/4}/12\sqrt{2}\pi^{3/2})[\Gamma(\tfrac{1}{4})]^2 (H')^{3/4}. \tag{62}$$

Thus our procedure gives the same results of Bohr–Sommerfeld method. This is true for all differentiable one-dimensional Hamiltonians $H = \frac{1}{2}p^2 + V(q)$ which admit only periodic orbits.

B. n-Dimensional Hydrogen Atom

As is well known, the maximal symmetry group for $E < 0$ is $SO(n+1)$; each energy surface is a homogeneous space with stability subgroup $SO(n-1)$; therefore, its topological structure is

$$\Sigma_E \approx \frac{SO(n+1)}{SO(n-1)} \equiv V_{n+1,2}, \tag{63}$$

(Stiefel manifold as defined, e.g., in Ref. 32, Sec. 7.7) while the entire space is

$$\Omega_{2n} = V_{n+1,2} \times R. \tag{64}$$

The realization of $SO(n+1)$ is well known in the cases $n = 2, 3$ (see Refs. 45–48, 52); for generic n it can be obtained by a straightforward generalization. Unlike the oscillator case, the Hamiltonian flow itself does not define a global action of $SO(2)$ in Ω_{2n} since the period of the motion is a function of the energy. A global action of $SO(2)$ in Ω_{2n} is generated by $X_{\mathfrak{F}^{-1}} = X_H/T_E = (dI)^{\#}$, where X_H is the Hamiltonian vector field, T_E is the period, and I is the quadratic invariant of $SO(n+1)$[56]

$$I = \sqrt{\tfrac{1}{2}\Sigma_{\mu,\nu} M_{\mu\nu} M_{\mu\nu}}, \qquad \mu, \nu = 1, \ldots, n+1. \tag{65}$$

The group $\tilde{\mathcal{G}}$ must contain the direct product $SO(n+1) \otimes SO(2)$ in such a way that the commutant of $SO(2)$ be exactly $SO(n+1)$: We are thus led to $\tilde{\mathcal{G}} = SO(n+1, 2)$. This group possesses just a symplectic singular realization within the homogeneous space

$$\frac{SO(n+1,2)}{[SO(n-1) \otimes SO(2,1)] \circledS [\mathcal{T}_{2n-2} \otimes \mathcal{T}_1]} \approx \frac{SO(n+1)}{SO(n-1)} \times R \equiv \Omega_{2n}, \tag{66}$$

where $\mathcal{T}_{2n-2} \circledS \mathcal{T}_1$ is a $(2n-1)$-dimensional Lie group, having a Euclidean topology; we denote by $\circledS$ the semidirect topological product. Let us see in a greater detail this realization. Call X_{ij}, $i,j = 1, \ldots, n+3$, the generators of $SO(n+1, 2)$, $M_{\mu\nu}$ the generators of the $SO(n+1)$ subgroup, $S = X_{n+2,n+3}$ the generator of the Hamiltonian group $\mathfrak{D}$, finally $Z_\mu = X_{\mu,n+2}$, and $W_\mu = X_{\mu,n+3}$. The above-mentioned realization corresponds to the orbit of the coadjoint representation of $\tilde{\mathcal{G}}$ through the point $\mathfrak{g} \in \tilde{\mathfrak{G}}^*$:

$$\mathfrak{g} = [X^{ij}] = \begin{bmatrix} 0 & 0 & 0 & \cdots & 0 & 0 & 0 \\ & 0 & 0 & \cdots & 0 & 0 & 0 \\ & & \cdot & & \cdot & \cdot & \cdot \\ & & & \cdot & \cdot & \cdot & \cdot \\ & & & & \cdot & \cdot & \cdot \\ & & & & 1 & 1 & 0 \\ & & & & & 0 & 1 \\ & & & & & & 1 \end{bmatrix} \tag{67}$$

X^{ij} being the dual base of X_{ij}. This orbit is a submanifold of $\tilde{\mathfrak{G}}^*$ defined by the following relations:

$$\begin{aligned} &SM^{\mu\nu} = Z^\mu W^\nu - Z^\nu W^\mu, \\ &W^\lambda Z^\lambda = 0, \\ &Z^\lambda Z^\lambda = W^\lambda W^\lambda = S^2. \end{aligned} \tag{68}$$

A basis for the Lie algebra of the stability subgroup is given by

M_{ab}, $a, b = 1, \ldots, n-1$, generators of $SO(n-1)$,

$Z_{n+1} + W_n$, $Z_n - S$, $W_{n+1} - S$, $M_{n,n+1} + S - Z_n - W_{n+1}$, generators of $SO(2,1) \otimes \mathcal{T}_1$,

$M_{in} + W_i, M_{i,n+1} - Z_i$, generators of $\mathcal{T}_{2n-2}$.

The explicit global structure of the stability subgroup has been easily derived in the most interesting cases $n = 2, 3$, where the special isomorphism $SO(4,2) \sim SU(2,2)$ greatly simplifies the calculations. The geometrical structure of Ω_{2n} is made more apparent by means of the following considerations: A point of the orbit is characterized by two mutually orthogonal $(n+1)$-vectors [with respect to $SO(n+1)$] of magnitude S; for S fixed the possible choices of Z^μ and W^μ are in a one-to-one correspondence with the maximal Abelian subalgebras of P in the Cartan decomposition $\mathfrak{so}(4,2) = [\mathfrak{so}(4) \oplus \mathfrak{so}(2)] \oplus \mathfrak{p}$ so that the transitivity of $SO(n+1) \otimes SO(2)$ on this set follows from a general theorem.[55]

That this realization is actually a global canonical one is guaranteed by a theorem on co-adjoint representations (see Refs. 30, 18, 12). An explicit atlas of canonical coordinates in Ω_{2n} can be constructed using the techniques of Ref. 29. The Hamiltonian flow, generated by[57] $S = mK/\sqrt{-2mH}$, in the coordinates given above, is

$$\begin{aligned} &M^{\mu\nu} = \text{const}, \\ &Z^\mu(t) = Z^\mu(0)\cos t + W^\mu(0)\sin t, \\ &W^\mu(t) = -Z^\mu(0)\sin t + W^\mu(0)\cos t. \end{aligned} \tag{69}$$

Let us remark that if we restrict ourselves to the noncompact subgroup $SO(n+1,1)$, the action in Ω_{2n} is still transitive.

Now, the Souriau prequantization of this realization provides a unitary reducible representation in the Hilbert space $L_2(\Omega_{2n})$. For the physically interesting case $n=3$, the selection made according to condition (ϵ) yields the so-called R_0 irreducible unitary representation of $SO(4,2)$, which has just the following remarkable properties[58]:

(i) When restricted to the subgroup $\mathcal{G}_0 = SO(4)$, the representation R_0 splits into the direct sum of tensor representations $\mathfrak{D}(l,l)$ each $\mathfrak{D}(l,l)$ appearing with multiplicity one.

(ii) The l^2-dimensional subspace $\mathfrak{H}(l)$, invariant under $SO(4)$, is an eigenspace for the generator S of the Hamiltonian flow which commutes with $SO(4)$: $f_l \in \mathfrak{H}(l) \to Sf_l = lf_l$.

(iii) R_0 remains irreducible when restricted to $SO(4,1)$; this has the already mentioned classical counterpart.

Let us stress the following points:

(1) This irreducible representation R_0 is already well known to describe the negative energy states for the three-dimensional hydrogen atom.[49,50] In Ref. 50 the quantization group $\tilde{\mathcal{G}}$ is called the "dynamical group" of the hydrogen atom in the sense that all the electromagnetic bound-bound transitions can be described in terms of operators belonging to $SO(4,2)$ and to its enveloping algebra. The subgroup $SO(4,1)$ which is irreducibly represented in the same space is called there the "quantum numbers" group, emphasizing the fact that it is sufficient to describe the whole spectrum of the Hamiltonian and of the conserved quantities. From our point of view, the dynamical character of $SO(4,2)$ in front of $SO(4,1)$ is connected with the fundamental fact that $SO(4,2)$ contains the Hamiltonian directly as a generator. On the contrary, within the canonical realization of $SO(4,1)$, the Hamiltonian must be defined in the enveloping algebra of the invariance group G_0 with the consequence that an intrinsic quantization based on $SO(4,1)$ would not be a Hamiltonian quantization. On the other hand, the fact that both the canonical realization and the unitary representation of $SO(4,2)$ are irreducible with respect to $SO(4,1)$ accounts for the use of this group as a suitable "noninvariance" group for the hydrogen atom.[47,48,51-53]

(2) As shown by Barut and Kleinert (see Ref. 50), within the representation R_0, it is possible to define canonical operators, irreducibly represented, in terms of which the Hamiltonian recovers its correspondential structure. From our point of view, this fact accounts for the success of the correspondential quantization of the hydrogen atom.

(3) If we consider the positive energy portion of phase space, we have a maximal symmetry group $\mathcal{G}_0'$ which is an "analytic continuation" of $\mathcal{G}_0$, actually $SO(3,1)$.[47] The dynamical quantization procedure should still be possible within the dynamical group $SO(4,2)$ but with the Hamiltonian identified with a non-compact generator e.g.with X_{46}, through "analytic continuation" (see the discussion given in Sec. 3).

ACKNOWLEDGMENTS

We thank warmly Professor F. Duimio for his kind interest in this work.

Note added in proof: More recent results show that questions of analyticity play an effective role in the present theory, as anticipated at the end of Sec. 3. Actually the choice of $\tilde{\mathcal{G}}$ *and* the selection of $\mathcal{R}$ can be reduced to the choice of a suitable *complex structure* in the phase space: E. Onofri and M. Pauri, "Analyticity and Quantization", to appear in "Lettere al Nuovo Cimento."

[1] W. Pauli, "Die Allgemeinen Prinzipien der Wellenmechanik," *Handbuch der Physik* (Springer-Verlag, Berlin, 1958), Bd. V, Teil 1, Sec. A. 5.
[2] L. D. Dureau, Amer. J. Phys. **33**, 895 (1965).
[3] K. Simon, Amer. J. Phys. **33**, 60 (1965).
[4] H. J. Groenewold, Physica **12**, 405 (1946).
[5] J. R. Shewell, Amer. J. Phys. **27**, 16 (1959).
[6] D. C. Rivier, Phys. Rev. **83L**, 862 (1951).
[7] J. P. Amiet, Nuovo Cimento **29**, 69 (1963).
[8] R. Arens and D. Babbitt, J. Math. Phys. **6**, 1071 (1965).
[9] A. Joseph, Commun. Math. Phys. **17**, 210 (1970).
[10] L. van Hove, Acad. Roy. Belg. Bull. Classe Sci. Mem. **37**, 610 (1951).
[11] U. Uhlhorn, Arkiv Fysik **11**, 87 (1956).
[12] J. M. Souriau, Commun. Math. Phys. **1**, 374 (1966); Ann. Inst. Henri Poincaré **4**, 311 (1967); *Structure des systèmes dynamiques* (Dunod, Paris, 1970).
[13] R. Hermann, *Lie Groups for Physicists* (Benjamin, New York, 1966), Chap. 16.
[14] I. E. Segal, Mat.-Fys. Medd. Dansk. Vid. Selsk. **31**, no. 12 (1959); Can. J. Math. **13**, 1 (1961); J. Math. Pures Appl. **44**, 71 (1965); J. Math. Phys. **1**, 468 (1960).
[15] R. F. Streater, Commun. Math. Phys. **2**, 354 (1966).
[16] N. E. Hurt, Nuovo Cimento **55A**, 534 (1968); **58B**, 361 (1968); J. Math. Phys. **11**, 539 (1970).
[17] Actually it covers the completely degenerate systems at most. It is a peculiar fact that in this way an intrinsic quantization can be performed for just those systems for which the classical Bohr-Sommerfeld method ran into logical difficulties.
[18] B. Kostant, *Lecture Notes in Mathematics*, edited by C. T. Taam (Springer, New York, 1970), Vol. 170.
[19] L. Auslander and B. Kostant, Bull. Amer. Math. Soc. **73**, 692 (1967).
[20] R. Abraham, *Foundations of Mechanics* (Benjamin, New York, 1967).
[21] V. I. Arnold and A. Avez, *Problèmes ergodiques de la mécanique classique* (Gauthier-Villars, Paris, 1967).
[22] An example from van Hove: $f(q,p) = -pq^m$, m integer greater than 1. The transformed variable $q' = q[1-(m-1)q^{m-1}\tau]^{1/1-m}$ becomes infinite for $(m-1)q^{m-1}\tau = 1$.
[23] (0) Of course, the precise meaning of the first Eq. (14) is that whenever the commutator is self-adjoint (essentially) its maximal extension is given by the rhs operator.
[24] P. A. M. Dirac, *Quantum Mechanics* (Oxford U.P., Oxford, 1947), 3rd ed. Chap. VI.
[25] H. Weyl, *The Theory of Groups and Quantum Mechanics* (Dover, New York, 1950), Chap. IV, D.
[26] C. Freifield, *Battelle Rencontres 1967*, edited by C. M. DeWitt and J. A. Wheeler (Benjamin, New York, 1968), Chap. XXII.
[27] R. S. Palais, Mem. Amer. Math. Soc. **22**, Chap. IV (1957).
[28] I. M. Gel'fand and M. I. Graev, Amer. Math. Soc. Transl. **37**, 351 (1964).
[29] M. Pauri and G. M. Prosperi, J. Math. Phys. **7**, 366 (1966); **8**, 2256 (1967); **9**, 1146 (1968).
[30] A. A. Kirillov, Usp. Mat. Nauk **17**, 57 (1962).
[31] J. Moser, Commun. Pure Appl. Math. **23**, 609 (1970).
[32] N. Steenrod, *The Topology of Fibre Bundles* (Princeton U.P., Princeton, N.J., 1951).
[33] G. Lugarini and M. Pauri, Ann. Phys. (N.Y.) **38**, 299 (1966)
[34] W. Pauli, "Continuous Groups in Quantum Mechanics," CERN Preprint 56-31, Geneva, 1956, Chap. III.
[35] The problem of the intrinsic prequantization of a homogeneous symplectic manifold has been already dealt with by Hurt[16];

however, some of the hypotheses assumed by this author are somewhat restrictive for our purposes; altogether, the example of $SO(3)$ is included in his classification.
[36] S. Sternberg, *Lectures on Differential Geometry* (Prentice-Hall, Englewood Cliffs, N.J., 1964).
[37] N. Ja. Vilenkin, *Fonctions spéciales et théorie de la representations des groupes* (Dunod, Paris, 1969), Chap. III, Sec. 6.4.
[38] The present scheme of assumptions has a provisional character; in particular, the relative independence of the various hypotheses, as stated here, is rather dubious.
[39] E. Onofri and M. Pauri, University of Parma, Preprint IFPR-T-010, 1970, J. Math. Phys. (to be published)
[40] L. P. Eisenhart, *Continuous Groups of Transformations* (Dover, New York, 1961).
[41] S. A. Dunne, J. Math. Phys. **10**, 860 (1969).
[42] R. F. Streater, Commun. Math. Phys. **4**, 217 (1967).
[43] R. L. Ingraham, ICTP, Preprint ICTP, no. 9, 1967.
[44] V. Fock, Z. Phys. **98**, 145 (1935).
[45] H. Bacry, H. Ruegg, and J. M. Souriau, Commun. Math. Phys. **3**, 323 (1966).
[46] E. Onofri and M. Pauri, Lett. Nuovo Cimento **1**, 607 (1967).
[47] M. Bander and C. Itzykson, Rev. Mod. Phys. **38**, 330, 346 (1966).
[48] H. Bacry, CERN Report TH 579, Geneva, 1965.
[49] C. Fronsdal, Phys. Rev. **156**, 1665 (1967).
[50] H. M. Kleinert, in *Lectures in Theoretical Physics*, edited by W. E. Brittin and A. O. Barut (Gordon and Breach, New York, 1968); Fortschr. Physik **16**, 1 (1968).
[51] B. Vitale, "'Invariance' and 'Noninvariance' Dynamical Groups," lecture notes, Institute of Mathematical Sciences, Madras, 1966.
[52] N. Mukunda, L. O'Raifeartaigh, and E. C. G. Sudarshan, Phys. Rev. Letters **15**, 1041 (1965).
[53] G. Cicogna, Nuovo Cimento **49**, 291 (1967).
[54] F. Duimio and M. Pauri, Nuovo Cimento **51A**, 1141 (1967).
[55] S. Helgason, *Differential Geometry and Symmetric Spaces* (Academic, New York, 1962).
[56] The other homogeneous invariants are all zero, so that the realization satisfies condition (β).
[57] For the meaning of these parameters, see Eq. (20).
[58] I. T. Todorov, "Derivation and Solution of an Infinite-Component Wave Equation for the Relativistic Coulomb Problem," in *Group Representations in Mathematics and Physics*, Battelle Seattle 1969 Rencontres, edited by V. Bargmann (Springer-Verlag, Berlin, 1970).

Note on the Angular Momentum and Mass of Gravitational Geons

Jeffrey M. Cohen
Physics Department, University of Pennsylvania, Philadelphia, Pennsylvania, 19104
and
Robert M. Wald
Physics Department, Princeton University, Princeton, New Jersey 08540
(Received 1 November 1971)

It is shown that (1) the angular momentum of a gravitational geon must be zero if it is axisymmetric and (2) the mass of a gravitational geon must be zero if it is stationary, i.e., if the space–time possesses a Killing vector which is timelike at infinity. Here angular momentum and mass are defined in terms of the asymptotic form of the metric at large distances; they are physical quantities which can be experimentally measured by distant observers. Since the gravitational geons previously considered are highly dynamical on a small scale, our result on the vanishing mass of a stationary geon does not conflict with previous analyses showing that gravitational geons can have mass. Similarly, our results do not exclude the possibility of gravitational geons having nonvanishing angular momentum if they are not strictly axisymmetric.

1. INTRODUCTION

A gravitational geon may be described physically as a localized region of pure space–time curvature. More precisely, we define a gravitational geon to be a solution of the vacuum Einstein field equations,

$$G_{\mu\nu} = 0 \tag{1}$$

which is (1) nonsingular, (2) topologically Euclidean, (3) asymptotically flat, i.e., there exist coordinates x^μ such that on the hypersurfaces $x^0 = \text{const}$ the metric takes the form $g_{\mu\nu} = \eta_{\mu\nu} + O(1/r)$ at large distances, where $\eta_{\mu\nu} = \text{diag}(-1, 1, 1, 1)$ and r is a radial parameter, and (4) approximately stationary in the asymtotically flat region, i.e., for sufficiently large r, derivatives of the metric with respect to x^0 can be neglected compared with derivatives with respect to the spacelike coordinates x^i. Gravitational geons as well as electromagnetic and neutrino geons have been studied as models for material bodies free from the uncertainty about any equations of state.[1]

In this paper, we prove that a gravitational geon cannot have a nonvanishing angular momentum if it is axisymmetric. We also show that the mass of a gravitational geon must vanish if it is stationary.

In Sec. 2 we review the definition of angular momentum and mass used in this paper. We obtain expressions for these quantities in Sec. 3 which are used in Sec. 4 to prove our results on gravitational geons.

2. DEFINITION OF ANGULAR MOMENTUM AND MASS

The discussion of this section follows closely that of Misner, Thorne, and Wheeler.[2]

The space–time metric of any asymptotically flat solution of Einstein's equations which is approximately stationary in the asymptotically flat region can be put in the following form[2] for large r:

$$dS^2 = -\left(1 - \frac{2m}{r} + \frac{2m^2}{r^2}\right)dt^2 - 4\epsilon_{jkl}J^k\left(\frac{x^l}{r^3}\right)dt\,dx^j + \left(1 + \frac{2m}{r} + \frac{3m^2}{2r^2}\right)\delta_{jk}dx^j dx^k + O\left(\frac{1}{r^3}\right)dx^\mu dx^\nu. \tag{2}$$

Here Roman indices run from 1 to 3, Greek indices run from 0 to 3, and ϵ_{jkl} is the completely antisymmetric tensor. The parameters m and $J = ((J^1)^2 + (J^2)^2 + (J^3)^2)^{1/2}$ of a space–time are uniquely defined by Eq. (2), i.e., their values cannot be changed by a coordinate transformation which preserves the form, Eq. (2), of the metric. If the gravitational field is weak throughout the space–time, the linearized theory of gravity yields the following expressions[2] for m and J^k:

$$m = \int T^{oo}d^3x^i, \tag{3}$$

$$J^k = \epsilon_{kln}\int (x^l T^{no} - x^n T^{lo})d^3x^i, \tag{4}$$

however, some of the hypotheses assumed by this author are somewhat restrictive for our purposes; altogether, the example of $SO(3)$ is included in his classification.
[36] S. Sternberg, *Lectures on Differential Geometry* (Prentice-Hall, Englewood Cliffs, N.J., 1964).
[37] N. Ja. Vilenkin, *Fonctions spéciales et théorie de la representations des groupes* (Dunod, Paris, 1969), Chap. III, Sec. 6.4.
[38] The present scheme of assumptions has a provisional character; in particular, the relative independence of the various hypotheses, as stated here, is rather dubious.
[39] E. Onofri and M. Pauri, University of Parma, Preprint IFPR-T-010, 1970, J. Math. Phys. (to be published)
[40] L. P. Eisenhart, *Continuous Groups of Transformations* (Dover, New York, 1961).
[41] S. A. Dunne, J. Math. Phys. **10**, 860 (1969).
[42] R. F. Streater, Commun. Math. Phys. **4**, 217 (1967).
[43] R. L. Ingraham, ICTP, Preprint ICTP, no. 9, 1967.
[44] V. Fock, Z. Phys. **98**, 145 (1935).
[45] H. Bacry, H. Ruegg, and J. M. Souriau, Commun. Math. Phys. **3**, 323 (1966).
[46] E. Onofri and M. Pauri, Lett. Nuovo Cimento **1**, 607 (1967).
[47] M. Bander and C. Itzykson, Rev. Mod. Phys. **38**, 330, 346 (1966).
[48] H. Bacry, CERN Report TH 579, Geneva, 1965.
[49] C. Fronsdal, Phys. Rev. **156**, 1665 (1967).
[50] H. M. Kleinert, in *Lectures in Theoretical Physics*, edited by W. E. Brittin and A. O. Barut (Gordon and Breach, New York, 1968); Fortschr. Physik **16**, 1 (1968).
[51] B. Vitale, "'Invariance' and 'Noninvariance' Dynamical Groups," lecture notes, Institute of Mathematical Sciences, Madras, 1966.
[52] N. Mukunda, L. O'Raifeartaigh, and E. C. G. Sudarshan, Phys. Rev. Letters **15**, 1041 (1965).
[53] G. Cicogna, Nuovo Cimento **49**, 291 (1967).
[54] F. Duimio and M. Pauri, Nuovo Cimento **51A**, 1141 (1967).
[55] S. Helgason, *Differential Geometry and Symmetric Spaces* (Academic, New York, 1962).
[56] The other homogeneous invariants are all zero, so that the realization satisfies condition (β).
[57] For the meaning of these parameters, see Eq. (20).
[58] I. T. Todorov, "Derivation and Solution of an Infinite-Component Wave Equation for the Relativistic Coulomb Problem," in *Group Representations in Mathematics and Physics*, Battelle Seattle 1969 Rencontres, edited by V. Bargmann (Springer-Verlag, Berlin, 1970).

Note on the Angular Momentum and Mass of Gravitational Geons

Jeffrey M. Cohen
Physics Department, University of Pennsylvania, Philadelphia, Pennsylvania, 19104
and
Robert M. Wald
Physics Department, Princeton University, Princeton, New Jersey 08540
(Received 1 November 1971)

It is shown that (1) the angular momentum of a gravitational geon must be zero if it is axisymmetric and (2) the mass of a gravitational geon must be zero if it is stationary, i.e., if the space–time possesses a Killing vector which is timelike at infinity. Here angular momentum and mass are defined in terms of the asymptotic form of the metric at large distances; they are physical quantities which can be experimentally measured by distant observers. Since the gravitational geons previously considered are highly dynamical on a small scale, our result on the vanishing mass of a stationary geon does not conflict with previous analyses showing that gravitational geons can have mass. Similarly, our results do not exclude the possibility of gravitational geons having nonvanishing angular momentum if they are not strictly axisymmetric.

1. INTRODUCTION

A gravitational geon may be described physically as a localized region of pure space–time curvature. More precisely, we define a gravitational geon to be a solution of the vacuum Einstein field equations,

$$G_{\mu\nu} = 0 \tag{1}$$

which is (1) nonsingular, (2) topologically Euclidean, (3) asymptotically flat, i.e., there exist coordinates x^μ such that on the hypersurfaces $x^0 = \text{const}$ the metric takes the form $g_{\mu\nu} = \eta_{\mu\nu} + O(1/r)$ at large distances, where $\eta_{\mu\nu} = \text{diag}(-1, 1, 1, 1)$ and r is a radial parameter, and (4) approximately stationary in the asymtotically flat region, i.e., for sufficiently large r, derivatives of the metric with respect to x^0 can be neglected compared with derivatives with respect to the spacelike coordinates x^i. Gravitational geons as well as electromagnetic and neutrino geons have been studied as models for material bodies free from the uncertainty about any equations of state.[1]

In this paper, we prove that a gravitational geon cannot have a nonvanishing angular momentum if it is axisymmetric. We also show that the mass of a gravitational geon must vanish if it is stationary.

In Sec. 2 we review the definition of angular momentum and mass used in this paper. We obtain expressions for these quantities in Sec. 3 which are used in Sec. 4 to prove our results on gravitational geons.

2. DEFINITION OF ANGULAR MOMENTUM AND MASS

The discussion of this section follows closely that of Misner, Thorne, and Wheeler.[2]

The space–time metric of any asymptotically flat solution of Einstein's equations which is approximately stationary in the asymptotically flat region can be put in the following form[2] for large r:

$$dS^2 = -\left(1 - \frac{2m}{r} + \frac{2m^2}{r^2}\right)dt^2 - 4\epsilon_{jkl}J^k\left(\frac{x^l}{r^3}\right)dt\,dx^j + \left(1 + \frac{2m}{r} + \frac{3m^2}{2r^2}\right)\delta_{jk}dx^j dx^k + O\left(\frac{1}{r^3}\right)dx^\mu dx^\nu. \tag{2}$$

Here Roman indices run from 1 to 3, Greek indices run from 0 to 3, and ϵ_{jkl} is the completely antisymmetric tensor. The parameters m and $J = ((J^1)^2 + (J^2)^2 + (J^3)^2)^{1/2}$ of a space–time are uniquely defined by Eq. (2), i.e., their values cannot be changed by a coordinate transformation which preserves the form, Eq. (2), of the metric. If the gravitational field is weak throughout the space–time, the linearized theory of gravity yields the following expressions[2] for m and J^k:

$$m = \int T^{oo}d^3x^i, \tag{3}$$

$$J^k = \epsilon_{kln}\int (x^l T^{no} - x^n T^{lo})d^3x^i, \tag{4}$$

where $T^{\mu\nu}$ is the stress–energy tensor of matter. Thus, in the weak field limit, m and J may be identified, respectively, as the total mass and angular momentum. In the strong field case, Eqs. (3) and (4) are, of course, no longer valid, but the expansion of the metric, Eq. (2), still holds in the asymptotically flat region. In the strong field case, we *define* the total (active gravitational) mass to be m and the total angular momentum to be J. Both m and J have direct physical significance: A distant observer can measure m by a study of Keplerian orbits and can measure J by observation of gyroscope precession resulting from the dragging of inertial frames.

Transforming from the symptotically Minkowskian coordinates of Eq. (2) to asymptotically spherical polar coordinates and aligning the z axis in the direction of J, we put the metric of Eq. (2) into the following form which is more useful for our purposes:

$$g_{\mu\nu} = \begin{pmatrix} \underline{t} & \underline{\varphi} & \underline{r} & \underline{\theta} \\ -[1 - 2m/r + O(1/r^2)] & [(-2J\sin^2\theta)/r + O(1/r^2)] & O(1/r^3) & O(1/r^2) \\ & r^2\sin^2\theta(1 + O(1/r)) & O(1/r^2) & O(1/r) \\ \text{SYM} & & 1 + O(1/r) & O(1/r^2) \\ & & & r^2[1 + O(1/r)] \end{pmatrix} \tag{5}$$

Note that, comparing Eq. (5) with the Kerr metric,

$$dS^2 = -\left(1 - \frac{2mr}{\Sigma}\right)dt^2 - \frac{4mar\sin^2\theta}{\Sigma}\,dt\,d\varphi + \left((r^2 + a^2)\sin^2\theta + \frac{2ma^2r\sin^4\theta}{\Sigma}\right)d\varphi^2 + \Sigma\left(d\theta^2 + \frac{dr^2}{\Delta}\right), \tag{6}$$

where

$$\Sigma = r^2 + a^2\cos^2\theta, \tag{7}$$

$$\Delta = r^2 - 2mr + a^2, \tag{8}$$

we can immediately see that the angular momentum of the Kerr metric is given by[3]

$$J = ma. \tag{9}$$

3. FORMULAS FOR J AND m

We now obtain formulas for J and m which will be used in Sec. 4.

Let ξ^μ be any vector field which reduces to the vector field $\partial/\partial\varphi$ at large distances, where φ is the angular coordinate defined at large distances by the form of the metric, Eq. (5). (The definition of ξ^μ is left arbitrary in the nonasymptotically flat region.) Then, we have[3]

$$J = \frac{1}{16\pi}\lim_{r\to\infty}\int_{r,t\,\text{const}} {}^*d\xi = \frac{1}{16\pi}\lim_{r\to\infty}\int_{r,t\,\text{const}}(-g)^{1/2}(\xi^{r;t} - \xi^{t;r})d\theta d\varphi, \tag{10}$$

where t, φ, r, θ are the coordinates of Eq. (5). To prove (10), we note that since ξ^μ agrees with $\partial/\partial\varphi$ for large r, we obtain by direct calculation from the metric of Eq. (5) that

$$\xi^{r;t} = g^{t\alpha}\Gamma^r_{\varphi\alpha} = 3J\sin^2\theta/r^2 + O(1/r^3), \tag{11}$$

$$\xi^{t;r} = g^{r\alpha}\Gamma^t_{\varphi\alpha} = -3J\sin^2\theta/r^2 + O(1/r^3), \tag{12}$$

$$(-g)^{1/2} = r^2\sin\theta\,[1 + O(1/r)]. \tag{13}$$

Equation (10) then follows immediately.

A similar calculation establishes the following formula for m. Let ψ^μ be any vector field which agrees with $\partial/\partial t$ for large values of r. Then we have[4]

$$m = -\frac{1}{8\pi}\lim_{r\to\infty}\int_{r,t\,\text{const}} {}^*d\psi = \frac{1}{8\pi}\int(-g)^{1/2}(\psi^{t;r} - \psi^{r;t})d\theta d\varphi \tag{14}$$

4. APPLICATION TO GRAVITATIONAL GEONS

The results of this paper now follow from Eqs. (10) and (14).

The assumption that the space–time is topologically Euclidean implies that the 2-surface of constant r and t over which the integral, Eq. (10), is to be taken is the boundary of the interior part of the hypersurface, t = const. Hence, the divergence theorem implies

$$J = \frac{1}{16\pi}\int_{t=\text{const}} {}^*d\xi = \frac{1}{16\pi}\int_{t=\text{const}}(-g)^{1/2}(\xi^{\mu;t} - \xi^{t;\mu})_{;\mu}\,dr d\theta d\varphi. \tag{15}$$

If the space–time is axisymmetric, we may take ξ^μ to be the axisymmetric Killing vector. Now, for a Killing vector, we have

$$\xi_{\mu;\nu} + \xi_{\nu;\mu} = 0. \tag{16}$$

Hence, we have

$$(\xi^{\mu;t} - \xi^{t;\mu})_{;\mu} = 2\xi^{\mu;t}{}_{;\mu}. \tag{17}$$

In addition,

$$\xi^\mu{}_{;\mu} = 0, \tag{18}$$

and by the Ricci identity we thus obtain

$$\xi^{\mu;\nu}{}_{;\mu} = \xi^{\mu;\nu}{}_{;\mu} - \xi^\mu{}_{;\mu}{}^{;\nu} = R_\mu{}^\nu\xi^\mu. \tag{19}$$

Thus, for an axisymmetric, topologically Euclidean space–time, we have

$$J = \frac{1}{8\pi}\int_{t=\text{const}}(-g)^{1/2}R_{\varphi}{}^{t}d^{3}x^{i}. \tag{20}$$

[Note that, in Eq. (20), φ is fixed by the symmetry but the choice of t is arbitrary except in the asymptotically flat region.] For a gravitational geon $R_{\mu}{}^{\nu}$ is zero by virtue of the field equation (1), and so for an axisymmetric gravitational geon Eq. (20) yields

$$J = 0. \tag{21}$$

An identical calculation starting from Eq. (14) shows that for a stationary, topologically Euclidean space-time,

$$m = -\frac{1}{4\pi}\int_{t=\text{const}}(-g)^{1/2}R_{t}{}^{t}d^{3}x^{i}. \tag{22}$$

(Here t is fixed throughout the space-time by the stationary symmetry.) Hence, for a stationary gravitational geon,

$$m = 0. \tag{23}$$

Since $G_{\varphi}{}^{t} = R_{\varphi}{}^{t}$, it follows from Eq. (20) that J is a conserved quantity associated with axial symmetry and arising from the conservation law $(G_{\mu}{}^{\nu}\xi^{\mu})_{;\nu} = 0$. This quantity was used in Ref. 3 to define the angular momentum of an axisymmetric space-time. However, since $G_{t}{}^{t} = R_{t}{}^{t} - \frac{1}{2}R$, we see from Eq. (22) that in the stationary case m is *not* (in general) equal to the conserved quantity arising from $(G^{\nu}{}_{\mu}\psi^{\mu})_{;\nu} = 0$, except for space-times with vanishing scalar curvature R (e.g., electiovac space-times).

Note that if matter is present in the interior, one obtains $m > 0$ for a stationary, topologically Euclidean space-time if $R_{t}{}^{t} = T_{t}{}^{t} - \frac{1}{2}T < 0$. For perfect fluid matter this condition becomes $\rho + 3p > 0$.

The proof of Eqs. (21) and (23) does not apply to black holes because for black holes the spacelike hypersurfaces either contain a singularity or have non-Euclidean ("wormhole") topology. In either case, Eq. (15) does not follow from Eq. (10). For further discussion see, e.g., Ref. 5 and the references cited there.

It should be emphasized that, in the proof of Eq. (21), it is required that the geon be strictly axisymmetric, i.e., axisymmetric on a small scale, not merely approximately axisymmetric when averaged over some region. Similarly, in the proof of Eq. (23), the geon must be strictly stationary. Since the gravitational geons previously considered are highly dynamical on a small scale, our results do not conflict with analyses which find them to have positive mass.[1,6] Nor do our results exclude the existence of geons having angular momentum which are not axisymmetric on a small scale, e.g., on account of gravitational waves traveling in the φ direction.

ACKNOWLEDGMENTS

For reading the manuscript and making helpful comments concerning the exposition we are indebted to Professor D. R. Brill, Professor J. A. Wheeler, and Professor J. W. York. One of us (JMC) would like to thank Dr. Carl Kaysen for his hospitality at the Institute for Advanced Study, where part of this work was done. This work was supported in part by the National Science Foundation under Grants GP-16147, GP-7669, and GP-30799X and by the U.S. Atomic Energy Commission under grant 2171T.

---

[1] J. A. Wheeler, *Geometrodynamics* (Academic, New York, 1962).
[2] C. W. Misner, K. S. Thorne, and J. A. Wheeler, *Gravitation*, (to be published). For alternative approaches, see, e.g., L. D. Landau and E. M. Lifshitz, *Classical Theory of Fields* (Addison-Wesley, Reading, Mass., 1962), E. Lubkin, Rosen Festschrift (1968), A. Trautman, in *Gravitation*, edited by L. Witten (Wiley, New York, 1962), L. Tamburino and J. Winicour, Phys. Rev. **150**, 1039 (1966).
[3] J. M. Cohen, J. Math. Phys. **9**, 905 (1968). Note that for an axisymmetric space-time the definition of angular momentum used in this reference is equivalent to the one given here. In Footnotes 5 and 9 a coordinate independent procedure is given.
[4] A. Komar, Phys. Rev. **127**, 1411 (1962). For a discussion of Komar's approach to the definition of energy, see C. W. Misner, Phys. Rev. **130**, 1590.
[5] J. M. Cohen, Rosen Festschrift (1968), notes published in *Relativity and Gravitation*, edited by G. Kuper and A. Peres (Gordon and Breach, New York, 1971).
[6] D. Brill and J. Hartle, Phys. Rev. **135B**, 271 (1964).

The Einstein Equations of Evolution—A Geometric Approach

Arthur E. Fischer* and Jerrold E. Marsden†
Department of Mathematics, University of California, Berkeley, California 94720
(Received 12 July 1971)

In this paper the exterior Einstein equations are explored from a differential geometric point of view. Using methods of global analysis and infinite-dimensional geometry, we answer sharply the question: "In what sense are the Einstein equations, written as equations of evolution, a Lagrangian dynamical system?" By using our global methods, several aspects of the lapse function and shift vector field are clarified. The geometrical significance of the shift becomes apparent when the Einstein evolution equations are written using Lie derivatives. The evolution equations are then interpreted as evolution equations as seen by an observer in space coordinates. Using the notion of body-space transitions, we then find the relationship between solutions with different shifts by finding the flow of a time-dependent vector field. The use of body and space coordinates is shown to be somewhat analogous to the use of such coordinates in Euler's equations for a rigid body and the use of Eulerian and Lagrangian coordinates in hydrodynamics. We also explore the geometry of the lapse function, and show how one can pass from one lapse function to another by integrating ordinary differential equations. This involves integrating what we call the "intrinsic shift vector field." The essence of our method is to extend the usual configuration space $\mathfrak{M} = \mathrm{Riem}(M)$ of Riemannian metrics to $\mathcal{T} \times \mathcal{D} \times \mathfrak{M}$, where $\mathcal{T} = C^\infty(M, \mathbf{R})$ is the group of relativistic time translations and $\mathcal{D} = \mathrm{Diff}(M)$ is the group of spatial coordinate transformations of M. The lapse and shift then enter the dynamical picture naturally as the velocities canonically conjugate to the configuration fields $(\xi_t, \eta_t) \in \mathcal{T} \times \mathcal{D}$. On this extended configuration space, a degenerate Lagrangian system is constructed which allows precisely for the arbitrary specification of the lapse and shift functions. We reinterpret a metric given by DeWitt for $\mathfrak{M}$ as a degenerate metric on $\mathcal{D} \times \mathfrak{M}$. On $\mathcal{D} \times \mathfrak{M}$, however, the metric is quadratic in the velocity variables. The groups $\mathcal{T}$ and $\mathcal{D}$ also serve as symmetry groups for our dynamical system. We establish that the associated conserved quantities are just the usual "constraint equations." A precise theorem is given for a remark of Misner that in an empty space-time we must have $\mathcal{H} = 0$. We study the relationship between the evolution equations for the time-dependent metric g_t and the Ricci flat condition of the reconstructed Lorentz metric g^L. Finally, we make some remarks about a possible "superphase space" for general relativity and how our treatment on $\mathcal{T} \times \mathcal{D} \times \mathfrak{M}$ is related to ordinary superspace and superphase space.

1. INTRODUCTION: THE EINSTEIN EQUATIONS OF EVOLUTION

Our aim in this paper is to study the Einstein equations of evolution as a dynamical system, to explore the geometrical meaning of the lapse function N and shift vector field X introduced by Wheeler,[1] to introduce the gauge groups $\mathcal{T}$ and $\mathcal{D}$ appropriate for the dynamical formulation of general relativity, and to study the interrelationships of the evolution equations with N and X, with the four-dimensional empty space condition $R_{\alpha\beta} = 0$, and with the gauge groups $\mathcal{T}$ and $\mathcal{D}$. The gauge groups $\mathcal{T}$ and $\mathcal{D}$ and the meaning of N and X are explained below.

Basic work on the problem of regarding the Einstein equations of evolution as a dynamical system has been done by Arnowitt, Deser, and Misner,[2] by DeWitt,[3] and by Wheeler.[1] We shall reformulate this work using the general theory of Lagrangian systems and exploiting differential geometric ideas. We shall also be extending this work and viewing it in a way which differs from the original approaches in several essential ways. Perhaps the most fundamental of these ways may be explained as follows: One usually fixes a three-dimensional manifold M (taken to be a spacelike hypersurface in the final space-time) and uses as configuration space the space $\mathfrak{M}$ of all Riemannian metrics on M. To incorporate the lapse function N and shift vector field X in what we believe is a natural way, we have found it necessary to enlarge $\mathfrak{M}$ to $\mathcal{T} \times \mathcal{D} \times \mathfrak{M}$, where

$\mathcal{T} = C^\infty(M; \mathbf{R})$ = all smooth real-valued functions, $\xi : M \to \mathbf{R}$, which one can think of as the "relativistic time translation group"
and

$\mathcal{D} = \mathrm{Diff}(M)$ = all diffeomorphisms $\eta : M \to M$, which one can think of as the "active" coordinate transformations or the "rotation group" of M.

The sense in which $\mathcal{T}$ is the relativistic time translation group and in which $\mathcal{D}$ is the rotation group of M is described briefly below and in detail in Secs. 7 and 5, respectively. The groups $\mathcal{T}$ and $\mathcal{D}$ are closely related to the lapse and shift as we shall explain shortly.

The lapse N, a real valued function, represents the clock rates for an observer relative to a reference system of clocks. The clock rates N depend on the space–time point for the observer. The fact that we change our clock rates, that is, allow an N not identically one, changes the equations of motion for the 3-metric g_{ij} which describes the geometry of the space M (the equations are written out below). Similarly a shift X is a vector field on the 3-manifold M which represents two observers in relative motion with velocity described by X. Again, a choice $X \neq 0$ will change the equations of motion.

The introduction of $\mathcal{T} \times \mathcal{D}$ is essentially the introduction of the configuration variables (ξ, η), whose canonically conjugate velocities are the lapse N and shift X (when the tangent space to $\mathcal{T} \times \mathcal{D}$ is "pulled back" to the identity; see Secs. 4 and 7 for a description of this process). On $\mathcal{T} \times \mathcal{D} \times \mathfrak{M}$ we construct an infinite-dimensional degenerate Lagrangian system $L : T(\mathcal{T} \times \mathcal{D} \times \mathfrak{M}) \to \mathbf{R}$. The degeneracy is, roughly speaking, in the direction of $\mathcal{T} \times \mathcal{D}$. The degeneracy allows precisely for the arbitrary specification of the lapse function and shift vector field.

In our approach, we also consider the geometrical significance of the lapse and shift in the equations of evolution. In the treatment of Arnowitt, Deser, and Misner,[2] the lapse and shift are incorporated into the Lagrangian on $\mathfrak{M}$ as Lagrange multipliers. The constraint equations (see below) are then obtained by varying the lapse and shift. In our formulation this situation is rather different. We consider the lapse and shift as velocities canonically conjugate to some configuration field variables rather than as Lagrange multipliers. The degeneracy of our Lagrangian on $\mathcal{T} \times \mathcal{D} \times \mathfrak{M}$ allows an arbitrary lapse and shift to be consistent with the equations of motion.

We consider the two basic constraints of the field equations, namely the divergence constraint

$$\delta((\mathrm{Tr}k)g - k)\mu_g = 0 \tag{1.1}$$

and the Hamiltonian constraint

$$\mathcal{H}(g,k)\,\mu_g \equiv \tfrac{1}{2}((\mathrm{Tr}k)^2 - (k\cdot k))\mu_g + 2R(g)\mu_g = 0 \quad (1.2)$$

(see below for definitions) as conservation laws rather than as "constraints." We shall show (see Sec. 6) that conservation in time of (1.1) is a result of the invariance of the evolution equations under the coordinate symmetry group $\mathcal{D}$. (1.2) is first established in Sec. 3 under the hypothesis that $\delta\pi = 0$ by a straightforward computation. What is interesting, however, is that we shall show that the *pointwise* conservation of $\mathcal{H}\mu_g$ is a *necessary* consequence of the "full relativistic invariance" of the theory (in a sense made precise in Sec. 7). Our theorem is a rigorous version of remarks of Misner[4,2] that a "topologically invariant" theory must have an identically zero Hamiltonian.

Note: Our Lagrangian on $\mathcal{T}\times\mathcal{D}\times\mathcal{M}$ is homogeneous and degenerate and our assertion about $\mathcal{H}\mu_g$ is distinct from and not to be confused with the elementary remark (see Ref. 5) that the energy of a homogeneous Lagrangian is always zero, as $\mathcal{H}$ is the energy of the Lagrangian before it is made homogeneous. We also remark that the infinite dimensionality of the invariance groups leads to pointwise integrals of the motion rather than integrated conserved quantities which one normally obtains.

We now formally write out the Einstein system for a given lapse N and shift X. It is important for the later geometrical development that certain combinations of the terms be recognized as Lie derivatives and Hessians.

Note: In the following, t occurring as a subscript indicates the variable t; it is never used to denote differentiation. Often the time-dependence of a field will be implicit.

The Einstein System (E): Let X_t be a time-dependent vector field on a fixed compact orientable three-dimensional manifold M, and let N_t be a time-dependent positive real valued (scalar) field on M, that is, $N_t(m) > 0$ for all $m \in M$ and $t \in \mathbf{R}$. The Einstein system is the system of evolution equations

$$\begin{cases} \dfrac{\partial g_t}{\partial t} = N_t k_t - L_{X_t} g_t, \\ \dfrac{\partial k_t}{\partial t} = N_t S_{g_t}(k_t) - 2N_t \mathrm{Ric}(g_t) + 2\mathrm{Hess}(N_t) - L_{X_t} k_t \end{cases} \quad (1.3)$$

with the supplementary conditions

$$\begin{cases} \delta((\mathrm{Tr}k_t)g_t - k_t) = 0, \\ \mathcal{H}(g_t,k_t) = \tfrac{1}{2}((\mathrm{Tr}k_t)^2 - k_t\cdot k_t) + 2R(g_t) = 0. \end{cases} \quad (1.4)$$

Our notation is the following:

g_t is a time-dependent metric on M,

$L_{X_t} g_t$ = Lie derivative of g_t with respect to the time-dependent vector field X_t (in coordinates, $L_{X_t} g_t = X_{i|j} + X_{j|i}$, with $|_j$ denoting the covariant derivative with respect to the time-dependent metric),

$L_{X_t} k_t$ = Lie derivative of $k_t = X^l k_{ij|l} + k_{il}X^l{}_{|j} + k_{jl}X^l{}_{|i}$,

$\mathrm{Ric}(g_t)$ = (Ricci curvature tensor formed from g_t) $= R_{ij} = \Gamma^k_{ij,k} - \Gamma^k_{ki,j} + \Gamma^k_{ij}\Gamma^l_{kl} - \Gamma^l_{ik}\Gamma^k_{lj}$,

$R(g_t)$ = scalar curvature $= R^k_k$,

$\mathrm{Hess}(N)$ = Hessian of N = double covariant derivative $= N_{|i|j}$,

δk = divergence of $k = (\delta k)_i = -k_i{}^j{}_{|j}$,

$\mathrm{Tr}k$ = trace $k = g^{ij}k_{ij} = k^i_i$,

$k\cdot k$ = dot product for symmetric tensors $= k_{ij}k^{ij}$,

$k \times k$ = cross product for symmetric tensors $= k_{il}k^l_j$,

$S_g(k)$ $= k\times k - \tfrac{1}{2}(\mathrm{Tr}k)k = k_{il}k^l_j - \tfrac{1}{2}(g^{mn}k_{mn})k_{ij}$.

Note: We have assumed M compact only to simplify the discussions. It is surely not essential.

The Lorentz geometry on $(-\epsilon, \epsilon)\times M$ corresponding to a solution $g_t = {}^3g_t$ of the Einstein system is given by

$${}^4g_{\alpha\beta}dx^\alpha dx^\beta = (X^iX_i - N^2)dt^2 - 2X_i dx^i dt + g_{ij}dx^i dx^j,$$

where $x^\alpha = (t, x^i)$ and $X^iX_i = {}^3g(X,X)$. For this to be Lorentz, we require ${}^3g(X,X) < N^2$ initially (and hence for a short t interval).

Our conventions in this paper will be that, for expressions written in coordinates, Latin indices will run from 1 to 3 and Greek indices from 0 to 3. Our Lorentz metrics will have signature $(-,+,+,+)$.

In the Einstein system, all the geometric quantities (such as Ric, Hess, δ, and Tr) are computed with respect to the time-dependent metric g_t. These equations appear in coordinates in Arnowitt, Deser, and Misner (see Ref. 2, p. 236 or Ref. 6, p. 1325) in terms of the tensor density $\pi = ((\mathrm{Tr}k)g - k)\sqrt{\det g}\, dx^1\wedge dx^2\wedge dx^3$ (using the momentum rather than the velocity variable). k, our energy density (the second of the supplementary conditions), and the second evolution equation are minus twice the corresponding quantities in Ref. 6, and our shift is minus theirs. Our change of numerics makes the system more manifestly a second-order system with $\partial g_t/\partial t = k_t$ when $N = 1$, $X = 0$ (see Sec. 3); changes the energy to the form $K + V$, where V is now the integrated scalar curvature (and not its negative; see DeWitt[3] and Sec. 3); and introduces a factor of $\frac{1}{2}$ into the kinetic energy part $\mathcal{K}$ of the energy density. The reason for changing the sign of the shift is explained at the end of Sec. 4.

Unfortunately, the Einstein system, when written in coordinates, obscures the central role played by the presence of the Lie derivative in the evolution equations. In fact, the apparent complexity of the equations as they appear in Refs. 2 or 6 dissolves when it is recognized that the last five terms in Ref. 6 or the last three terms in Ref. 2 are just the Lie derivative of either k or π (see also remarks in Sec. 6). Thus, when written intrinsically, the Einstein system is geometrically simplified.

The Lie derivative terms have a natural geometric interpretation related to changing from "space" to "body" coordinates in a manner similar to that of the rigid body and hydrodynamics (cf. Ref. 7). This may

be more specifically described as follows. The shift X_t is a time-dependent vector field on M, and as such it has an integral, or flow $\eta_t \in \mathfrak{D}$ with $\eta_0 =$ (identity diffeomorphism of M) $= id_M$. The solutions of the Einstein system with the shift X and those with the shift zero are related by the active coordinate transformation η_t in the usual way one transforms metrics. Moreover, we consider the manifold M to be the "body" and the flow η_t of the shift vector field as being a motion or "*rotation*" of M. Then if we assume that the time-dependent metric field g_t is "dragged along" by the rotation of M, the Einstein system can be interpreted as the equations of evolution as seen by an observer fixed in space, taken to be "off" the rotating manifold. This interpretation is worked out in Secs. 4 and 5.

The lapse N_t enters the evolution equations in a slightly more subtle way, as it involves a system of clocks on M whose rates may be different at different points of M. The complication due to the possible space-dependence of N_t is reflected in the Hessian term in the Einstein system (E).

As for the shift vector field we show how to solve the Einstein equations for a general lapse N, given the solution for $N = 1$, again by integrating a system of *ordinary* differential equations. Conversely, given a solution g for a given lapse N, we shall show how to construct a vector field on M, called the *intrinsic shift* of the lapse N, whose flow, together with a proper time function, brings us into a Gaussian normal coordinate system in which $N = 1$. The intrinsic shift vector field may be interpreted as a "sliding effect" due to the fact that N_t is not constant in the space variable.

Finally, we remark that on $\mathfrak{D} \times \mathfrak{M}$, our Lagrangian is quadratic in the velocities $(X, \partial g/\partial t)$ and is therefore of the classical form—kinetic energy minus potential energy, with the kinetic energy being derived from a degenerate metric on $\mathfrak{D} \times \mathfrak{M}$. In fact the evolution equation $\partial k/\partial t = k \times k - \frac{1}{2}(\mathrm{Tr}k)k - L_X k$, along a solution for which the kinetic energy $= 0$, is just the geodesic equation on $\mathfrak{D} \times \mathfrak{M}$ with respect to the aforementioned metric. When written just on $\mathfrak{M}$, the full Lagrangian does not have this classical form.

We now summarize the topics treated in this paper:

1. A treatment of infinite-dimensional degenerate Lagrangian systems (Sec. 2). A basic conservation law is given, similar to the nondegenerate case, which generalizes the classical conservation laws.[8,9,11]

2. The introduction of the gauge groups $\mathcal{T}$ and $\mathfrak{D}$ and of the space $\mathcal{T} \times \mathfrak{D} \times \mathfrak{M}$ as the configuration space for the Einstein system. The gauge groups $\mathcal{T}$ and $\mathfrak{D}$ are the analog for the dynamical formulation of the coordinate gauge group of the four-dimensional geometry.

3. A geometrical interpretation of the lapse and shift functions as the velocities canonically conjugate to the new configuration variables $(\xi, \eta) \in \mathcal{T} \times \mathfrak{D}$.

4. A treatment of the Einstein system as an infinite-dimensional degenerate Lagrangian system on a suitable subset of $\mathcal{T} \times \mathfrak{D} \times \mathfrak{M}$ (Secs. 2 and 7). On $\mathfrak{D} \times \mathfrak{M}$ we construct a Lagrangian $L = K - V$, where K is quadratic in the velocities (X, h), and on $\mathcal{T} \times \mathfrak{D} \times \mathfrak{M}$ a Lagrangian homogeneous in the velocities (N, X, h). The quadratic nature of the Lagrangian on $\mathfrak{D} \times \mathfrak{M}$ results in evolution equations quadratic in the velocities (X, h) which is analagous to the quadratic nature of the evolution equations for geodesics on a manifold and for hydrodynamics and the rigid body.

5. A derivation of the supplementary conditions for the Einstein system as conserved quantities (Secs. 3 and 10), using conservation laws for degenerate Lagrangian systems and infinite-dimensional symmetry groups ($\mathcal{T}$ and $\mathfrak{D}$), together with a precise explanation of why the energy density for the empty space–time equations must be identically zero; cf. Misner.[4]

6. The recognition of the central role played by the Lie derivative terms in the evolution equations. The use of Lie derivatives shows that these terms in the evolution equations are geometrically simple and makes their geometric meaning transparent.

7. An interpretation of the shift vector field X_t as generating a "rotation" of M and of the evolution equations as being the equations as seen by an observer in space coordinates. Using the notion of body-space transitions, we then show that if we can solve the equations for $X = 0$, then they can be solved for any X by integrating a system of *ordinary* differential equations (Corollary 4.1). Similarly, we give a geometrical derivation of how if one can solve the Einstein equations for $N = 1$, then they can be solved for an arbitrary N by integrating geodesic equations, again *ordinary* differential equations (Theorem 10.3).

8. A geometrical derivation of the Hessian term Hess(N) in the Einstein system using generalized Gauss–Codazzi equations in coordinates which are not necessarily normal Gaussian (Sec. 8).

9. The interpretation of the lapse N_t as related to the tangent of the curve $\tau_t \in \mathcal{T}$ where τ_t, the proper time function, can be interpreted as a change of time parameter for each point $m \in M$ from the canonical parameter of evolution to an arbitrary parameter of evolution (Sec. 10).

10. The introduction of a new object, the *intrinsic shift Y of N*, whose integration gives the rest of the Gaussian coordinate system (Theorems 10.1 and 10.2). The intrinsic shift is interpreted as the "tilting effect" of the coordinate system due to the spatial dependence of N.

2. DEGENERATE LAGRANGIAN SYSTEMS

In this section we study the notion of a degenerate Lagrangian system in the spirit of Ref. 8. Degenerate Lagrangian systems have been used in some previous analyses of general relativity and are fairly common in classical mechanics (cf. Ref. 9). Here we shall treat such systems from the coordinate independent or global point of view.

We are going to be working with spaces of maps, for example, the space of all Riemannian metrics $\mathfrak{M}$. For simplicity we assume that all such objects are C^∞. Properly one should work with Sobolev spaces, but the modifications needed are fairly routine and do not involve any new physical ideas (see Refs. 10 and 11).

For our purposes we find it convenient to use the general ideas about Hamiltonian and Lagrangian systems as developed in Ref. 8 rather than variational

principles. The chief difference with Ref. 8 is that we must use infinite-dimensional configuration spaces. We, therefore, shall assume that the reader has some aquaintance with calculus in infinite dimensional spaces and manifolds, as expounded for example in Ref. 12. As mentioned above, the spaces are usually spaces of maps; in the physics literature the derivative of a function defined on such a space is often called its "functional derivative."

Let B be a manifold, possibly infinite-dimensional (modeled on a Fréchet or Banach space). Let TB denote its tangent bundle and let $\tau: TB \to B$ be the natural projection map. For $b \in B$, let T_bB be the tangent space at b, that is, the fiber over b; $T_bB = \tau^{-1}(b)$.

Let $L: TB \to \mathbf{R}$ be a mapping (called the Lagrangian or action integral) and let L_b be the restriction of L to T_bB. The derivative of L_b, $DL_b(v)$, at a point $v \in T_bB$ defines a map $FL: TB \to T^*B$ (T^*B is the contangent bundle), called the *fiber derivative* or *Legendre transformation*. As in Ref. 8, using FL, one obtains a closed 2-form ω_L (i.e., $d\omega_L = 0$) on TB by pulling back the canonical symplectic structure on T^*B. If B is modeled on a linear space $\mathbf{E}$, so locally TB looks like $U \times \mathbf{E}$ where $U \subset \mathbf{E}$ is open, then $\omega_L(u, e)$, for $(u, e) \in U \times \mathbf{E}$, is a skew-symmetric bilinear form on $\mathbf{E} \times \mathbf{E}$ given by

$$\begin{aligned}2\omega_L(u, e)\cdot((e_1, e_2), (e_3, e_4)) &= D_1(D_2L(u, e)\cdot e_1)\cdot e_3 - D_1(D_2L(u, e)\cdot e_3)\cdot e_1 \\ &\quad + D_2D_2L(u, e)\cdot e_4\cdot e_1 - D_2D_2L(u, e)\cdot e_2\cdot e_3,\end{aligned}$$

where D_1, D_2 denote the partial derivatives of L.

We say ω_L is (weakly) *nondegenerate* if $\omega_L(u, e)\cdot((e_1, e_2), (e_3, e_4)) = 0$ for all $e_3, e_4 \in \mathbf{E}$ implies that $(e_1, e_2) = 0$. However we will want to allow for degenerate ω_L, so we do not make this assumption. We say that L is *degenerate* if ω_L is degenerate (as a 2-form). It is easy to see that ω_L is nondegenerate iff $D_2D_2L(u, e)$ is nondegenerate.

The *action* of L is defined by $A: TB \to \mathbf{R}$, $A(v) = FL(v)\cdot v$, and the *energy* of L is $E = A - L$. In charts,

$$E(u, e) = D_2L(u, e)\cdot e - L(u, e)$$

and in finite dimensions it is the usual expression

$$E(q, \dot{q}) = \frac{\partial L}{\partial \dot{q}^i}\dot{q}^i - L(q, \dot{q}).$$

Now given L, we say that a vector field Z on TB is a *Lagrangian vector field* or a *Lagrangian system* for L if the *Lagrangian condition* holds:

$$2\omega_L(v)(Z(v), w) = dE(v)\cdot w$$

for all $v \in T_bB$, $w \in T_v(TB)$. Here dE denotes the differential of E.

If ω_L were a (weak) symplectic form, i.e., nondegenerate, there would be at most one such Z. The fact that ω_L is degenerate however means that Z is not uniquely determined by L so that there is some arbitrariness in what we may choose for Z. For the Einstein system, this degeneracy will correspond to an arbitrary choice of lapse and shift.

It should also be stressed that, in general, L need not have a corresponding Z. If there is one, we say that we can find *consistent equations of motion* for L. As above, there can be several equations of motion consistent with L.

The dynamics is obtained by finding the integral curves of Z; that is, the curves $v(t)$ such that $v(t) \in TB$ satisfies $(dv/dt)(t) = Z(v(t))$. From the Lagrangian condition, it is trivial to check that energy is conserved even though L may be degenerate.

Proposition 2.1: Let Z be a Lagrangian vector field for L and let $v(t) \in TB$ be an integral curve of Z. Then $E(v(t))$ is constant in t.

Proof: By the chain rule,

$$\begin{aligned}\frac{d}{dt}E(v(t)) &= dE(v(t))\cdot v'(t) - dE(v(t))\cdot Z(v(t)) \\ &= 2\omega_L(v(t))(Z(v(t)), Z(v(t))) \\ &= 0 \text{ by the skew symmetry of } \omega_L.\blacksquare\end{aligned}$$

For a general degenerate Lagrangian system, Lagrange's equations also hold, if we assume that Z is second order. That Z is second order means that, in a chart $U \times \mathbf{E}$, Z has the form $Z(u, e) = (e, Z_2(u, e))$. (See Refs. 8 and 12 for the intrinsic definition of second order.)

Proposition 2.2: Let Z be a Lagrangian system for L and suppose Z is a second order equation. Then in the chart $U \times \mathbf{E}$, an integral curve $(u(t), v(t)) \in U \times \mathbf{E}$ of Z satisfies Lagrange's equations:

$$\begin{cases}\dfrac{du}{dt}(t) = v(t), \\ \dfrac{d}{dt}(D_2L(u(t)), v(t))\cdot\omega) = D_1L(u(t), v(t))\cdot\omega\end{cases}$$

for all $\omega \in \mathbf{E}$. In case L is nondegenerate,

$$\frac{dv}{dt} = \{D_2D_2L(u, v)\}^{-1}\{D_1L(u, v) - D_1D_2L(u, v)\cdot v\}$$

Proof: We work in a chart $U \times \mathbf{E}$ so that $Z(u, e) = (e, Z_2(u, e))$. From the definition of E, we have $dE(u, e)\cdot(e_1, e_2) = D_1(D_2L(u, e)\cdot e)\cdot e_1 + D_2D_2L(u, e)\cdot e\cdot e_2 - D_1L(u, e)\cdot e_1$. Using the formula for ω_L, the condition on Z_2 may be written, after a short computation, as

$$D_1L(u, e)\cdot e_1 = D_1(D_2L(u, e)\cdot e_1)\cdot e + D_2(D_2L(u, e)\cdot Z_2(u, e))\cdot e_1$$

for all $e_1 \in \mathbf{E}$.

If $(u(t), v(t))$ is an integral curve of Z, we obtain (where the dot means d/dt)

$$\begin{aligned}D_1L(u, v)\cdot e_1 &= D_1D_2L(u, \dot{u})\cdot e_1\cdot\dot{u} + D_2D_2L(u, \dot{u})\cdot\ddot{u}\cdot e_1 \\ &= \frac{d}{dt}D_2L(u, \dot{u})\cdot e_1\end{aligned}$$

by the chain rule.$\blacksquare$

We wish to emphasise a special case of Proposition 2.2 for later use. Suppose that $\langle , \rangle$ is a symmetric bilinear form defined on each tangent space of B; we shall refer to $\langle , \rangle$ as a *metric*. Define $L: TB \to \mathbf{R}$ to

be the *kinetic energy* $L(v) = \frac{1}{2}\langle v, v\rangle$. Suppose that the metric is nondegenerate, which implies that the Lagrangian L is also nondegenerate. Then Z is called the *geodesic spray* and one can easily check from Proposition 2.2 that in the finite-dimensional case,

$$L(v) = \tfrac{1}{2} g_{ij} v^i v^j \quad \text{and} \quad Z_2(u, e) = -\Gamma^k_{ij}(u) e^i e^j$$

where Γ^k_{ij} are the Christoffel symbols of the metric g_{ij}.

Thus the integral curves of the geodesic spray are given by $v(t) = dx(t)/dt$, where $x(t)$ is a geodesic on B. Now let $V: B \to \mathbf{R}$ be given and consider the Lagrangian $L(v_b) = \frac{1}{2}\langle v_b, v_b\rangle - V(b)$. Then from Proposition 2.2 we see that the Lagrangian vector field for L is given by $Z(u, e) = (e, S_2(u, e) - \text{grad}V(u))$, where S is the spray of the metric and where grad$V$ is the gradient of V, a vector field on B defined by

$$\langle \text{grad}V(b), v_b\rangle = dV(b) \cdot v_b,$$

where $dV(b)$ is the differential of V evaluated at b. Again, in the finite-dimensional case, grad$V = g^{ij}\,\partial V/\partial x^j$.

We remark that if ω_L were nondegenerate, Z would automatically be a second order equation (cf. Ref. 8). But for a general ω_L, a Lagrangian vector field Z need not be second order; if it is not, Lagrange's equations may fail—and do in some important examples, such as when a quantum mechanical system is regarded as a Lagrangian system.

Often L is obtained in the form

$$L(u, \dot{u}) = \int_M \mathcal{L}\left(u, \frac{\partial u}{\partial x^k}, \dot{u}\right) dx.$$

Here M is some fixed manifold, say $\mathbf{R}^n$, u is a scalar or possibly tensor field on M and $\mathcal{L}$ is a given scalar function with the indicated arguments. The space of the u's forms the manifold B and the Lagrange's equations can be converted to the usual Lagrange density form.

We next give the basic conservation law for Lagrangian mechanics. A key point is that the validity of the result is not affected by the fact that L may be degenerate.

Proposition 2.3: Let Z be a Lagrangian vector field for $L: TB \to \mathbf{R}$, and suppose Z is a second-order equation.

Let Φ_t be a one-parameter group of diffeomorphisms of B generated by the vector field $Y: B \to TB$. Suppose that for each real number t, $L \circ T\Phi_t = L$. Then the function $P(Y): TB \to \mathbf{R}$, $P(Y)(v) = FL(v) \cdot Y$ is a constant along integral curves of Z.

Proof: Let $v(t)$ be an integral curve for Z. Then we shall show that $(d/dt)P(Y)(v(t)) = 0$. Indeed, in a coordinate chart, if $(u(t), v(t))$ is the integral curve, we get from the chain rule

$$\frac{d}{dt}\{FL(v(t)) \cdot Y\} = \frac{d}{dt}\{D_2 L(u(t), v(t)) \cdot Y(u(t))\}$$
$$= D_1 D_2 L(u(t), v(t)) \cdot Y(u(t)) \cdot u'(t) + D_2 D_2 L(u(t), v(t)) \cdot Y(u(t)) \cdot v'(t) + D_2 L(u(t), v(t)) \cdot DY(u(t)) \cdot u'(t).$$

Now the condition that Z be the Lagrangian vector field of L means exactly that the first two terms equal $D_1 L(u(t), v(t)) \cdot Y(u(t))$ (see the proof of Proposition 2.2 above). However if we differentiate $L \circ T\Phi_t$ with respect to t we obtain for any point (u, v),

$$0 = \frac{d}{dt} L(\Phi_t(u), D\Phi_t(u) \cdot v)\big|_{t=0}$$
$$= D_1 L(u, v) \cdot Y(u) + D_2 L(u, v) \cdot DY(u) \cdot v.$$

Comparing this with the above gives $(d/dt)\{FL(v) \cdot Y\} = 0$ and proves the assertion. ■

Proposition 2.4: Let $L: TB \to \mathbf{R}$ be a Lagrangian, possibly degenerate and let Z be a Lagrangian vector field for L. Suppose $\Phi: B \to D$ is a diffeomorphism. Set $\tilde{L}: TD \to \mathbf{R}$, $\tilde{L} = L \circ T\Phi^{-1}$ where $T\Phi: TB \to TD$ is the tangent (derivative) of Φ. Then a Lagrangian vector field for $\tilde{L}$ is given as follows:

$$\tilde{Z}: TD \to T^2 D, \qquad \tilde{Z} = T(T\Phi) \circ Z \circ T\Phi^{-1} \equiv (T\Phi)_* Z$$

($\tilde{Z}$ is the "push-forward" of Z by the diffeomorphism $T\Phi$).

This is a straightforward check using the definitions.

3. THE EINSTEIN SYSTEM ON $\mathfrak{M}$ (NO SHIFT AND NO LAPSE)

In this section we consider the Einstein system in Gaussian coordinates. Given this coordinate choice, the system is described by a nondegenerate Lagrangian formalism (cf. Refs. 2 and 3) which we now globalize using the language of Sec. 2. We choose as our configuration space $\mathfrak{M}$, the space of all Riemannian metrics on a fixed compact three-dimensional manifold M. Thus $\mathfrak{M}$ is an open convex cone in $S_2(M)$, the space of two symmetric covariant tensor fields on M. Therefore for $g \in \mathfrak{M}$, $T_g\mathfrak{M} = S_2(M)$, and we can write

$$T\mathfrak{M} = \mathfrak{M} \times S_2(M).$$

We equip $\mathfrak{M}$ with an indefinite metric $\mathcal{G}$, referred to as the *DeWitt metric*, by setting for $g \in \mathfrak{M}$,

$$\mathcal{G}_g: T_g\mathfrak{M} \times T_g\mathfrak{M} \approx S_2(M) \times S_2(M) \to \mathbf{R},$$
$$\mathcal{G}_g(h, k) = \int_M ((\text{Tr}h)(\text{Tr}k) - h \cdot k)\,\mu_g,$$

where $h \cdot k = h_{ij} k^{ij}$, the dot product of h and k, and μ_g $(= \sqrt{\det g}\; dx^1 \wedge dx^2 \wedge \cdots \wedge dx^n)$ is the usual volume element associated with the metric g. Note that the space $\mathfrak{M}$ has the feature, appropriate for general relativity, of having metric structures which do not depend on any particular Riemannian metric for the underlying manifold M. In other classical field theories, such as electrodynamics and hydrodynamics, it is necessary to specify *a priori* a metric on M. In general relativity such a specification is, of course, unnatural as it is precisely the metric structure of M that is evolving.

Proposition 3.1: The Lagrangian $L_0(g, h) = \frac{1}{2}\mathcal{G}_g(h, h)$ is nondegenerate and the associated Lagrangian vector field exists and is given by

$$Z: \mathfrak{M} \times S_2(M) \to S_2(M) \times S_2(M)$$
$$Z(g, k) = (k, k \times k - \tfrac{1}{2}(\text{Tr}k)k + \tfrac{1}{8}[(\text{Tr}k)^2 - k \cdot k]g)$$
$$\equiv (k, S_g(k) + \tfrac{1}{4}\mathcal{K}g),$$

where $(k \times k)_{ij} = k_{il}k^l_j$, the cross product of k and k, $S_g(k) = k \times k - \frac{1}{2}(\mathrm{Tr}k)k$, and $\mathcal{K} = \frac{1}{2}((\mathrm{Tr}k)^2 - k\cdot k)$, the kinetic energy scalar.

For each $(\mathring{g}, \mathring{k}) \in \mathfrak{M} \times S_2(M)$, there exists a unique smooth curve $(g_t, k_t) \in \mathfrak{M} \times S_2(M)$ with initial conditions $(g_0, k_0) = (\mathring{g}, \mathring{k})$ and which satisfies Z, for t in some interval about 0.

Proof: That $\mathcal{G}$ is nondegenerate follows from the fact that if $\mathcal{G}_g(k, h) = 0$ for all $h \in S_2(M)$, then in particular, setting $h = \frac{1}{2}(\mathrm{Tr}k)g - k$, $0 = \mathcal{G}_g(k, \frac{1}{2}(\mathrm{Tr}k)g - k) = \int_M k\cdot k\, \mu_g$, which implies $k = 0$ as the integrand is positive.

We first consider Lagrange's equation in the form $(d/dt)D_2L_0(g,k)\cdot\omega = D_1L_0(g,k)\cdot\omega$, $\omega \in S^2(M)$. Note that L_0: $\mathfrak{M} \times S_2(M) \to \mathbf{R}$, so that $D_1L_0(g,k)$: $S_2(M) \to \mathbf{R}$ and $D_2L_0(g,k)$: $S_2(M) \to \mathbf{R}$. Let g_t be a curve in $\mathfrak{M}$ that satisfies

$$g_0 = g \quad \text{and} \quad \frac{dg}{dt}\Big|_{t=0} = \omega \in S_2(M).$$

Then from the rule for taking partial derivatives,

$$D_1L_0(g,k)\cdot\omega = \frac{d}{dt}L_0(g_t, k)\Big|_{t=0}.$$

Similarly,

$$D_2L_0(g,k)\cdot\omega = \frac{d}{dt}L_0(g,k_t)\Big|_{t=0} \quad \text{where} \left(k_t, \frac{dk_t}{dt}\right)\Big|_{t=0} = (k,\omega).$$

By applying this rule, Lagrange's equation becomes

$$\frac{d}{dt}\int_M [(\mathrm{Tr}k)g - k]\cdot\omega\,\mu_g, = \int_M [-(\mathrm{Tr}k)k + k\times k + \tfrac{1}{2}\mathcal{K}g]\cdot\omega\,\mu_g, \quad (3.1)$$

where we have used the fact that the derivative of the map

$$\mu\colon \mathfrak{M} \to \Omega(M), \qquad g \mapsto \mu_g$$

[$\Omega(M)$ = smooth volume elements on M] is given by

$$D\mu(g)\colon S_2(M) \to \Lambda^n(M), \qquad g \mapsto D\mu(g)\cdot\omega = \tfrac{1}{2}(\mathrm{Tr}\omega)\mu_g$$

[$\Lambda^n(M)$ = the space of smooth n-forms on M].

In coordinates this result follows from the formula

$$\frac{d}{dt}\det(g_{ij} + t\omega_{ij})\Big|_{t=0} = \mathrm{Tr}(\omega_{ij})\det(g_{ij})$$

Since $\omega \in S_2(M)$ is arbitrary, the integrands of (3.1) are equal and we get

$$\frac{d}{dt}[(\mathrm{Tr}k)g - k]^{-1}\mu_g = [k\times k - (\mathrm{Tr}k)k + \tfrac{1}{2}\mathcal{K}g]^{-1}\mu_g, \quad (3.2)$$

where we are using the superscript^{-1} to mean that all indices are contravariant, and juxtaposition, like $k^{-1}\mu_g$, to mean the tensor product $k^{-1}\mu_g = k^{-1}\otimes\mu_g$ (classically, expressions like $k^{-1}\mu_g$ are tensor densities, written in coordinates as $k^{ij}\sqrt{\det g}$).

Equation (3.2) is Lagrange's equation on $T^*\mathfrak{M}$. To get Lagrange's equation on $T\mathfrak{M}$, we pull (3.2) back using the fiber derivative of L_0. This is equivalent to solving (3.2) directly for dk/dt. Using the facts that $dg/dt = k$, $dg^{-1}/dt = -k^{-1}$ (in coordinates this is $dg^{ij}/dt = -g^{ia}g^{jb}k_{ab}$) and $d\mu_g/dt = \frac{1}{2}(\mathrm{Tr}k)\mu_g$ (in coordinates this is $(d/dt)\sqrt{\det g} = \frac{1}{2}g^{ab}(dg_{ab}/dt)\sqrt{\det g}$), (3.2) becomes

$$\left(\frac{d}{dt}[(\mathrm{Tr}k)g - k]^{-1}\right)\mu_g + \tfrac{1}{2}(\mathrm{Tr}k)[(\mathrm{Tr}k)g - k]^{-1}\mu_g = [(k\times k - (\mathrm{Tr}k)k + \tfrac{1}{2}\mathcal{K}g]^{-1}\mu_g.$$

Eliminating μ_g gives

$$\frac{d}{dt}[(\mathrm{Tr}k)g - k]^{-1} = [k\times k - \tfrac{1}{2}(\mathrm{Tr}k)k]^{-1} - \tfrac{1}{2}(\mathrm{Tr}k)2g^{-1} + \tfrac{1}{2}\mathcal{K}g^{-1}. \quad (3.3)$$

From $(d/dt)(\mathrm{Tr}k) = -(\mathrm{Tr}k)^2 + \mathrm{Tr}(dk/dt)$ and taking the trace in (3.3), we find

$$\frac{d}{dt}(\mathrm{Tr}k) = -\tfrac{1}{2}(\mathrm{Tr}k)^2 + \tfrac{3}{4}\mathcal{K}$$

and thus

$$\frac{dk^{-1}}{dt} = -[k\times k - \tfrac{1}{2}(\mathrm{Tr}k)k]^{-1} - (\mathrm{Tr}k)k^{-1} + \tfrac{1}{4}\mathcal{K}g^{-1}.$$

Lowering the indices on k^{-1} by using $(dk/dt)^{-1} = (dk^{-1}/dt) + 2(k\times k)^{-1}$ (in coordinates this is $g^{ai}g^{bj}(dk_{ab}/dt) = (dk^{ij}/dt) + 2k^i_a k^{aj}$) gives the result,

$$\frac{dk}{dt} = k\times k - \tfrac{1}{2}(\mathrm{Tr}k)k + \tfrac{1}{4}\mathcal{K}g.$$

Finally we remark that Z is simply algebraic in g and k and is thus a smooth vector field on $T\mathfrak{M}$ and on $T\mathfrak{M}^k$, $0 \le k < \infty$, where $\mathfrak{M}^k$ is the manifold of C^k-Riemannian metrics on M in the C^k topology (uniform convergence of derivatives up to order k). The manifolds $T\mathfrak{M}^k$, $0 \le k < \infty$ are Banach manifolds and hence the usual Picard method for ordinary differential equations is sufficient to establish that Z has a smooth flow on $T\mathfrak{M}^k$, $0 \le k < \infty$, defined for a short time interval and a simple limit argument shows that $k = \infty$ is also allowed. Indeed, what one must show is that the time of existence ϵ_k does not go to zero as $k \to \infty$. To show this, suppose we have a solution (g,k) in $T\mathfrak{M}^k$ with initial data (g_0, k_0) which are C^{k+1}. Then we assert the solution is C^{k+1} as long as it is defined in C^k. This comes from examining the linear differential equation for the spatial derivative of (g,k), as in Refs. 13 and 14. The result then follows. ■

The second order equation $Z(g,k) = (k, Z_2(g,k)) = (k, k\times k - \frac{1}{2}[(\mathrm{Tr}k)k + \frac{1}{8}(\mathrm{Tr}k)^2 - k\cdot k]g)$ is quadratic in the second or velocity variable k; that is, for $\lambda \in \mathbf{R}$, $Z_2(g,\lambda k) = \lambda^2 Z_2(g,k)$. A second order equation with this property is called a *spray* (see Lang[12], p. 67). If Z is the spray of a metric it is called a *geodesic spray* and its base integral curves are the geodesics of the metric. Thus the Z of Proposition 3.1 is the *geodesic spray of the DeWitt metric*. However, Propositions 3.3 and 3.4 below show that the truncated quantity $S_g(k) = k\times k - \frac{1}{2}(\mathrm{Tr}k)k$ also enters the equations in a fundamental way. We will refer to $S_g(k)$ as the *DeWitt spray*.

We remark that the geodesics of $\mathcal{G}$ (that is, the base integral curves of Z) can in fact be found explicitly as has been done by Dewitt[3] and in a different context by Eardley, Liang, and Sachs.[15] From these explicit formulas, it is seen that the geodesics exist for short time only, as they eventually run out of $\mathfrak{M}$ into $S_2(M)$.

When this happens, the equations break down and become singular since they involve taking the inverse of g.

Since $\mathcal{G}$ is nondegenerate, the bundle map $\mathcal{G}^\flat$: $T\mathfrak{M} \to T^*\mathfrak{M}$ defined by $\mathcal{G}_g^\flat(h)\cdot k = \mathcal{G}_g(h,k)$ is injective. This map in the classical case $g^\flat$: $TM \to T^*M$ corresponds to lowering of indices by a metric tensor g_{ij}. Note, however, that since the model space $S_2(M)$ is the space of covariant tensor fields, $\mathcal{G}^\flat(k) \in T_g^*\mathfrak{M}$ has contravariant indices.

The terminology weakly nondegenerate comes from the fact that the map $\mathcal{G}^\flat$, while injective, need not be an isomorphism, cf. Ref. 11. We should also point out that in contrast to strong metrics (as in Ref. 12), weak metrics need not have geodesics. For the DeWitt metric, however, the existence of geodesics was checked directly in Proposition 3.1.

Now $\mathcal{G}_g^\flat(k)\cdot h = \mathcal{G}_g(k,h) = \int_M [(\mathrm{Tr}k)(\mathrm{Tr}h) - k\cdot h]\mu_g$ $= \int_M [(\mathrm{Tr}k)g - k]^{-1}\cdot h\,\mu_g$. If we consider $S^2(\mathrm{M})\otimes\mu_g$ (= space of two-symmetric contravariant tensor densities) as a subspace of $T_g^*\mathfrak{M}$, the continuous linear functionals on $S_2(\mathrm{M}) = T_g\mathfrak{M}$ in the C^∞-topology of uniform convergence of derivatives of all orders, then we can set $\mathcal{G}_g^\flat(k) = [(\mathrm{Tr}k)g - k]^{-1}\mu_g$. The space $T_g^*\mathfrak{M}$ is a space of tensor-valued distributions.

We will denote by π the momentum canonically conjugate to the velocity k, $\pi = \mathcal{G}_g^\flat(k) \in S^2(M)\otimes\mu_g$. In coordinates, $\pi^{ij} = [(g^{ab}k_{ab})g^{ij} - k^{ij}]\sqrt{\det g}$ which is the expression for π that appears in Ref. 2. Thus π can be interpreted as k with its indices raised by the DeWitt metric.

If we consider $S^2(M)\otimes\mu_g$ as the range of $\mathcal{G}_g^\flat$ rather than $T_g^*\mathfrak{M}$, then $\mathcal{G}_g^\flat$ is onto, and its inverse

$$[\mathcal{G}_g^\flat]^{-1} = \mathcal{G}_g^\#\colon S^2(M)\otimes\mu_g \to S_2(M)$$

is given by

$$\mathcal{G}_g^\#(\pi'\otimes\mu_g) = \tfrac{1}{2}(\mathrm{Tr}\pi')g - (\pi')^\flat \quad [= \tfrac{1}{2}(\mathrm{Tr}\pi')g_{ij} - \pi'_{ij}],$$

where π' is the tensor part of $\pi = \pi'\otimes\mu_g$, and $(\pi')^\flat$ means π' with its indices lowered $[(\pi')^\flat_{ij} = g_{ik}g_{jl}(\pi')^{kl} = \pi'_{ij}]$. The factor $\frac{1}{2}$ enters so that $\mathcal{G}_g^\#[\mathcal{G}_g^\flat(k)] = \mathcal{G}^\#\{[(\mathrm{Tr}k)g - k]^{-1}\otimes\mu_g\} = \frac{1}{2}\{\mathrm{Tr}[(\mathrm{Tr}k)g - k]\}g - [(\mathrm{Tr}k)g - k] = k$, as we expect.

We now proceed to consider the gravitational potential of DeWitt and to compute its gradient with respect to the metric $\mathcal{G}$; see the discussion following Proposition 2.2. Although the spray Z of $\mathcal{G}$ was simply algebraic, the gradient of the potential will be a nonlinear differential operator.

Let

$$V\colon \mathfrak{M}\to\mathbf{R},\quad g\mapsto 2\int_M R(g)\mu_g,$$

where $R(g)$ is the scalar curvature of g, and set

$$L(g,k) = \tfrac{1}{2}\mathcal{G}_g(k,k) - V(g).$$

Our k and L is minus twice DeWitt's[3] so that our L is of the form kinetic energy minus potential energy and the kinetic energy enters with the classical factor $\frac{1}{2}$.

Adding this potential to L_0 in Proposition 3.1 adds a forcing or gradient term to the equations of motion, which we now compute. We include a positive scalar function N: $M\to\mathbf{R}$ in the potential for later use (see Sec. 8).

Proposition 3.2: Let N: $M\to\mathbf{R}$ be a smooth positive scalar function on M and let

$$V\colon \mathfrak{M}\to\mathbf{R},\quad g\mapsto 2\int_M NR(g)\mu_g.$$

Then

$$-\mathrm{grad}V = -2N\,\mathrm{Ric}(g) + \tfrac{1}{2}NR(g)g + 2\,\mathrm{Hess}N,$$

where $\mathrm{Ric}(g)$ = Ricci tensor of g, $\mathrm{Hess}N = N_{|i|j}$ = double covariant derivative, and where

$$\mathrm{grad}V = \mathcal{G}^\#(dV)\colon \mathfrak{M}\to S_2(M)$$

is computed with respect to the DeWitt metric $\mathcal{G}$.

Proof: We first compute $dV(g) = DV(g)\colon S_2(M)\to\mathbf{R}$. Let $g_t\in\mathfrak{M}$ be a curve in $\mathfrak{M}$ with $(g_t, dg_t/dt)|_{t=0} = (g,\omega)\in\mathfrak{M}\times S_2$. Then since $d/dt\,\mu_{g_t} = \frac{1}{2}\mathrm{Tr}(dg/dt)\mu_{g_t}$, as in the proof of Proposition 3.1, we get

$$\begin{aligned} dV(g)\cdot\omega &= \frac{d}{dt}V(g_t)\Big|_{t=0} \\ &= 2\int_M N\frac{d}{dt}(R(g_t)\mu_{g_t})\Big|_{t=0} \\ &= 2\int_M\left[\frac{dR(g_t)}{dt}\mu_{g_t} + R(g_t)\tfrac{1}{2}\mathrm{Tr}\left(\frac{dg_t}{dt}\right)\mu_{g_t}\right]\Big|_{t=0}. \end{aligned}$$

A classical computation gives[16]

$$\frac{d}{dt}R(g_t)\Big|_{t=0} = \Delta(\mathrm{Tr}\omega) + \delta\delta\omega - \mathrm{Ric}(g)\omega,$$

where $\Delta f = -g^{ij}f_{|i|j}$ = Laplace–Beltrami operator on scalars, and where $\delta\delta\omega = \omega^{ij}{}_{|i|j}$ = the double covariant divergence. Thus

$$dV(g)\cdot\omega = 2\int_M N[\Delta(\mathrm{Tr}\omega) + \delta\delta\omega - \mathrm{Ric}(g)\cdot\omega + R(g)\tfrac{1}{2}(\mathrm{Tr}\omega)]\mu_g.$$

Since M is compact without boundary, an integration by parts yields

$$dV(g)\cdot\omega = 2\int_M\{g\Delta N + \mathrm{Hess}N - N[\mathrm{Ric}(g) - \tfrac{1}{2}R(g)g]\}\cdot\omega\,\mu_g$$

and since ω is arbitrary

$$dV(g) = 2\{g\Delta N + \mathrm{Hess}H - N[\mathrm{Ric}(g) - \tfrac{1}{2}R(g)g]\}^{-1}\mu_g. \tag{3.4}$$

Using $\mathcal{G}_g^\#$ to pull $DV(g)\in T_g^*\mathfrak{M}$ back to $T\mathfrak{M}$, we find

$$\begin{aligned} &\mathcal{G}_g^\#[(g\Delta N + \mathrm{Hess}N)^{-1}\mu_g] \\ &\quad = \tfrac{1}{2}g\mathrm{Tr}(g\Delta N + \mathrm{Hess}N) - (g\Delta N + \mathrm{Hess}N) \\ &\quad = \tfrac{1}{2}g(3\Delta N - \Delta N) - (g\Delta N + \mathrm{Hess}N) \\ &\quad = -\mathrm{Hess}N \end{aligned}$$

and

$$\begin{aligned} &\mathcal{G}_g^\#\{-N[\mathrm{Ric}(g) - \tfrac{1}{2}R(g)g]^{-1}\mu_g\} \\ &\quad = -\tfrac{1}{2}Ng\mathrm{Tr}[\mathrm{Ric}(g) - \tfrac{1}{2}gR(g)] \\ &\qquad + N[\mathrm{Ric}(g) - \tfrac{1}{2}gR(g)] = N[\mathrm{Ric}(g) - \tfrac{1}{4}gR(g)]. \end{aligned}$$

Hence, $-\mathrm{grad}V(g) = -2N\,\mathrm{Ric}(g) + \frac{1}{2}NgR(g) + 2\,\mathrm{Hess}N$. ■

Combining Propositions 3.1 and 3.2 (and taking $N = 1$) gives

Proposition 3.3: The vector field equations on $T\mathfrak{M}$ associated with L are

$$\begin{cases} \dfrac{\partial g}{\partial t} = k, \\ \dfrac{\partial k}{\partial t} = k \times k - \frac{1}{2}(\mathrm{Tr}k)k + \frac{1}{8}[(\mathrm{Tr}k)^2 - k\cdot k]g - 2\mathrm{Ric}(g) \\ \qquad + \frac{1}{2}R(g)g \\ \qquad = S_g(k) - 2\mathrm{Ric}(g) + \frac{1}{4}\mathcal{K}(g,k)g, \end{cases}$$

where $\mathcal{K}(g,k) = \frac{1}{2}[(\mathrm{Tr}k)^2 - k\cdot k] + 2R(g)$, and S_g is the DeWitt spray.

The basic conservation laws for this system of equations are

Proposition 3.4: Let (g_t, k_t) satisfy the equations of Proposition 3.3:

$$\begin{cases} \dfrac{\partial g}{\partial t} = k, \\ \dfrac{\partial k}{\partial t} = S_g(k) - 2\mathrm{Ric}(g) + \frac{1}{4}\mathcal{K}(g,k). \end{cases}$$

Let $\pi = [(\mathrm{Tr}k)g - k]^{-1}\mu_g$. Then

$$\frac{\partial}{\partial t}(\delta\pi)^\flat = 0$$

and

$$\frac{\partial}{\partial t}[\mathcal{K}(g,k)\mu_g] + 2\delta\delta\pi = 0,$$

where $(\delta\pi)^\flat = -\pi_i{}^j{}_{|j}$ and

where $\delta\delta\pi = \pi^{ij}{}_{|i|j}$ = the covariant double divergence.

If at $t = 0$, $\delta\pi = 0$, then this condition is maintained for all t for which the solution is defined, say $|t| < \epsilon$, and $(\partial/\partial t)[\mathcal{K}(g,k)\mu_g] = 0$. If $\delta\pi = 0$ and $\mathcal{K}(g,k) \equiv 0$ at $t = 0$, then $\mathcal{K}(g,k) \equiv 0$ is maintained for all t, $|t| < \epsilon$.

Remark: Note that $(\delta\pi)^\flat$ is conserved whether or not $(\delta\pi)^\flat = 0$ initially, but that $\mathcal{K}(g,k)\mu_g$ is independent of time only if $\delta\pi = 0$. It is conceptually best to derive these conservation laws from general symmetry principles. We do that for the divergence condition in Sec. 6 using the symmetry group $\mathfrak{D} = \mathrm{Diff}(M)$. In Proposition 8.1 we show that $(\partial/\partial t)[\mathcal{K}(g,k)\mu_g] = 0$ if the theory is invariant under the relativistic time-translation group $\mathcal{T} = C^\infty(M;\mathbf{R})$, which is the case for the empty space field equations. Here we show directly that $(\partial/\partial t)(\delta\pi)^\flat = 0$ and $(\partial/\partial t)[\mathcal{K}(g,k)\mu_g] + 2\delta\delta\pi = 0$ for any solution (g_t, k_t) of the equations in Proposition 3.4. Of course such a continuity equation is a general feature of Lagrangian field theories. Proposition 3.4, in essence, goes back to Arnowitt-Deser-Misner.[2]

Proof of Proposition 3.4: First we show $(\partial/\partial t)[\mathcal{K}(g,k)\mu_g] + 2\delta\delta\pi = 0$: Let $\mathcal{K} = \frac{1}{2}[(\mathrm{Tr}k)^2 - k\cdot k]$ be the kinetic energy scalar. Then from Proposition 3.3 and the proofs of Propositions 3.1 and 3.2 we have

$$\begin{aligned} \frac{d}{dt}\mathcal{K}\mu_g &= \tfrac{1}{2}\frac{dk}{dt}\cdot(g\mathrm{Tr}k - k)^{-1}\mu_g + \tfrac{1}{2}k\cdot\frac{d}{dt}[(g\mathrm{Tr}k - k)^{-1}\mu_g] \\ &= \tfrac{1}{2}[k\times k - \tfrac{1}{2}(\mathrm{Tr}k)k + \tfrac{1}{4}g\mathcal{K} - 2\mathrm{Ric}(g) \\ &\quad + \tfrac{1}{2}R(g)g]\cdot(g\mathrm{Tr}k - k)^{-1}\mu_g \\ &\quad + \tfrac{1}{2}k\cdot\{k\times k - (\mathrm{Tr}k)k + \tfrac{1}{2}g\mathcal{K} \\ &\quad + 2[\mathrm{Ric}(g) - \tfrac{1}{2}R(g)g]\}^{-1} \\ &= 2[\mathrm{Ric}(g) - \tfrac{1}{2}R(g)g]\cdot k. \end{aligned}$$

Also,

$$\frac{d}{dt}[2R(g)\mu_g] = 2\{\Delta(\mathrm{Tr}k) + \delta\delta k - [\mathrm{Ric}(g) - \tfrac{1}{2}R(g)g]\cdot k\}\mu_g$$

so that

$$\begin{aligned} \frac{d}{dt}[\mathcal{K}\mu_g + 2R(g)\mu_g] &= 2(\Delta\mathrm{Tr}k + \delta\delta k)\mu_g \\ &= -2\delta\delta\{((\mathrm{Tr}k)g - k)\mu_g\} \\ &= -2\delta\delta\pi. \end{aligned}$$

Thus $(\partial/\partial t)\mathcal{K}\mu_g + 2\delta\delta\pi = 0$.

To show that $(\partial/\partial t)(\delta\pi)^\flat = 0$ we proceed as follows. From Eqs. (3.2) and (3.4) we have

$$\frac{\partial\pi}{\partial t} = \{k\times k - (\mathrm{Tr}k)k + \tfrac{1}{2}\mathcal{K}g + 2[\mathrm{Ric}(g) - \tfrac{1}{2}R(g)g]\}^{-1}\mu_g. \tag{3.5}$$

Note that (3.5) is the system of Proposition 3.3 in Hamiltonian form, namely

$$\frac{d}{dt}D_2L(g,k) = D_1L(g,k) = -D_1H(g,\pi),$$

where $H(g,\pi) = \frac{1}{2}\mathcal{G}_g(\mathcal{G}^\#(\pi), \mathcal{G}^\#(\pi)) + V(g)$.

A computation in coordinates shows that

$$\frac{\partial}{\partial t}(\delta\pi)^\flat = \left[\delta\left(\frac{\partial\pi}{\partial t}\right)^\flat + \delta((\mathrm{Tr}k)k - k\times k - \tfrac{1}{2}\mathcal{K}g)\mu_g\right]. \tag{3.6}$$

Substitution of (3.5) into (3.6) yields our result $(\partial/\partial t)(\delta\pi)^\flat = 0$. ■

In the case of electromagnetic fields E, B, $\mathcal{K} = [g(E,E) + g(B,B)]$ and $-\delta\pi = *(E\wedge B)\mu_g$ so that

$$\begin{aligned} \frac{d}{dt}(\tfrac{1}{2}\mathcal{K}\mu_g) + \delta\delta\pi &= \frac{d}{dt}(\tfrac{1}{2}[g(E,E) + g(B,B)]\mu_g) \\ &\quad - \delta*(E\wedge B) = 0. \end{aligned}$$

which is just Poynting's theorem. Here $*$ is the Hodge star operator which maps k-forms into $(n-k)$-forms. In this case we do not have $(\delta\pi)^\flat = 0$ conserved; cf. Secs. 6, 7, and 8.

From Propositions 3.3 and 3.4 we find that a solution (g_t, k_t) of

$$(Z)\quad \begin{cases} \dfrac{\partial g}{\partial t} = k, \\ \dfrac{\partial k}{\partial t} = S_g(k) - 2\mathrm{Ric}(g) + \frac{1}{4}\mathcal{K}(g,k)g, \end{cases}$$

whose Cauchy data (g_0, k_0) at $t = 0$ satisfies

$$(C)\quad \begin{cases} \delta[(\mathrm{Tr}k)g - k] = 0, \\ \mathcal{K}(g,k) = 0, \end{cases}$$

satisfies (C) for all time t for which the solution exists, say $|t| < \epsilon$, and hence also satisfies the truncated system of evolution equations

$$(\tilde{E}) \begin{cases} \dfrac{\partial g}{\partial t} = k, \\ \dfrac{\partial k}{\partial t} = S_g(k) - 2\mathrm{Ric}(g). \end{cases}$$

Solutions to the system $(\tilde{E})$ however need not be solutions to (Z), nor is it *a priori* obvious that solutions to $(\tilde{E})$ preserve the conditions (C) on the Cauchy data. Our next proposition shows, in fact, that the simpler system $(\tilde{E})$ is in fact equivalent to (Z) under the hypothesis that the Cauchy data satisfies (C).

Proposition 3.5: Let (g_t, k_t) satisfy the system $(\tilde{E})$ with Cauchy data (g_0, k_0) that satisfies (C). Then $\mathcal{H}(g_t, k_t) \equiv 0$ so that (g_t, k_t) also satisfies (Z).

Proof: If (g_t, k_t) satisfies $(\tilde{E})$, then a computation as in Proposition 3.4 shows that

$$\frac{\partial \pi}{\partial t} = \{k \times k - (\mathrm{Tr}k)k + 2[\mathrm{Ric}(g) - \tfrac{1}{2}R(g)g] - R(g)g\}^{-1}\mu_g. \quad (3.7)$$

Also, a direct verification shows that

$$\frac{\partial}{\partial t}(\delta\pi)^\flat = \left[\delta\left(\frac{\partial \pi}{\partial t}\right)^\flat + \delta((\mathrm{Tr}k)k - k \times k - \tfrac{1}{2}\mathcal{H}g)\mu_g\right]. \quad (3.8)$$

Combining (3.7) and (3.8) gives

$$\frac{\partial}{\partial t}(\delta\pi)^\flat + \tfrac{1}{2}\delta[(\mathcal{H}\mu_g)g] = 0. \quad (3.9)$$

A computation similar to the one in Proposition 3.4 gives

$$\frac{\partial(\mathcal{H}\mu_g)}{\partial t} + 2\delta\delta\pi + \tfrac{1}{2}(\mathrm{Tr}k)\mathcal{H}\mu_g = 0. \quad (3.10)$$

Consider (3.9) and (3.10) as a first order linear homogeneous system of partial differential equations for $(\mathcal{H}\mu_g)$ and $(\delta\pi)^\flat$. Then if $\mathcal{H}\mu_g = 0$ and $(\delta\pi)^\flat = 0$ at $t = 0$, (3.9) and (3.10) imply $\mathcal{H}\mu_g = 0$, $(\delta\pi)^\flat = 0$ for all t for which (g_t, k_t) satisfies (E). Hence (g_t, k_t) satisfies (Z). ■

We remark that the proposition also follows if we assume that solutions to $(\tilde{E})$ are unique.[14] Let (g_t, k_t) be a solution of $(\tilde{E})$ with Cauchy data that satisfies (C), and let $(\bar{g}_t, \bar{k}_t)$ be a solution to (Z) with this same Cauchy data. Then $(\bar{g}_t, \bar{k}_t)$ is a solution to $(\tilde{E})$ and by the uniqueness assumption, $(g_t, k_t) = (\bar{g}_t, \bar{k}_t)$ is also a solution to (Z). Unfortunately, there is not as yet a *direct* existence or uniqueness proof for the system $(\tilde{E})$. Using a four-dimensional formulation and a theorem of Leray, Lichnerowicz[17] proves that Cauchy data

$$\left(g_{\mu\nu}(x^i), \frac{\partial g_{\mu\nu}}{\partial t}(x^i)\right), \quad 0 \le \mu, \nu \le 3,$$

of Sobolev class (H^s, H^{s-1}) evolves for short time into a space-time of class H^{s-1}. Using an improvement of the Leray theorem by Dionne, Choquet-Bruhat[13] shows in fact that the space-time is of class H^s. In a forthcoming paper we give a simple direct proof that (H^s, H^{s-1}) Cauchy data evolves for short time into a space-time of class H^s, $s \ge 4$ (see Refs. 18 and 19).

Eardley, Liang, and Sachs[15] give conditions when the Ricci term may be neglected, called *velocity-dominated solutions*. This condition prevails when the metric is changing very fast as compared with its curvature, as for example near a singular hypersurface. As $\mathrm{Ric}(g)$ is the only term involving spatial derivatives, neglecting $\mathrm{Ric}(g)$ reduces the equations to the geodesic equations on $\mathfrak{M}$ which can then be solved explicitly.

As explained in Ref. 1, the evolution equations $(\tilde{E})$ plus the initial conditions (C) are equivalent to the statement that the Lorentz metric g^L given in a neighborhood $(-\epsilon, \epsilon) \times M$ of the initial hypersurface $\{0\} \times M$ by

$$g^L(t, m) \cdot ((r, v_m), (s, w_m)) = g_t(m)\,(v_m, w_m) - rs,$$

where $(r, v_m), (s, w_m) \in T_{(t,m)}(\mathbf{R} \times M) \approx \mathbf{R} \times T_mM$ and g_t is the time-dependent metric with interval of existence $(-\epsilon, \epsilon)$, is Ricci flat; that is, Einstein's empty space field equations hold. In coordinates the formula reads

$$g^L_{\alpha\beta}dx^\alpha dx^\beta = -dt^2 + g_{ij}dx^i dx^j,$$

where $x^\alpha = (t, x^i)$. As we shall explain in Sec. 7, there are compelling reasons why we want to restrict our solutions to satisfy (C) in addition to the fact that only then do the solutions correspond to Ricci flat space-times.

Note that we are not postulating that the whole space-time is of the form $\mathbf{R} \times M$; rather it is of the form $(-\epsilon, \epsilon) \times M$ only in a tubular neighborhood of the initial hyperspace. As the metric evolves in time, the topology of the space-time structure could become more intricate. This global aspect in time is a difficult problem, closely related to the singularity problem, about which little is known.[17,19,20]

The above construction is for a space-time in Gaussian coordinates $g_{00} = -1$, $g_{0i} = 0$. To get the most general space-time, we must modify the equations of evolution to include the shift and lapse function.

We shall deal with the shift and lapse separately since their geometrical meanings are quite different; but they can be handled simultaneously or in succession. When done together, one uses the semidirect product group structure on $\mathcal{T} \times \mathcal{D}$.

4. EINSTEIN'S EQUATIONS WITH A SHIFT

If g_{0i} is nonzero and we write $X = X^i = -g^{ij}g_{0j}$ (g^{ij} is the inverse of the time-dependent 3-metric g_{ij}) so that X is a time-dependent vector field on M, then the evolution equations corresponding to the metric

$$g_{\alpha\beta}dx^\alpha dx^\beta = -(1 - X^iX_i)dt^2 - 2X_i dt dx^i + g_{ij}dx^i dx^j$$

are

$$\begin{cases} \dfrac{\partial g}{\partial t} = k - L_X g, \\ \dfrac{\partial k}{\partial t} = S_g(k) - 2\mathrm{Ric}(g) - L_X k. \end{cases}$$

Because of the presence of the Lie derivative in the first term, k is no longer the canonical velocity associated with the configuration field g, but is now a supplementary variable, defined to simplify the evolution equations; that is, the equation for $\partial k/\partial t$ is simpler than the equation for the acceleration $\partial^2 g/\partial t^2 = \partial k/\partial t - (\partial/\partial t)(L_X g)$. We let $h = \partial g/\partial t$ denote the velocity canonically conjugate to the field g.

In Ref. 3 the above equations are considered as Lagrange's equations on $\mathfrak{M}$. However in this approach the kinetic energy $K(g,h) = \frac{1}{2}\mathcal{G}_g(h + L_X g, h + L_X g)$ fails to be a quadratic function of the velocity h (although it is a quadratic function of the supplementary variable k).

In order to have a kinetic energy term which is quadratic, and to incorporate the shift vector field into the theory in a natural way, we enlarge the configuration space $\mathfrak{M}$ to $\mathfrak{D} \times \mathfrak{M}$. We recall that $\mathfrak{D} = \mathrm{Diff}(M)$ is the group of all smooth orientation-preserving diffeomorphisms of M. We can still regard the equations of Proposition 3.3 on $\mathfrak{D} \times \mathfrak{M}$ by just ignoring the factor $\mathfrak{D}$; namely, for $X \in T\mathfrak{D}$, $(g,h) \in T\mathfrak{M}$ we have

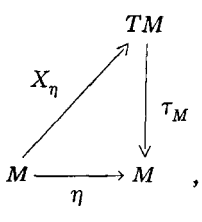
$$\begin{cases} \dfrac{\partial \eta}{\partial t} = X, & \dfrac{\partial g}{\partial t} = h, \\ \dfrac{\partial X}{\partial t} = \dot{X}, & \dfrac{\partial h}{\partial t} = S_g(h) - 2\mathrm{Ric}(g) + \frac{1}{4}\mathfrak{K}(g,h). \end{cases}$$

These equations come from a degenerate Lagrangian on $\mathfrak{D} \times \mathfrak{M}$, $\bar{L}(X,g,h) = L(g,h)$. The degeneracy is clear because the Lagrangian $\bar{L}$: $T(\mathfrak{D} \times \mathfrak{M}) \to \mathbf{R}$ does not depend on X; thus, X can be specified arbitrarily. We give a less trivial extension of the Lagrangian L to $\mathfrak{D} \times \mathfrak{M}$ shortly.

At this point it will be necessary to set out a few properties of the diffeomorphism group $\mathfrak{D}$ of our manifold M. We shall need only the most elementary aspects of this group, which can all be understood rather easily, as we shall explain. For the more detailed analysis, consult Refs. 7 and 10 and related references.

To begin with, $\mathfrak{D}$ is an infinite-dimensional manifold. It is not a linear space, as M does not have a linear structure, but $\mathfrak{D}$ is locally like C^∞ functions; hence it is plausible that $\mathfrak{D}$ has the structure of a manifold modeled on a Fréchet space of C^∞ (vector) functions.

What we would like to demonstrate is that the tangent space $T_\eta\mathfrak{D}$ at a point $\eta \in \mathfrak{D}$ is the set of smooth maps $X_\eta : M \to TM$ which cover η; that is, such that the following diagram commutes:



where τ_M denotes the canonical projection of TM to M. To see that X_η is of the form described, let $\eta_t \in \mathfrak{D}$ be a curve in $\mathfrak{D}$, $\eta_0 = \eta$, so that $(d\eta_t/dt)|_{t=0}$ represents a tangent vector in $T_\eta\mathfrak{D}$.[8] But for fixed $m \in M$, $\sigma(t) = \eta_t(m)$ is a curve in M with $\sigma(0) = \eta(m)$ and with tangent vector

$$\sigma'(0) = \frac{d\eta_t}{dt}(m)\Big|_{t=0} \in T_{\eta(m)}M.$$

Thus $d\eta_t/dt$ is a map from M to TM covering η.

We refer to X_η *as a vector field which covers* η, so that $T\mathfrak{D}$ is the *manifold of vector fields covering diffeomorphisms*. In particular, $T_e\mathfrak{D} = \mathfrak{X}(M) = \{$the vector space of smooth vector fields on $M\} = \{$the Lie algebra of $\mathfrak{D}\}$. There is a natural projection $\bar{\tau}$: $T\mathfrak{D} \to \mathfrak{D}$ defined by $\bar{\tau}(X_\eta) = \tau_M \circ X_\eta = \eta \in \mathfrak{D}$. For $X_\eta \in T_\eta\mathfrak{D}$, $X_\eta \circ \eta^{-1}$ is an "ordinary" vector field on M. The assertion $d\eta_t/dt = X_{\eta_t}$ means that η_t is the *flow* of the time-dependent vector field $X_t = X_{\eta_t} \circ \eta_t^{-1} = d\eta_t/dt \circ \eta_t^{-1}$. In other words, η_t gives the integral of the ordinary differential equations defined by X_t, or

$$\frac{d\eta_t(x)}{dt} = X_t(\eta_t(x)).$$

Now we introduce a new manifold $\mathfrak{A}$ of maps of the form $g \circ \eta$ where $g \in \mathfrak{M}$ and $\eta \in \mathfrak{D}$. This is isomorphic to $\mathfrak{D} \times \mathfrak{M}$ by mapping $(\eta, g) \mapsto g \circ \eta$. This map may be viewed as realizing $\mathfrak{A}$ as $\mathfrak{D} \times \mathfrak{M}$ by *right translation* or as $\mathfrak{A}$ in *body coordinates*. We can also realize $\mathfrak{A}$ in *space coordinates* using *"left translation"*:

$$g \circ \eta \mapsto ((\eta^{-1})^* g, \eta),$$

where $(\eta^{-1})^* g$ represents a new metric obtained by "actively changing the coordinates" by the diffeomorphism η. If we let $X_m, Y_m \in T_m M$, then

$$(\eta^{-1})^* g(m) \cdot (X_m, Y_m) = g \circ \eta^{-1}(m) \cdot (T_m\eta^{-1}(X_m), T_m\eta^{-1}(Y_m)).$$

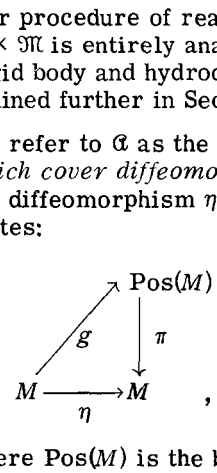
In a coordinate system, $x = (x^1, \ldots, x^n)$, this operation $(\eta^{-1})^* g$ reads as follows: Let $\bar{x}^i$ be the ith coordinate of $\bar{x} = \eta(x)$; we suppose for simplicity that the coordinate chart is so large that η maps it to itself. Then the new g has coordinates in this system given by

$$\bar{g}_{ij}(\bar{x}) = \frac{\partial x^k}{\partial \bar{x}^i}(x) \cdot \frac{\partial x^l}{\partial \bar{x}^j}(x) g_{kl}(x).$$

Our conventions on the placing of the stars agrees with the convention in Ref. 12 but is the opposite of that in Ref. 8. For example, $(\eta \circ \zeta)^* g = \zeta^* \eta^* g$.

Our procedure of realizing $\mathfrak{A}$ in these two ways as $\mathfrak{D} \times \mathfrak{M}$ is entirely analogous to what occurs in the rigid body and hydrodynamics, cf. Ref. 7. This is explained further in Sec. 5 below.

We refer to $\mathfrak{A}$ as the *manifold of Riemannian metrics which cover diffeomorphisms*, that is, $g_\eta \in \mathfrak{A}$ covers the diffeomorphism η if the following diagram commutes:



where $\mathrm{Pos}(M)$ is the bundle of positive-definite symmetric two-covariant tensors (not tensor fields) on M. In the above realizations of $\mathfrak{A}$ as $\mathfrak{D} \times \mathfrak{M}$, the space $\mathfrak{M}$ plays the role of the "Lie algebra" of $\mathfrak{A}$, so that right pullback to the "Lie algebra" is given by $g_\eta \mapsto g_\eta \circ \eta^{-1}$, and left pullback is given by $g_\eta \mapsto (\eta^{-1})^*(g_\eta \circ \eta^{-1})$; $g_\eta \circ \eta^{-1}$ is the *representation of* g_η *in body coordinates*, $(\eta^{-1})^*$

($g_\eta \circ \eta^{-1}$) *is the representation of* g_η *in space coordinates*, and the transition from body to space coordinates is given by $g_\eta \circ \eta^{-1} \mapsto (\eta^{-1})^*(g_\eta \circ \eta^{-1})$; that is, this map represents the transition from the "body representation" to the "space representation." This can also be interpreted more mundanely as the transition from a stationary to a moving frame.

To transfer the Lagrangian on $\mathfrak{M}$ to $\mathfrak{D} \times \mathfrak{M}$, we consider the map

$$\Phi\colon \mathfrak{D} \times \mathfrak{M} \to \mathfrak{M}, \qquad (\eta, g) \mapsto \eta^* g$$

(Φ is the standard left action of the group $\mathfrak{D}$ on $\mathfrak{M}$).

The tangent (or derivative) of Φ (not the tangent action) is easily computed. We shall prove that

$$T\Phi : T\mathfrak{D} \times T\mathfrak{M} \to T\mathfrak{M} \approx \mathfrak{M} \times S_2(\mathfrak{M}),$$
$$(X_\eta, (g,h)) \mapsto (\eta^* g, \eta^*(h + L_{X_\eta \circ \eta^{-1}} g)).$$

Note that $X_\eta \circ \eta^{-1}$ is an ordinary vector field so that $L_{X_\eta \circ \eta^{-1}} g$ is the usual Lie derivative. The proof that $T\Phi$ is as given follows from a lemma from geometry.

Lemma 4.1: If Y_t is a time-dependent vector field with flow $\eta_t (\eta_0 = id_M$ = identity), then for $g \in \mathfrak{M}$

$$\frac{d}{dt}(\eta_t^* g) = \eta_t^*(L_{Y_t} g).$$

This is the usual fundamental theorem connecting flows and Lie derivatives.[8]

To prove the formula for $T\Phi$, we may proceed as follows: Let η_t be the flow of the vector field $X_\eta \circ \eta^{-1}$, η_0 = identity, so that as a curve in $\mathfrak{D}$, η_t is tangent to $X_\eta \circ \eta^{-1} \circ \eta_t$, and $\eta_t \circ \eta$ is tangent at $t = 0$ to X_η. Let g_t be tangent at $t = 0$ to h and $g_0 = g$. Then by definition of the tangent[8]

$$T\Phi(X_\eta, (g,h)) = \frac{d}{dt}\,\Phi(\eta_t \circ \eta, g_t)\Big|_{t=0}.$$

Using the definition of Φ and Lemma 4.1, this becomes

$$\frac{d}{dt}(\eta_t \circ \eta)^* g_{t\ t=0} = \frac{d}{dt}(\eta)^* \eta_t^* g_t\Big|_{t=0} = \eta^* \left\{\frac{d}{dt}\,\eta_t^* g_t\right\}\Big|_{t=0}$$
$$= \eta^* \left\{\eta_t^* L_{X_\eta \circ \eta^{-1}} g_t + \eta_t^* \frac{dg_t}{dt}\right\}\Big|_{t=0},$$

which proves our assertion.

Note: If we had used the right action $\Psi(\eta, g) = \eta_* g = (\eta^{-1})^* g$, the formula for $T\Psi$ would be $(\eta_* g, \eta_* h + L_{X_\eta \circ \eta^{-1}}(\eta_* g))$, which is not as convenient for later purposes.

By composing L with $T\Phi$, we can extend our Lagrangian $L\colon T\mathfrak{M} \to \mathbf{R}$ to a degenerate Lagrangian $\bar{L}$: $T(\mathfrak{D} \times \mathfrak{M}) \to \mathbf{R}$ given by $\bar{L} = L \circ T\Phi$, that is, $\bar{L}(X_\eta, g, h) = L(\eta^* g, \eta^*(h + L_{X_\eta \circ \eta^{-1}} g)) = L(g, h + L_{X_\eta \circ \eta^{-1}} g)$, where the last equality follows from the invariance of L by the pullback action of $\mathfrak{D}$ (see Sec. 6).

We now explain why the Lagrangian $\bar{L}$ is quadratic in the velocities when viewed as a Lagrangian on $\mathfrak{D} \times \mathfrak{M}$, whereas it is not when viewed as a Lagrangian on $T\mathfrak{M}$. We write out $\bar{L}$ as

$$\bar{L}(X_\eta, g, h) = \tfrac{1}{2}\mathcal{G}_g(h + L_{X_\eta \circ \eta^{-1}} g, h + L_{X_\eta \circ \eta^{-1}} g) - 2\int_M R\mu_g.$$

To test whether $\bar{L}$ is quadratic, we must consider the transformation of the velocities $v \mapsto \lambda v$, $\lambda \in \mathbf{R}$. On $\mathfrak{D} \times \mathfrak{M}$, the velocities are both X_η and h (rather than just h), so that we must consider the transformation $(X_\eta, h) \mapsto (\lambda X_\eta, \lambda h)$. It then follows immediately that

$$\bar{L}(\lambda X_\eta, g, \lambda h) = \lambda^2 \bar{L}(X_\eta, g, h)$$

[whereas $\bar{L}(X_\eta, g, \lambda h) \neq \lambda^2 \bar{L}(X_\eta, g, h)$]. Working on $\mathfrak{M}$ alone, the fact that the shift is also a velocity (but not determined by evolution equations) is obscured; the nonquadratic nature of the Lagrangian on $\mathfrak{M}$ is trying to tell us that the shift is a velocity variable or equivalently that $\mathfrak{D}$ should be considered as part of the configuration space.

We also remark that $\bar{L}$ is now a degenerate Lagrangian on $\mathfrak{D} \times \mathfrak{M}$, as the metric term is now degenerate. Thus we have achieved a quadratic Lagrangian but only at the expense of giving up a nondegenerate one and also giving up well-defined equations of motion (see Sec. 2). Although it might appear that the price we have had to pay for the exchange is too great, this is not true. The degeneracy leads to an arbitrariness in the evolution equations which allows precisely for an arbitrary specification of a motion of M, that is a curve $\eta_t \in \mathfrak{D}$, or its generator, the shift vector field X_t.

Using $\bar{L}$ and Proposition 2.4, the equations of motion may also be transferred to $\mathfrak{D} \times \mathfrak{M}$.

Theorem 4.1: Consider on $\mathfrak{D} \times \mathfrak{M}$ the degenerate Lagrangian

$$L(\eta, X, g, h) = \tfrac{1}{2}\mathcal{G}_g(h + L_{X_\eta \circ \eta^{-1}} g, h + L_{X_\eta \circ \eta^{-1}} g) - 2\int_M R(g)\mu_g.$$

For any curve η_t with $d\eta_t/dt = X_t \circ \eta_t$, a possible Lagrangian vector field for L is given by the equations [at a point $(Y_\zeta, g, h) \in T(\mathfrak{D} \times \mathfrak{M})$]

$$\begin{cases} \dfrac{d\zeta_t}{dt} = X_t \circ \eta_t, & \dfrac{\partial g}{\partial t} = k - L_X g, \\[2ex] \dfrac{d(X_t \circ \eta_t)}{dt} = \dfrac{dY_\zeta}{\partial t}, & \dfrac{\partial k}{\partial t} = S_g(k) - 2\mathrm{Ric}(g) + \tfrac{1}{4}\mathcal{K}(g,k) - L_X k. \end{cases}$$

Note: The expression for L may be written (intrisically) directly on the manifold $\mathfrak{A}$.

As a corollary, we get a simple method for solving the equations with a general shift X_t if the solution for $X = 0$ is known.

Corollary 4.1: Let g_t, k_t be a solution of the Einstein system with $N = 1$, $X = 0$. Let X_t be a given time-dependent vector field with flow η_t, $\eta_0 = id$. Then the solution of the Einstein system with $N = 1$ and shift X_t and the same initial conditions (g_0, k_0) is given by

$$\bar{g}_t = (\eta_t^{-1})^* g_t, \qquad \bar{k}_t = (\eta_t^{-1})^* k_t.$$

To prove the corollary, we need a lemma.

Lemma 4.2: Let X_t be a time-dependent vector field and let the flow of X_t be η_t. Then $-(\eta_t^{-1})_* X_t$ has flow η_t^{-1}, where $(\eta_t^{-1})_* X_t = T\eta_t^{-1} \circ X_t \circ \eta_t$ is the "pullback"

of vector fields [$(\eta_t)_*$ is "push forward" of vector fields]. Let $g \in \mathfrak{M}$ and $k \in S_2(M)$. Then

$$\frac{d}{dt}(\eta_t^{-1})^* g = -L_{X_t}(\eta_t^{-1})^* g$$

and

$$\frac{d}{dt}(\eta_t^{-1})^* k = -L_{X_t}(\eta_t^{-1})^* k.$$

Proof: Let X_t be the generator of η_t and Y_t the generator of η_t^{-1}. By differentiating $\eta_t \circ \eta_t^{-1} = id_M$, we find

$$\frac{d}{dt}(\eta_t \circ \eta_t^{-1}) = \frac{d\eta_t}{dt} \circ \eta_t^{-1} + T\eta_t \circ \frac{d\eta_t^{-1}}{dt} = X_t + T\eta_t \circ Y_t \circ \eta_t^{-1}$$
$$= X_t + (\eta_t)_* Y_t = 0,$$

where $(\eta_t)_* Y_t = T\eta_t \circ Y_t \circ \eta_t^{-1}$. Thus $Y_t = -(\eta_t^{-1})_* X_t$ is the generator of the flow η_t^{-1}. From Lemma 4.1, for $g \in \mathfrak{M}$, or $k \in S_2(M)$,

$$\frac{d}{dt}(\eta_t^{-1})^* g = (\eta_t^{-1})^* L_{Y_t} g = L_{(\eta_t)_* Y_t}(\eta_t^{-1})^* g = -L_{X_t}(\eta_t^{-1})^* g,$$

where the second equality follows from the fact that $L_{Y_t} g$ is a tensor and hence commutes with push forward in each of its arguments; in coordinates this is just covariance with respect to coordinate transformations.

Note: We have given these details because care is required when working with time-dependent vector fields X_t. Indeed X_t is *not* invariant under its own flow. ■

Proof of Corollary 4.1: By the lemma we get

$$\frac{\partial \bar{g}_t}{\partial t} = (\eta_t^{-1})^* \frac{\partial g_t}{\partial t} - L_{X_t}\bar{g}_t = \bar{k}_t - L_{X_t}\bar{g}_t,$$

since $\partial g_t/\partial t = k_t$ when $X_t = 0$.

Similarly,

$$\frac{\partial \bar{k}_t}{\partial t} = (\eta_t^{-1})^* \frac{\partial k_t}{\partial t} - L_{X_t}\bar{k}_t$$
$$= (\eta_t^{-1})^*(S_{g_t}(k_t)) - 2(\eta_t^{-1})^*(\mathrm{Ric}(g_t)) - L_{X_t}\bar{k}_t$$
$$= S_{\bar{g}_t}(\bar{k}_t) - 2\mathrm{Ric}(\bar{g}_t) - L_{X_t}\bar{k}_t,$$

where the last equality again follows from the fact that $S(\cdot,\cdot)$ and $\mathrm{Ric}(\cdot)$ are tensor operators and hence commute with push forward (again, in coordinates, this is just covariance with respect to coordinate transformations), and we have used the equations for $\partial k_t/\partial t$ when $X_t = 0$. ■

Corollary 4.1 shows that even though the evolution equations with a shift involve extra nonlinear, second order terms [since $L_X k = L_X(h + L_X g)$ is quadratic in the velocities], the more general system can be solved merely by solving an ordinary differential equation; that is, by finding the flow of X_t.

The above geometry also makes it transparent how the space–time in the presence of a shift is to be constructed. Namely we have a diffeomorphism

$$\Psi: \mathbf{R} \times M \to \mathbf{R} \times M, \qquad \Psi(t,m) = (t, \eta_t(m)),$$

which rotates the space M. It transforms, by Corollary 4.1, the old solution to the new solution. Thus it transforms the old space–time to the new one. It is easy to check that the new space–time metric is the one stated at the beginning of this section. Thus the space–time with a shift is isometric to the space–time without a shift.

We now explain why we have changed the sign of the shift (see also the next section). If the first evolution equation, for example, were $\partial g_t/\partial t = k_t + L_{X_t} g_t$, then we would consider $\bar{g}_t = \eta_t^* g_t$ and $\bar{k}_t = \eta_t^* k_t$ as the solution with shift X_t if (g_t, k_t) is the solution with shift zero. But then

$$\frac{\partial \bar{g}_t}{\partial t} = \eta_t^* \frac{\partial g_t}{\partial t} + \eta_t^* L_{X_t} g_t = \bar{k}_t + L_{(\eta_t^{-1})_* X_t}(\eta_t^* g_t)$$
$$= \bar{k}_t + L_{(\eta_t^{-1})_* X_t}\bar{g}_t,$$

so that now the equations depend on η_t explicitly, which is not natural.

5. THE EINSTEIN SYSTEM IN SPACE AND BODY COORDINATES

Interestingly enough, it is possible to interpret the Einstein system (E) in terms analogous to the concepts of space and body coordinates used to describe the motion of a rigid body or of a fluid in hydrodynamics. The basis for this interpretation is the two identifications of $\mathcal{G}$, the manifold of Riemannian metrics which cover diffeomorphisms, with $\mathfrak{D} \times \mathfrak{M}$; the right identification $g_\eta \mapsto (\eta, g_\eta \circ \eta^{-1}) \in \mathfrak{D} \times \mathfrak{M}$ leading to what we loosely call "body coordinates," and the left identification $g_\eta \mapsto (\eta, (\eta^{-1})^*(g_\eta \circ \eta^{-1})) \in \mathfrak{D} \times \mathfrak{M}$ leading to "space coordinates." Here $\mathfrak{M}$ plays the role of the "Lie algebra" of $\mathcal{G}$ and is therefore analogous to the velocity phase space $T_e SO(3) \approx \mathbf{R}^3$ for the rigid body or $T_{id_M}\mathfrak{D}_\mu = \mathfrak{X}_0(M)$ (the space of all divergence free vector fields) for hydrodynamics (see Ref. 7).

If we transpose the Lie derivative terms in (E) to the left-hand side, we see that the operator $(\partial/\partial t) + L_{X_t}$ enters in the evolution equations for both g_t and k_t. The derivatives $(\partial g_t/\partial t) + L_{X_t} g_t$ and $(\partial k_t/\partial t) + L_{X_t} k_t$ are entirely analogous to the material or Eulerian derivative $(\partial X_t/\partial t) + \nabla_{X_t} X_t$ which appears in hydrodynamics or the time derivative $(d(I\cdot\omega)/dt) + \omega \times (I\cdot\omega)$ of the angular momentum $L = I \cdot \omega$ of a rigid body as observed in space coordinates.

The Eulerian derivative is the total time derivative of the fluid velocity as the fluid moves around in space. Although it is the time derivative of the velocity with respect to an observer who is moving with the fluid, it is expressed in terms of quantities referring to points fixed in space; that is, it is the total time derivative of the fluid as seen by an observer fixed in space. We say that an observer moving with the fluid is in *body coordinates*, or is "on" the fluid, and an observer fixed in space is in *space coordinates*, or is "off" the fluid.

We now wish to investigate further this analogy of general relativity with hydrodynamics in which the derivatives $(\partial g_t/\partial t) + L_{X_t} g_t$ and $(\partial k_t/\partial t) + L_{X_t} k_t$ can be interpreted as the total derivative of a time-dependent metric field g_t (or of k_t) as seen by an observer in *space coordinates*. We let the manifold M be the *body*. We consider a curve $\eta_t \in \mathfrak{D}(M)$, $\eta_0 = id_M$, as describing a *rotation* of the body M. Thus we consi-

der the points of the manifold M to be *moving* according to the rule that a point which at time $t = 0$ is at $m \in M$ is at $\eta_t(m)$ after time t. We then make the convention that an observer is in *body coordinates* if he is *on* the manifold, and is in *space coordinates* if he is *off* the manifold. An observer in body coordinates, as he moves with the manifold, detects no motion of the manifold.

Now let g_t be a time-dependent metric field on M. We assume that this field is rigidly attached to the body M as it moves according to the curve η_t, so that we set $g_t = g_{\text{body}}$ (time-dependence implied). An observer in body coordinates (who detects no motion of M) then finds $(\partial g_{\text{body}}/\partial t) = k_{\text{body}}$ as the "velocity" of the metric.

An observer in space coordinates (who is off the manifold) sees the metric field g_{body} as it is dragged past him by the moving manifold. He sees the (time-dependent) metric field $g_{\text{space}} = (\eta_t^{-1})^* g_{\text{body}}$ and computes as the "velocity" of g_{space};

$$\frac{\partial g_{\text{space}}}{\partial t} = k_{\text{space}} - L_X g_{\text{space}}, \tag{5.1}$$

where $k_{\text{space}} = (\eta_t^{-1})^* k_{\text{body}} = (\eta_t^{-1})^* (\partial g_{\text{body}}/\partial t)$ and X is the shift vector field which generates the motion η_t of M. Similarly he computes

$$\frac{\partial k_{\text{space}}}{\partial t} = S_{g_{\text{space}}}(k_{\text{space}}) - 2\text{Ric}(g_{\text{space}}) - L_X k_{\text{space}}. \tag{5.2}$$

But (5.1) and (5.2) are just the evolution equations of (E) with lapse $N = 1$.

6. CONSERVATION OF THE DIVERGENCE CONDITION

In Proposition 3.4 we saw that the divergence condition is maintained by the Einstein equations. Now we want to give a more natural geometric proof of this fact using general symmetry methods.

The idea is extremely simple and goes as follows. Group $\mathfrak{D}$ acts on $\mathfrak{M}$ by $g \mapsto (\eta^{-1})^* g$ as we have seen before. We assert that this action is a symmetry for our Lagrangian $L(g, h) = \frac{1}{2}\mathcal{G}_g(h, h) - V(g)$ and that the corresponding conserved quantities, computed according to Proposition 2.3 give us the desired conservation law.

Of course by Corollary 4.1 it is enough to show this for $X = 0$; we get the corresponding result with a shift immediately (and for a lapse too using the results below).

Let us denote for fixed $\eta \in \mathfrak{D}$, the map $g \mapsto (\eta^{-1})^* g$ by θ_η. First we assert that θ_η is an isometry for the DeWitt metric. This is almost obvious since $\mathcal{G}_g$ is defined intrinsically and everything transforms properly (see Refs. 5 or 10 for an analogous result). Secondly, θ_η leaves invariant the potential. Indeed, since R is a tensor,

$$\int_M R((\eta^{-1})^* g)\mu_{(\eta^{-1})^* g} = \int_M [R(g) \circ \eta^{-1}](\eta^{-1})^* \mu_g = \int_M R(g)\mu_g$$

by the change of variables formula [$(\eta^{-1})^* \mu_g$ is just the Jacobian of η^{-1} times μ_g].

We can therefore compute the conserved quantities using Proposition 2.3. Let ζ_t be a one-parameter group of diffeomorphisms generated by a vector field Z on M. Then θ_{ζ_t} is a one-parameter group of motions of $\mathfrak{M}$ leaving L fixed and θ_{ζ_t} is generated by $g \mapsto -L_Z g$, a Killing vector field on $\mathfrak{M}$. The corresponding conserved quantity is therefore the real-valued function on $T\mathfrak{M}$ given by

$$(g, k) \mapsto \mathcal{G}_g(k, -L_Z g).$$

Lemma 6.1: $\int_M (L_Z g \cdot k)\mu_g = 2 \int_M (Z \cdot \delta k)\mu_g$.

Proof: As was stated in the introduction, and as is easy to check, we have the formula

$$L_Z g = Z_{i|j} + Z_{j|i}.$$

From this we see that $\delta(k \cdot Z) = (\delta k) \cdot Z - k \cdot \nabla Z = (\delta k) \cdot Z - \frac{1}{2} k \cdot L_Z g$.

By Stokes theorem we have that $\int_M \delta(k \cdot Z)\mu_g = 0$, and so the lemma follows. Recall that $\delta k = -k_{i|j}$. ■

Now $\mathcal{G}_g(k, L_Z g) = \int_M (L_Z g \cdot \pi')\mu_g$, where $\pi' = [(\text{Tr}k)g - k]^{-1}$ is the tensor part of $\pi = \pi' \otimes \mu_g$. From Lemma 6.1 we get the fact that

$$\int_M (Z \cdot \delta\pi')\mu_g \text{ is conserved.}$$

Since Z is arbitrary, $(\delta\pi')^\flat \otimes \mu_g = (\delta\pi)^\flat$ is conserved, where, as above, the symbol $(\ )^\flat$ indicates that the index is lowered by the time-dependent metric g_{ij}, that is $(\delta\pi')^\flat = -(\pi')_{ij|j}$. Thus if $\delta[(\text{Tr}k)g - k] = 0$ at $t = 0$ then this condition is maintained in time, thereby proving our conservation law.

Note that $\delta\pi'$ must be taken with its index down, that is, regarded as a one form, in order that its contration with Z does not involve the metric. Note that we have shown $(\delta\pi)^\flat$ is conserved even if $(\delta\pi)^\flat$ is *not* zero at $t = 0$, although this is not true for $\delta((\text{Tr}k)g - k)$ because of the μ_g term (see also Ref. 19). This is for the *full* set of Hamiltonian equations as in Proposition 3.3. For the truncated system it is necessary to require that $\delta\pi$ and $\mathcal{H}$ both be zero at $t = 0$ as we saw in Proposition 3.4.

Geometrically, $\delta((\text{Tr}k)g - k) = 0$ on $\mathfrak{M}$ means that k is perpendicular to the orbit of $\mathfrak{D}$ through g (perpendicular in the DeWitt metric). This is exactly a restatement of the condition $\mathcal{G}_g(k, L_Z g) = 0$ for all Z.

Thus the conservation of $\delta((\text{Tr}k)g - k)$ means that if (g, k) starts off perpendicular to the orbit through g, it must evolve in such a manner that it remains perpendicular to the orbit. If a shift is present, we measure perpendicularity by the DeWitt metric on $\mathfrak{D} \times \mathfrak{M}$. In the metric $\mathcal{G}_g(\cdot, \cdot)$, this means that g_t, rather than proceeding perpendicular to the orbits, also "slides along" the orbits; this sliding is determined by the flow of the shift vector field.

We also remark that

$$L_X \pi = L_X(\pi' \otimes \mu_g) = L_X \pi' \otimes \mu_g + \pi' \otimes L_X \mu_g = L_X \pi' \otimes \mu_g + (\text{div}X)\pi' \otimes \mu_g,$$

where $L_X \mu_g = (\text{div}X)\mu_g$, $\text{div}X = -\delta X = X^i{}_{|i}$. Thus the Lie derivative of a tensor density has the extra

divergence term $(\mathrm{div}X)\pi'$ in it. Comparing this with the evolution equations in Arnowitt, Deser, and Misner[2] reveals the last three terms in the equation for $\partial\pi/\partial t$ as $L_X\pi$.

In summary, we have proven:

Theorem 6.1: Let L be a Lagrangian system $L(g,k) = \frac{1}{2}\mathcal{G}_g(k,k) - V(g)$ on $T\mathfrak{M}$ with Lagrangian vector field Z. Suppose V is invariant under the action of $\mathfrak{D}$ on $\mathfrak{M}$. Then if (g,k) is an integral curve of Z, $\partial/\partial t\{\delta[(\mathrm{Tr}k)g - k]\mu_g\} = 0$.

Thus if $\delta[(\mathrm{Tr}k)g - k] = 0$ holds at $t = 0$, it holds for all time.

In other words, the divergence condition results precisely from the spatial coordinate invariance of our Lagrangian. One can similarly work out laws for other coordinate invariant theories which are built on tensor or vector bundles other than $S_2(M)$.

7. THE EINSTEIN EQUATIONS WITH A LAPSE

We have just seen that the shift vector field X has a simple geometric interpretation and the solutions to the modified Einstein equations are related to those with zero shift in a very simple and geometrically transparent way. The lapse function is more interesting and a bit more intricate. In dealing with the lapse we may assume the shift is zero.

If one has a Lagrangian function on TB, there is a standard way (see Ref. 9, p. 133) of associating a homogeneous Lagrangian on $T(\mathbf{R}\times B)$. Namely set

$$\overline{L}: T\mathbf{R}\times TB \approx \mathbf{R}\times\mathbf{R}\times TB \to \mathbf{R},$$
$$\overline{L}(t,\lambda,v) = \lambda L(v/\lambda),\quad \lambda\neq 0.$$

Here $\lambda \in T_t\mathbf{R}$ is now a velocity, and so $\overline{L}$ is homogeneous of first order in the velocities. $\overline{L}$ is defined just on the subset where $\lambda \neq 0$. This $\overline{L}$ is degenerate and its base integral curves are obtained from those of L by suspending them in $\mathbf{R}\times B$ with an arbitrary change of parametrization (reflecting the degeneracy). Physically, the time has changed roles from being the evolution parameter of the system to a coordinate in the extended configuration space; one is then free to choose an arbitrary evolution parameter for the system.

With this classical example in mind, we thus extend our Lagrangian L on $T\mathfrak{M}$, $L(g,h) = \frac{1}{2}\mathcal{G}_g(h,h) - 2\int_M \mathrm{R}(g)\mu_g$ to $\overline{L}: T(\mathcal{T}\times\mathfrak{M})\to\mathbf{R}$ by setting for $(\xi,N)\in\mathcal{T}\times\mathcal{T}$, $N>0$,

$$\overline{L}(\xi,N,g,h) = \int_M N\mathcal{L}\left(g,\frac{h}{N}\right)\mu_g$$
$$= \tfrac{1}{2}\int_M N\left(\left(\mathrm{Tr}\frac{h}{N}\right)^2 - \frac{h}{N}\cdot\frac{h}{N}\right)\mu_g - 2\int_M N\mathrm{R}(g)\mu_g.$$

Recall $\mathcal{T}$ is the space of smooth functions $\xi: M\to\mathbf{R}$, $\mathcal{T} = C^\infty(M;R)$. Since $\mathcal{T}$ is a linear space, $T\mathcal{T} = \mathcal{T}\times\mathcal{T}$ and we denote elements in the tangent space by $(\xi,N)\in\mathcal{T}\times\mathcal{T}$. Note that the constant functions form a subgroup of $\mathcal{T}$ naturally isomorphic to $\mathbf{R}$; restricting to these functions we recover the classical extension of L. For the relativistic case, the introduction of $\mathcal{T}$ instead of $\mathbf{R}$ is quite natural as it allows for observers at different points of M to have different clock rates.

For the classical extension of L to $T(\mathbf{R}\times B)$, there is no problem about the existence of its Lagrangian vector field. One can easily check that the most general such second order vector field is given by

$$\overline{Z}(t,\lambda,v) = \alpha(t,\lambda)\oplus\lambda Z(v),$$

where α is any second order equation on $T\mathbf{R}$ (cf. Ref. 8, p. 136) and Z is the Lagrangian vector field for L on TB.

Now we come to a somewhat surprising result. This is that, when we extend L to $T(\mathcal{T}\times\mathfrak{M})$ as above, the Lagrangian vector field need not exist at every point of $T(\mathcal{T}\times\mathfrak{M})$. In fact, in the next theorem we shall see that we are *forced* to restrict to the set on which $\mathcal{H}$ is identically constant. The result is quite general for any Lagrangian system, although we deal explicitly with the case at hand. This provides the explanation of why $\mathcal{H}$ *must* be identically constant (generally taken zero) rather than just the total integrated energy being conserved; cf. Misner.[4,2]

Sachs[15] has pointed out that in some dust models, $\mathcal{H}$ can be a nonzero constant. Observe that $\mathcal{H}$ is the total Hamiltonian governing the evolution of all quantities in the theory. For instance in the presence of an electromagnetic field, $\mathcal{H} = -G_0{}^0 + T_0{}^0$ is the energy governing the evolution of *both* the gravitational and electromagnetic fields. For a physical solution $\mathcal{H} = 0$; the hypothesis (ii) in the theorem below means physically that our system is relativistic in the sense that one cannot physically distinguish between the various spacelike hypersurfaces. If there are given *a priori* sources or other "painted on" external fields present, such as the velocity field of a fluid or an electromagnetic field, one can physically distinguish the various hypersurfaces and the hypothesis (ii) will not hold.

Theorem 7.1: (i) If a Lagrangian vector field $\overline{Z}$ for $\overline{L}$ [defined on $T(\mathcal{T}\times\mathfrak{M})$ above] exists, then it must be a second-order equation provided that it is second order in either ξ or g. (ii) In order that $\overline{Z}$ should exist as a second-order equation at (ξ,N,g,h) and that N be arbitrarily specifiable, that is, that the "degenerate direction" is all of $\mathcal{T}$, it is necessary that for any curve $(\xi(t),N(t),g(t),h(t))$ tangent to $\overline{Z}$, we have $(\partial/\partial t)\{\mathcal{H}(g,h/N)\mu_g\} = 0$, where $\mathcal{H}(g,h/N) = \frac{1}{2}[(\mathrm{Tr}(h/N))^2 - (h/N)\cdot(h/N)] + 2\mathrm{R}(g)$.

Proof: (i) In general the relation between Z and L is the Lagrangian condition

$$2\omega_L(v)[Z(v),w] = dE(v)\cdot w$$

on TB (see Sec. 2). If we let $Z = (Z_1, Z_2)$ locally on TB, this condition reads as follows: For all e_1, e_2 we have

$$D_1D_2L(u,e)\cdot e_1\cdot e - D_1L(u,e)\cdot e_1 + D_2D_2L(u,e)\cdot e_2\cdot e$$
$$= D_2D_2L(u,e)\cdot e_1\cdot Z_1 - D_1D_2L(u,e)\cdot Z_1\cdot e_1$$
$$+ D_2D_2L(u,e)\cdot Z_1\cdot e_2 - D_2D_2L(u,e)\cdot Z_2\cdot e_1.$$

These split up into two conditions:

$$D_2D_2L(u,e)\cdot Z_1\cdot e_2 = D_2D_2L(u,e)\cdot e_2\cdot e \qquad (7.1)$$

and

$$D_1D_2L(u,e)\cdot e_1\cdot e - D_1L(u,e)\cdot e_1$$
$$= D_1D_2L(u,e)\cdot e_1\cdot Z_1 - D_1D_2L(u,e)\cdot Z_1\cdot e_1$$
$$- D_2D_2L(u,e)\cdot Z_2\cdot e_1 \qquad (7.2)$$

In general, we cannot conclude from (7.1) that $Z_1(u,e) = e$ because L is degenerate.

Now let us turn to the case at hand. Let us incorporate μ_g into $\mathfrak{L}$ so we can briefly just write

$$\overline{L}(\xi, g, N, h) = \int N\mathfrak{L}(g, h/N).$$

We also suppress the fact that $\mathfrak{L}$ depends explicitly on Dg, D^2g, which is irrelevant for the present discussion.

Using obvious notation, the derivatives of $\overline{L}$ are easily worked out to be the following:

$$D_1L(\xi,g,N,h)\cdot(\tilde{\xi},\tilde{g})$$
$$= \text{derivative of } L \text{ with respect to } (\xi,g) \text{ in direction } (\tilde{\xi},\tilde{g})$$
$$= \int N\partial_g\mathfrak{L}(g,h/N)\cdot\tilde{g};$$

$$D_2L(\xi,g,N,h)\cdot(\tilde{N},\tilde{h})$$
$$= \text{derivative with respect to the velocity variables } (N,h) \text{ in direction } (\tilde{N},\tilde{h})$$
$$= \int\tilde{N}\mathfrak{L}\left(g,\frac{h}{N}\right) - \int\tilde{N}\partial_h\mathfrak{L}\left(g,\frac{h}{N}\right)\cdot\frac{h}{N} + \int\partial_h\mathfrak{L}\left(g,\frac{h}{N}\right)\cdot\tilde{h};$$

$$D_1D_2L(\xi,g,N,h)\cdot(\tilde{\xi},\tilde{g})\cdot(\tilde{N},\tilde{h})$$
$$= \int\tilde{N}\partial_g\mathfrak{L}\left(g,\frac{h}{N}\right)\cdot\tilde{g} - \int\tilde{N}\partial_g\partial_h\mathfrak{L}\left(g,\frac{h}{N}\right)\cdot\tilde{g}\cdot\frac{h}{N}$$
$$+ \int\partial_g\partial_h\mathfrak{L}\left(g,\frac{h}{N}\right)\cdot\tilde{g}\cdot\tilde{h};$$

$$D_2D_2L(\xi,g,N,h)\cdot(\tilde{N},\tilde{h})\cdot(\tilde{\tilde{N}},\tilde{\tilde{h}})$$
$$= \int\tilde{N}\tilde{\tilde{N}}\partial_h^2\mathfrak{L}\left(g,\frac{h}{N}\right)\cdot\frac{h}{N}\cdot\frac{h}{N^2} - \int\tilde{\tilde{N}}\partial_h^2\mathfrak{L}\left(g,\frac{h}{N}\right)\cdot\tilde{h}\cdot\frac{h}{N^2}$$
$$- \int\tilde{N}\partial_h^2\mathfrak{L}\left(g,\frac{h}{N}\right)\cdot\frac{\tilde{\tilde{h}}}{N}\cdot\frac{h}{N} + \int\partial_h^2\mathfrak{L}\left(g,\frac{h}{N}\right)\cdot\tilde{h}\cdot\frac{\tilde{\tilde{h}}}{N}.$$

Note that in the computation of the second derivative of L with respect to the velocity variables, two pairs of terms canceled out. Now let us use this expression to write out condition (7.1). Let us write $Z_1(\xi,g,N,h) = (\dot{\xi},\dot{g})$ for convenience. Condition (7.1) splits into two conditions, taking respectively $e_2 = (\tilde{N},0)$ and $e_2 = (0,\tilde{h})$. We get

$$0 = \int\tilde{N}\dot{\xi}\partial_h^2\mathfrak{L}\left(g,\frac{h}{N}\right)\cdot\frac{h}{N}\cdot\frac{h}{N^2} - \int\tilde{N}\partial_h^2\mathfrak{L}\left(g,\frac{h}{N}\right)\cdot\frac{\dot{g}}{N}\cdot\frac{h}{N} \qquad (7.1'a)$$

and

$$0 = -\int\dot{\xi}\partial_h^2\mathfrak{L}\left(g,\frac{h}{N}\right)\cdot\tilde{h}\cdot\frac{h}{N^2} + \int\partial_h^2\mathfrak{L}\left(g,\frac{h}{N}\right)\cdot\tilde{h}\cdot\frac{g}{N}. \qquad (7.1'b)$$

Each of these conditions is equivalent to the single condition $\dot{\xi}h = N\dot{g}$. Thus if $\dot{\xi} = N$, then $h = \dot{g}$ and vice versa. Hence (i) follows.

To establish (ii), we write out condition (7.2) which now becomes

$$D_1L(u,e)\cdot e_1 = D_1D_2L(u,e)\cdot e\cdot e_1 + D_2D_2L(u,e)\cdot Z_2\cdot e_1.$$

Again we have a split into two separate conditions taking, respectively, $e_1 = (\tilde{N},0)$ and $e_1 = (0,\tilde{h})$. By letting $Z_2 = (\dot{N},\dot{h})$ we get

$$0 = \int\tilde{N}\partial_g\mathfrak{L}\left(g,\frac{h}{N}\right)\cdot h - \int\tilde{N}\partial_g\partial_h\mathfrak{L}\left(g,\frac{h}{N}\right)\cdot h\cdot\frac{h}{N}$$
$$+ \int\tilde{N}\dot{N}\partial_h^2\mathfrak{L}\left(g,\frac{h}{N}\right)\cdot\frac{h}{N}\cdot\frac{h}{N^2} - \int\tilde{N}\partial_h^2\mathfrak{L}\left(g,\frac{h}{N}\right)\cdot\frac{\dot{h}}{N}\cdot\frac{h}{N}$$

and (7.2'a)

$$\int N\partial_g\mathfrak{L}\left(g,\frac{h}{N}\right)\tilde{h}$$
$$= \int\partial_g\partial_h\mathfrak{L}\left(g,\frac{h}{N}\right)\cdot h\cdot\tilde{h} - \int\dot{N}\partial_h^2\mathfrak{L}\left(g,\frac{h}{N}\right)\cdot\tilde{h}\cdot\frac{h}{N^2}$$
$$+ \int\partial_h^2\mathfrak{L}\left(g,\frac{h}{N}\right)\cdot\tilde{h}\cdot\left(\frac{\dot{h}}{N}\right). \qquad (7.2'b)$$

Condition (7.2b) is just the condition for Lagrange's equation for $h/N = k$ which we work out in Theorem 8.1 below.

For now we want to focus our attention on the nontrivial condition (7.2a). Since we are supposed to have complete degeneracy in $\mathcal{T},\tilde{N}$ is arbitrary, so (7.2a) is equivalent to

$$0 \equiv \partial_g\mathfrak{L}\left(g,\frac{h}{N}\right)\cdot h - \partial_g\partial_h\mathfrak{L}\left(g,\frac{h}{N}\right)\cdot h\cdot\frac{h}{N}$$
$$+ \dot{N}\partial_h^2\mathfrak{L}\left(g,\frac{h}{N}\right)\cdot\frac{h}{N}\cdot\frac{h}{N^2} - \partial_h^2\mathfrak{L}\left(g,\frac{h}{N}\right)\cdot\frac{\dot{h}}{N}\cdot\frac{h}{N}$$

Setting $k = h/N$, this becomes

$$0 \equiv \partial_g\mathfrak{L}\left(g,\frac{h}{N}\right)\cdot h - \partial_g\partial_h\mathfrak{L}\left(g,\frac{h}{N}\right)\cdot h\cdot\frac{h}{N} - \partial_h^2\mathfrak{L}\left(g,\frac{h}{N}\right)\cdot\dot{k}\cdot\frac{h}{N} \qquad (7.2'c)$$

Let us take a curve $(\xi(t),g(t),N(t),h(t))$ tangent to $\overline{Z}$ which we suppose exists. Then (7.2c) just says that

$$0 \equiv \frac{\partial}{\partial t}[\partial_h\mathfrak{L}(g,k)\cdot k - \mathfrak{L}(g,k)]$$

or

$$\frac{\partial}{\partial t}[\mathfrak{K}(g,k)\mu_g] \equiv 0.$$

This proves the theorem. ■

We shall continue this investigation by showing how to construct $\overline{Z}$ in the next section.

8. CONSTRUCTION OF THE EQUATIONS FOR A GENERAL LAPSE

In view of the results of Sec. 7 and the discussion of Sec. 1, we introduce the following "constraint" subset of $T(\mathcal{T}\times\mathfrak{M})$:

$$\mathcal{C} \subset T(\mathcal{T}\times\mathfrak{M}),$$
$$\mathcal{C} = \left\{(\xi,N,g,h):\ N>0 \text{ and } \delta\left[\left(\mathrm{Tr}\frac{h}{N}\right)g - \frac{h}{N}\right] = 0\right.$$
$$\left.\text{and } \mathfrak{K}\left(g,\frac{h}{N}\right) = 0\right\}.$$

Since ultimately N will be specified in advance, it is useful to think of $\mathcal{C}$ as a subset of $T\mathfrak{M}$. Unfortunately, at points g of $\mathcal{C}$ which admit a nontrivial isometry group, $\mathcal{C}$ does not seem to be a manifold. This is

analogous to the nonmanifold structure of superspace[5]; see also Sec. 11. However, we shall not require smoothness of $\mathcal{C}$ in what follows:

Theorem 8.1: Let $\bar{L}$ on $T(\mathcal{T}\times\mathfrak{M})$ be defined as in Sec. 7. Then at points of $\mathcal{C}$, $\bar{L}$ has a Lagrangian vector field $\bar{Z}$. The most general such second order vector field is as follows: Specify an arbitrary curve $\xi(t)\in\mathcal{T}$ and set $N_t = d\xi(t)/dt$, assuming $N_t > 0$. By writing

$$\bar{Z}(\xi,N,g,h) = (N,\dot{N},\dot{g},\dot{h}), \qquad k = \frac{h}{N},$$

then

$$\begin{cases} \dot{g} = \dfrac{\partial g_t}{\partial t} = N_t k_t, \\ \dot{k} = \dfrac{\partial k_t}{\partial t} = N_t S_{g_t}(k_t) - 2N_t\ \mathrm{Ric}(g_t) + \tfrac{1}{4}N_t\mathcal{H}(g_t,k_t) \\ \qquad\qquad + 2\ \mathrm{Hess}N_t. \end{cases}$$

Moreover, an integral curve of these equations which begins in $\mathcal{C}$ remains in $\mathcal{C}$.

Proof: With $\bar{Z}$ defined as in the statement of the theorem, we must verify the Lagrangian condition for $\bar{L}$. Referring to the proof of condition 7.1, this amounts to showing that (7.2b) holds *and* that $(\partial/\partial t)[\mathcal{H}(g,k)\mu_g] = 0$ for any integral curve of $\bar{Z}$ starting in $\mathcal{C}$.

From (7.2b) it is clear that $\dot{k}$ must be the spray of the DeWitt metric minus the gradient of the potential $V(g) = 2\int NR(g)\mu_g$. The equations then follow from Proposition 3.2.

To complete the verification of the Lagrange condition we must show that $(\partial/\partial t)[\mathcal{H}(g,k)\mu_g] = 0$. To do this we first observe that the condition $\delta[(\mathrm{Tr}k)g - k] = 0$ is maintained in time; this can be proved exactly as in Sec. 6 by invariance of the Lagrangian under the action of $\mathfrak{D}$. $\mathfrak{D}$ must now be considered to act on $\mathcal{T}$ in the natural way, $\mathfrak{D}\times\mathcal{T}\to\mathcal{T}; \eta\times\xi\mapsto\xi\circ\eta$. We then prove, under the hypothesis $\delta[(\mathrm{Tr}k)g - k] = 0$, that $(\partial/\partial t)(\mathcal{H}\mu_g) = 0$ by a direct verification, analogous to the proof of Proposition 3.4, using the stated equations defined for $\bar{Z}$. ■

Note: In general, $\bar{Z}$ will not exist at points other than those in $\mathcal{C}$.

The Hessian term is a nonlinear coupling between N and g. However, we again assert that the solution for a general N may be obtained from a solution for $N = 1$ by integrating a system of ordinary differential equations. This is explained in Sec. 10.

There is another interesting way to see that one has $(\partial/\partial t)(\mathcal{H}\mu_g) = 0$ for any theory invariant under the full relativistic time translation group $\mathcal{T}$. This is an alternative approach to that used in 7.1 although it is not detailed enough to allow for the construction of the equations of motion. It does, however, provide a group theoretical argument for the relationship between $(\partial/\partial t)(\mathcal{H}\mu_g) = 0$ and time translation invariance (in the relativistic sense).

Proposition 8.1: Let $\mathfrak{L}$ be any Lagrangian density on $\mathfrak{M}$ (or any function-space for that matter) with extension to $\bar{\mathfrak{L}}$ on $T(\mathcal{T}\times\mathfrak{M})$ as defined in Sec. 7. Suppose

$$\bar{L} = \int N\mathfrak{L}\left(g,\frac{h}{N}\right)\mu_g$$

has a Lagrangian vector field $\bar{Z}$ on some subspace $\mathcal{C}\subset T(\mathcal{T}\times\mathfrak{M})$. Let $\mathcal{C}$ be invariant under relativistic time translations (see below), and let integral curves of $\bar{Z}$ map $\mathcal{C}$ to $\mathcal{C}$. Then along such integral curves, $(\partial/\partial t)(\mathcal{H}\mu_g) \equiv 0$.

Proof: $\mathcal{T}$ is a vector space and as an additive group, acts on $\mathcal{T}\times\mathfrak{M}$ and $\mathcal{C}$ in a natural way. For $\xi_0\in\mathcal{T}$ we get a map of $\mathcal{T}\times\mathfrak{M}\to\mathcal{T}\times\mathfrak{M}$ by $(\xi,g)\mapsto(\xi+\xi_0,g)$. There is a corresponding one parameter group $\Phi_t(\xi,g) = (\xi + t\xi_0, g)$. This is generated by the vector field $(\xi,g)\mapsto(\xi_0,0)$.

Now the tangent action of Φ_t leaves $\bar{L}$ invariant, since $T\Phi_t(\xi,N,g,h) = (\xi+t\xi_0,N,g,h)$ and $\bar{L}$ depends only on N, not on ξ. Thus we may apply Proposition 2.3. By a straightforward computation, we find that the fiber derivative is given as follows:

$$F\bar{L}(\xi,N,g,h)\colon T_{(\xi,g)}(\mathcal{T}\times\mathfrak{M})\to\mathbf{R},$$

$$(\bar{N},\bar{h})\mapsto -\int_M \bar{N}\mathcal{H}\left(g,\frac{h}{N}\right)\mu_g + \int_M\left[\partial_h\mathfrak{L}\left(g,\frac{h}{N}\right)\cdot\bar{h}\right]\mu_g.$$

Thus with $(\bar{N},\bar{h}) = (\xi_0,0)$ we conclude from proposition 2.3 that

$$\int_M \xi_0\mathcal{H}\mu_g$$

is a constant of the motion. Since ξ_0 is arbitrary, the result follows. ■

Observe that $\mathcal{H}$ is not the energy density for $\bar{L}$ but rather is that for L. Since $\bar{L}$ is homogeneous, its associated energy function is identically zero; since $\omega_{\bar{L}}$ is degenerate, this does not imply trivial equations of motion. Finally note that the requirement $(\delta\pi)^b = 0$ is buried in Proposition 8.1 through the assumptions that $\bar{Z}$ exists and integral curves stay in $\mathcal{C}$. Thus Proposition 8.1 is just illustrative, with the main results in Theorems 7.1 and 8.1.

We prefer the proofs we have given for the maintenance of the supplementary conditions since they are natural consequences of the Hamiltonian structure of the evolution equations and their dynamical symmetries. Moreover, in this approach we need not rely on identities in the corresponding four geometry.

9. RELATIONSHIP WITH THE FOUR GEOMETRY

In this section we establish the equivalence between the Einstein system (E), with a given lapse N_t and shift X_t, for the evolving three geometry g_{ij} and the Ricci flatness of the Lorentz metric g^L constructed on $I\times M$ $[I = (-\epsilon,\epsilon)]$; the metric g^L is obtained by decreeing that $(1/N, X/N)$ be a unit timelike vector field on $I\times M$ orthogonal to the $\{t\}\times M$ hypersurfaces. To satisfy this condition, we construct g^L from g_t, X_t, and N_t as follows:

$$g^L(t,m)\cdot((r,v_m),(s,w_m)) = g_t(m)\cdot(v_m - rX_t(m), w_m - sX_t(m)) - rsN_t^2(M).$$

In coordinates, this formula reads

$${}^4g_{\alpha\beta}dx^\alpha dx^\beta = (X^iX_i - N^2)(dt)^2 - 2X_i dx^i dt + g_{ij}dx^i dx^j,$$

where $x^\alpha = (t, x^i)$ and $X_i = g_{ij}X^j$. We are assuming that X_t has length less than N_t which means that our observer has velocity less than that of light, relative to a Gaussian reference system.

Theorem 9.1: Let X_t and N_t be a given lapse and shift. Then a curve $g_t \in \mathfrak{M}$ satisfies the system (E) with lapse and shift N_t and X_t if and only if the Lorentz metric g^L constructed above is Ricci flat; i.e., $R_{\alpha\beta} = 0$.

Obviously this theorem is basic to the whole program and is in Ref. 1 for $N = 1$ and $X = 0$. Here we are interested in the situation for arbitrary N. One interesting feature is to see how the Hessian term of N in the equations of evolution arises. We have seen in Secs. 3 and 8 how it arose in the Lagrangian formulation.

The proof of Theorem 9.1 is based on a decomposition of the Riemann–Christoffel curvature tensor in terms of quantities associated with an embedded hypersurface. Four of the equations are the Gauss–Codazzi equations which relate the curvature tensor of $I \times M$ to the curvature tensor and second fundamental form S of the embedded hypersurface M. The other six equations involve more than the geometry of M and S; they depend also on a family of embeddings. A convenient reference for this result is Yano[21] Chap. 5; see also, Abraham,[22] Sec. 9. For the purposes of this paper we shall translate the formulas into coordinate notation. In doing this we choose a coordinate system in which the t-axis is normal to the hypersurface M; in other words, we assume that the unit timelike normal is of the form $Z = (Z^0, 0)$ so that $g_{0k} = 0$.

Thus $g_{\alpha\beta}$ is of the form $-N^2dt^2 + g_{ij}dx^idx^j$ and $Z = (1/N, 0)$. The case of an arbitrary shift X_t may be dealt with by the methods explained in Secs. 4 and 5.

Lemma 9.1: Let M_t be a family of three manifolds embedded as spacelike hypersurfaces in a Lorentz manifold V. Let ${}^4R_{\alpha\beta\gamma\delta}$ be the curvature tensor on V and ${}^3R_{ijkl}$ that on M. Let S_{ij} be the second fundamental form ("extrinsic curvature") of M and Z the unit normal to M. Then in a coordinate system in which $Z = (1/N, 0)$, we have the following decomposition of ${}^4R_{\alpha\beta\gamma\delta}$:

(i) $${}^4R_{0i0j} = N^2\left\{\frac{1}{N}\frac{\partial S_{ij}}{2t} - (S \times S)_{ij} - \frac{1}{N}N_{|i|j}\right\}$$

(ii) $${}^4R_{ijkl} = {}^3R_{ijkl} + S_{il}S_{jk} - S_{ik}S_{jl}$$, and

(iii) $${}^4R_{ijk0} = S_{ki|j} - S_{kj|i}$$,

where all covariant derivatives are taken in the metric on M.

Proof: The decomposition (ii) is the Gauss equation (Yano,[21] p. 94)

$${}^4R(X, Y, U, W) = {}^3R(X, Y, U, W) + [S(X, W)S(Y, U) - S(Y, W)S(X, U)]$$

written in coordinates. There is a change of sign over what is in Yano because $g(Z, Z) = -1$ rather than $+1$. The decomposition (ii) holds generally for any hypersurface and is a direct consequence of the relationship between the induced connection ${}^3\nabla$ and the second fundamental form S,

$${}^4\nabla_X Y = {}^3\nabla_X Y + S(X, Y)Z \tag{9.1}$$

and the definition of the Riemann–Christoffel curvature tensor. Similarly (iii) is the Codazzi equation (Yano,[21] p. 95)

$$R(X, Y, U, Z) = \nabla_Y S(X, U) - \nabla_X S(Y, U).$$

The decomposition (i) involves the geometry of M, S, and the family of embeddings (otherwise $\partial S_{ij}/\partial t$ has no meaning). The decomposition (i) may be deduced (by a long computation) from Abraham[22] Sec. 9, but we can also give a direct proof as follows. Now we can write

$$g_{\alpha\beta}dx^\alpha dx^\beta = -N^2dt^2 + g_{ij}dx^idx^j$$

and compute directly from $\Gamma^\alpha_{\mu\nu} = \frac{1}{2}g^{\lambda\alpha}(g_{\mu\lambda,\nu} + g_{\nu\lambda,\mu} - g_{\mu\nu,\lambda})$ that

$$\Gamma^0_{00} = N_{,0}/N, \qquad \Gamma^0_{i0} = N_{,i}/N,$$
$$\Gamma^0_{ij} = g_{ij,0}/2N^2, \qquad \Gamma^k_{0j} = \tfrac{1}{2}g^{lk}g_{lj,0},$$
$$\Gamma^i_{00} = Ng^{ji}N_{,j}, \qquad {}^4\Gamma^i_{jk} = {}^3\Gamma^i_{jk}.$$

Contracting (9.1) with Z gives

$$S(X, Y) = -({}^4\nabla_X Y, Z) = ({}^4\nabla_X Z, Y).$$

Thus in coordinates,

$$S_{ij} = Z_{i|j} = Z_{i,j} - \Gamma^\alpha_{ij}Z_\alpha = \Gamma^0_{ij}N = g_{ij,0}/2N,$$

since $Z_\alpha = (-N, 0)$.

Now by definition

$$R(Z, X, Z, Y) = \langle\nabla_Z\nabla_X Z, Y\rangle - \langle\nabla_X\nabla_Z Z, Y\rangle - \langle\nabla_{[X,Z]}Z, Y\rangle,$$

so that

$$\frac{1}{N^2}R_{0i0j} = R(Z, X, Z, Y),$$

where $Z = ((1/N)(\partial/\partial t), 0)$, $X = (0, \partial/\partial x^i)$, $Y = (0, \partial/\partial x^j)$. Now one easily computes the following:

$$(\nabla_X Z)^0 = S^j_i,$$
$$(\nabla_X Z)^0 = -\frac{1}{N^2}N_{,i} + \Gamma^0_{i0}\frac{1}{N} = 0,$$
$$(\nabla_Z W)^k = W^k_{,0}/N + S^k_j W^j + g^{jk}N_{,j}W^0.$$

So we get

$$\langle\nabla_Z\nabla_X Z, Y\rangle = g_{jl}(\nabla_Z\nabla_X Z)^l$$
$$= g_{jl}S^l_{i,0}/N + S^l_k S^k_i g_{jl}$$
$$= \frac{1}{N}\frac{\partial}{\partial t}(S_{ij}) - (S \times S)_{ij}. \tag{9.2}$$

Similarly,

$$\langle\nabla_X\nabla_Z Z, Y\rangle = (N_{,j}/N)_{|i} = N_{|j|i}/N - N_{,j}N_{,i}/N^2 \tag{9.3}$$

and finally since $[X, Z]^k = 0$, $[X, Z]^0 = N_{,i}/N^2$, we get

$$\langle\nabla_{[X,Z]}Z, Y\rangle = N_{,i}N_{,j}/N^2. \tag{9.4}$$

Thus adding up (9.2), (9.3) and (9.4) yields (i). ■

In terms of $k_{ij} = (1/N)(\partial g_{ij}/\partial t) = 2S_{ij}$, we have

$$\begin{cases} {}^4R_{0i0j} = N^2\left(\frac{1}{2N}\frac{\partial k_{ij}}{\partial t} - \frac{1}{4}(k\times k)_{ij} - \frac{N_{|i|j}}{N}\right), \\ {}^4R_{ijkl} = {}^3R_{ijkl} + \frac{1}{4}(k_{il}k_{jk} - k_{ik}k_{jl}), \\ {}^4R_{ijk0} = \frac{1}{2}(k_{ki|j} - k_{kj|i}). \end{cases}$$

Proof of Theorem 9.1: We prove Theorem 9.1 for the case of $X_t = 0$. The general case is handled using the methods of Secs. 4 and 5. So suppose that $g_{\alpha\beta}dx^\alpha dx^\beta = -N^2dt^2 + g_{ij}dx^idx^j$ is Ricci flat. Using Yano's conventions, the Ricci tensor is

$${}^4R_{\alpha\beta} = {}^4g^{\gamma\delta}\,{}^4R_{\gamma\alpha\beta\delta} = -{}^4g^{\gamma\delta}R_{\gamma\alpha\delta\beta}.$$

We have

$$0 = {}^4R_{ij} = -{}^4g^{\alpha\beta}\,{}^4R_{\alpha i\beta j} = \frac{1}{N^2}\,{}^4R_{0i0j} - g^{kl}\,{}^4R_{kilj}$$

$$= \left(\frac{1}{2N}\frac{\partial k}{\partial t} - \frac{1}{4}k\times k - \frac{1}{N}(\mathrm{Hess}N)\right) + \{\mathrm{Ric}(g) - \tfrac{1}{4}[k\times k - (\mathrm{Tr}k)k]\},$$

where we have used decomposition (ii) as well as (i). This gives the required equation for $\partial k/\partial t$.

Similarly using decomposition (iii) we have $0 = {}^4R_{0i} = -{}^4g^{\alpha\beta}\,{}^4R_{\alpha i\beta 0} = -{}^3g^{kl}\,{}^4R_{kil0} = \frac{1}{2}(k^j_{i|j} - k^j_{j|i}) = -\frac{1}{2}\delta(k - (\mathrm{Tr}k)g)$, which gives the divergence condition.

Finally using decomposition (i) again, we have

$$0 = {}^4R_{00} = -{}^4g^{\alpha\beta}\,{}^4R_{\alpha0\beta0} = -{}^3g^{kl}\,{}^4R_{0k0l} = -{}^3g^{kl}N^2\left(\frac{1}{2N}\frac{\partial k_{kl}}{\partial t} - \frac{1}{4}(k\times k)_{kl} - \frac{(\mathrm{Hess}N)_{kl}}{N}\right).$$

If we now substitute the equation for $\partial k/\partial t$ in this expression, it simplifies down to $\frac{1}{2}N^2\,\mathcal{H}(g,k) = \frac{1}{2}N^2[\mathcal{H} + 2R(g)]$ so we get the energy condition.

The converse of the theorem is proved by retracing the steps. ■

10. THE INTRINSIC SHIFT VECTOR FIELD

In this section we study the relationship between solutions of the Einstein system (E) with the same Cauchy data but with different prescribed lapse functions. We suppose that we have a solution (g_t, k_t), $|t| < \epsilon$ of (E) for a given lapse N_t and shift vector field $X_t = 0$, and we wish to find the solution $(\bar g_t, \bar k_t)$ to (E) with lapse $\bar N = 1$ and $\bar X = 0$ such that $(\bar g_0, \bar k_0) = (g_0, k_0)$. The converse problem of finding (g_t, k_t) for an arbitrary lapse N given the solution $(\bar g_t, \bar k_t)$ for $\bar N = 1$ proceeds similarly (see Theorem 10.3 below).

The above problem is well known to be equivalent to finding the Gaussian normal coordinates for the Lorentz metric $g_{\alpha\beta}dx^\alpha dx^\beta = -N^2dt^2 + g_{ij}dx^idx^j$. What we wish to do is geometrize this situation a bit.

Lemma 10.1: Let N_t be an arbitrary lapse function, let $g_t \in \mathfrak{M}$, $|t| < \epsilon$, be a one-parameter curve of metrics, and let $g^L(t,m)\cdot[(r,v_m)(s,w_m)] = -N_t^2(m)rs + g_t(m)\cdot(v_m,w_m)$ be the associated Lorentz metric on $I\times M$, $I = (-\epsilon,\epsilon)$. Then there exists a unique curve $\tau_t \in \mathcal{T}$, $|t| < \epsilon' \le \epsilon$ with $\tau_0 = 0$ such that

$$\frac{d\tau}{dt} = N_t\sqrt{1 + \|\mathrm{grad}\tau_t\|}, \tag{10.1}$$

where $(\mathrm{grad}\tau_t)^j = g^{ij}(d\tau/dx^i)$ is computed with respect to the time-dependent metric g_t. The function $\tau\colon I\times M \to \mathbf{R}$, $\tau(t,m) \mapsto \tau_t(m)$ is the proper time from (t,m) to $\{0\}\times M$ measured backwards along a unit timelike geodesic normal to $\{0\}\times M$.

Proof: Equation (10.1) is just the eikonal equation

$$-\frac{1}{N_t^2}\left(\frac{\partial\tau}{\partial t}\right)^2 + \|\mathrm{grad}\tau\|^2 = -1 = g^{\alpha\beta}\frac{\partial\tau}{\partial x^\alpha}\frac{\partial\tau}{\partial x_\beta}, \tag{10.2}$$

which is a single first order nonlinear partial differential equation. By the Cauchy method of characteristics, this single equation can be reduced to a system of eight ordinary differential equations in Hamiltonian form, namely the geodesic equations of $g_{\mu\nu}$:

$$\begin{cases} \dfrac{dx^u}{dt} = g^{\mu\nu}p_\nu = \dfrac{\partial H}{\partial p_\mu}, \\ \dfrac{dp_\mu}{d\tau} = -\frac{1}{2}\dfrac{\partial g^{\alpha\beta}}{\partial x^\mu}p_\alpha p_\beta = -\dfrac{\partial H}{\partial x^\mu}, \end{cases} \tag{10.3}$$

where the Hamiltonian $H = \frac{1}{2}g^{\mu\nu}p_\mu p_\nu$.

From the initial condition $\tau(0,m) = 0$ and Eq. (10.2) itself, we conclude that $(d\tau/dt)(0,m) = N_0(m)$. System (10.3) can be integrated for short time $0 < \epsilon' \le \epsilon$ subject to the initial conditions $x^\mu(0,m) = (0,m)$ and $p_\mu(0,m) = (N_0(m), 0)$ to give $(x^\mu(\tau,m), p_\mu(\tau,m))$. Since the hypersurface $\{0\}\times M$ is noncharacteristic, $x^\mu(\tau,m)$ can be inverted for $|t| < \epsilon'' \le \epsilon' \le \epsilon$ to give a function $\tau(x^\mu)$ which satisfies $\tau(0,m) = 0$, Eq. (10.2), and $\partial\tau/\partial x^\mu = p_\mu$. That the geodesics are unit timelike geodesics follows from conservation of $2H = g^{\alpha\beta}p_\alpha p_\beta = -1$ and since $dx^\mu/dt(0,m) = g^{\mu\nu}p_\nu(0,m) = g^{\mu\nu}(N_0(m),0) = (-(1/N_0(m)),0)$, they are normal to $\{0\}\times M$. ■

The factor $\sqrt{1 + \|\mathrm{grad}\tau\|^2}$ in the expression for $d\tau/dt$ takes into account the fact that, in general, the lapse depends on space coordinates and therefore "pushes" up the hypersurface $\{0\}\times M$ unevenly along $\tau =$ constant hypersurfaces in $I\times M$.

There is another way of looking at the lapse function which has been given by Wheeler[1]: Namely it is trivial to check that

$$\xi_t(m) = \int_0^t N_\lambda(m)d\lambda$$

is the proper time from $\{0\}\times M$ to $\{t\}\times M$ measured along the curve $\lambda \mapsto (\lambda, m)$ for m fixed. This differs from $\tau_t(m)$ in that $\tau_t(m)$ is the proper time for an observer following a geodesic; that is, one in free fall. The curve $\lambda \mapsto (\lambda, m)$ is not a geodesic because of the spatial dependence of N.

For the rest of this section, we will refer to τ_t as computed from Lemma 10.1 as the *proper time function associated with N_t and g_t*. Note that τ_t is just the time part $\bar t(t,x')$ of the coordinate transformation $\bar x^\alpha(x^\mu)$ which transforms $g_{\alpha\beta}dx^\alpha dx^\beta = -N^2dt^2 + g_{ij}dx^idx^j$ to Gaussian coordinates, as can be seen from the eikonal equation for $g_{\alpha\beta}$,

$$-1 = -\frac{1}{N^2}\frac{\partial\bar t}{\partial t}^2 + g^{kl}\frac{\partial\bar t}{\partial x^k}\frac{\partial\bar t}{\partial x^l}.$$

The question naturally arises if we can construct the rest of the Gaussian coordinate system from τ_t alone. Let $\phi_t(m) = \bar x^i(t, x^j)$ denote the spatial part of the

transformation to Gaussian coordinates, so that $(\tau_t(m), \phi_t(m)) = \bar{x}^\mu(x^\alpha)$ is the coordinate transformation leading to Gaussian coordinates. Since ϕ_t is a diffeomorphism, it is the flow of a time-dependent vector field on M, so that it behaves just like a shift η_t generated by a shift vector field X_t. We call ϕ_t the *intrinsic shift* and its generator Y_t as the *intrinsic shift vector field of* N. Theorem 10.1 below gives a way to compute ϕ_t or Y_t from the proper time function τ_t alone.

In the transformation to Gaussian coordinates, the hypersurface $\{t\} \times M$ is mapped into the hypersurface $\tau^{-1}(t)$, so $\phi_t(m)$ is the spatial coordinate of (t, m) in the Gaussian coordinate system. The intrinsic shift thus describes the shifting of the spatial coordinates in the $\{t\} \times M$ hypersurface due to the fact that the lapse N_t depends on the space variable so that each point of $\{t\} \times M$ does not have the same proper time coordinate $\tau_t(m)$. Thus the hypersurface $\{t\} \times M$ is tilted when it is stretched to fit the $\tau =$ constant contours. This tilting causes a shifting of the spatial coordinates in $\{t\} \times M$ which is described by the intrinsic shift ϕ_t.

Theorem 10.1: Let N_t and g_t be given and let τ_t be determined from Lemma 10.1. Let Z_t be the time-dependent vector field defined by

$$Z_t = -\frac{N_t}{\sqrt{1 + \|\operatorname{grad}\tau_t\|^2}} \operatorname{grad}\tau_t$$

and let $\psi_t, \psi_0 = id_M$ be its flow. Then the intrinsic shift ϕ_t is given by $\phi_t = \psi_t^{-1}$ and the intrinsic shift vector field is

$$Y_t = -(\psi_t^{-1})_* Z_t = -(\phi_t)_* Z_t$$

Proof: Let ϕ_t be the space part of the transformation to the Gaussian coordinate system. Then the condition on ϕ_t comes from requiring that the g^{0i} components of g^L remain zero in the Gaussian coordinate system. This condition is

$$0 = -\frac{1}{N^2}\frac{\partial\tau}{\partial t}\frac{\partial\bar{x}^i}{\partial t} + g^{kl}\frac{\partial\tau}{\partial x^k}\frac{\partial\bar{x}^i}{\partial x^l}.$$

This condition is rewritten as

$$\frac{\partial\tau_t}{\partial t}\frac{\partial\phi_t^i}{\partial t} = N_t^2 g_t^{kl}\frac{\partial\phi_t^i}{\partial x^l}\frac{\partial\tau_t}{\partial x^k} = N_t^2 T\phi_t \cdot \operatorname{grad}\tau_t = \left(\frac{d\tau_t}{dt}\right)^2 \frac{1}{1 + \|\operatorname{grad}\tau_t\|^2} T\phi_t \cdot \operatorname{grad}\tau_t,$$

so that

$$\frac{d\phi_t}{dt} = \left(\frac{d\tau_t}{dt}\right)\left(\frac{1}{1 + \|\operatorname{grad}\tau_t\|^2}\right) T\phi_t \cdot \operatorname{grad}\tau_t = \frac{N_t}{\sqrt{1 + \|\operatorname{grad}\tau_t\|^2}} T\phi_t \cdot \operatorname{grad}\tau_t = -T\phi_t \circ Z_t.$$

Thus ϕ_t satisfies

$$\frac{d\phi_t}{dt} \circ \phi_t^{-1} = -T\phi_t \circ Z_t \circ \phi_t^{-1} = -(\phi_t)_* Z_t$$

so that $-(\phi_t)_* Z_t$ is the generator of ϕ_t. Let $\psi_t, \psi_0 = id_M$ be the flow of Z_t. Then from Lemma 4.2, $-(\psi_t^{-1})_* Z_t$ has the flow ψ_t^{-1} so that $\phi_t = \psi_t^{-1}$. Thus $-(\phi_t)_* Z_t = -(\psi_t^{-1})_* Z_t = Y_t$ is the intrinsic shift vector field associated with N. ■

We now consider how solutions (g_t, k_t), $|t| < \epsilon$ of the Einstein system (E) with an arbitrary lapse N_t are related to solutions $(\bar{g}_t, \bar{h}_t)$, $|t| < \epsilon' \le \epsilon$ of (E) with the same Cauchy data and $\bar{N} = 1$. As before, this is equivalent to finding how the space part of $g_{\alpha\beta}dx^\alpha dx^\beta = -N^2dt^2 + g_{ij}dx^i dx^j$ transforms when we transform to Gaussian coordinates. This is also equivalent to finding the metric $\bar{g}_\tau$ induced on the $\tau =$ constant hypersurfaces in the space $I \times M$ with Lorentz metric $g_{\alpha\beta}$.

The fact that g_t is a solution of (E) is again peripheral as we are just computing how the space part of a Lorentz metric of the form $g_{\alpha\beta}dx^\alpha dx^\beta = -N^2dt^2 + g_{ij}dx^i dx^j$ transforms when we transform to Gaussian coordinates.

Theorem 10.2: Let N_t and g_t be given, $|t| < \epsilon$ and let τ_t be determined from Lemma 10.1. Let $\psi_t, \psi_0 = id_M$ be the flow of $Z_t = -N_t(1 + \|\operatorname{grad}\tau_t\|^2)^{-1/2} \operatorname{grad}\tau$; let $\phi_t = \psi_t^{-1}$ be the intrinsic shift; and let $Y_t = -(\phi_t)_* Z_t$ be the instrinsic shift vector field associated with N_t.

Let $g_t^{-1} = g^{ij}$ be g_t in contravariant form and let

$$\bar{g}^{-1}(\tau_t(m), \phi_t(m)) = T\phi_t \otimes T\phi_t \left(g_t^{-1} - \frac{Z_t}{N_t} \otimes \frac{Z_t}{N_t}\right)(m) = T\phi_t \otimes T\phi_t \left(g_t^{-1} - \frac{\operatorname{grad}\tau_t}{\sqrt{1 + \|\operatorname{grad}\tau_t\|^2}} \otimes \frac{\operatorname{grad}\tau_t}{\sqrt{1 + \|\operatorname{grad}\tau_t\|^2}}\right)(m)$$
$$= T\phi_t \otimes T\phi_t(g_t^{-1})(m) - \frac{Y_t}{N_t} \otimes \frac{Y_t}{N_t}(\phi_t(m)).$$

Then $\bar{g}_\tau$ is the metric induced on $\tau =$ constant hypersurfaces by the Lorentz metric $-N^2dt^2 + g_{ij}dx^i dx^j$. As above, denoting the Gaussian coordinates by $(\tau(t, x^i), \phi^i(t, x))$ we have the coordinate expression

$$\bar{g}^{ij} = \frac{\partial\phi^i}{\partial x^k}\frac{\partial\phi^j}{\partial x^l} \times \left(g^{kl} - \left(1 + g^{ab}\frac{\partial\tau}{\partial x^a}\frac{\partial\tau}{\partial x^b}\right)^{-1} g^{km}g^{ln}\frac{\partial\tau}{\partial x^m}\frac{\partial\tau}{\partial x^n}\right)(t, x).$$

Remark: Note that in order that $\bar{g}^{-1}$ remain positive definite,

$$g_t(X, X) > \left(\frac{d\tau \cdot X}{\sqrt{1 + \|\operatorname{grad}\tau_t\|^2}}\right)^2.$$

This holds at $t = 0$ and so will hold for some t-interval around 0.

Proof: Let $-N^2dt^2 + g_{ij}dx^i dx^j$ be the Lorentz metric associated with g_t and N_t. Transforming this metric to Gaussian coordinates gives the transformation law for the 3-metric g^{ij} as

$$\bar{g}^{ij}(\tau(t, x^k), \bar{x}^i(t, x^k)) = \frac{\partial\bar{x}^i}{\partial x^m}(t, x^k)\frac{\partial\bar{x}^j}{\partial x^n}(t, x^k)g^{mn}(t, x^k) - \frac{1}{N^2}\frac{\partial\bar{x}^i}{\partial t}(t, x^k)\frac{\partial\bar{x}^j}{\partial t}(t, x^k).$$

By using

$$\frac{1}{\bar{N}}\frac{\partial \bar{x}^i}{\partial t} = \frac{N}{\partial \tau/dt}\frac{\partial \bar{x}^i}{\partial x^l} g^{kl}\frac{\partial \tau}{\partial x^k}$$

from the proof of Theorem 10.1,

we have

$$\bar{g}^{ij}(\tau(t, x^k), \bar{x}^i(t, x^k))$$

$$= \frac{\partial \bar{x}^i}{\partial x^m}\frac{\partial \bar{x}^j}{\partial x^m}\left(g^{mn} - \frac{1}{1 + \|\operatorname{grad}\tau\|^2} g^{mk}\frac{\partial \tau}{\partial x^k} g^{nl}\frac{\partial \tau}{\partial x^l}\right)(t, x^k)$$

which is written as

$$\bar{g}^{-1}(\tau_t(m), \phi_t(m)) = T\phi_t \otimes T\phi_t \times \left(g_t^{-1}(m) - \frac{\operatorname{grad}\tau_t(m)}{\sqrt{1 + \|\operatorname{grad}\tau_t\|^2}} \otimes \frac{\operatorname{grad}\tau_t(m)}{\sqrt{1 + \|\operatorname{grad}\tau_t(m)\|^2}}\right).$$

As a corollary we get an expression for the solution to the Einstein system (E) for $\bar{N} = 1$, given a solution (g_t, k_t) for arbitrary lapse N_t.

Corollary 10.1: Let (g_t, k_t) be a solution to the Einstein system (E) with N_t given and $X_t = 0$. Let τ_t, $\tau_0 = 0$ be determined from Lemma 10.1, let $Z_t = -N_t/(1 + \|\operatorname{grad}\tau_t\|^2)^{-1/2} \operatorname{grad}\tau_t$ and let ϕ_t be the intrinsic shift. Let g_t^{-1} be g_t in contravariant indices; let

$$Y_t = -(\phi_t)_* Z_t,$$

$$\bar{g}^{-1}(\tau_t(m), \phi_t(m)) = T\phi_t \otimes T\phi_t\left(g_t^{-1} - \frac{Y_t}{\bar{N}_t} \otimes \frac{Y_t}{\bar{N}_t}\right)(m),$$

and let $\bar{k}_\tau = \partial \bar{g}_\tau/\partial\tau$. Then $(\bar{g}_\tau, \bar{k}_\tau)$ is a solution to (E) with $\bar{N}_t = 1$ and

$$(\bar{g}_0, \bar{k}_0) = (\bar{g}_0, k_0).$$

Proof: From the theorem, $\bar{g}_\tau$ is the metric induced on $\tau =$ constant hypersurfaces and so is the solution of (E) in Gaussian coordinates; that is, with $\bar{N}_t = 1$.

Since $\tau_0 = 0$, $\bar{g}_0 = g_0$. From the chain rule,

$$\frac{\partial \bar{g}^{-1}}{\partial \tau}\frac{\partial \tau}{\partial t}\bigg|_{t=0} = \frac{\partial \bar{g}^{-1}}{\partial t}\bigg|_{t=0}.$$

Since $Y_0 = 0$ and $\phi_0 = id_M$, one sees from the above expression for $\bar{g}^{-1}$ that

$$\frac{\partial \bar{g}^{-1}}{\partial t}\bigg|_{t=0} = \frac{\partial g^{-1}}{\partial t}\bigg|_{t=0}.$$

Thus

$$\frac{\partial \bar{g}^{-1}}{\partial \tau}\frac{\partial \tau}{\partial t}\bigg|_{t=0} = \frac{\partial g^{-1}}{\partial t}\bigg|_{t=0},$$

or $N_0\bar{k}_0 = N_0 k_0$, so that $\bar{k}_0 = k_0$. ■

In case the lapse $N_t = N(t)$ does not depend on the space coordinate, then $N_t = d\tau_t/dt$ and the relation of solutions (g_t, k_t) to Einstein's equations with $N(t)$, $X = 0$, and $\bar{N}_t = 1$, $X = 0$ is particularly simple. In fact, if we define $\tau(t) = \int_0^t N(\lambda)d\lambda$, then the solutions to the Einstein equations with $\bar{N} = 1$, $X = 0$ are just reparameterizations by $\tau(t)$ of the solutions (g_t, k_t) of the Einstein system with $N_t = N(t)$, $X = 0$. We check this formally as follows:

Proposition 10.1: Let $N_t = N(t)$ be a function of t alone and let $(g_t, k_t) \in \mathfrak{M} \times S_2(M)$ be a solution of the Einstein system with $N_t = N(t)$, $X = 0$. Let $\tau(t) = \int_0^t N(\lambda)d\lambda$ and let $\tau^{-1}(t)$ be its inverse. Then

$$\bar{g}_t = g \circ \tau^{-1}(t) \quad \text{and} \quad \bar{k}_t = k \circ \tau^{-1}(t)$$

is the solution to (E) with $N_t = 1$ and

$$(\bar{g}_0, \bar{k}_0) = (g_0, k_0).$$

Remark: $\bar{g}_{\tau(t)} = g_t$ and $\bar{k}_{\tau(t)} = k_t$.

Proof: That $(\bar{g}_0, \bar{k}_0) = (g_0, k_0)$ follows from $\tau^{-1}(0) = 0$. Also,

$$\frac{d\bar{g}(t)}{dt} = \frac{dg}{dt}(\tau^{-1}(t))\frac{d\tau^{-1}}{dt}(t) = (Nk)(\tau^{-1}(t))\left(\frac{d\tau}{dt}(\tau^{-1}(t))\right)^{-1}$$

$$= N(\tau^{-1}(t))k(\tau^{-1}(t))\frac{1}{N(\tau^{-1}(t))} = \bar{k}(t)$$

and

$$\frac{d\bar{k}(t)}{dt} = \frac{dk}{dt}(\tau^{-1}(t))\frac{d\tau^{-1}}{dt}(t) = S_{g_{(\tau^{-1}(t))}}[k(\tau^{-1}(t))] - 2\operatorname{Ric}[g(\tau^{-1}(t))] = S_{\bar{g}_t}(\bar{k}(t)) - 2\operatorname{Ric}(\bar{g}(t))$$

so that $(\bar{g}_t, \bar{k}_t)$ is a solution to (E) with $\bar{N}_t = 1$. ■

Now we briefly consider the converse program; namely, given the solution (g_t, k_t) to (E) with lapse $N_t = 1$, and given an arbitrary lapse N_t, find the solution $(\bar{g}_t, \bar{k}_t)$ to (E) with lapse N_t and such that $(g_0, k_0) = (\bar{g}_0, \bar{k}_0)$.

We claim that by a simple trick this program can be carried out by solving for the Gaussian coordinates of a suitably altered space-time.

Theorem 10.3: Let $(g_t, k_t) \in \mathfrak{M} \times S_2(M), |t| < \epsilon$ be a solution of the Einstein system (E) with lapse $N_t = 1$, $X_t = 0$. Let N_t be a given lapse function. Construct a Lorentz metric on $I \times M$ by setting

$$l_{\alpha\beta} = \frac{-dt^2}{N^2} + \frac{g_{ij}}{N^2} dx^i dx^j$$

and let $\bar{x}^\mu(x^\alpha)$, $\bar{t}(0, x^i) = 0$, $\bar{x}^i(0, x^j) = x^i$ be the transformation of $l_{\alpha\beta}$ to Gaussian coordinates. In these new coordinates, the metric $g_{\alpha\beta}dx^\alpha dx^\beta = -dt^2 + g_{ij}dx^i dx^j$ is transformed to $\bar{g}_{\alpha\beta}d\bar{x}^\alpha d\bar{x}^\beta = -N^2 d\bar{t}^2 + \bar{g}_{ij}d\bar{x}^i d\bar{x}^j$, so that $(\bar{g}_t, (1/N_t)(\partial \bar{g}_t/\partial t))$, $|t| < \epsilon' \le \epsilon$, $\bar{g}_t = \bar{g}_{ij}$ solves (E) with lapse N_t and $(\bar{g}_0, \bar{k}_0) = (g_0, k_0)$.

Proof: The conditions that the new coordinates $\bar{x}^\mu(x^\alpha)$ transform $l_{\alpha\beta}$ to Gaussian coordinates are

$$\begin{cases} -1 = -N^2\left(\dfrac{\partial \bar{t}}{\partial t}\right)^2 + N^2 g^{ij}\dfrac{\partial \bar{t}}{\partial x^i}\dfrac{\partial \bar{t}}{\partial x^j} \\ 0 = -N^2\dfrac{\partial \bar{t}}{\partial t}\dfrac{\partial \bar{x}^i}{\partial t} + N^2 g^{kl}\dfrac{\partial \bar{t}}{\partial x^k}\dfrac{\partial \bar{x}^i}{\partial x^l}. \end{cases} \tag{10.4}$$

For any g_{ij}, we can solve these partial differential equations for $\bar{x}^\mu(x^\alpha)$ with initial conditions $\bar{x}^\mu(0, x^i) = (0, x^i)$ by Lemma 10.1 and Theorem 10.1.

The conditions that a coordinate transformation $\bar{y}^\mu(x^\alpha)$ transform $g_{\alpha\beta}dx^\alpha dx^\beta = -dt^2 + g_{ij}dx^i dx^j$ to $\bar{g}_{\alpha\beta}d\bar{x}^\alpha d\bar{x}^\beta = -N^2 d\bar{t}^2 + \bar{g}_{ij}d\bar{x}^i d\bar{x}^j$ are

$$\begin{cases} -\dfrac{1}{N^2} = -1\left(\dfrac{\partial \bar{y}^0}{\partial t}\right)^2 + g^{ij}\dfrac{\partial \bar{y}^0}{\partial x^i}\dfrac{\partial \bar{y}^0}{\partial x^j}, \\ 0 = -1\dfrac{\partial \bar{y}^0}{\partial t}\dfrac{\partial \bar{y}^i}{\partial t} + g^{il}\dfrac{\partial \bar{y}^0}{\partial x^k}\dfrac{\partial \bar{y}^i}{\partial x^l}. \end{cases} \tag{10.5}$$

But Eqs. (10.5) are equivalent to (10.4). ■

We remark that by our above work, we know that when transforming $l_{\alpha\beta}$ to Gaussian coordinates, the functions $\tau_t(m) = \bar{t}(t, x^i)$ and $\phi_t(m) = \bar{x}^i(t, x^j)$ satisfy

$$\begin{cases} \dfrac{d\phi_t}{dt} = \sqrt{\dfrac{1}{N^2} + \|\operatorname{grad}\tau_t\|^2}, \\ \dfrac{d\phi_t}{dt}\cdot\phi_t^{-1} = -\phi_*\left(-\dfrac{\operatorname{grad}\tau}{d\tau/dt}\right), \end{cases} \tag{10.6}$$

so that if ψ_t is the flow of $-\operatorname{grad}\tau/(d\tau/dt)$, $\phi_t = \psi_t^{-1}$. Equations (10.6) are just (10.4) rewritten.

Also, as in Theorem 10.2 the equations for $\bar{g}_t$ are given by

$$\bar{g}^{-1}(\tau_t(m), \phi_t(m)) = T\phi_t \otimes T\phi_t\left(g_t^{-1} - \frac{\operatorname{grad}\tau}{d\tau/dt} \otimes \frac{\operatorname{grad}\tau}{d\tau/dt}\right),$$

which follows from the coordinate expression

$$\bar{g}^{ij}(\bar{x}^\lambda(x^\alpha)) = \frac{\partial \bar{x}^i}{\partial x^k}(x^\alpha)\frac{\partial \bar{x}^j}{\partial x^l}(x^\alpha)g^{kl}(x^\alpha) - \frac{\partial \bar{x}^i}{\partial t}(x^\alpha)\frac{\partial \bar{x}^j}{\partial t}(x^\alpha)$$

and (10.4).

11. THE RELATIONSHIP OF THE MANIFOLD $\mathcal{G}$ TO SUPERSPACE AND SOME REMARKS ON SUPERPHASE SPACE

Let

$$\Phi: \mathfrak{D} \times \mathfrak{M} \to \mathfrak{M}, \qquad (\eta, g) \mapsto (\eta^{-1})^* g$$

be the left action of $\mathfrak{D}$ on $\mathfrak{M}$. Then superspace $\mathcal{S}(M)$, or the space of all geometries of M, is defined as the orbit space $\mathfrak{M}/\mathfrak{D}$ of this action. This is explained briefly as follows: For fixed $g \in \mathfrak{M}$, let

$$\mathcal{O}_g = \{(\eta^{-1})^* g \mid \eta \in \mathfrak{D}\} \subset \mathfrak{M}$$

be the orbit of $\mathfrak{D}$ through g. Then $\mathcal{O}_g$ is the set of all metrics isometric to g. Since $\mathcal{S}(M)$ is the set of all orbits in $\mathfrak{M}$,

$$\mathcal{S}(M) = \mathfrak{M}/\mathfrak{D} = \{\mathcal{O}_g \mid g \in \mathfrak{M}\},$$

$\mathcal{S}(M)$ is the space of all isometry classes of Riemannian metrics or geometries on M.

The importance of $\mathcal{S}(M)$ is that it is the natural configuration space for a dynamical theory of general relativity. The reason for this is that isometric Riemannian metrics are physically indistinguishable; thus a physical state determines only an isometry class of Riemannian metrics. In the language of the classical physicist, the metric representing the physical state is determined only up to a coordinate transformation.

Unfortunately $\mathcal{S}(M)$ is not a differentiable manifold. This is because the isometry group $I_g = \{\eta \in \mathfrak{D} \mid (\eta^{-1})^* g = g\}$ of a metric $g \in \mathfrak{M}$ is different for different g. As the isometry group I_g is the isotropy group of the action Φ at g, the resulting orbit space is not a manifold; in other words, the symmetric geometries do not have neighborhoods homeomorphic to neighborhoods of geometries which have no symmetries whatsoever. $\mathcal{S}(M)$ can, however, be stratified into differentiable manifolds, each strata $\mathcal{S}_{(G)}(M)$ being all those geometries whose isometry groups I_g determine equivalent group actions on M. This, however, is a rather long story; for details see Ref. 5.

Since $\mathcal{S}(M)$ is not a manifold, it is awkward to use $\mathcal{S}(M)$ as the configuration space for a dynamical system. This difficulty becomes apparent as soon as we try to construct the tangent bundle $T(\mathfrak{M}/\mathfrak{D})$, the velocity phase space. It is probably possible to give meaning to $T(\mathfrak{M}/\mathfrak{D})$ by taking limits of tangent spaces and using the notion of tangent cones. However, the singularities of $\mathcal{S}(M)$ would then be severely compounded.

A way to short circuit this approach is to define as the superphase space not $T(\mathfrak{M}/\mathfrak{D})$, but rather

$$(T\mathfrak{M})/\mathfrak{D} = \frac{\mathfrak{M} \times S_2(M)}{\mathfrak{D}} \neq \frac{\mathfrak{M}}{\mathfrak{D}} \times \frac{S_2(M)}{\mathfrak{D}}$$

the orbit space of the action

$$\Phi': \mathfrak{D} \times T\mathfrak{M} \to T\mathfrak{M}, \qquad (\eta, g, k) \mapsto ((\eta^{-1})^* g, (\eta^{-1})^* k).$$

Note that $T\mathfrak{M}/\mathfrak{D}$ is not equivalent to $T(\mathfrak{M}/\mathfrak{D})$. $T\mathfrak{M}/\mathfrak{D}$ is perhaps a more likely candidate for *superphase space* as the tensor fields $(g, k) \in T\mathfrak{M}$ are subjected only to the same active coordinate transformation; that is, the pair (g, k) must transform together. From Corollary 4.1 we see that this is appropriate from the dynamical point of view.

Unfortunately, $T\mathfrak{M}/\mathfrak{D} = (\mathfrak{M} \times S_2(M))/\mathfrak{D}$ suffers from all of the pathologies that $\mathcal{S}(M)$ does, as well as the added difficulties related to the structure of $S_2(M)/\mathfrak{D}$. Note for example that $\mathfrak{D}$ leaves invariant the 0-tensor field in $S_2(M)$, so that the isotropy subgroup of 0 need not even be finite-dimensional (as it is for Φ, Φ'). Thus the construction of equations of motion on $T\mathfrak{M}/\mathfrak{D}$ directly does not seem feasible at this time. Nevertheless, the dynamics on $T\mathfrak{M}$, followed by a projection onto $T\mathfrak{M}/\mathfrak{D}$ does recapture all the essential elements that a dynamical system on $T\mathfrak{M}/\mathfrak{D}$ would have to possess. In fact, $T\mathfrak{M}/\mathfrak{D}$ inherits a continuous flow from the flow on $T\mathfrak{M}$. Thus we have a C^0 or topological dynamical system.

To incorporate the shift vector field into the dynamics we have in the course of this paper chosen $T(\mathfrak{D} \times \mathfrak{M})$ as the velocity phase space. The degeneracy in our Lagrangian allows one to specify arbitrarily a curve $\eta_t, \eta_0 = id_M$, in the factor $\mathfrak{D}$ or equivalently a shift vector field X_t. The solution (g_t, k_t) of (E) with shift vector field X_t is then related to a reference solution $(\bar{g}_t, \bar{k}_t)$ of (E) with shift $\bar{X}_t = 0$ by $(g_t, k_t) = ((\eta_t^{-1})^* \bar{g}_t, (\eta_t^{-1})^* \bar{k}_t)$. Thus (g_t, k_t) can be thought of as $(\bar{g}_t, \bar{k}_t)$ sliding along the orbits $\mathcal{O}_{(g_t, k_t)} = \{((\eta^{-1})^* g, (\eta^{-1})^* k) \mid \eta \in \mathfrak{D}\} \subset \mathfrak{M} \times S_2(M)$. The factor $\mathfrak{D}$ then keeps track of the amount of sliding relative to the reference curve $(\bar{g}_t, \bar{k}_t)$.

Note that the curve (g_t, k_t) and $(\bar{g}_t, \bar{k}_t)$ are projected onto the same curve in $T\mathfrak{M}/\mathfrak{D}$ as we have divided out by $\mathfrak{D}$, so that solutions to (E) which differ only by a shift vector field map to the same curve in $T\mathfrak{M}/\mathfrak{D}$. Thus if we could construct a dynamical system on $T\mathfrak{M}/\mathfrak{D}$ directly, it would be independent of the shift.

We also remark that if we define $(\eta_1, g_1) \in \mathfrak{D} \times \mathfrak{M}$ to be equivalent (= ~) to $(\eta_2, g_2) \in \mathfrak{D} \times \mathfrak{M}$ if $(\eta_1^{-1})^* g_1 = (\eta_2^{-1})^* g_2$, then $\mathfrak{D} \times \mathfrak{M}/\sim \cong \mathfrak{M}/\mathfrak{D}$. Similarly, if on $\mathcal{T}(\mathfrak{D} \times \mathfrak{M})$ we define $(X_{\eta_1}, g_1, k_1) \sim (Y_{\eta_1}, g_2, k_2)$ if $((\eta_1^{-1})^* g_1, (\eta_1^{-1})^* k_1) = ((\eta_2^{-1})^* g_2, (\eta_2^{-1})^* k_2)$, then $T(\mathfrak{D} \times \mathfrak{M})/\sim = T\mathfrak{M}/\mathfrak{D}$. Thus we recapture the space of *physically indistinguishable states* $T\mathfrak{M}/\mathfrak{D}$ by defining a suitable equivalence relation on $T\mathfrak{A} \approx T(\mathfrak{D} \times \mathfrak{M})$. Of course this equivalence relation is natural from the point of view of the dynamical development of the states.

We now wish to incorporate the lapse function into the picture. For nonrelativistic classical field theories, there is a canonical parameter of evolution, namely the time t. For covariant relativistic field theories in general, and for general relativity in particular, the proper time τ plays the role of a canonical parameter of evolution. In order to maintain covariance, however, one must allow for an arbitrary reparameterization of this evolution parameter. This reparameterization may also depend on the space points of the field. It is because of this possible space dependence of the change of parameter that Wheeler refers to time as a many fingered entity; this is associated with the Dirac–Tomonaga–Schwinger many time formalism for quantum field theory.

Another well-known implication of covariance is that a covariant field theory when expressed in a dynamical formulation must be degenerate. This situation comes about because the resulting dynamics must be able to be summed up as a tensor field on a four dimensional manifold V. Each slicing of V, therefore, gives rise to a different dynamical system all of which are equivalent in the sense that they lead to the same tensor field on V. As the dynamical formalism must take into account this arbitrary slicing, it must be degenerate. In this paper the introduction of $\mathcal{T}$ accounts for this arbitrary slicing of a space–time; $\mathfrak{D}$ takes into account the possible coordinates in each slice. Let $V = I \times M$, and let $\{t\} \times M$ be the t = constant hypersurfaces. Let τ_t be a curve in $\mathcal{T}$ and let $\tau\colon I \times M \to \mathbf{R}$, $(t, m) \mapsto \tau_t(m)$. Then $\tau^{-1}(t)$ are the τ = constant hypersurfaces. Thus each curve $\tau_t \in \mathcal{T}$ maps t = constant to τ = constant hypersurfaces and thus can be considered as an arbitrary slicing of V with respect to some reference slicing which represents V as $I \times M$.

Let $(\bar{g}_t, \bar{k}_t)$ be a solution of (E) with $\overline{X} = 0$, $\overline{N} = 1$. We consider the curve $\tau_t \in \mathcal{T}$ defined by $\bar{\tau}_t(m) = 1$ as a reference curve. We construct the space–time $g_{\alpha\beta} dx^\alpha dx^\beta = -dt^2 + g_{ij} dx^i dx^j$ on $I \times M$. Now suppose that we are given an arbitrary curve $\tau_t \in \mathcal{T}$, $\tau_0 = 0$. Alternately, by Lemma 10.1 we may suppose we are given some lapse function N_t and then find τ_t associated with it and $\bar{g}_t$. The evolution for this new lapse or new τ_t is determined by finding the metric on the τ = constant hypersurfaces. Of course, these evolutions are equivalent in that they determine isometric Lorentz metrics or, in other words, are summed up by the same space–time. Nevertheless, given a curve $(\bar{g}_t, \bar{k}_t) \in T\mathfrak{M}$, each curve τ_t determines a dynamical curve $(g_t, k_t) \in T\mathfrak{M}$. Thus we map curves in $\mathcal{T}$ to curves in $T\mathfrak{M}/\mathfrak{D} = T(\mathfrak{D} \times \mathfrak{M})/\sim$. The image in $S(M) \times S_2(M)/\mathfrak{D}$ of all curves in $\mathcal{T}$, projected onto $S(M)$, is just the sheaf in superspace which summarizes the space–time. This idea is described by DeWitt[23] without the use of the space $\mathcal{T}$.

12. CONCLUSIONS AND FURTHER WORK

In this paper we have attempted to clarify the Hamiltonian structure of the Einstein equations and to achieve a clear understanding of the geometrical roles played by the lapse and shift functions. We feel that we have gained a more natural form for the phase space of general relativity by introducing the groups $\mathfrak{D}$ and $\mathcal{T}$. For example, by enlarging the configuration space from $\mathfrak{M}$ to $\mathcal{T} \times \mathfrak{D} \times \mathfrak{M}$, the lapse and shift functions may be incorporated into the dynamics as dynamical velocities. Moreover, we showed explicitly how one can obtain solutions for any lapse or shift from the trivial ones $N = 1$, $X = 0$ by integrating a system of ordinary differential equations. In connection with the lapse, we introduced a new object, the intrinsic shift, which takes into account the spatial shifting of the $\{t\} \times M$ hypersurfaces when mapped into τ = constant hypersurfaces.

We feel that the introduction of the groups $\mathfrak{D}$ and $\mathcal{T}$ helps to properly understand the basic conservation laws for $\delta\pi$ and $\mathcal{H}$ as a consequence of dynamical symmetries. On the other hand, we are forced to accept a degenerate Lagrangian system. This degeneracy is present and is perfectly natural when one considers any covariant field theory from a dynamical point of view.

Some work which remains to be done is to explore whether or not the procedure presented here helps to clarify any of the difficult quantization problems. However, preliminary indications are that quantization problems run much deeper. For example, in the usual quantum theory of fields one deals with equations of the form $\Box\phi + F(\phi) = 0$, for definiteness say $\Box\phi + \lambda\phi^3 = m^2\phi$. As a classical partial differential equation, this equation is semilinear, as the highest order derivatives occur linearly, the nonlinearity occurring only in the ϕ^3 term. As is well known, a rigorous and complete quantization of such equations is very difficult and, in fact, has not yet been achieved for four-dimensional space–times. In relativity, the basic structure of the evolution equations is quite different. Let us, for example, neglect the fact that the equations for a space–time are a *system* of partial differential equations. Then, roughly speaking, the "scalar analog" of the evolution equations is the quasilinear equation $\phi\Box\phi + \|\text{grad}\phi\|^2 = 0$, $\text{grad}\phi = g^{\alpha\beta}(\partial\phi/\partial x^\alpha)$. Now ϕ itself is involved in the coefficients of the operator $\phi\Box$. Also, the equations involves nonlinear derivative coupling terms. Very little is known about the quantization of such an equation. Moreover, the equations for relativity are much more involved, as they involve a system of quasilinear equations, the components of which are very badly mixed in the highest order (unless one chooses the harmonic coordinate condition) and first-order derivative terms. Thus, a complete quantum theory of general relativity seems quite far away.[3,24]

In a forthcoming paper[18,14] we shall be focussing our attention on problems of existence and uniqueness of

solutions using the theory of quasilinear *first*-order symmetric hyperbolic systems, rather than the usual methods using the considerably more complicated theory of second-order strictly hyperbolic systems. Our existence proof will follow simply and directly from such a first-order treatment. Moreover, we will be able to give a more intrinsic treatment by making use of a global version of the symmetric hyperbolic theory.

ACKNOWLEDGMENTS

We thank P. Chernoff, D. Ebin, H. Kunzle, K. Kuchař, R. Sachs, and A. Taub for a variety of helpful suggestions.

* Partially supported by AEC Contract AT(04-3)-34.
† Partially supported by Navy Contract N00014-69-A-0200-1002 and NSF Contract GP-8257. Part of this author's work was done while at the University of Toronto, Toronto, Canada, and the Institute for Advanced Study, Princeton, New Jersey.
1 J. A. Wheeler, "Geometrodynamics and the Issue of the Final State," in *Relativity, Groups and Topology*, edited by B. DeWitt and C. DeWitt (Gordon and Breach, New York, 1964).
2 R. Arnowitt, S. Deser, and C. W. Misner, "The Dynamics of General Relativity," in *Gravitation: An Introduction to Current Research*, edited by L. Witten (Wiley, New York, 1962).
3 B. DeWitt, Phys. Rev. **160**, 1113 (1967).
4 C. W. Misner, Rev. Mod. Phys. **29**, 497 (1957).
5 A. Fischer, "The Theory of Superspace," in *Relativity*, edited by M. Carmeli, S. Fickler, and L. Witten (Plenum, New York, 1967).
6 R. Arnowitt, S. Deser, and C. W. Misner, Phys. Rev. **116**, 1322 (1959).
7 V. Arnold, Ann. Inst. Fourrer (Grenoble) **16**, 319 (1966); D. Ebin and J. Marsden, Ann. Math. **92**, 102 (1970).
8 R. Abraham, *Foundations of Mechanics* (Benjamin, New York, 1967).
9 C. Lanczos, *The Variational Principles of Mechanics* (U. of Toronto Press, Toronto, 1966), 3rd ed.
10 D. Ebin, "The Manifold of Riemannian Metrics," in *Proceedings of the Symposium on Pure Mathematics* (Amer. Math. Soc., Providence, R. I., 1970), Vol. 15.
11 J. Marsden, Arch. Ratl. Mech. Anal. **28**, 362 (1968).
12 S. Lang, *Introduction to Differentiable Manifolds* (Wiley, New York, 1962).
13 Y. Choquet-Bruhat, Compt. Rend. **272**, 386 (1971); Ann. Inst. H. Poincaré **8**, 327 (1968).
14 A. Fischer and J. Marsden, "The Einstein Evolution Equations as a Quasi-Linear First Order Symmetric Hyperbolic System. I" (to be published).
15 D. Eardley, E. Liang, and R. Sachs, J. Math. Phys. **13**, 99 (1972).
16 S. Deser, Ann. Inst. H. Poincaré **7**, 146 (1967); M. Berger and D. Ebin, J. Diff. Geom. **3**, 379 (1969).
17 A. Lichnerowicz, *Relativistic Hydrodynamics and Magnetohydrodynamics* (Benjamin, New York, 1967).
18 A. Fischer and J. Marsden, "General Relativity, Partial Differential Equations, and Dynamical Systems," in *Proceedings of the Symposium on Pure Mathematics* (Amer. Math. Soc. Providence, R. I., 1972), Vol. 23.
19 Y. Choquet-Bruhat and R. Geroch, Commun. Math. Phys. **14**, 328 (1969).
20 D. Brill and S. Deser, Ann. Phys. (N.Y.) **50**, 548 (1968); J. Marsden and A. Fischer, "The Existence of Complete Spacetimes" (in preparation).
21 K. Yano, *Integral Formulas in Riemannian Geometric* (Decker, New York, 1971).
22 R. Abraham, J. Math. Mech. **11**, 553 (1962).
23 B. DeWitt, "Space–Time as a Sheaf of Geodesics in Superspace," in *Relativity*, edited by M. Carmeli, S. Fickler, and L. Witten (Plenum, New York, 1970).
24 L. D. Faddeev, Act. Cong. Intern. Math. **3**, 35 (1970).

States, Minimizing the Uncertainty Product of the Oscillator Phase Operators

Evangelos K. Ifantis

Nuclear Research Center "Democritos", Aghia Paraskevi, Attikis, Athens, Greece

(Received 8 October 1971; Revised Manuscript Received 29 November 1971)

The normalizable states that minimize the uncertainty product of the oscillator phase operators are determined and some of their physical properties are discussed. A physical classification of these states has been made and the class of "analogous" states to the well-known coherent states is physically defined.

1. INTRODUCTION

Quantum mechanically it is convenient to define the phase ϕ of the harmonic oscillator indirectly by defining the "cosine" and "sine" operators C and S, which correspond to $\cos\phi$ and $\sin\phi$ in the classical limit. The operators C and S found do not commute, i.e., the $\cos\phi$ and $\sin\phi$ cannot be measured simultaneously. It is therefore interesting to find the normalizable states, which minimize the uncertainty product $(\Delta C)^2\cdot(\Delta S)^2$ of C and S.

It was proved in Ref. 1 and was noted in further research[2,3] on the quantum mechanical oscillator phase problem that there exist no normalizable states that minimize the uncertainty product $(\Delta C)^2\cdot(\Delta S)^2$. This result is correct in the sense that for these states the inequality $(\Delta C)^2(\Delta S)^2 \geqslant \frac{1}{4}\langle(1-p)/2\rangle^2$ becomes an equality and, moreover, $\langle 1-p\rangle^2$ becomes a greatest lower bound.

We have shown in a previous work,[4] using methods of the spectral theory of bounded operators, that normalizable states minimizing the uncertainty product $(\Delta C)^2\cdot(\Delta S)^2$ do exist in the sense that the above inequality becomes an equality.

In the present work we determine these states and find some of their physical properties. Moreover, we classify the normalizable minimal uncertainty states and characterize, both mathematically and physically, the states that have properties analogous to those of the well known coherent states.

In Sec. 2 we present some general properties of the normalizable minimal uncertainty states and the "minimal uncertainty sequences" of states for arbitrary noncompatible observables A and B. We note that the knowledge of the point spectrum and the continuous spectrum of the non-self-adjoint operator $A + i\gamma B$ is sufficient for the determination of the expectation values of A and B in the corresponding states. The determination is exact in the case of the point spectrum and approximate in the case of the continuous spectrum. In addition we characterize mathematically the states that have properties analogous to those of the coherent states.

solutions using the theory of quasilinear *first*-order symmetric hyperbolic systems, rather than the usual methods using the considerably more complicated theory of second-order strictly hyperbolic systems. Our existence proof will follow simply and directly from such a first-order treatment. Moreover, we will be able to give a more intrinsic treatment by making use of a global version of the symmetric hyperbolic theory.

ACKNOWLEDGMENTS

We thank P. Chernoff, D. Ebin, H. Kunzle, K. Kuchař, R. Sachs, and A. Taub for a variety of helpful suggestions.

* Partially supported by AEC Contract AT(04-3)-34.
† Partially supported by Navy Contract N00014-69-A-0200-1002 and NSF Contract GP-8257. Part of this author's work was done while at the University of Toronto, Toronto, Canada, and the Institute for Advanced Study, Princeton, New Jersey.
1 J. A. Wheeler, "Geometrodynamics and the Issue of the Final State," in *Relativity, Groups and Topology*, edited by B. DeWitt and C. DeWitt (Gordon and Breach, New York, 1964).
2 R. Arnowitt, S. Deser, and C. W. Misner, "The Dynamics of General Relativity," in *Gravitation: An Introduction to Current Research*, edited by L. Witten (Wiley, New York, 1962).
3 B. DeWitt, Phys. Rev. **160**, 1113 (1967).
4 C. W. Misner, Rev. Mod. Phys. **29**, 497 (1957).
5 A. Fischer, "The Theory of Superspace," in *Relativity*, edited by M. Carmeli, S. Fickler, and L. Witten (Plenum, New York, 1967).
6 R. Arnowitt, S. Deser, and C. W. Misner, Phys. Rev. **116**, 1322 (1959).
7 V. Arnold, Ann. Inst. Fourrer (Grenoble) **16**, 319 (1966); D. Ebin and J. Marsden, Ann. Math. **92**, 102 (1970).
8 R. Abraham, *Foundations of Mechanics* (Benjamin, New York, 1967).
9 C. Lanczos, *The Variational Principles of Mechanics* (U. of Toronto Press, Toronto, 1966), 3rd ed.
10 D. Ebin, "The Manifold of Riemannian Metrics," in *Proceedings of the Symposium on Pure Mathematics* (Amer. Math. Soc., Providence, R. I., 1970), Vol. 15.
11 J. Marsden, Arch. Ratl. Mech. Anal. **28**, 362 (1968).
12 S. Lang, *Introduction to Differentiable Manifolds* (Wiley, New York, 1962).
13 Y. Choquet-Bruhat, Compt. Rend. **272**, 386 (1971); Ann. Inst. H. Poincaré **8**, 327 (1968).
14 A. Fischer and J. Marsden, "The Einstein Evolution Equations as a Quasi-Linear First Order Symmetric Hyperbolic System. I" (to be published).
15 D. Eardley, E. Liang, and R. Sachs, J. Math. Phys. **13**, 99 (1972).
16 S. Deser, Ann. Inst. H. Poincaré **7**, 146 (1967); M. Berger and D. Ebin, J. Diff. Geom. **3**, 379 (1969).
17 A. Lichnerowicz, *Relativistic Hydrodynamics and Magnetohydrodynamics* (Benjamin, New York, 1967).
18 A. Fischer and J. Marsden, "General Relativity, Partial Differential Equations, and Dynamical Systems," in *Proceedings of the Symposium on Pure Mathematics* (Amer. Math. Soc. Providence, R. I., 1972), Vol. 23.
19 Y. Choquet-Bruhat and R. Geroch, Commun. Math. Phys. **14**, 328 (1969).
20 D. Brill and S. Deser, Ann. Phys. (N.Y.) **50**, 548 (1968); J. Marsden and A. Fischer, "The Existence of Complete Spacetimes" (in preparation).
21 K. Yano, *Integral Formulas in Riemannian Geometric* (Decker, New York, 1971).
22 R. Abraham, J. Math. Mech. **11**, 553 (1962).
23 B. DeWitt, "Space–Time as a Sheaf of Geodesics in Superspace," in *Relativity*, edited by M. Carmeli, S. Fickler, and L. Witten (Plenum, New York, 1970).
24 L. D. Faddeev, Act. Cong. Intern. Math. **3**, 35 (1970).

States, Minimizing the Uncertainty Product of the Oscillator Phase Operators

Evangelos K. Ifantis

Nuclear Research Center "Democritos", Aghia Paraskevi, Attikis, Athens, Greece

(Received 8 October 1971; Revised Manuscript Received 29 November 1971)

The normalizable states that minimize the uncertainty product of the oscillator phase operators are determined and some of their physical properties are discussed. A physical classification of these states has been made and the class of "analogous" states to the well-known coherent states is physically defined.

1. INTRODUCTION

Quantum mechanically it is convenient to define the phase ϕ of the harmonic oscillator indirectly by defining the "cosine" and "sine" operators C and S, which correspond to $\cos\phi$ and $\sin\phi$ in the classical limit. The operators C and S found do not commute, i.e., the $\cos\phi$ and $\sin\phi$ cannot be measured simultaneously. It is therefore interesting to find the normalizable states, which minimize the uncertainty product $(\Delta C)^2\cdot(\Delta S)^2$ of C and S.

It was proved in Ref. 1 and was noted in further research[2,3] on the quantum mechanical oscillator phase problem that there exist no normalizable states that minimize the uncertainty product $(\Delta C)^2\cdot(\Delta S)^2$. This result is correct in the sense that for these states the inequality $(\Delta C)^2(\Delta S)^2 \geqslant \frac{1}{4}\langle(1-p)/2\rangle^2$ becomes an equality and, moreover, $\langle 1-p\rangle^2$ becomes a greatest lower bound.

We have shown in a previous work,[4] using methods of the spectral theory of bounded operators, that normalizable states minimizing the uncertainty product $(\Delta C)^2\cdot(\Delta S)^2$ do exist in the sense that the above inequality becomes an equality.

In the present work we determine these states and find some of their physical properties. Moreover, we classify the normalizable minimal uncertainty states and characterize, both mathematically and physically, the states that have properties analogous to those of the well known coherent states.

In Sec. 2 we present some general properties of the normalizable minimal uncertainty states and the "minimal uncertainty sequences" of states for arbitrary noncompatible observables A and B. We note that the knowledge of the point spectrum and the continuous spectrum of the non-self-adjoint operator $A + i\gamma B$ is sufficient for the determination of the expectation values of A and B in the corresponding states. The determination is exact in the case of the point spectrum and approximate in the case of the continuous spectrum. In addition we characterize mathematically the states that have properties analogous to those of the coherent states.

In Sec. 3 we determine all the normalizable minimal uncertainty states for the phase operators C and S: in which both C and S have vanishing expectation values. These states are called Y_0-states.

In Sec. 4 we characterize and determine the class of Y-states, which are the complement to the Y_0-states with respect to all normalizable minimal uncertainty states. Among the Y-states are the f_z-states, which are eigenstates of the adjoint of the unilateral shift operator. In these states the operators C and S have the same mean square deviation. This property is a characteristic property of the f_z-states. All the physical quantities in the Y_0- and Y-states are determined and give reasonable physical results.

In Sec. 5, we find the physical quantities of C and S in the "minimal uncertainty sequences." It is found that in these sequences of states the expectation values of C and S tend to the classical quantities $\cos\phi$ and $\sin\phi$ and the mean square deviations tend to zero.

Finally in Appendixes A and B we clarify the method of obtaining normalizable minimal uncertainty states and "minimal uncertainty sequences."

2. GENERAL PROPERTIES OF MINIMAL UNCERTAINTY STATES

Normalizable States

The normalizable minimal uncertainty states for the noncompatible observables A and B, satisfying the commutation relation $AB - BA = iC$, are obtained as eigenstates of the non-self-adjoint operator

$$L = A + i\gamma B, \tag{2.1}$$

where γ is a nonvanishing free real parameter (see Appendix A).

From (2.1), setting $M = A - iB$, $M^* = A + iB$, we obtain

$$L = [(1+\gamma)/2]\cdot T,$$

where

$$T = rM + M^* \tag{2.2}$$

and

$$r = (1-\gamma)/(1+\gamma). \tag{2.3}$$

Since $\gamma \neq 0$, we have from (2.3)

$$r \neq 1.$$

Thus, the minimal uncertainty states for A and B are obtained as eigenstates of the operator (2.2).

In Ref. 4 we have studied the spectrum of T for a large class of bounded non-self-adjoint operators M. Here we note that the knowledge of the spectrum of T has the following physical importance.

The expectation values of A and B can be determined only from the eigenvalues of T.

We start from the eigenvalue equation

$$(rM + M^*)f = \lambda f, \qquad \|f\| = 1 \tag{2.4}$$

and write λ in the form:

$$\lambda = r\mu e^{i\phi} + \mu e^{-i\phi}, \qquad \mu \geq 0,\ 0 \leq \phi \leq 2\pi. \tag{2.5}$$

Thus, every normalizable minimal uncertainty state f is characterized from three real parameters r, μ, ϕ. Write $f(r,\mu,\phi)$ and denote by Y_0 the class of states $f(r,0,0)$ and by Y the states $f(r,\mu,\phi)$ with $\mu \neq 0$. The class Y_0 consists from states, which are eigenstates of T with eigenvalue zero. These states are characterized from one real parameter. The states Y are obtained from (2.4) with $\lambda \neq 0$. Obviously, in the following we consider the case in which the sets Y_0 and Y are not empty.

From the eigenvalue equation (2.4), using the identities:

$$\begin{aligned} \langle A\rangle &= (Af,f) = \frac{\lambda}{2} - \frac{r-1}{2}(Mf,f), \\ \langle B\rangle &= (Bf,f) = \frac{\lambda}{2i} - \frac{r+1}{2i}(Mf,f), \end{aligned} \tag{2.6}$$

or directly from the equation

$$\langle A\rangle + i\gamma\langle B\rangle = [(1+\gamma)/2]\cdot\lambda,$$

combining real and imaginary parts, we derive easily the following statements:

(1) The states Y_0 are the only normalizable minimal uncertainty states in which both A and B have vanishing expectation values.

(2) The expectation values of A and B in the states $f(r,\mu,\phi) \in Y$ are obtained from the following formulas:

$$\begin{aligned} \langle A\rangle &= (\mathrm{Re}\lambda)/(1+r), \\ \langle B\rangle &= (\mathrm{Im}\lambda)/(1-r), \end{aligned} \tag{2.7}$$

or writing λ in the form (2.5),

$$\langle A\rangle = \mu\cos\phi, \tag{2.8}$$

$$\langle B\rangle = -\mu\sin\phi. \tag{2.9}$$

Normalizable Sequences of States

In Ref. 4 we have introduced the concept of "minimal uncertainty sequences." They are normalizable sequences of states, which tend to minimize the uncertainty product $(\Delta A)\cdot(\Delta B)$ of A and B. As we note in Appendix B, the minimal uncertainty sequences of states correspond to the continuous spectrum of the operator (2.2), i.e., for every λ in the continuous spectrum of T there exists a minimal uncertainty sequence of states $\{f_n\}_1^\infty$, satisfying the relation

$$\lim \|(T - \lambda I)f_n\| = 0 \quad \text{as } n \to \infty. \tag{2.10}$$

From the identities

$$\begin{aligned} A &= \frac{\lambda I}{2} - \frac{r-1}{2}M + \frac{T-\lambda I}{2}, \\ B &= \frac{\lambda I}{2i} - \frac{r+1}{2i}M + \frac{T-\lambda I}{2i}, \end{aligned}$$

we obtain the relations

$$\begin{aligned} \langle A\rangle &= (Af_n,f_n) = \frac{\lambda + R(f_n)}{2} - \frac{r-1}{2}(Mf_n,f_n), \\ \langle B\rangle &= (Bf_n,f_n) = \frac{\lambda + R(f_n)}{2i} - \frac{r+1}{2i}(Mf_n,f_n), \end{aligned} \tag{2.11}$$

where

$$R(f_n) = (\{T - \lambda I\}f_n,f_n). \tag{2.12}$$

The relations (2.11) are analogous to the relations (2.6). Thus, we obtain

$$\langle A\rangle = \frac{\mathrm{Re}[\lambda + R(f_n)]}{1+r} = \frac{\mathrm{Re}\lambda}{1+r} + \frac{\mathrm{Re}R(f_n)}{1+r}, \qquad \langle B\rangle = \frac{\mathrm{Im}[\lambda + R(f_n)]}{1-r} = \frac{\mathrm{Im}\lambda}{1-r} + \frac{\mathrm{Im}R(f_n)}{1-r}. \tag{2.13}$$

From (2.12) we have $|R(f_n)| \leq \|(T-\lambda I)f_n\|$ and due to (2.10) $\lim|R(f_n)| = 0$, as $n \to \infty$, i.e., the expectation values (2.13) tend to the expectation values (2.7).

States Mathematically Analogous to the Coherent States

Consider the position and momentum operators Q and P, satisfying the commutation relation $QP - PQ = iI$.

For the operators Q and P the operator (2.2) has the form

$$T = ra^* + a, \tag{2.14}$$

where $a^* = 2^{-1/2}(Q - iP)$ and $a = 2^{-1/2}(Q + iP)$ are the creation and annihilation operators.

As is well known, the coherent states form a class of minimal uncertainty states for Q and P. They are the eigenstates of the annihilation operator a.

We observe that the coherent states are eigenstates of the operator (2.14) for $r = 0$. Therefore, generally, the states which are mathematically analogous to the coherent states are the eigenstates of the operator (2.2) for $r = 0$. They belong to the Y class.

It is easy to see that the so-defined, mathematically analogous states to the coherent states have the following physical property.

Both observables A and B have in these states the same mean square deviation.

3. THE Y_0-STATES OF THE OSCILLATOR PHASE OPERATORS

The oscillator phase operators C and S, studied in Refs. 1 and 3 are the following[5]:

$$C = (V^* + V)/2, \qquad S = (V^* - V)/2i, \tag{3.1}$$

where V is the unilateral shift operator on an abstract Hilbert space with the orthonormal basis $\{e_n\}_1^\infty$, i.e., V is defined as follows:

$$Ve_n = e_{n+1}, \quad n = 1, 2, \cdots .$$

The operator (2.2) in case of the operators (3.1) has the form

$$T_r = rV + V^*. \tag{3.2}$$

It is well known[4] that for $|r| < 1$ every point in the interior of the ellipse

$$z = re^{i\varphi} + e^{-i\varphi}, \qquad 0 \leq \phi \leq 2\pi,$$

is an eigenvalue of the operator (3.2).

We note that the eigenstates of the operator (3.2) for r negative are related with the eigenstates for r positive by a unitary transformation and a rotation, i.e.,

$$T_{-r} = e^{i\pi/2} W T_r,$$

where the unitary operator W is defined as follows:

$$We_n = e^{-i\pi n/2} e_n \qquad n = 1, 2.$$

In fact we have

$$T_{-r} = -rV + V^* = re^{i\pi}V + V^* = e^{i\pi/2}W^*(rV + V^*)W.$$

For this reason, in the following we consider always the operator (3.2) with r positive.

For $r = 0$ we obtain from (3.2) the well known eigenstates of the operator V^*. These states, as we have noted in Sec. 2, are the analogous to the coherent states and play an important role in many problems of the operator theory.[6,7]

The eigenstates of V^* belong to the class Y and we shall study their physical properties in Sec. 4. Only the state e_1 (ground state) belongs to the class Y_0, because $V^*e_1 = 0$.

Determination of the Y_0-States

We determine the eigenstates of the operator (3.2) for $r \neq 0$, corresponding to the eigenvalue zero. Let ψ_r be such a state, i.e.,

$$(rV + V^*)\psi_r = 0. \tag{3.3}$$

Setting $(\psi_r, e_1) = a$, we obtain from (3.3)

$$(\psi_r, e_2) = (\psi_r, e_4) = \cdots = 0$$

and $(\psi_r, e_3) = -ra$, $(\psi_r, e_5) = r^2 a, \ldots (\psi_r, e_{2n-1}) = [a\cdot(-r)^{n-1}]$, i.e., $\psi_r = a\sum_{n=1}^{\infty}(-r)^{n-1}\cdot e_{2n-1}$. The normalization of these states gives $a = (1-r^2)^{1/2}$. Thus

$$\psi_r = (1-r^2)^{1/2}\cdot\sum_{n=1}^{\infty}(-r)^{n-1}\cdot e_{2n-1}. \tag{3.4}$$

From (3.4) we have

$$C\psi_r = \frac{(1-r^2)^{1/2}\cdot(1-r)}{2}\cdot(e_2 - re_4 + r^2e_6 - \cdots), \tag{3.5}$$

$$S\psi_r = \frac{(1-r^2)^{1/2}\cdot(1+r)}{2}\cdot(-e_2 + re_4 - r^2e_6 + \cdots). \tag{3.6}$$

Thus

$$(C\psi_r, \psi_r) = (S\psi_r, \psi_r) = 0.$$

From (3.5) and (3.6) we obtain the mean deviations of C and S in the states (3.4)

$$\Delta C = \|C\psi_r\| = (1-r)/2, \tag{3.7}$$

$$\Delta S = (1+r)/2. \tag{3.8}$$

Thus

$$(\Delta C)\cdot(\Delta S) = (1-r^2)/4. \tag{3.9}$$

Observe that $\Delta C + \Delta S = 1$.

The operators C and S satisfy the commutation relation

$$CS - SC = i\cdot(I-P)/2, \tag{3.10}$$

where $I - P$ is a projection, which projects on the sub-

space, spanned by the element $\{e_1\}$, i.e., $I-P$ acts as follows:

$$(I-P)\psi_r = (\psi_r, e_1)e_1 = (1-r^2)^{1/2}e_1.$$

Thus

$$\langle (I-P)/2\rangle = \tfrac{1}{2}\cdot(\{I-P\}\psi_r, \psi_r) = (1-r^2)/2. \qquad (3.11)$$

From (3.9) and (3.11) we have

$$(\Delta C)\cdot(\Delta S) = \tfrac{1}{2}\langle (I-P)/2\rangle .$$

This is the minimal uncertainty relation for the commutation relation (3.10).

The Number Operator N Acting on the Set Y_0

The states (3.4) belong to the definition domain of the oscillator number operator N. In fact we have

$$\|N\psi_r\|^2 = (1-r^2)\cdot\sum_{n=1}^{\infty} r^{2n-2}\cdot(2n-2)^2 < \infty.$$

(Note that the oscillator number operator in the abstract form is defined[5] as follows: $Ne_n = (n-1)e_n$, $n = 1, 2, \cdots$.)

The expectation values of N in the states (3.4) are determined as follows:

$$\langle N+I\rangle = (\{N+I\}\psi_r, \psi_r) = (1-r^2)\cdot\sum_{n=1}^{\infty} r^{2(n-1)}\cdot(2n-1)$$
$$= (1-r^2)\cdot\frac{d}{dr}\left(\frac{r}{1-r^2}\right) = \frac{1+r^2}{1-r^2},$$

i.e.,

$$\langle N\rangle = \langle N+I\rangle - 1 = 2r^2/(1-r^2). \qquad (3.12)$$

From (3.4) we have

$$\langle (N+I)^2\rangle = \|(N+I)\psi_r\|^2$$
$$= (1-r^2)\cdot(1+3^2r^2+5^2r^4+\cdots)$$
$$= (1-r^2)\cdot\frac{d}{dr}\left[r\cdot\frac{d}{dr}\left(\frac{r}{1-r^2}\right)\right]$$
$$= \langle N+I\rangle^2 + \frac{4r^2}{(1-r^2)^2}.$$

Thus from the relation $(\Delta(N+I))^2 = \langle (N+I)^2\rangle - \langle N+I\rangle^2$, we obtain

$$\Delta(N+I) = 2r/(1-r^2)$$

and, since $\Delta(N+I) = \Delta N$, we have

$$\Delta N = 2r/(1-r^2). \qquad (3.13)$$

From (3.7), (3.8), and (3.13) it follows that

$$(\Delta N)(\Delta C) = r/(1+r)$$
$$(\Delta N)(\Delta S) = r/(1-r).$$

4. Y-STATES FOR THE OPERATORS C AND S

Determination

We shall use the method of the representation of the Hardy-Lebesgue space by means of the operator V^*, in order to determine the eigenvalues $\lambda \neq 0$. According to this method[6] λ is an eigenvalue of the operator (3.2) if and only if the equation

$$rzf(z) + (1/z)[f(z) - f(0)] = \lambda f(z) \qquad (4.1)$$

has a solution $f(z) = \sum_{n=1}^{\infty} c_n z^{n-1}$ analytic in the unit disk and satisfying the condition $\sum_{n=1}^{\infty} |c_n|^2 < \infty$.

From (4.1) we have

$$f(z) = f(0)\cdot(rz^2 - \bar{\lambda}z + 1)^{-1}, \qquad (4.2)$$

i.e., λ is an eigenvalue if and only if the roots of the equation

$$rz^2 - \bar{\lambda}z + 1 = 0 \qquad (4.3)$$

lie outside the unit disk, or, equivalently, the roots of the equation

$$W^2 - \bar{\lambda}W + r = 0 \quad (W = 1/z) \qquad (4.4)$$

lie inside the unit disk [note[6] that, for the operator (3.2), $f(0) \neq 0$].

A way to see that the roots of Eq. (4.4) lie inside the unit disk for every λ in the interior of the ellipse $z = r\exp(i\phi) + \exp(-i\phi)$, i.e., for every λ of the form

$$\lambda = r\mu e^{i\phi} + \mu e^{-i\phi}, 0 \leq \mu < 1,\ 0 \leq \phi \leq 2\pi, \qquad (4.5)$$

is the following:

We set $W = \rho\exp(i\vartheta)$ and obtain from (4.4) $\bar{\lambda} = \rho\exp(i\vartheta) + (r/\rho)\exp(-i\vartheta)$ and

$$|\lambda|^2 = \rho^2 + 1/\rho^2 + 2r\cos 2\vartheta. \qquad (4.6)$$

From (4.6) it follows that

$$|\lambda|^2 - 2r\cos 2\vartheta > 0 \qquad (4.7)$$

and

$$\rho^2 = \tfrac{1}{2}[|\lambda|^2 - 2r\cos 2\vartheta$$
$$\pm (|\lambda|^2 - 2r\cos 2\vartheta)^2 - 4r^2)^{1/2}]. \qquad (4.8)$$

Due to (4.7) we observe from (4.8) that the bigger root of Eq. (4.4) decreases with $|\lambda|$, and $|\lambda|$, as follows from (4.5), decreases with μ. For $\mu = 1$ we obtain from (4.4) $w_1 = r\exp(-i\phi)$, $w_2 = \exp(i\phi)$, i.e., since $r < 1$ the bigger root has the value $|w_2| = 1$.

For $\mu < 1$ we must have $|w_1| \leq |w_2| < 1$, i.e., Eq. (4.4) has roots inside the unit disk. Therefore, Eq. (4.3) has roots outside the unit disk for every λ of the form (4.5).

Now let ρ_1, ρ_2 ($|\rho_1| > 1$, $|\rho_2| > 1$) be the roots of Eq. (4.3). Then from (4.2) we obtain

$$f(z) = f(0)\cdot\frac{1}{r(z-\rho_1)\cdot(z-\rho_2)} = f(0)\cdot\frac{1}{(1-z/\rho_1)(1-z/\rho_2)}$$
$$= f(0)\cdot\sum_{n=1}^{\infty}\frac{1}{\rho_1^{n-1}}\cdot z^{n-1}\cdot\sum_{n=1}^{\infty}\frac{1}{\rho_2^{n-1}}\cdot z^{n-1}$$
$$= f(0)\sum_{n=1}^{\infty}\left(\frac{1}{\rho_1^{n-1}} + \frac{1}{\rho_1^{n-2}}\cdot\frac{1}{\rho_2} + \cdots + \frac{1}{\rho_2^{n-1}}\right)z^{n-1}$$
$$= f(0)\sum_{n=1}^{\infty}\frac{1}{\rho_1^{n-1}}(1 + \rho_1^2 r + \rho_2^4 r^2$$
$$+ \cdots + \rho_1^{2(n-1)}r^{n-1})\cdot z^{n-1}$$

$$= f(0)\cdot\sum_{n=1}^{\infty}\frac{1}{\rho_1^{n-1}}\cdot\frac{\rho_1^{2n}r^n-1}{\rho_1^2 r-1}z^{n-1}$$
$$= \frac{f(0)\rho_1}{\rho_1^2 r-1}\cdot\sum_{n=1}^{\infty}\left(\frac{1}{\rho_2^n}-\frac{1}{\rho_1^n}\right)z^{n-1}.$$

The corresponding to $f(z)$ eigenelement of the operator (3.2) is[6]

$$f = A\cdot\sum_{n=1}^{\infty}\left(\frac{1}{\bar\rho_2^n}-\frac{1}{\bar\rho_1^n}\right)e_n, \tag{4.9}$$

where A is the normalization constant and $\bar\rho_1$ and $\bar\rho_2$ are the complex conjugates of ρ_1 and ρ_2.

With some manipulations we find

$$A^2 = \frac{[1+r^2-r^2\cdot(|\rho_1|^2+|\rho_2|^2)]\cdot[1+r^2-r^2(\bar\rho_1\rho_2+\bar\rho_2\rho_1)]}{r^2(1-r^2)\cdot(|\rho_1|^2+|\rho_2|^2-\bar\rho_1\rho_2-\bar\rho_2\rho_1)}. \tag{4.10}$$

We summarize here our results:

For every λ in the interior of the ellipse

$$z = r\cdot e^{i\phi}+e^{-i\phi}, \quad r<1,\ 0\le\phi<2\pi,$$

i.e., for every λ in the form

$$\lambda = r\mu e^{i\phi}+\mu e^{-i\phi}, \quad 0\le\mu<1,$$

we have two roots of the Eq. (4.3) $\rho_1(r,\mu,\phi)$ and $\rho_2(r,\mu,\phi)$ such that $|\rho_1|>1$ and $|\rho_2|>1$ and a normalizable eigenstate of the operator (3.2) $f(r,\mu,\phi)$ given from (4.9). This eigenstate belongs for $\mu\neq 0$ to the class Y.

Note that for every eigenvalue we have only one eigenstate. This statement can be proved[6] easily, for tridiagonal operators, as is the case of the operator (3.2).

The class of Y-states is very large in the sense that only for $r=0$ we obtain from (4.2) the well-known[6] f_z-states, which form a complete system of states and which we shall study below.

f_z-States

We call the normalized eigenstates of the operator $V^* f_z$-states. These are the following:

$$f_z = (1-|z|^2)^{1/2}\cdot\sum_{n=1}^{\infty}z^{n-1}e_n, \quad |z|<1. \tag{4.11}$$

The expectation values of C and S in the states (4.11) can be found from the general formulas (2.8) and (2.9). We get

$$\langle C\rangle = \mu\cos\phi, \quad \langle S\rangle = -\mu\sin\phi, \quad \mu = |z|. \tag{4.12}$$

The mean deviations of C and S are obtained as follows:

From the eigenvalue equation $V^* f_z = z f Z$ we have

$$(V+V^*)f_z = zf_z + Vf_z$$

or

$$2\cdot Cf_z - zf_z = Vf_z$$

and, since V is an isometry,

$$\|2Cf_z - zf_z\| = \|Vf_z\| = 1. \tag{4.13}$$

From (4.13) we obtain

$$4\|Cf_z\|^2 - 4|z|\langle C\rangle\cos\phi + |z|^2 = 1$$

and due to (4.12)

$$\|Cf_z\|^2 = (1-|z|^2)/4 + \langle C\rangle^2. \tag{4.14}$$

From (4.14) it follows that

$$(\Delta C)^2 = (1-|z|^2)/4. \tag{4.15}$$

In the same way we find

$$(\Delta S)^2 = (1-|z|^2)/4. \tag{4.16}$$

From (4.15) and (4.16) we have

$$(\Delta C)\cdot(\Delta S) = (1-|z|^2)/4. \tag{4.17}$$

Observe that $(I-P)f_z = (1-|z|^2)^{1/2}\cdot e_1$ and

$$\langle (I-P)/2\rangle = (1-|z|^2)/2. \tag{4.18}$$

From (4.17) and (4.18) we obtain the minimum uncertainty relation for the operators C and S.

The Action of the Number Operator N on the States f_z

From (4.11) we have

$$\langle N\rangle = (Nf_z, f_z)$$
$$= (1-|z|^2)\cdot|z|^2(1+2|z|^2+3|z|^4+\cdots)$$
$$= (1-|z|^2)\cdot|z|^2\cdot\left(\frac{1}{1-|z|^2}+\frac{|z|^2}{1-|z|^2}+\frac{|z|^4}{1-|z|^2}+\cdots\right),$$

i.e., $$\langle N\rangle = \frac{|z|^2}{1-|z|^2}. \tag{4.19}$$

And

$$\|Nf_z\|^2 = (1-|z|^2)\cdot|z|^2\cdot(1+2^2|z|^2+3^2|z|^4+\cdots).$$

Setting $|z|^2 = w$, we obtain

$$\|Nf_z\|^2 = (1-w)\cdot w\frac{d}{dw}\left[w\cdot\frac{d}{dw}\left(\frac{w}{1-w}\right)\right] = \frac{w(w+1)}{(1-w)^2}.$$

Thus

$$\langle N^2\rangle = |z|^2(1+|z|^2)/(1-|z|^2)^2. \tag{4.20}$$

From (4.19) and (4.20) it follows that

$$\Delta N = |z|/(1-|z|^2).$$

It is interesting to note here that

$$\langle C^2 + S^2\rangle \to \langle C\rangle^2 + \langle S\rangle^2 \to 1 \quad \text{as } |z| \to 1$$

and that $|z| \to 1$ means, due to (4.19), that $\langle N\rangle \to \infty$.

Moreover, since in case $\mu = 1$ the spectrum of the operator (3.2) is purely continuous,[4] we observe from (4.12) that there exist no normalizable states, minimizing the uncertainty product $(\Delta C)\cdot(\Delta S)$ and satisfying the relation

$$\langle C\rangle^2 + \langle S\rangle^2 = 1.$$

For minimal uncertainty sequences for which we have $\mu = 1$ the expectation values of C and S are not given exactly from the formulas (4.12).

From (4.15) and (4.16) we observe that both C and S in the f_z-states have the same mean deviation. As we shall see later, this is a characteristic property of the f_z-states.

Moreover, it is interesting to note that the f_z-states are obtained from one monoparametric class of invertible operators acting on the ground state e_1, i.e., we can easily verify that

$$(I - zV)f_z = (1 - |z|^2)^{1/2}\cdot e_1, \quad z : |z| < 1,$$

or

$$f_z = (1 - |z|^2)^{1/2}\cdot(I - zV)^{-1}e_1,$$

where V is the unilateral shift operator.

Physical Quantities in Y-states

Let the eigenvalue equation

$$(rV + V^*)f = \lambda f, \quad \lambda = r\mu e^{i\phi} + \mu e^{-i\phi}, \quad 0 < \mu < 1,\ 0 \leq \phi \leq 2\pi. \tag{4.21}$$

The expectation values of C and S in the state f are given from the general formulas (2.8) and (2.9).

From (4.21) we have

$$2Cf - \lambda f = (1 - r)Vf \tag{4.22}$$

and from (4.22), using the same technique as before, we obtain

$$\langle C^2\rangle = [(1 - r)^2 - |\lambda|^2]/4 + \mu^2(1 + r)\cos^2\phi \tag{4.23}$$

and

$$\langle S^2\rangle = [(1 + r)^2 - |\lambda|^2]/4 + \mu^2(1 - r)\sin^2\phi, \tag{4.24}$$

where

$$|\lambda|^2 = \mu^2(1 + r^2) + 2\mu^2 r\cos 2\phi. \tag{4.25}$$

From (4.23) and (4.24) we obtain the mean square deviations of C and S:

$$(\Delta C)^2 = [(1 - r)^2 - |\lambda|^2]/4 + \mu^2 r\cos^2\phi,$$

$$(\Delta S)^2 = [(1 + r)^2 - |\lambda|^2]/4 - \mu^2 r\sin^2\phi,$$

or because of the relation (4.25)

$$(\Delta C)^2 = (1 - r)^2(1 - \mu^2)/4, \tag{4.26}$$

$$(\Delta S)^2 = [(1 + r)^2\cdot(1 - \mu^2)]/4. \tag{4.27}$$

Also we obtain

$$\langle C^2 + S^2\rangle = [\mu^2(1 - r^2) + 1 + r^2]/2 \to 1 \quad \text{for } \mu \to 1.$$

From (4.26) and (4.27) it follows

$$(\Delta S)^2 - (\Delta C)^2 = r(1 - \mu^2), \tag{4.28}$$

i.e., for $r \neq 0$ only in the limit $\mu \to 1$, C and S have the same mean square deviations.

The Action of the Number Operator on the States $f(r, \mu, \phi)$ with $r \neq 0, \mu \neq 0$

We set $f_1 = A\sum_{n=1}^{\infty}(1/\bar{\rho}_2)^n e_n, f_2 = A\sum_{n=1}^{\infty}(1/\bar{\rho}_1)^n e_n$. Then, from (4.9), $f(r, \mu, \phi) = f_1 - f_2$. Starting from the identities

$$\sum_{n=1}^{\infty}(n - 1)x^n = \frac{x^2}{(1 - x)^2}, \quad |x| < 1,$$

$$\sum_{n=1}^{\infty}(n - 1)^2 x^n = \frac{x^2(1 + x)}{(1 - x)^3}, \quad |x| < 1,$$

we determine the quantities

$$(Nf_1, f_1) = A^2\cdot(|\rho_2|^2 - 1)^{-2}, \tag{4.29}$$

$$(Nf_2, f_2) = A^2\cdot(|\rho_1|^2 - 1)^{-2}, \tag{4.30}$$

$$(Nf_1, f_2) = A^2\cdot(\bar{\rho}_2\rho_1 - 1)^{-2} \tag{4.31}$$

$$(Nf_2, f_1) = A^2\cdot(\rho_2\bar{\rho}_1 - 1)^{-2}, \tag{4.32}$$

$$(N^2 f_1, f_1) = A^2\cdot[(|\rho_2|^2 + 1)/(|\rho_2|^2 - 1)^3], \tag{4.33}$$

$$(N^2 f_2, f_2) = A^2\cdot[(|\rho_1|^2 + 1)/(|\rho_1|^2 - 1)^3], \tag{4.34}$$

$$(N^2 f_1, f_2) = A^2\cdot[(\bar{\rho}_2\rho_1 + 1)/(\bar{\rho}_2\rho_1 - 1)^3], \tag{4.35}$$

$$(N^2 f_2, f_1) = A^2\cdot[(\rho_2\bar{\rho}_1 + 1)/(\rho_2\bar{\rho}_1 - 1)^3]. \tag{4.36}$$

From (4.29), (4.30), (4.31), and (4.32) we determine (Nf, f) and from (4.33), (4.34), (4.35), and (4.36) we determine $(N^2 f, f)$.

A Characteristic Property of the States f_z

From (4.9) we have all the normalizable minimal uncertainty states for the operators C and S. For $\lambda = 0$, i.e., for $\rho_1 = ir^{-1/2}$ and $\rho_2 = -ir^{-1/2}$ we obtain the Y_0-states and for $\lambda \neq 0$ and $r = 0$ the class of f_z-states. If we require that in the minimal uncertainty states both C and S must have the same mean deviation, then we obtain from (4.28) $r = 0$, i.e., we obtain the states f_z. Thus the equality of the mean square deviations is a characteristic property of the states f_z. The so important for its applications[7] class of f_z-states can be also physically defined as follows: f_z-states are the minimal uncertainty states for the oscillator phase operators C and S, in which both C and S have the same mean square deviation.

5. MINIMAL UNCERTAINTY SEQUENCES FOR THE OPERATORS C AND S

A sequence $\{f_n\}$, $n = 1, 2, \cdots$, is called a minimal uncertainty sequence if it corresponds to the continuous spectrum of the operator (3.2), i.e., f_n has the property $\lim\|(T - \lambda I)f_n\| = 0$ as $n \to \infty$.

In general, we have found in Sec. 2 that the expectation values of A and B tend to the values $\mu\cos\phi$ and $-\mu\sin\phi$, where μ is given from the relation $\lambda = r\mu\exp(i\phi) + \mu\exp(-i\phi)$ and λ belongs to the continuous spectrum of the operator (2.2).

The continuous spectrum of the operator (3.2) is obtained for $\mu = 1$. Thus, the expectation values of C and S in the sequence of states f_n tend to the values $\cos\phi$ and $-\sin\phi$, $0 \leq \phi \leq 2\pi$.

Setting $(rV + V^* - \lambda I)f_n = \xi_n$, where $\lim\|\xi_n\| = 0$ as $n \to \infty$ we can determine as in Sec. 4 the mean square deviation of C and S in the sequence of states f_n. These mean deviations tend to the values given from (4.26) and (4.27) as $n \to \infty$. But in this case, we have $\mu = 1$ and therefore, as follows from (4.26) and (4.27), we have

$$(\Delta C)^2 \to 0, \qquad (\Delta S)^2 \to 0.$$

Thus, the minimal uncertainty sequences for the operator (3.2) have the property that the expectation values of C and S tend to the limit $\cos\phi$ and $-\sin\phi$ and the mean square deviations tend to zero.

From a general theorem, proved in Ref. 4, it follows that the operators (3.2) and V have the same minimal uncertainty sequences. Setting $f_n = A\cdot\sum_{k=1}^{\infty} a_k e_k$, we can determine the coefficients a_k and the constant A in order that f_n be a minimal uncertainty sequence for the operator V, corresponding to the continuous spectrum of V, $\lambda = \exp(-i\phi)$, $0 \leq \phi \leq 2\pi$. We find $a_k = \exp(ik\phi)$ and $A = n^{-1/2}$. Thus

$$f_n = n^{-1/2}\cdot\sum_{k=1}^{n} e^{ik\phi} e_k. \tag{5.1}$$

It is not difficult to find the physical quantities in the sequence of states (5.1). A straightforward calculation gives

$$\begin{aligned}
\langle C\rangle_n &= (1 - 1/n)\cdot\cos\phi,\\
\langle S\rangle_n &= -(1 - 1/n)\cdot\sin\phi,\\
\langle C^2\rangle_n &= \cos^2\phi + (3 - 8\cos^2\phi)/4n,\\
\langle S^2\rangle_n &= \sin^2\phi + (3 - 8\sin^2\phi)/4n,\\
(\Delta C)^2 &= (1/n)\cdot[\tfrac{3}{4} - (\cos^2\phi)/n],\\
(\Delta S)^2 &= (1/n)\cdot[\tfrac{3}{4} - (\sin^2\phi)/n],\\
\langle N\rangle_n &= (n - 1)/2,\\
\langle N^2\rangle_n &= (n - 1)(2n - 1)/6,\\
(\Delta N)^2 &= (n^2 - 1)/12.
\end{aligned}$$

APPENDIX A: NORMALIZABLE MINIMAL UNCERTAINTY STATES

Let A, B be two self-adjoint operators, satisfying the commutation relation $AB - BA = iC$. Let $D(A), D(B)$, and $D(C)$ be the definition domains of A, B, and C, respectively. Assume that $A_0 f \neq 0$, $B_0 f \neq 0$, $\langle C\rangle = (Cf,f) \neq 0$, where $A_0 = A - \langle A\rangle$, $B_0 = B - \langle B\rangle$. Then for every real γ we have[8]

$$\|(A_0 \pm i\gamma B_0)f\|^2 = \|A_0 f\|^2 + \gamma^2\|B_0 f\|^2 \mp \gamma\langle C\rangle. \tag{A1}$$

For $\gamma = \frac{1}{2}\langle C\rangle\cdot\|B_0 f\|^2$ we have from (A1)

$$\left\|\left(A_0 + i\frac{\langle C\rangle}{2\|B_0 f\|^2}B_0\right)f\right\|^2 = \frac{4\|A_0 f\|^2\cdot\|B_0 f\|^2 - \langle C\rangle^2}{4\|B_0 f\|^2}, \tag{A2}$$

From (A2) we obtain the uncertainty relation for A and B

$$4\|A_0 f\|^2\cdot\|B_0 f\|^2 - \langle C\rangle^2 \geq 0. \tag{A3}$$

Assume that there exists an element f such that the equality in the relation (A3) holds. Then

$$A_0 f + i\frac{\langle C\rangle}{2\|B_0 f\|^2}B_0 f = 0$$

or

$$Af + i\frac{\langle C\rangle}{2\|B_0 f\|^2}Bf = \left(\langle A\rangle + i\frac{\langle C\rangle}{2\|B_0 f\|^2}\langle B\rangle\right)f,$$

i.e., (since $Af \neq \langle A\rangle f$) there exists a real $\gamma \neq 0$ such that f is an eigenelement of the operator

$$L = A + i\gamma B. \tag{A4}$$

If, conversely, there exists a real $\gamma \neq 0$ such that f be an eigenelement of the operator (A4) with the eigenvalue K, then we must have $K = \langle A\rangle + i\gamma\langle B\rangle$, i.e.,

$$A_0 f + i\gamma B_0 f = 0. \tag{A5}$$

From (A5) we obtain $\gamma = \|A_0 f\|/\|B_0 f\|$ in case $\gamma > 0$ and $\gamma = -\|A_0 f\|/\|B_0 f\|$ in case $\gamma < 0$. From the identity (A1) it follows always that

$$2\|A_0 f\|\cdot\|B_0 f\| = |\langle C\rangle|,$$

i.e., the equality in the relation (A3) holds. Thus we conclude that the state f minimize the relation (A3) if and only if there exists a real $\gamma \neq 0$ such that f is an eigenelement of the operator (A4).

Remark: Since γ can take positive and negative values, the residual spectrum of the operator L for $\gamma > 0$ is the point spectrum of L for $\gamma < 0$. This is the role of the residual spectrum of the operator L. The continuous spectrum of L plays also an important role, which we shall examine below.

APPENDIX B: MINIMAL UNCERTAINTY SEQUENCES OF STATES

Let K belongs to the continuous spectrum of the operator (A4) for a real $\gamma \neq 0$. This means that there exists a normalizable sequence f_n such that

$$(A + i\gamma B)f_n = g_n \tag{B1}$$

and

$$\lim\|g_n\| = 0 \quad \text{as} \quad n \to \infty. \tag{B2}$$

We call the sequence f_n minimal uncertainty sequence for the operator L.

From (B1) we obtain

$$A_0^{(n)} f_n + i\gamma B_0^{(n)} f_n = \xi_n, \tag{B3}$$

where

$$\begin{aligned}
\xi_n &= g_n - (g_n, f_n)f_n,\\
A_0^{(n)} &= A - (Af_n, f_n),\\
B_0^{(n)} &= B - (Bf_n, f_n).
\end{aligned} \tag{B4}$$

From (B4) we have

$$\|\xi_n\| \leq 2\|g_n\| \to 0 \quad \text{as} \quad n \to \infty \tag{B5}$$

and from (B3)

$$\gamma = \|\xi_n - A_0^{(n)} f_n\| / \|B_0^{(n)} f_n\|.$$

[We consider the case $\gamma > 0$, i.e., due to (A1), $\langle C\rangle_n = (Cf_n, f_n) > 0$.]

If $\lim\|A_0^{(n)} f_n\| = 0$ or $\lim\|B_0^{(n)} f_n\| = 0$ as $n \to \infty$, we say that the sequence f_n tends to minimize the relation (A3) in the sense that both $\langle C\rangle_n$ and $2\|A_0^{(n)} f_n\| \cdot \|B_0^{(n)} f_n\|$ tend to zero.

If $\lim\|A_0^{(n)} f_n\| \neq 0$ and $\lim\|B_0^{(n)} f_n\| \neq 0$ as $n \to \infty$, we may assume that from a large n on $\|\xi_n\| < \|A_0^{(n)} f_n\|$.

Thus we can set

$$\|A_0^{(n)} f_n\| - \|\xi_n\| \leq \|\xi_n - A_0^{(n)} f_n\| \leq \|\xi_n\| + \|A_0^{(n)} f_n\|. \tag{B6}$$

On the other hand, equality (A1) gives

$$\|\xi_n\|^2 = \|A_0^{(n)} f_n\|^2 + \|\xi_n - A_0^{(n)} f_n\|^2 - \frac{\|\xi_n - A_0^{(n)} f_n\|}{\|B_0^{(n)} f_n\|}\langle C\rangle_n,$$

$$\langle C\rangle_n > 0. \tag{B7}$$

Due to (B6) we obtain from (B7)

$$\|\xi_n\|^2 \geq \|A_0^{(n)} f_n\|^2 + (\|A_0^{(n)} f_n\| - \|\xi_n\|)^2 - \frac{\|\xi_n\| + \|A_0^{(n)} f_n\|}{\|B_0^{(n)} f_n\|}\langle C\rangle_n$$

or

$$2\|A_0^{(n)} f_n\| \cdot \|B_0^{(n)} f_n\| - \langle C\rangle_n \leq \|\xi_n\|^2 \cdot \frac{2\|A_0^{(n)} f_n\| \cdot \|B_0^{(n)} f_n\| + \langle C\rangle_n}{\|A_0^{(n)} f_n\|},$$

i.e., due to (B5) the sequence f_n tends to minimize the relation (A3).

1 R. Jackiw, J. Math. Phys. **9**, 339 (1968).
2 E. C. Lerner, H. W. Huang, and G. E. Walters, J. Math. Phys. **11**, 1679 (1970), p. 1683.
3 P. Carruthers and M. Nieto, Rev. Mod. Phys. **40**, 411 (1968), p. 435.
4 E. K. Ifantis, J. Math. Phys. **12**, 2512 (1971).
5 E. K. Ifantis, J. Math. Phys. **12**, 1021 (1971).
6 E. K. Ifantis, J. Math. Phys. **11**, 3138 (1970).
7 E. K. Ifantis, J. Math. Phys. **12**, 1961 (1971).
8 From the identity (A1) it follows that if $C \geq 0$ ($C \leq 0$) and $\gamma < 0$ ($\gamma > 0$), then the point spectrum of $A + i\gamma B$ is empty. In particular, if $C = I$ and $\gamma < 0$, the point spectrum of $A + i\gamma B$ is empty. This is the case of the position and momentum operators Q and P.

Erratum: Weight Lowering Operators and the Multiplicity-Free Isoscalar Factors for the Group R_5

[J. Math. Phys. **12**, 594 (1971)]

S. J. Ališauskas and A. P. Jucys

Institute of Physics and Mathematics of the Academy of Sciences of Lithuanian SSR, Vilnius, USSR

(Received 24 November 1971)

The last entry on the right of Eq. (30b) should be $-K_1 - 2$. In the next-to-last line of the right-hand side of Eq. (38), the subscript 1 was omitted. The symbol $\langle K0\rangle$ on the left in Eq. (43) should be $\langle k\ 0\rangle$. The left-hand side of Eq. (44) should read

$$\begin{bmatrix} \langle K\Lambda\rangle & \langle k\ 0\rangle & \langle \lambda\lambda\rangle \\ ij & ij & 00 \end{bmatrix}.$$

The factor $(2I_2 - \beta)!$ in the numerator on the right of Eq. (27) was omitted. Instead of the last y in the denominator of the sum in Eq. (A. 2) should be z.

Erratum: Solution of the Schrödinger Equation in the Hardy-Lebesgue Space

[J. Math. Phys. **12**, 1961 (1971)]

E. K. Ifantis

Nuclear Research Center "Democritos, Aghia Paraskevi Attikis, Athens, Greece

(Received 22 November 1971)

Proposition 4 is not correct. The error is in the counter example. The eigenelement $x = \sum_{L=0}^{\infty} x_L e_L$ was normalized by taking $x_0 = 1$ while in this case $x_0 = 0$. In Proposition 2 the operator $\phi(V)$ is bounded because $\phi(z) \cdot f(z) \in \mathcal{K}_2(\mathfrak{D})$ for every $f(z) \in \mathcal{K}_2(\mathfrak{D})$. The statement $\phi(z) \cdot f(z) \in \mathcal{K}_2(\mathfrak{D})$ follows easily from the integral condition that characterizes the elements of the space $\mathcal{K}_2(\mathfrak{D})$ (See Ref. 1).

1 E. K. Ifantis, J. Math. Phys. **11**, 3138 (1970).

From (B4) we have

$$\|\xi_n\| \leq 2\|g_n\| \to 0 \quad \text{as} \quad n \to \infty \tag{B5}$$

and from (B3)

$$\gamma = \|\xi_n - A_0^{(n)} f_n\| / \|B_0^{(n)} f_n\|.$$

[We consider the case $\gamma > 0$, i.e., due to (A1), $\langle C\rangle_n = (Cf_n, f_n) > 0$.]

If $\lim\|A_0^{(n)} f_n\| = 0$ or $\lim\|B_0^{(n)} f_n\| = 0$ as $n \to \infty$, we say that the sequence f_n tends to minimize the relation (A3) in the sense that both $\langle C\rangle_n$ and $2\|A_0^{(n)} f_n\| \cdot \|B_0^{(n)} f_n\|$ tend to zero.

If $\lim\|A_0^{(n)} f_n\| \neq 0$ and $\lim\|B_0^{(n)} f_n\| \neq 0$ as $n \to \infty$, we may assume that from a large n on $\|\xi_n\| < \|A_0^{(n)} f_n\|$.

Thus we can set

$$\|A_0^{(n)} f_n\| - \|\xi_n\| \leq \|\xi_n - A_0^{(n)} f_n\| \leq \|\xi_n\| + \|A_0^{(n)} f_n\|. \tag{B6}$$

On the other hand, equality (A1) gives

$$\|\xi_n\|^2 = \|A_0^{(n)} f_n\|^2 + \|\xi_n - A_0^{(n)} f_n\|^2 - \frac{\|\xi_n - A_0^{(n)} f_n\|}{\|B_0^{(n)} f_n\|}\langle C\rangle_n,$$

$$\langle C\rangle_n > 0. \tag{B7}$$

Due to (B6) we obtain from (B7)

$$\|\xi_n\|^2 \geq \|A_0^{(n)} f_n\|^2 + (\|A_0^{(n)} f_n\| - \|\xi_n\|)^2 - \frac{\|\xi_n\| + \|A_0^{(n)} f_n\|}{\|B_0^{(n)} f_n\|}\langle C\rangle_n$$

or

$$2\|A_0^{(n)} f_n\| \cdot \|B_0^{(n)} f_n\| - \langle C\rangle_n \leq \|\xi_n\|^2 \cdot \frac{2\|A_0^{(n)} f_n\| \cdot \|B_0^{(n)} f_n\| + \langle C\rangle_n}{\|A_0^{(n)} f_n\|},$$

i.e., due to (B5) the sequence f_n tends to minimize the relation (A3).

1 R. Jackiw, J. Math. Phys. **9**, 339 (1968).
2 E. C. Lerner, H. W. Huang, and G. E. Walters, J. Math. Phys. **11**, 1679 (1970), p. 1683.
3 P. Carruthers and M. Nieto, Rev. Mod. Phys. **40**, 411 (1968), p. 435.
4 E. K. Ifantis, J. Math. Phys. **12**, 2512 (1971).
5 E. K. Ifantis, J. Math. Phys. **12**, 1021 (1971).
6 E. K. Ifantis, J. Math. Phys. **11**, 3138 (1970).
7 E. K. Ifantis, J. Math. Phys. **12**, 1961 (1971).
8 From the identity (A1) it follows that if $C \geq 0$ ($C \leq 0$) and $\gamma < 0$ ($\gamma > 0$), then the point spectrum of $A + i\gamma B$ is empty. In particular, if $C = I$ and $\gamma < 0$, the point spectrum of $A + i\gamma B$ is empty. This is the case of the position and momentum operators Q and P.

Erratum: Weight Lowering Operators and the Multiplicity-Free Isoscalar Factors for the Group R_5

[J. Math. Phys. **12**, 594 (1971)]

S. J. Ališauskas and A. P. Jucys

Institute of Physics and Mathematics of the Academy of Sciences of Lithuanian SSR, Vilnius, USSR

(Received 24 November 1971)

The last entry on the right of Eq. (30b) should be $-K_1 - 2$. In the next-to-last line of the right-hand side of Eq. (38), the subscript 1 was omitted. The symbol $\langle K0\rangle$ on the left in Eq. (43) should be $\langle k\ 0\rangle$. The left-hand side of Eq. (44) should read

$$\begin{bmatrix} \langle K\Lambda\rangle & \langle k\ 0\rangle & \langle \lambda\lambda\rangle \\ ij & ij & 00 \end{bmatrix}.$$

The factor $(2I_2 - \beta)!$ in the numerator on the right of Eq. (27) was omitted. Instead of the last y in the denominator of the sum in Eq. (A. 2) should be z.

Erratum: Solution of the Schrödinger Equation in the Hardy-Lebesgue Space

[J. Math. Phys. **12**, 1961 (1971)]

E. K. Ifantis

Nuclear Research Center "Democritos, Aghia Paraskevi Attikis, Athens, Greece

(Received 22 November 1971)

Proposition 4 is not correct. The error is in the counter example. The eigenelement $x = \sum_{L=0}^{\infty} x_L e_L$ was normalized by taking $x_0 = 1$ while in this case $x_0 = 0$. In Proposition 2 the operator $\phi(V)$ is bounded because $\phi(z) \cdot f(z) \in \mathcal{K}_2(\mathfrak{D})$ for every $f(z) \in \mathcal{K}_2(\mathfrak{D})$. The statement $\phi(z) \cdot f(z) \in \mathcal{K}_2(\mathfrak{D})$ follows easily from the integral condition that characterizes the elements of the space $\mathcal{K}_2(\mathfrak{D})$ (See Ref. 1).

1 E. K. Ifantis, J. Math. Phys. **11**, 3138 (1970).