A double angle sum formula for Gegenbauer polynomials

John V. Shebalin
Naval Surface Weapons Center, Dahlgren, Virginia

(Received 4 December 1978; revised manuscript received 15 March 1979)

A formula is given by which a Gegenbauer polynomial whose argument is the cosine of twice an angle can be equated with a finite, alternating series of products of Gegenbauer polynomials whose arguments are all cosines of the angle.

Gegenbauer polynomials are a particularly useful set of functions in mathematical physics. They include, as specific cases, the Legendre and Chebyshev polynomials, and are related to other special functions, such as the associated Legendre functions and the Hermite and Jacobi polynomials. Definitions and results concerning Gegenbauer polynomials appear in numerous sources and have been collected and compiled in various invaluable reference texts.[1-4]

Among the known results concerning the Gegenbauer polynomials $C_n^\lambda(x)$ are the following ($\lambda \neq 0$; $m = 0, 1$):

$$C_{2n+m}^\lambda(x) = \frac{2^{n+m} n! \Gamma(\lambda + n + m)}{(2n+m)! \Gamma(\lambda)} \times x^m P_n^{\lambda - 1/2, m - 1/2}(2x^2 - 1) \tag{1}$$

and

$$C_n^\lambda(x) = (-1)^n \frac{\Gamma(n + 2\lambda)}{n! \Gamma(2\lambda)} \times {}_2F_1\left(-n, n + 2\lambda; \lambda + \frac{1}{2}; \frac{1+x}{2}\right), \tag{2}$$

where $P_n^{\alpha,\beta}(y)$ is a Jacobi polynomial in y, ${}_2F_1(a,b;c;y)$ is a hypergeometric function, and the $\Gamma(y)$ are gamma functions.

If the substitution $x = \cos\frac{1}{2}\theta$ is made in (1) and $x = \cos 2\theta$ is made in (2), then these equations are seen to be "half-angle" and "double-angle" formulas, respectively. They represent Gegenbauer polynomials of the corresponding argument by other special functions which are functions of $\cos\theta$. It may prove useful, however, to have half-angle and double-angle formulas which express Gegenbauer polynomials in terms of themselves. Although no particularly elegant formula presents itself for the half-angle case, it is possible to obtain such a result for the double-angle case, as will now be done.

In order to proceed, the following results will be needed. First, a generating function for Gegenbauer polynomials,

$$(1 - 2xz + z^2)^{-\lambda} = \sum_{n=0}^{\infty} C_n^\lambda(x) z^n, \quad |z| < 1, \quad \lambda \neq 0, \tag{3}$$

and second, a statement of parity,

$$C_n^\lambda(-x) = (-1)^n C_n^\lambda(x). \tag{4}$$

An essential decomposition

$$(1 - 2\cos 2\theta z^2 + z^4) = (1 - 2\cos\theta z + z^2)(1 + 2\cos\theta z + z^2) \tag{5}$$

can easily be verified by using the trigonometric identity $\cos 2\theta = 2\cos^2\theta - 1$.

Using (3) and (5), the following can be written,

$$\sum_{n=0}^{\infty} C_n^\lambda(\cos 2\theta) z^{2n} = \sum_{l=0}^{\infty} \sum_{k=0}^{\infty} C_l^\lambda(\cos\theta) C_k^\lambda(\cos\theta) z^{l+k}. \tag{6}$$

The product of series on the right-hand side of (6) can be rewritten as a "Cauchy product"[5] so that (6) becomes

$$\sum_{n=0}^{\infty} C_n^\lambda(\cos\theta) z^{2n} = \sum_{m=0}^{\infty} \sum_{k=0}^{m} C_{m-k}^\lambda(\cos\theta) C_k^\lambda(-\cos\theta) z^m. \tag{7}$$

Equating coefficients of like powers of z in (7) and using (4) gives, for $m = 2n$

$$C_n^\lambda(\cos 2\theta) = \sum_{k=0}^{2n} (-1)^k C_k^\lambda(\cos\theta) C_{2n-k}^\lambda(\cos\theta), \tag{8}$$

and for $m = 2n + 1$

$$0 = \sum_{k=0}^{2n+1} (-1)^k C_k^\lambda(\cos\theta) C_{2n+1-k}^\lambda(\cos\theta). \tag{9}$$

The result (9) is not particularly noteworthy since its right-hand side is identically zero (the first half of the summation cancels the second half); (8), however, is a new and perhaps more notable result: a double-angle sum formula for Gegenbauer polynomials ($\lambda \neq 0$).

In closing, I would like to thank the referee for his useful comments.

[1] M. Abramowitz and I.A. Stegun, Editors, *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables*, Applied Mathematics Series 55 (National Bureau of Standards, U.S. Government Printing Office, Washington, 1970), Chap. 22.
[2] A. Erdélyi, *et al*, *Higher Transcendental Functions* (McGraw-Hill, New York, 1953), Vol. 2, Chap. 10.
[3] I.S. Gradshteyn and I.M. Ryzhik, *Tables of Integrals, Series, and Products* (Academic, New York, 1965), Sec. 8.93.
[4] W. Magnus, F. Oberhettinger, and R.P. Soni, *Formulas and Theorems for the Special Functions of Mathematical Physics* (Springer-Verlag, New York, 1966), Sec. 5.3.
[5] W. Rudin, *Principles of Mathematical Analysis*, 2nd ed. (McGraw-Hill, New York, 1964), pp. 63–4.

 0022-2488/79/091837-01$01.00 © 1979 American Institute of Physics

On the applicability of the variation of action method to some one-field solitons

E. W. Laedke and K. H. Spatschek
Fachbereich Physik Universität Essen D-4300 Essen, Federal Republic of Germany

(Received 26 February 1979)

The applicability of the widely used method of variation of action is discussed by investigating the transverse stability of the kink solution of the nonlinear Schrödinger equation and the bell solution of the Kadomtsev–Petviashvili equation. An exact calculation shows that neither instability nor stability can be predicted correctly by the variation of action method. The reasons for that defect are discussed and the correct stability regions are presented.

I. INTRODUCTION

In many areas of nonlinear physics [1–3] field equations occur which allow solitary wave solutions. Some of these systems are completely integrable and the initial value problems have been solved by the inverse scattering technique. [4] In one dimension, emerging from multisoliton collisions the solitary wave solutions have the same shapes and velocities with which they entered, thus satisfying the requirements for considering them as solitons (kinks). [1]

However, the so far undivided picture of stability breaks down when a second space dimension is allowed for. Calculations show that the soliton solutions of the nonlinear Schrödinger (NS) equation are transversely (perpendicular to the soliton motion) unstable [5] and that the kink solution of the sine-Gordon equation is transversely stable. [1] For two-dimensional Korteweg–de Vries (KdV) equations, the stability behavior depends on parameters which discriminate between different types of generalizations. [6] Beyond that, the situation is not very clear for other solutions or different field equations. For example, one is tempted to conjecture that the shock solutions of the NS equation are transversely stable, in agreement with the claim for the Higgs field equation. [7] The latter statement as well as calculations for other field equations are based on a widely used method, the so called variation of action method (VAM). [8]

In this paper, we discuss the applicability of this method by investigating the NS and KdV equation, which describe a broad class of nonlinear physical systems. [9]

We show that some previous proofs of transverse stability of kink solutions were incomplete. In the case of the NS equation, instead of stability instability occurs, in agreement with the small k prediction by Zakharov and Rubenchik. [5] On the other hand, the VAM can predict instability in some region where a more exact calculation proves stability. The latter failure of the VAM will be demonstrated for the example of the Kadomtsev–Petviashvili (KP) form of the KdV equation. We shall therefore conclude that neither instability nor stability can be precisely predicted by the VAM.

The paper is organized as follows: In Sec. II, we review the VAM and summarize the results for the kink solution of the NS equation which follow by that method. In Sec. III, we reconsider the transverse stability of the kink solutions of the NS equation. Deriving a general energy principle we prove instability of the kink solution. In Sec. IV we reinvestigate the transverse instability of the KP equation and compare the results with the predictions by the VAM. [10] We explicitely show that previously unstable regions are actually stable. Possible generalizations for other field equations are discussed in Sec. V.

II. THE VARIATION OF ACTION METHOD

To demonstrate the VAM [8] we consider the NS equation

$$i\psi_t + \nabla^2\psi - \psi^2\psi^* = 0, \tag{1}$$

and discuss the transverse stability of the plane kink-type solution

$$\psi = G(x)\exp(-2i\eta^2 t), \tag{2}$$

where

$$G(x) = \sqrt{2}\eta\tanh\eta x. \tag{3}$$

The longitudinal stability of this solution has been investigated previously. [11]

For the NS equation, the variation of action can be written in the form

$$\delta S = \delta\int dx\,dy\,dt\,(\mathscr{L} - \mathscr{L}_\infty) = 0. \tag{4}$$

Here, the vacuum has been subtracted and the Lagrangian $\mathscr{L}$ is given by

$$\mathscr{L} = -[\tfrac{1}{2}i(\psi\psi_t^* - \psi_t\psi^*) + \nabla\psi\cdot\nabla\psi^* + \tfrac{1}{2}(\psi\psi^*)^2]. \tag{5}$$

In the VAM, the test functions are constructed from the solitary wave solution by perturbing its shape and phase.

In the present case, the appropriate choice, being identical to that used by Makhankov, [7] is

$$\psi = A\tanh Bx\exp(i\Phi). \tag{6}$$

Here, the coefficients A, B, and Φ depend on the transverse coordinate y and the time t; the unperturbed values are $A_0 = (2)^{1/2}\eta$, $B_0 = \eta$, and $\Phi_0 = -2\eta^2$.

Inserting Eq. (6) into the Lagrangian (5) and performing the x-integration in the action integral, we obtain

$$S = 2\int dy\,dt\Big[A^2(\Phi_t + \Phi_y^2)/B + \tfrac{2}{3}A^4/B - \tfrac{2}{3}A^2B + A_y^2/B + \tfrac{1}{2}(A^2)_y(1/B)_y - IA^2B_y^2/B^3\Big], \tag{7}$$

where $I = \int_0^\infty dz\,z^2\operatorname{sech}^4 z, > 0.$,

Taking the variations with respect to Φ, A, and B we obtain the Euler equtions

$$(A^2/B)_t + 2(A^2\Phi_y/B)_y = 0, \tag{8}$$

 © 1979 American Institute of Physics

$$(\Phi_t + \Phi_y^2) + \tfrac{4}{3}A^2 - \tfrac{2}{3}B^2 - (A_y/B)_y B/A - \tfrac{1}{2}(1/B)_{yy}B - IB_y^2/B^2 = 0, \quad (9)$$

$$(\Phi_t + \Phi_y^2) + \tfrac{2}{3}A^2 + \tfrac{2}{3}B^2 + A_y^2/A^2 - \tfrac{1}{2}(A^2)_{yy}/A^2 - 2I(A^2B_y/B^3)_y B^2/A^2 - 3IB_y^2/B^2 = 0. \quad (10)$$

Linearizing in the form

$$A = A_0 + \sqrt{2}A_1 \exp(iky + \gamma t),$$
$$B = B_0 + B_1 \exp(iky + \gamma t), \quad (11)$$
$$\Phi = \Phi_0 + \Phi_1 \exp(iky + \gamma t),$$

we get the dispersion relation

$$(\tfrac{4}{3} + \kappa^2/2 + 2\kappa^2 I)\Gamma^2 = -\Big[(\tfrac{16}{3} + \kappa^2)(\tfrac{4}{3} + 2\kappa^2 I) + (\tfrac{4}{3} + \kappa^2/2)(\tfrac{8}{3} + \kappa^2)\Big], \quad (12)$$

where the abbreviations $\kappa = k/\eta$ and $\Gamma = \gamma/k\eta$ have been used. Equation (12) clearly shows

$$\Gamma^2 < 0, \quad (13)$$

corresponding to stable perturbations.

In principle, the more general ansatz $\psi = A\tanh(Bx + C)\exp(i\phi)$ should be used instead of (6). However, for the present case, the linearized Euler equations yield $C_1 = \text{const}$, and the dispersion relation (12) will not be changed.

At this stage one comment is in order: By this method, in general, stability cannot be concluded since it restricts the perturbed states to a certain subclass, here the translation mode and odd functions in $Bx + C$. And indeed, in Sec. III, we derive an energy principle from which instability can be concluded in the present case.

III. AN ENERGY PRINCIPLE FOR THE NS EQUATION

We now reconsider the transverse stability of kink solutions of the NS equation and derive a sufficient criterion for instability. Here, we only report the linear calculation; the generalization to the full nonlinear treatment is similar to that presented previously by the authors. [5]

Writing a perturbed solution of Eq. (1) in the form

$$\psi = (G + a + ib)\exp(-2i\eta^2 t), \quad (14)$$

where G is the zeroth order solution (3), we obtain

$$a_t = H_- b, \quad (15)$$
$$b_t = -H_+ a. \quad (16)$$

The operators H_+ and H_- are defined by

$$H_- = -\nabla^2 + G^2 - \eta^2, \quad (17)$$
$$H_+ = -\nabla^2 + 3G^2 - 2\eta^2. \quad (18)$$

After Fourier transformation, they can be written in the form

$$H_+ = -\frac{d^2}{dx^2} + k^2 - 6\eta^2\text{sech}^2\eta x + 4\eta^2, \quad (19)$$

$$H_- = -\frac{d^2}{dx^2} + k^2 - 2\eta^2\,\text{sech}^2\eta x, \quad (20)$$

where k is the transverse wavenumber.

The spectra of these operators are well known [12]: H_- has only one discrete eigenvalue $k^2 - \eta^2$ and the continuum starts at k^2. H_+ is positive definite (for $k \neq 0$) with discrete eigenvalues k^2 and $3\eta^2 + k^2$; the continuum starts at $4\eta^2 + k^2$. We note that the eigenfunction of H_- corresponding to the eigenvalue $k^2 - \eta^2$, i.e., $\text{sech}\eta x$, is even.

H_+ can be inverted and the functional

$$L = \tfrac{1}{2}\int dx\, \varphi^* H_+^{-1}\varphi, \quad (21)$$

where $\varphi = H_+ a_k$, satisfies

$$(L_t/L)_t \geqslant 0. \quad (22)$$

This result is similar to an energy principle derived by Laval *et al.* [13] in a completely different connection. The growth rate is

$$\gamma^2 = \sup_\varphi \frac{-\int dx \varphi^* H_- \varphi}{\int dx \varphi^* H_+^{-1}\varphi}. \quad (23)$$

As long as the discrete eigenvalue of H_-, i.e., $k^2 - \eta^2$, is negative, transverse instability occurs. This sufficient instability criterion suggests a cutoff at $k = k_c = \eta$, and indeed a Ljapunov functional for stability proves the result $k < \eta$ being a necessary and sufficient criterion for transverse instability of kink solutions.

We want to emphasize that the method proposed in this section has several advantages: It clearly shows which modes cause instability, i.e., those modes must have a component in the direction of the eigenfunction belonging to the negative eigenvalue of H_-. Since that eigenfunction is $\text{sech}\eta x$, i.e., an even function in x, it is obvious why the VAM fails. Furthermore, Eq. (23) is the result of a variational principle for the growth rate and allows the determination of the values of γ in the whole unstable k region by standard numerical methods. The derivation of a necessary and sufficient instability criterion was beyond the scope of previous instability calculations by Zakharov and Rubenchik, [5] who investigated only the small k limit.

IV. A LJAPUNOV FUNCTIONAL FOR THE KP EQUATION

We now turn to the question whether at least the predictions of instability by the VAM are correct. For this discussion we choose the KP form of the KdV equation as a concrete example.

The transverse stability of soliton solutions of the KP equation,

$$\varphi_{tx} + (\varphi\varphi_x)_x + \varphi_{xxxx} - \varphi_{yy} = 0, \quad (24)$$

has been already investigated analytically [6,10] and numerically. [14] The result of the calculation by the VAM [10] is that instability exists below a certain cutoff, $k < k_c$, where

$$k_c = 6\eta^2. \quad (25)$$

Here, k is the transverse wave number (normalized by the inverse electron Debye length) and the parameter η follows from the zeroth order plane soliton solution,

$$\varphi = 12\eta^2\text{sech}^2\eta(x - x_0 - 4\eta^2 t). \quad (26)$$

The analytical result of Kadomtsev and Petviashvili [6] is restricted to the small k limit. Zakharov[6] has presented a sufficient criterion for instability in the regime $k < (3)^{1/2}\eta^2$; the stability outside that region has only be treated by the VAM. We shall show that the VAM result of instability is not correct in the region

$$\sqrt{3}\eta^2 < k < 6\eta^2. \tag{27}$$

This actually means that a stable physical system with a periodicity length $L_y < 2\pi/(3^{1/2}\eta^2)$, is found to be unstable by the VAM.

In order to prove the assertion that in the region (27) the system is stable we derive a sufficient criterion for stability using Ljapunov's lemma.

For a reversible system, a Ljapunov functional for stability should be constructed from the constants of motion. This leads to the ansatz

$$L = \int d\Gamma \left\{ \left(\frac{\partial\varphi}{\partial x}\right)^2 + \left[\left(\frac{\partial}{\partial x}\right)^{-1}\frac{\partial\varphi}{\partial y}\right]^2 + 4\eta^2\varphi^2 - \frac{1}{3}\varphi^3 - \left(\frac{\partial\varphi_0}{\partial x}\right)^2 - 4\eta^2\varphi_0^2 + \frac{1}{3}\varphi_0^3\right\}. \tag{28}$$

Now we discuss whether L fulfills all conditions of Ljapunov's stability lemma: The condition $dL/dt \leqslant 0$ is trivially satisfied; it remains to show that an upper and a lower bound of L can be constructed in terms of the norm $\|\varphi - \varphi_0\|^2$.

We define the norm by

$$\|\varphi\|^2 = \int d\Gamma \left\{ \left(\frac{\partial\varphi}{\partial x}\right)^2 + \left[\left(\frac{\partial}{\partial x}\right)^{-1}\frac{\partial\varphi}{\partial y}\right]^2 + 4\eta^2\varphi^2\right\}. \tag{29}$$

Introducing the even and odd parts of the perturbed state φ, i.e.,

$$a' = \tfrac{1}{2}[\varphi(x,y) + \varphi(-x,y)] \equiv \frac{da}{dx},$$
$$b' = \tfrac{1}{2}[\varphi(x,y) - \varphi(-x,y)] \equiv \frac{db}{dx}, \tag{30}$$

we can, to lowest order in the norm of the perturbation, express the functional L in the form

$$L_1 = \int d\Gamma\,(aHa + bHb). \tag{31}$$

The fourth order operator

$$H = \frac{d^4}{dx^4} + 12\eta^2\frac{d}{dx}\text{sech}^2\eta x\frac{d}{dx} - 4\eta^2\frac{d^2}{dx^2} + k^2, \tag{32}$$

has the lowest eigenvalue $-3\eta^4 + k^2$. The corresponding eigenfunction is sech ηx tanh ηx. Thus, for $k > 3^{1/2}\eta^2$, the functional L_1 is positive and can be estimated in terms of the norm. The higher order contribution $L - L_1$ can also be estimated by the norm using standard mathematical tools, e.g., Schwarz's and Sobolev's inequalities. Therefore, L is indeed a Ljapunov functional for $k > (3)^{1/2}\eta^2$ and in contrast to the previous result[10] perturbations with wavenumbers k in the region

$$\sqrt{3}\eta^2 \leqslant k \leqslant 6\eta^2, \tag{33}$$

are stable. This yields to the additional conclusion that even the instability predictions by the VAM are sometimes doubtful. Somehow this is expected: The variational principle is for the action and not for the growth rate. Therefore, trial functions approximating the action do not necessarily imply a proper dispersion relation. In other words, it is by no means clear that those trial functions are approximate solutions of the dynamic equations, when the whole x-dependence is retained.

We note that it is possible to derive an energy principle analogous to that found in Sec. III. The evaluation of this principle leads—in agreement with findings of Zakharov[6]—to the result that for $k < 3^{1/2}\eta^2$ instability occurs. The derivation of a necessary and sufficient criterion for stability was beyond the scope of previous works.[6]

V. SUMMARY AND DISCUSSION

In this paper, we discussed the transverse stability of some one-field solitons. We showed that conclusions based on the VAM should be used with care: In some cases, neither stability nor instability can be predicted correctly. We believe that this defect is somehow inherent to this method since the variational principle is for the action and not for the growth rate. As the interrelation between these two different problems is not clear, the proper choice of the trial functions is somehow a game of chance.

We have exemplified the failure of the VAM for two systems: On the one hand, we demonstrated for the KP solitons that a physically stable system can be VAM unstable. We showed this by constructing a Ljapunov functional for stability. It is noteworthy that this calculation can be extended to find a necessary and sufficient criterion for stability. The reason why the unstable trial functions approximating the action do not yield the correct dispersion relation is that they are not approximate solutions of the dynamic equations when the whole x-dependence is retained.

On the other hand, we showed that the kink solution of the NS equation are actually unstable but VAM stable. An explicit expression for the growth rate was obtained in the whole k range which can be evaluated by standard numerical methods. It is important to note that this procedure can also be extended to give a necessary and sufficient instability criterion. The reason why the prediction of stability by the VAM cannot be accepted in general is obvious. By the latter method the possible forms of the perturbed states are restricted in an unjustified manner.

We conclude with the remark that also for other field equations, e.g., the complex Higgs field equation, the previous predictions based on the VAM should be critically reinvestigated. Our results of the NS equation indicate that the kink solution of the complex Higgs field equation should be also transversely unstable,[15] in contrast to the result of the VAM.[7] The reason is that in the limit of a weak time dependence (i.e., small phase shifts and growth rates) the Higgs field equation reduces to the NS equation.

ACKNOWLEDGMENT

This work was supported in part by the Deutsche Forschungsgemeinschaft through SFB 162.

[1]A.C. Scott, F.Y.F. Chu, and D.W. McLaughlin, Proc. IEEE **61**, 1443 (1973).
[2]V.G. Makhankov, Phys. Rep. C **35**, 1 (1978).
[3]A. Hasegawa, *Plasma Instabilities and Nonlinear Effects* (Springer, Berlin, 1975), Chapter 4. 4.
[4]V.E. Zakharov, *Inverse Scattering Method*, in Lectures in Physics (Springer, Berlin) to be published.
[5]V.E. Zakharov and A.M. Rubenchik, Zh. Eksp. Teor. Fiz. **65**, 997 (1973) [Sov. Phys.-JETP **38**, 494 (1974)]; N. Yajima, Prog. Theor. Phys. Jpn. **52**, 1066 (1974); E.W. Laedke and K.H. Spatschek, Phys. Rev. Lett. **41**, 1798 (1978).
[6]B.B. Kadomtsev and V.I. Petviashvili, Dokl. Akad. Nauk SSSR **192**, 753 (1970) [Sov. Phys.-Dokl. **11**, 539 (1970)]; V.E. Zakharov, Pisma Zh. Eksp. Teor. Fiz. **22**, 364 (1975) [JETP Lett. **22**, 172 (1975)]; H.H. Chen, J. Math. Phys. **16**, 2382 (1975).
[7]See p. 66 ff of Ref. 2.
[8]See, e.g., Chapter 4 of the review article by V.G. Makhankov, Ref. 2; also L.M. Degtyarev, V.E. Zakharov, and L.I. Rudakov, Zh. Eksp. Teor. Fiz. **68**, 115 (1975) [Sov. Phys.-JETP **41**, 57 (1975)].
[9]T. Taniuti and N. Yajima, J. Math. Phys. **10**, 1369, 2020 (1969); **14**, 1389 (1973).
[10]Yu. V. Katyshev and V.G. Makhankov, Phys. Lett. A **57**, 10 (1976).
[11]A. Hasegawa and F.D. Tappert, Appl. Phys. Lett. **23**, 142, 171 (1973).
[12]P.M. Morse and H. Feshbach, *Methods of Theoretical Physics* (McGraw-Hill, New York, 1953), Chapter 6.
[13]G. Laval, C. Mercier, and R. Pellat, Nucl. Fusion **5**, 156 (1965).
[14]F.D. Tappert, Bull. Am. Phys. Soc. **23**, 818 (1978).
[15]A.M. Polyakov, Pisma Zh. Eksp. Teor. Fiz. **20**, 430 (1974) [JETP Lett. **20**, 194 (1974)]; A.M. Polyakov, Zh. Eksp. Teor. Fiz. **68**, 1975 (1975) [Sov. Phys.-JETP **41**, 988 (1975)].

When does a projective system of state operators have a projective limit?

Wilhelm Ochs

Fachbereich Physik der Universität München, München, Germany

(Received 8 May 1978)

In analogy to Kolmogorov's classical *extension theorem*, we establish necessary and sufficient conditions under which a family $\{W_{(t_1,\cdots,t_n)}\}$ of state operators, defined on the finite tensor products $H_{t_1}\otimes\cdots\otimes H_{t_n}$ of some family $\{H_t: t\in T\}$ of complex Hilbert spaces, extends to a state operator on the infinite tensor product $\otimes_{t\in T}H_t$.

1. INTRODUCTION

One of the fundamental theorems[1,2] of classical probability theory establishes the conditions under which a family $\{\mu_{(t_1,\ldots,t_n)}\}$ of probability measures, defined on the finite products $\otimes_{i=1}^{n}(\Omega_{t_i},\mathscr{A}_{t_i})$ of some family $\{(\Omega_t,\mathscr{A}_t): t\in T\}$ of measurable spaces, has a unique *projective limit*, that is an extension to a probability measure on $\otimes_{t\in T}(\Omega_t,\mathscr{A}_t)$. This theorem, originating with Kolmogorov,[3] is equally fundamental for the interpretation[4] and for the application[1] of the classical probability calculus. Accordingly, any systematic development of a non-Boolean probability theory has to investigate the analogous problem.

The most important and most studied case of a non-Boolean probability calculus is found in the Hilbert space formalism of quantum mechanics.[5-7] In this frame, the event structure is represented by the lattice $\mathscr{P}(\mathscr{H})$ of all orthogonal projection operators on a complex separable Hilbert space $\mathscr{H}$; the classical probability measure is replaced by the *quantum state*,[8] defined as a σ-orthoadditive functional m: $\mathscr{P}(\mathscr{H})\to[0,1]$ with $m(\mathbb{1}) = 1$; and the composition of the event structures of different systems is accomplished *via* the tensor product of the corresponding Hilbert spaces.[9] As shown by Gleason,[10] every quantum state m can be represented (in the case $\dim\mathscr{H}\geqslant 3$) as $m(P) = \operatorname{tr} WP$ by means of a unique positive linear operator W on $\mathscr{H}$ with $\operatorname{tr} W = 1$ which is therefore called the *state operator* or *statistical operator* (STO) corresponding to m. And since every positive linear operator V on $\mathscr{H}$ with $\operatorname{tr} V = 1$ yields a quantum state by $P\mapsto \operatorname{tr} VP$, it is natural and justified to regard the state operators themselves as the quantal counterparts of the classical probability measures.

These parallels between the classical and the quantum probability calculus raise the following question. Given a family $\{W_{(t_1,\ldots,t_n)}\}$ of STO's defined on the finite tensor products $\mathscr{H}_{t_1}\otimes\cdots\otimes\mathscr{H}_{t_n}$ of some family $\{\mathscr{H}_t: t\in T\}$ of complex separable Hilbert spaces: *On what conditions does the family* $\{W_{(t_1,\ldots,t_n)}\}$ *have a projective limit* (i.e., an extension to a STO on $\otimes_{t\in T}\mathscr{H}_t$) *and when is this limit unique*? The first part of this question has recently been answered by Christensen[11] for the special case that all given STO's are tensor products of "one-particle states." For this case, Christensen established a necessary and sufficient condition for the existence of a projective limit.

In the present paper, we give a general answer to the first part of the above question. In Sec. 2 we establish necessary and sufficient conditions under which a family $\{W_{(t_1,\ldots,t_n)}\}$ of STO's has a projective limit. In contrast to the classical result mentioned at the beginning, such a projective limit (if it exists at all) is in general *not unique*; this ambiguity will be investigated in a forthcoming paper.[12] In Sec. 3 we consider the case of product STO's and show how to recover Christensen's condition as a special case of our conditions of Sec. 2. Finally we add an Appendix where we compile the definition and some basic properties of the *partial trace*.

After this manuscript was completed, the author received a preprint by A. Bartoszewicz[13] which contains a related result on the extension of a family of STO's to a STO defined on an *incomplete* tensor product of given Hilbert spaces.

2. NECESSARY AND SUFFICIENT CONDITIONS

In the following, "Hilbert space" stands for "complex Hilbert space of dimension $\geqslant 1$." Let $\mathscr{K}$ be a Hilbert space. By $\mathscr{B}(\mathscr{K})$, $\mathscr{T}(\mathscr{K})$, $\langle\cdot,\cdot\rangle$, $\|\cdot\|$ and $\|\cdot\|_1$ we denote the algebra of all bounded linear operators on $\mathscr{K}$, the trace class in $\mathscr{B}(\mathscr{K})$, the scalar product in $\mathscr{K}$, the operator norm on $\mathscr{B}(\mathscr{K})$ and the trace norm on $\mathscr{T}(\mathscr{K})$, respectively.[14,15] To tackle our problem, we need some preliminaries on the infinite tensor product of Hilbert spaces (cf. Ref. 16): Let $\{\mathscr{H}_t: t\in T\}$ be a nonempty collection of Hilbert spaces. To every $\emptyset\neq M\subseteq T$, we associate the *complete* tensor product $\mathscr{H}^M := \otimes_{t\in M}\widehat{\mathscr{H}_t}$; for $\mathscr{H}^T$ we write $\widehat{\mathscr{H}}$. An element $\alpha = \otimes_{t\in T}\alpha_t$ in $\widehat{\mathscr{H}}$ with $\|\alpha_t\| = 1$ $(\forall t\in T)$ is called a *product unit vector* (PUV). Two PUV's α,β are called *equivalent* (written $\alpha\sim\beta$), if $\Sigma_{t\in M}|1-\langle\alpha_t,\beta_t\rangle| < \infty$. The equivalence relation $\sim$ divides the PUV's into equivalence classes $\boldsymbol{a}, \boldsymbol{b},\cdots$ the set of which we denote by Γ. The PUV's in $\boldsymbol{a}\in\Gamma$ span a closed subspace $\widehat{\mathscr{H}}_a$ of $\widehat{\mathscr{H}}$ which is called the *incomplete tensor product* (ICT) of the $\mathscr{H}_t$'s *with respect to* $\boldsymbol{a}$. If $\alpha = \otimes_t\alpha_t\in\boldsymbol{a}$, then the set of all PUV's $\otimes_t\beta_t$ such that $\alpha_t = \beta_t$ for all but a finite number of $t\in T$ contains many complete orthonormal systems of PUV's in $\widehat{\mathscr{H}}_a$ which we call α-*bases* for short. Thus $\widehat{\mathscr{H}}_a$ can be thought of as being "generated" by any PUV in $\boldsymbol{a}$; if we want to refer to a particular generating PUV in $\boldsymbol{a}$, say α, we also write $\widehat{\mathscr{H}}[\alpha]$ for $\widehat{\mathscr{H}}_a$. The complete tensor product $\widehat{\mathscr{H}}$ is the direct sum of all ICT's, $\widehat{\mathscr{H}} = \oplus_{a\in\Gamma}\widehat{\mathscr{H}}_a$. By Q_a we denote the orthogonal projection from $\widehat{\mathscr{H}}$ onto $\widehat{\mathscr{H}}_a$. Hence $\Sigma_{a\in\Gamma}Q_a = \mathbb{1}_T$ where $\mathbb{1}_M$ denotes the identity operator on $\mathscr{H}^M$. If $K\neq\emptyset$ is a finite subset of T,

 0022-2488/79/091842-06$01.00 © 1979 American Institute of Physics

then all operators of the form $Y \otimes \mathbb{1}_{T \setminus K}$ with $Y \in \mathscr{B}(\mathscr{H}^K)$ commute with Q_a for all $a \in \Gamma$. If $\{\epsilon_t : t \in D\}$, $D \subset T$, is a collection of unit vectors $\epsilon_t \in \mathscr{H}_t$, then we abbreviate $\otimes_{t \in D} \epsilon_t$ by $\epsilon[D]$; expressions like $\varphi \otimes \epsilon[\phi]$ or $Y \otimes \mathbb{1}_\emptyset$ are to be read as φ or Y, respectively.

For $\emptyset \neq A \subseteq B \subseteq T$, we denote the *partial trace* from $\mathscr{T}(\mathscr{H}^B)$ to $\mathscr{T}(\mathscr{H}^A)$ by $\Theta(A,B)$ and write Θ_A for $\Theta(A,T)$. The definition and several properties of the partial trace are complied in the Appendix.

After these preliminaries we can now state our problem more precisely. Consider a family $\{\mathscr{H}_t : t \in T\}$, $T \neq \emptyset$, of Hilbert spaces[17] and let $\mathscr{F}$ be the *directed set* of all finite nonempty subsets of T directed by inclusion.

Definition: A family $\{W_K : K \in \mathscr{F}\}$ of STO's $W_K \in \mathscr{T}(\mathscr{H}^K)$ is called a *projective system*, if $W_K = \Theta(K,H) W_H$ for all $K,H \in \mathscr{F}$ with $K \subset H$. A STO $V \in \mathscr{T}(\widehat{\mathscr{H}})$ is called a *projective limit* of $\{W_K\}$, if $W_K = \Theta_K V$ for all $K \in \mathscr{F}$.[18]

As in the classical case we search for a necessary and sufficient condition under which a family $\{W_K : K \in \mathscr{F}\}$ of STO's W_K on $\mathscr{H}^K$ has a projective limit. According to Corollary A.5(i), the consistency of $\{W_K\}$, i.e., the property $W_K = \Theta(K,H) W_H$ for $K \subset H$, is an obvious part of such a condition. Therefore, we need only consider projective systems.

Theorem 2.1: A projective system $\{W_K : K \in \mathscr{F}\}$ of STO's has a projective limit if and only if there exists a countable set $\{p_\sigma \in \mathbb{R} : \sigma \in S, p_\sigma > 0, \Sigma_\sigma p_\sigma = 1\}$ and, to every $\sigma \in S$, there exist families $\{\beta_t^\sigma : t \in T\}$, $\{\varphi_K^\sigma : K \in \mathscr{F}\}$ of unit vectors $\beta_t^\sigma \in \mathscr{H}_t$, and $\varphi_K^\sigma \in \mathscr{H}^K$ such that

$$\lim_K \left\| W_K - \sum_{\sigma \in S} p_\sigma P(\varphi_K^\sigma) \right\|_1 = 0, \tag{2.1}$$

and

$$\lim_K \left(\sup_{\substack{H \in \mathscr{F} \\ K \subseteq H}} \| \varphi_H^\sigma - \varphi_K^\sigma \otimes \beta^\sigma[H \setminus K] \| \right) = 0. \tag{2.2}$$

If these conditions are satisfied, then the nets $(\varphi_K^\sigma \otimes \beta^\sigma[T \setminus K])_K$ converge to unit vectors Φ^σ in $\widehat{\mathscr{H}}$ and the STO $\Sigma_{\sigma \in S} p_\sigma P(\Phi^\sigma)$ is a projective limit of $\{W_K\}$.

Proof: (I) Assume that the conditions (2.1) and (2.2) are satisfied. If we define $\tilde{\varphi}_K^\sigma := \varphi_K^\sigma \otimes \beta^\sigma[T \setminus K]$, then

$$\|\tilde{\varphi}_H^\sigma - \tilde{\varphi}_K^\sigma\| = \|\varphi_H^\sigma - \varphi_K^\sigma \otimes \beta^\sigma[H \setminus K]\|$$

for all $K,H \in \mathscr{F}$ with $K \subseteq H$, and by (2.2), $(\tilde{\varphi}_K^\sigma)_K$ is a Cauchy net in $\widehat{\mathscr{H}}$. As a consequence there exists, for every $\sigma \in S$, exactly one unit vector $\Phi^\sigma := \lim_K \tilde{\varphi}_K^\sigma$ in $\widehat{\mathscr{H}}$ and we get

$$\lim_K \langle \tilde{\varphi}_K^\sigma, X \tilde{\varphi}_K^\sigma \rangle = \langle \Phi^\sigma, X \Phi^\sigma \rangle \tag{2.3}$$

for all $X \in \mathscr{B}(\widehat{\mathscr{H}})$. For $H,K \in \mathscr{F}$, $K \subseteq H$, and $Y \in \mathscr{B}(\mathscr{H}^K)$,

$$\langle \tilde{\varphi}_H^\sigma, Y \otimes \mathbb{1}_{T \setminus K} \tilde{\varphi}_H^\sigma \rangle = \langle \varphi_H^\sigma, Y \otimes \mathbb{1}_{H \setminus K} \varphi_H^\sigma \rangle,$$

and so by (2.3)

$$\lim_H \langle \varphi_H^\sigma, Y \otimes \mathbb{1}_{H \setminus K} \varphi_H^\sigma \rangle = \langle \Phi^\sigma, Y \otimes \mathbb{1}_{T \setminus K} \Phi^\sigma \rangle.$$

By (A1) and the dominated convergence theorem, this extends to

$$\lim_H \operatorname{tr}_H \left[(Y \otimes \mathbb{1}_{H \setminus K}) \sum_{\sigma \in S} p_\sigma P(\varphi_H^\sigma) \right] = \operatorname{tr}_T \left[(Y \otimes \mathbb{1}_{T \setminus K}) \sum_{\sigma \in S} p_\sigma P(\Phi^\sigma) \right]. \tag{2.4}$$

With $U_K := W_K - \Sigma_\sigma p_\sigma P(\varphi_K^\sigma)$, condition (2.1) implies that

$$\lim_H \operatorname{tr}_H [(Y \otimes \mathbb{1}_{H \setminus K}) U_H] = 0. \tag{2.5}$$

From $\Theta(K,H) W_H = W_K$ we get

$$\begin{aligned} \operatorname{tr}_K [Y W_K] &= \operatorname{tr}_H [(Y \otimes \mathbb{1}_{H \setminus K}) W_H] \\ &= \operatorname{tr}_H [(Y \otimes \mathbb{1}_{H \setminus K}) U_H] + \operatorname{tr}_H \Big[(Y \otimes \mathbb{1}_{H \setminus K}) \\ &\qquad \times \sum_{\sigma \in S} p_\sigma P(\varphi_H^\sigma) \Big]. \end{aligned} \tag{2.6}$$

Combining (2.4)–(2.6) we obtain

$$\operatorname{tr}_K [Y W_K] = \operatorname{tr}_T \left[(Y \otimes \mathbb{1}_{T \setminus K}) \sum_{\sigma \in S} p_\sigma P(\Phi^\sigma) \right],$$

for all $Y \in \mathscr{B}(\mathscr{H}^K)$, $K \in \mathscr{F}$ which implies, by Theorem A.1, that

$$W_K = \Theta_K \sum_{\sigma \in S} p_\sigma P(\Phi^\sigma),$$

for all $K \in \mathscr{F}$. This proved the sufficiency of our conditions and the second assertion of the theorem.

(II) To prove the necessity, we begin with

Lemma 2.2: Let $\alpha = \otimes_t \alpha_t$ be a PUV in $\widehat{\mathscr{H}}$ and let Φ be a unit vector in $\widehat{\mathscr{H}}[\alpha]$. Then there exists a collection $\{\varphi_K : K \in \mathscr{F}\}$ of unit vectors φ_K in $\mathscr{H}^K$ such that

$$\lim_K \|\Theta_K P(\Phi) - P(\varphi_K)\|_1 = 0, \tag{2.7}$$

and

$$\lim_K \left(\sup_{\substack{H \in \mathscr{F} \\ K \subseteq H}} \|\varphi_H - \varphi_K \otimes \alpha[H \setminus K]\| \right) = 0. \tag{2.8}$$

Proof: Let $\{\xi^\nu : \nu \in I\}$ be an α-basis of PUV's $\xi^\nu = \otimes_t \xi_t^\nu$ in $\widehat{\mathscr{H}}[\alpha]$. Then there exists a countable subset $M \subseteq I$ such that

$$\Phi = \sum_{\nu \in M} x_\nu \xi^\nu, \quad \sum_{\nu \in M} |x_\nu|^2 = 1.$$

To every $\nu \in M$ we have a smallest finite set $B_\nu \subseteq T$ such that $\xi_t^\nu = \alpha_t$ for $t \notin B_\nu$. Hence $Z := \cup_{\nu \in M} B_\nu$ is countable and

$$\Phi = \Big(\sum_{\nu \in M} x_\nu \xi^\nu[Z] \Big) \otimes \alpha[T \setminus Z].$$

We fix some element τ in M with $x_\tau \neq 0$ and associate, to every $K \in \mathscr{F}$ with $B_\tau \subseteq K$, the unit vectors

$$\varphi_K := \sum_{B_\nu \subseteq K} \tilde{x}_\nu(K) \xi^\nu[K], \quad \text{in } \mathscr{H}^K,$$

and

$$\tilde{\varphi}_K := \varphi_K \otimes \alpha[T \setminus K] = \sum_{B_\nu \subseteq K} \tilde{x}_\nu(K) \xi^\nu, \quad \text{in } \hat{\mathscr{H}},$$

where

$$\tilde{x}_\nu(K) := x_\nu \Big(\sum_{B_\nu \subseteq K} |x_\nu|^2 \Big)^{-1/2} = : x_\nu d_K.$$

Hence

$$\lim_K d_K = 1 \tag{2.9}$$

and

$$\|\Phi - \tilde{\varphi}_K\| = \Big\| \sum_{B_\nu \subseteq K} (x_\nu - \tilde{x}_\nu(K)) \xi^\nu + \sum_{B_\nu \not\subseteq K} x_\nu \xi^\nu \Big\|$$

$$\leqslant \Big(\sum_{B_\nu \subseteq K} |x_\nu - \tilde{x}_\nu(K)|^2 \Big)^{1/2} + \Big(\sum_{B_\nu \not\subseteq K} |x_\nu|^2 \Big)^{1/2}$$

$$\leqslant (1 - d_K) + (1 - d_K^{-2})^{1/2}, \tag{2.10}$$

for all $K \in \mathscr{F}$ with $B_\tau \subseteq K$. By Corollary A.4(c),

$$\Theta_K P(\tilde{\varphi}_K) = P(\varphi_K). \tag{2.11}$$

From (2.9)–(2.11) and from Corollary A.4(f) we finally conclude that

$$\lim_K \|\Theta_K P(\Phi) - P(\varphi_K)\|_1 = 0.$$

In analogy to (2.10) one finds

$$\|\varphi_H - \varphi_K \otimes \alpha[H \setminus K]\| \leqslant d_H(1 - d_K) + d_H(1 - d_K^{-2})^{1/2},$$

which yields assertion (2.8). □

To finish the proof of Theorem 2.1, we have to extend the result of Lemma 2.2 to an arbitrary STO on $\hat{\mathscr{H}}$. We first consider a unit vector ψ in $\hat{\mathscr{H}}$ and set $\Gamma_\psi := \{a \in \Gamma : Q_a \psi \neq 0\}$, $y_a := \|Q_a \psi\|$. For $a \in \Gamma_\psi$, $\psi_a := y_a^{-1} Q_a \psi$ is a unit vector in $\mathscr{H}_a$ and

$$\psi = \sum_{a \in \Gamma_\psi} y_a \psi_a. \tag{2.12}$$

From $\Sigma_{a \in \Gamma} Q_a = \mathbb{1}$ and $[Q_a, Y \otimes \mathbb{1}_{T \setminus K}] = 0$ we obtain

$$\operatorname{tr}(Y \otimes \mathbb{1}_{T \setminus K}) P(\psi) = \sum_{a \in \Gamma} \langle \psi, Q_a (Y \otimes \mathbb{1}_{T \setminus K}) Q_a \psi \rangle$$

$$= \sum_{a \in \Gamma_\psi} y_a^2 \langle \psi_a, Y \otimes \mathbb{1}_{T \setminus K} \psi_a \rangle$$

$$= \operatorname{tr}\Big[(Y \otimes \mathbb{1}_{T \setminus K}) \sum_{a \in \Gamma_\psi} y_a^2 P(\psi_a) \Big]$$

for all $Y \in \mathscr{B}(\mathscr{H}^K)$, $K \in \mathscr{F}$ which implies that

$$\Theta_K P(\psi) = \Theta_K \sum_{a \in \Gamma_\psi} y_a^2 (\psi_a) \tag{2.13}$$

for all $K \in \mathscr{F}$.

Finally we consider an arbitrary STO W on $\hat{\mathscr{H}}$. If $W = \Sigma_{i \in I} c_i P(\zeta_i)$ is a diagonal representation of W with $c_i > 0$ $(\forall i \in I)$, then it follows from (2.12), (2.13), and Corollary A.4(h) that

$$\Theta_K W = \sum_{i \in I} \sum_{a \in \Gamma_i} c_i y_{ia}^2 \Theta_K P(\zeta_{ia}), \tag{2.14a}$$

where $\Gamma_i := \{a \in \Gamma : Q_a \zeta_i \neq 0\}$, $y_{ia} := \|Q_a \zeta_i\|$, and $\zeta_{ia} := y_{ia}^{-1} Q_a \zeta_i$. If we replace the double index i, a in (2.14a) by a single index σ and set $p_\sigma := c_i y_{ia}^2$, then (2.41a) reads

$$\Theta_K W = \sum_{\sigma \in S} p_\sigma \Theta_K P(\zeta_\sigma), \tag{2.14b}$$

with $\Sigma_{\sigma \in S} p_\sigma = 1$ and $p_\sigma > 0$ for all $\sigma \in S = \cup_{i \in I} \Gamma_i$. Here every unit vector ζ_σ is contained in an ICT generated by some PUV, say $\beta^\sigma = \otimes_t \beta_t^\sigma$. Applying Lemma 2.2 to each summand of (2.14b) we obtained, for every $\sigma \in S$, a family $\{\beta_t^\sigma : t \in T\}$ of unit vectors $\beta_t^\sigma \in \mathscr{H}_t$ and a family $\{\varphi_K^\sigma : K \in \mathscr{F}\}$ of unit vectors $\varphi_K^\sigma \in \mathscr{H}^K$ which satisfy (2.2) and

$$\lim_K \|\Theta_K P(\zeta_\sigma) - P(\varphi_K^\sigma)\|_1 = 0. \tag{2.15}$$

From

$$\Big\| \Theta_K W - \sum_{\sigma \in S} p_\sigma P(\varphi_K^\sigma) \Big\|_1$$

$$= \Big\| \sum_{\sigma \in S} p_\sigma \{\Theta_K P(\zeta_\sigma) - P(\varphi_K^\sigma)\} \Big\|_1$$

$$\leqslant \sum_{\sigma \in S} p_\sigma \|\Theta_K P(\zeta_\sigma) - P(\varphi_K^\sigma)\|_1 \leqslant 2,$$

and from (2.15) we finally obtain

$$\lim_K \Big\| \Theta_K W - \sum_{\sigma \in S} p_\sigma P(\varphi_K^\sigma) \Big\|_1 = 0,$$

which shows the necessity of our conditions. □

It can be seen from (2.13) or (2.14a) that even if $\{W_K\}$ has a projective limit, this limit is in general *not unique*, which is in striking contrast to the classical situation. In a forthcoming paper[12] we will investigate the structure of the set of all projective limits belonging to a given projective system of STO's.

Applications of theorem 2.1 will often be concerned with the special case $T = \mathbb{N}$.[19] In this case it is expedient to use $\mathbb{N}$ itself with its natural ordering, instead of $\mathscr{F}$, as the directed set: Let $(\mathscr{H}_i)_{i \in \mathbb{N}}$ be a sequence of Hilbert spaces and $(W_n)_{n \in \mathbb{N}}$ a sequence of STO's $W_n \in \mathscr{T}(\mathscr{H}_n)$, $\mathscr{H}^n := \otimes_1^n \mathscr{H}_i$. The sequence (W_n) is called *projective* if, in self-explanatory notation, $\Theta(n, n+1) W_{n+1} = W_n$ for all $n \in \mathbb{N}$. A STO V on $\hat{\mathscr{H}} = \otimes_1^\infty \mathscr{H}_i$ is called a *projective limit* of (W_n) if $W_n = \Theta_n V$ for all $n \in \mathbb{N}$. By an obvious adaptation of the proof of Theorem 2.1 we obtain

Corollary 2.3: A projective sequence (W_n) has a projective limit if and only if there exists a countable set $\{p_\sigma \in \mathbb{R} : \sigma \in S, p_\sigma > 0, \Sigma_\sigma p_\sigma = 1\}$ and, to every $\sigma \in S$, there exist sequences $(\beta_i^\sigma)_{i \in \mathbb{N}}$, $(\varphi_n^\sigma)_{n \in \mathbb{N}}$ of unit vectors $\beta_i^\sigma \in \mathscr{H}_i$, $\varphi_n^\sigma \in \mathscr{H}^n$ such that

$$\lim_{n \to \infty} \Big\| W_n - \sum_{\sigma \in S} p_\sigma P(\varphi_n^\sigma) \Big\|_1 = 0,$$

and

$$\lim_{n \to \infty} \Big(\sup_{m > n} \|\varphi_m^\sigma - \varphi_n^\sigma \otimes \otimes_{i=n+1}^m \beta_i^\sigma\| \Big) = 0.$$

In this case, the sequences $(\varphi_n^\sigma \otimes \otimes_{i>n} \beta_i^\sigma)$ converge strongly to unit vectors Φ^σ in $\hat{\mathscr{H}}$ and the STO $\Sigma_\sigma p_\sigma P(\Phi^\sigma)$ is a projective limit of (W_n).

3. SYSTEMS OF PRODUCT STATES

In this section we consider projective systems of product STO's. It turns out that, in this special case, the necessary and sufficient conditions of Theorem 2.1 take a particularly simple form.

Let $\{\mathscr{H}_t : t\in T\}$, $T\neq\emptyset$, be a family of Hilbert spaces and let a STO X_t on $\mathscr{H}_t$ be given for all $t\in T$. Each X_t has a diagonal representation

$$X_t = \sum_{i=1}^{\infty} q_{ti} P(\alpha_{ti}), \tag{3.1}$$

with $\langle\alpha_{ti},\alpha_{tj}\rangle = \delta_{ij}$, $q_{t1}\geqslant q_{t2}\geqslant\cdots\geqslant 0$ and $\Sigma_{i=1}^{\infty} q_{ti} = 1$. Let G be the set of all mappings $g{:}T\to\mathbb{N}$ such that $g(t) = 1$ for all but finitely many $t\in T$; in particular, G contains a mapping e with $e(t) = 1$ for all t. The family $\{X_t\}$ uniquely determines the nonnegative real numbers

$$q_g(K) := \prod_{t\in K} q_{tg(t)}, \quad q_g := q_g(T),$$

for all $g\in G$, $\emptyset\neq K\subseteq T$.

Lemma 3.1:

$$q_e\begin{cases} >0 \\ =0 \end{cases} \Longleftrightarrow B := \sum_{g\in G} q_g = \begin{cases} 1 \\ 0 \end{cases}.$$

Proof: Obviously, $B\geqslant q_e\geqslant 0$, and $q_e = 0$ implies that $B = 0$. Let H be any nonempty finite subset of G and define $T_g := \{t\in T{:}g(t)\neq 1\}$, $T(H) := \cup_{g\in H} T_g$ and $H_1 := \{g\in G{:}T_g\subseteq T(H)\}$. Then

$$\sum_{g\in G} q_g \leqslant \sum_{g\in H_1} q_g = \left\{\prod_{t\in T(H)} \sum_{i=1}^{\infty} q_{ti}\right\} q_e(T\setminus T(H))$$

$$= q_e(T\setminus T(H)) \leqslant 1.$$

Since H was arbitrary, $B\leqslant 1$. So it remains to show that $q_e > 0$ implies that $B\geqslant 1$. Assume $q_e > 0$. Then there exists an ascending sequence of finite subsets of T, $\emptyset\neq K_1\subseteq K_2\subseteq\cdots\subseteq T$, such that

$$\lim_{n\to\infty} q_e(T\setminus K_n) = 1.$$

With $G_n := \{g\in G : T_g\subseteq K_n\}$ we find

$$B\geqslant \sum_{g\in G_n} q_g = \left\{\prod_{t\in K_n} \sum_{i=1}^{\infty} q_{ti}\right\} q_e(T\setminus K_n) = q_e(T\setminus K_n)$$

for all $n\in\mathbb{N}$ which implies that $B\geqslant 1$. □

Theorem 3.2: Let $\{X_t{:}\ t\in T\}$, G and $q_g(K)$ be defined as above. Then $\{\otimes_{t\in K} X_t : K\in\mathscr{F}\}$ is a projective system of STO's; and this system has a projective limit if and only if $\Pi_{t\in T}\|X_t\| > 0$. In this case

$$V := \sum_{g\in G} q_g P\left(\otimes_{t\in T} \alpha_{tg(t)}\right) \tag{3.2}$$

is a projective limit of $\{\otimes_{t\in K} X_t\}$.

Proof: (I) Obviously, $\{\otimes_{t\in K} X_t\}$ is a projective system. Suppose that it has a projective limit. Then we infer from Theorem 2.1 that, in the notation of this theorem,

$$\lim_K \left\| \otimes_{t\in K} X_t - \sum_{\sigma\in S} p_\sigma P(\varphi_K^\sigma)\right\|_1 = 0.$$

By $\|X - Y\|_1\geqslant\|X - Y\|\geqslant|\,\|X\| - \|Y\|\,|$, this yields

$$\lim_K \left| \prod_{t\in K} \|X_t\| - \left\| \sum_{\sigma\in S} p_\sigma P(\varphi_K^\sigma)\right\|\right| = 0.$$

Since $\|\Sigma_\sigma p_\sigma P(\varphi_K^\sigma)\|\geqslant\max\{p_\sigma{:}\sigma\in S\} > 0$, we obtain

$$\prod_{t\in T}\|X_t\|\geqslant\max\{p_\sigma : \sigma\in S\} > 0.$$

(II) Suppose that $\Pi_{t\in T}\|X_t\| > 0$. By $\|X_t\| = q_{t1}$ and Lemma 3.1, this implies that $\Sigma_{g\in G}\, q_g = 1$. Hence the sum (3.2) contains only countably many nonzero terms and we conclude from Corollary A.4(h) that V is a STO on $\hat{\mathscr{H}}$. With

$$\varphi_K^g := \otimes_{t\in K} \alpha_{tg(t)},$$

we get

$$\left\|\varphi_H^g - \varphi_K^g \otimes \otimes_{t\in H\setminus K} \alpha_{tg(t)}\right\| = 0, \tag{3.3}$$

for all $K\subset H\in\mathscr{F}$.

$$\left\| \otimes_{t\in K} X_t - \sum_{g\in G} q_g P(\varphi_K^g)\right\|_1$$

$$\leqslant \left\| \sum_{g\in G_1(K)} \{q_g(K) - q_g\} P(\varphi_K^g)\right\|_1$$

$$+ \left\| \sum_{g\in G_2(K)} q_g P(\varphi_K^g)\right\|_1 \equiv S_1(K) + S_2(K), \tag{3.4}$$

where $G_1(K) := \{g\in G{:}\ g(t) = 1\,\forall t\notin K\}$, $G_2(K) := G\setminus G_1(K)$.

$$S_1(K) = \sum_{g\in G_1(K)} |q_g(K) - q_g|$$

$$= (1 - q_e(T\setminus K)) \sum_{g\in G_1(K)} q_g(K)$$

$$= (1 - q_e(T\setminus K)) \prod_{t\in K} \sum_{i=1}^{\infty} q_{ti} = 1 - q_e(T\setminus K);$$

$$S_2(K) = \sum_{g\in G_2(K)} q_g.$$

Since $\Sigma_{g\in G}\, q_g = 1$, the nets $(S_1(K))_K$ and $(S_2(K))_K$ converge to zero and we conclude from (3.4) that

$$\lim_K \left\| \otimes_{t\in K} X_t - \sum_{g\in G} q_g P(\varphi_K^g)\right\|_1 = 0. \tag{3.5}$$

Equations (3.3), (3.5), and Theorem 2.1 show that V is a projective limit of $\{\otimes_{t\in K} X_t\}$. □

Theorem 3.2 can be proven just as simply without recourse to Theorem 2.1. But the above proof shows explicitly how the conditions of both theorems are related in the special case of product STO's.

ACKNOWLEDGMENTS

It is a pleasure to thank D. Castrigiano, R.-J. Dümcke, and U. Oppel for their critical reading of the manuscript.

APPENDIX: THE PARTIAL TRACE

In this Appendix we give a short survey of some important properties of the partial trace.

Preliminaries: Let $\mathscr{H}$ be a Hibert space. For $A\in\mathscr{B}(\mathscr{H})$, we define $|A| := (A^*A)^{1/2}$ and $\|A\|_1 := \mathrm{tr}|A|$; hence $\|A\|\leqslant\|A\|_1$. The trace class $\mathscr{T}(\mathscr{H}) := \{A\in\mathscr{B}(\mathscr{H}):\|A\|_1 < \infty\}$ is a Banach space with norm $\|\cdot\|_1$ which contains AX and XA for $A\in\mathscr{B}(\mathscr{H})$, $X\in\mathscr{T}(\mathscr{H})$. The trace has the properties $\mathrm{tr}AX = \mathrm{tr}XA$ and

$$|\mathrm{tr}XA| \leqslant \|X\|_1\|A\| \tag{A1}$$

for all $X\in\mathscr{T}(\mathscr{H})$, $A\in\mathscr{B}(\mathscr{H})$. For further details, the reader is referred to Refs. 14 and 20.

We consider now two Hilbert spaces $\mathscr{K}_1$, $\mathscr{K}_2$ and set $\mathscr{K}_3 := \mathscr{K}_1\otimes\mathscr{K}_2$ and $\mathscr{T}_i := \mathscr{T}(\mathscr{K}_i)$, $\mathscr{B}_i := \mathscr{B}(\mathscr{K}_i)$ for $i = 1,2,3$. By tr_i and $\mathbb{1}_i$ we denote the trace on $\mathscr{T}_i$ and the identity operator on $\mathscr{K}_i$, respectively.

Theorem A.1[9]: To every X in $\mathscr{T}_3$, there exists exactly one $\tilde{X}$ in $\mathscr{T}_1$ such that

$$\mathrm{tr}_3X(A\otimes\mathbb{1}_2) = \mathrm{tr}_1\tilde{X}A$$

for all $A\in\mathscr{B}_1$.

This theorem establishes a mapping $\Theta:\mathscr{T}_3\rightarrow\mathscr{T}_1$ by $X\mapsto\tilde{X}$ which is called the *partial trace* (or *reduction*) *from* $\mathscr{T}_3$ *to* $\mathscr{T}_1$. This name indicates that the mapping $X\mapsto\mathrm{tr}_3X$ can be divided in two steps, $X\mapsto\Theta X\mapsto\mathrm{tr}_1\Theta X = \mathrm{tr}_3X$, at which the mapping $X\mapsto\Theta X$ appears as the partial evaluation of the trace on $\mathscr{T}_3$.

Theorem A.2[21]: Let $\{\varphi_i\}$ be a complete orthonormal system in $\mathscr{K}_2$. Then

$$\langle\alpha,\Theta X\beta\rangle = \sum_{\{\varphi_i\}}\langle\alpha\otimes\varphi_i,X\beta\otimes\varphi_i\rangle \tag{A2}$$

for all $\alpha,\beta\in\mathscr{K}_1$ and $X\in\mathscr{T}_3$. There relations determine the mapping Θ uniquely.

Theorem A.3: $\|\Theta X\|_1\leqslant\|X\|_1$ for all $X\in\mathscr{T}_3$. For $X\geqslant 0$, $\|\Theta X\|_1 = \|X\|_1$.

Proof[22]: $\|\Theta X\|_1 := \mathrm{tr}_1|\Theta X| = \mathrm{tr}_1U^*\Theta X = \mathrm{tr}_3(U^*\otimes\mathbb{1}_2)X \leqslant\|X\|_1\|U^*\otimes\mathbb{1}_2\| = \|X\|_1$ where U is the partial isometry in the polar decomposition $\Theta X = U|\Theta X|$.[14] The second assertion is an immediate consequence of (A2). □

We note in passing that, if $\dim\mathscr{K}_2 > 1$, Θ is not a contraction with respect to the norm $\|\cdot\|$. From Theorems A.1 to A.3 one easily derives a lot of interesting properties of Θ.

Corollary A.4: (a) Θ is linear.

(b) Θ preserves order, positivity and self-adjointness.

(c) $\Theta(X\otimes Y) = (\mathrm{tr}_2Y)X$ for all $X\in\mathscr{T}_1$, $Y\in\mathscr{T}_2$.

(d) Θ is surjective.

(e) If $\dim\mathscr{K}_2 = \infty$, then Θ is neither $\|\cdot\|$-bounded nor $\|\cdot\|$-continuous.

(f) $\|\Theta P(\varphi) - \Theta P(\psi)\|_1\leqslant 2\sqrt{2}\|\varphi-\psi\|$ for any two unit vectors $\varphi,\psi\in\mathscr{K}_3$.

(g) Θ is continuous in the $\|\cdot\|_1$-topology.

(h) Let $\Sigma_1^\infty c_i$ be an absolutely convergent series of complex numbers and let (X_i) be a uniformly $\|\cdot\|_1$-bounded sequence of $X_i\in\mathscr{T}_3$. Then $\Sigma_1^\infty c_iX_i$ is in $\mathscr{T}_3$ and

$$\Theta\sum_{i=1}^{\infty}c_iX_i = \sum_{i=1}^{\infty}c_i\Theta X_i$$

where the sums are understood in the $\|\cdot\|_1$ topology.

(i) $\Theta[(A\otimes B)W] = A\Theta[(\mathbb{1}_1\otimes B)W]$, $\Theta[W(A\otimes B)] = \Theta[W(\mathbb{1}_1\otimes B)]A$, for all $A\in\mathscr{B}_1$, $B\in\mathscr{B}_2$ and $W\in\mathscr{T}_3$. If $\dim\mathscr{K}_1<\infty$, then these relations hold also for all $A\in\mathscr{T}_1$, $B\in\mathscr{T}_2$ and $W\in\mathscr{B}_3$.

Proof: (a) follows from Theorem A.1. (b), (c) and (i) follow from (A2). (d) and (e) follow from (c). (g) follows from Theorem A.3 and (h) follows from (g).

Proof of (*f*): By Theorem A.3, $\|\Theta P(\varphi) - \Theta P(\psi)\|_1\leqslant\|P(\varphi)-P(\psi)\|_1$. $P(\varphi)-P(\psi)$ is a self-adjoint operator of rank two whose nonzero eigenvalues are easily calculated to be $c_{1,2} = \pm(1-|\langle\varphi,\psi\rangle|^2)^{1/2}$. Hence, $\|P(\varphi)-P(\psi)\|_1 = |c_1|+|c_2| = 2(1-|\langle\varphi,\psi\rangle|^2)^{1/2}\leqslant 2\sqrt{2}\|\varphi-\psi\|$. □

Finally we state two properties of the partial trace which apply only to tensor products of more than two Hilbert spaces. Let $\{\mathscr{K}_i : i\in M\}$ be a finite nonempty family of Hilbert spaces. For $\emptyset\neq A\subset B\subseteq M$, Theorem A.1 establishes the partial trace from $\mathscr{T}_B := \mathscr{T}(\otimes_{i\in B}\mathscr{K}_i)$ to $\mathscr{T}_A := \mathscr{T}(\otimes_{i\in A}\mathscr{K}_i)$ which we denote by $\Theta(A,B)$; $\Theta(A,A)$ is defined as the identity map on $\mathscr{T}_A$. Since the Hilbert space tensor product is not ordered, $\Theta(A,B)$ has all the properties established above for Θ. In addition, we find the following.

Corollary A.5: (i) For $\emptyset\neq A\subseteq B\subseteq C\subseteq M$,

$$\Theta(A,B)\circ\Theta(B,C) = \Theta(A,C). \tag{A3}$$

(ii) Let $M = C\cup D\cup E\cup F$ be a partition of M into disjoint nonempty subsets and set $A := C\cup D$, $B := E\cup F$ and $G := D\cup E$. Then

$$\begin{aligned}\Theta(G,M)(X\otimes Y) &= [\Theta(D,M)(X\otimes Y)]\\ &\quad\otimes[\Theta(E,M)(X\otimes Y)]\\ &= [\Theta(D,A)X]\otimes[\Theta(E,B)Y],\end{aligned}$$

for all $X\in\mathscr{T}_A$, $Y\in\mathscr{T}_B$.

Proof: Both assertions are easily derived from (A2). □

[1]H. Bauer, *Wahrscheinlichkeitstheorie und Grundzüge der Masstheorie* (de Gruyter, Berlin, 1968), §62.

[2]J.Neveu, *Mathematical Foundations of the Calculus of Probability* (Holden–Day, San Francisco, 1965), §III.3.

[3]A.N. Kolmogoroff, *Grundbegriffe der Wahrscheinlichkeitsrechnung* (Springer, Berlin, 1933), §III.4.

[4]H. Richter, *Wahrscheinlichkeitstheorie* (Springer, Berlin, 1956), §V.1.

[5]G.W. Mackey, *The Mathematical Foundations of Quantum Mechanics* (Benjamin, New York, 1963), §2.2.

[6]V.S. Varadarajan, *Geometry of Quantum Theory* (Van Nostrand, Princeton, 1968), Vol. 1, Chapter VIII.

[7]J.M. Jauch, Synthese **29**, 131 (1974).

[8]This notion of state is characteristic of the "lattice-theoretic approach" to abstract quantum theory (cf. Refs. 6 and 9). In the "algebraic approach," a different concept of state is used which corresponds to the *expectation* in classical probability theory [cf. G.G. Emch, *Algebraic Methods in Statistical Mechanics and Quantum Field Theory* (Wiley, New York, 1972)]. Both concepts of state are closely related and both lead, on corresponding conditions, to the representation of states by state operators.
[9]J.M. Jauch, *Foundations of Quantum Mechanics* (Addison–Wesley, Reading, Mass., 1968).
[10]A.M. Gleason, J. Math. Mech. **6**, 885 (1957).
[11]M.J. Christensen, J. Math. Phys. **18**, 113 (1977).
[12]W. Ochs, "The Set of All Projective Limits of a Projective System of State Operators," to appear.
[13]A. Bartoszewicz, "Kolmogoroff Consistency Theorem for Gleason Measures," to appear in Coll. Math.
[14]M. Reed and B. Simon, *Methods of Modern Mathematical Physics. Vol. I: Functional Analysis* (Academic, New York, 1972).
[15]For simplicity we omit the index $\mathscr{H}$ for scalar products and norms where the meaning is clear from the context.
[16]J. von Neumann, Composito Math. **6**, 1 (1938).
[17]After having chosen to consider the state operators as the counterpart of classical probability measures, there is no further reason to restrict the dimension of the Hilbert spaces $\mathscr{H}_t$ (aside from the trivial restriction $\dim \mathscr{H}_t \geqslant 1$ which guarantees that $\mathscr{H}_t$ contains at least one unit vector). This generalization, however, is primarily of mathematical interest as soon as we can no longer rely on Gleason's theorem.
[18]This definition follows Ref. 1 and Z. Takeda, Tohoku Math. J. 7, 67 (1955). The term "projective" is justified by the transitive law (A3) of the partial trace established in the appendix.
[19]W. Ochs, J. Phil. Logic **6**, 437 (1977).
[20]R. Schatten, *Norm Ideals of Completely Continuous Operators* (Springer, Berlin, 1960).
[21]J. von Neumann, *Mathematische Grundlagen der Quantenmechanik* (Springer, Berlin, 1932).
[22]The author owes this elegant proof to R.-J. Dümcke.

Birman–Schwinger bounds for the Laplacian with point interactions

Lawrence E. Thomas[a)]
Department of Mathematics, University of Virginia, Charlottesville, Virginia 22903

(Received 23 February 1979)

Birman–Schwinger bounds on the eigenvalues of $-\Delta$ are described, where $-\Delta$ is the self adjoint operator obtained from the three-dimensional Laplacian by imposing boundary conditions at N distinct points in $\mathbf{R}^3$.

I. INTRODUCTION

Let $-\Delta$ be the self-adjoint operator acting in $L^2(\mathbb{R}^3)$ obtained by imposing boundary conditions on the three-dimensional Laplacian at distinct fixed points $x_1, x_2, \ldots, x_N \in \mathbb{R}^3$

$$\lim_{r_i \downarrow 0}\left(\frac{\partial}{\partial r_i} r_i - \alpha r_i\right)\Psi = 0, \quad i = 1,2,\cdots,N, \quad r_i = |x - x_i|. \tag{1.1}$$

As is well known, the resolvent kernel for this operator can be expressed,

$$(-\Delta + k^2)^{-1}(x,y) = \frac{1}{4\pi}\frac{\exp(-k|x-y|)}{|x-y|} - \frac{1}{(4\pi)^2} \times \sum_{i,j}^{N} \frac{\exp(-k|x-x_i|)}{|x-x_i|} T_{ij}(k) \frac{\exp(-k|y-x_j|)}{|y-x_j|}, \tag{1.2}$$

where $T(k)$ is the matrix inverse of $A(k)$ defined by

$$\begin{aligned} A_{ij}(k) &= \frac{1}{4\pi}\frac{\exp(-k|x_i-x_j|)}{|x_i-x_j|}, \quad i\neq j \\ &= -\frac{1}{4\pi}(k+\alpha), \quad i=j. \end{aligned} \tag{1.3}$$

The operator $-\Delta$ provides a very simple model for scattering; in terms of the matrix T, the scattering amplitude for $-\Delta$ is given by

$$f(k_{\text{out}},k_{\text{in}}) = \frac{-1}{4\pi}\sum_{i,j}^{N} \exp(-ik_{\text{out}}\cdot x_i) T_{ij}(-ik) \times \exp(ik_{\text{in}}\cdot x_j), \tag{1.4}$$

where $k_{\text{in}}, k_{\text{out}}$ are the incoming and outgoing momenta. Two-body potential scattering with suitably scaled negative potential supported about $x_1, x_2, \ldots, x_N$ has a scattering amplitude asymptotic to (1.4) at low energy. For a discussion of the operator $-\Delta$ with a single "pin" (along with a discussion of a related three-body operator) see the survey article of Flamand.[1]

The purpose of this note is to provide Birman-Schwinger bounds[2,3] on the eigenvalues of $-\Delta$ (see Theorem 1). Estimate (2.1) of the theorem has the following statistical mechanical interpretation: If N non-self-interacting fermions are attracted to the N pins via the boundary conditions (1.1), their resulting energy exceeds a negative constant times N plus another negative constant $-c$ times the sum of the inverse square of the distances between pins. Thus if a potential between pins is added which is at least as repulsive as $c|x_i - x_j|^{-2}$, thermodynamic stability is attained in the sense that the resulting Hamiltonian H_N for the N fermions and pins satisfies $H_N \geqslant -\text{const}\, N$ with the constant independent of N and the pin configuration (cf. Ref. 4).

The derivation of inequality (2.1) is reminicent of that for the potential case; we first obtain a bound on the number of eigenvalues less than a given energy, inequality (2.2). We remark that this bound is not particularly applicable if, for example, the pins are placed in a rectangular lattice of fixed lattice spacing and we take $N\to\infty$; rather, the bound is more applicable in the "collapsed" situation, where more and more pins are added to a given fixed finite region $\mathbb{R}^3$.

II. BIRMAN–SCHWINGER BOUNDS

Let $N(k)$, $k\geqslant 0$, be the number of eigenvalues of $-\Delta$, including multiplicities, which are less than or equal to $-k^2$. Let $k_1, k_2, \ldots, k_{N-1}$ be the positive numbers defined as follows: Set

$$k_{N-1} = \sup_{1\leqslant i\leqslant N}\left(\sum_{j\neq i}^{N}|x_i - x_j|^{-2}\right)^{1/2}.$$

Assume, by relabeling the x_i's if necessary, that the supremum is attained for $i = N$. Next, set

$$k_{N-2} = \sup_{1\leqslant i\leqslant N-1}\left(\sum_{j\neq i}^{N-1}|x_i - x_j|^{-2}\right)^{1/2}.$$

Again by relabeling, assume that this supremum is attained for $i = N-1$. The remaining k_i's are defined by continuing in this manner. Note that $k_1^2 + k_2^2 + \cdots + k_{N-1}^2 = \Sigma_{i<j}^{N}|x_i - x_j|^{-2}$. For notational convenience, set $k_0 = 0$, $k_N = \infty$. Finally, let $\kappa = 0$ if $\alpha\geqslant 0$, $\kappa = -\alpha$ if $\alpha < 0$.

Theorem 1: *The eigenvalues* $\{e_i\}$ *of* $-\Delta$ *less than* $-\kappa^2$ *satisfy*

$$\sum_{e_i < -\kappa^2} |e_i| \leqslant 2\kappa^2(N-1) + c\sum_{i<j}^{N}|x_i - x_j|^{-2} \tag{2.1}$$

with c independent of N,α. For $k>\kappa$, $N(k)$ *satisfies the inequality*

$$N(k+\kappa)\leqslant 4\sum_{i<j}^{l}[\exp(-2(k+\kappa)|x_i - x_j|)] \times (k^2|x_i - x_j|^2) + \exp(-2(k+\kappa)|x_i - x_j|) + (N-l), \quad k_{l-1}\leqslant k < k_l, \quad l = 1,2,\ldots,N. \tag{2.2}$$

Remark: In the particular case $\alpha = 0$ ($\kappa = 0$) the eigenvalues of $-\Delta$ are homogeneous functions of degree -2 in

a)Partially supported by NSF MCS 74-07313-A03.

 © 1979 American Institute of Physics

the x_i's. For this case, with $N=2$, the only eigenvalue is $e \cong -0.322|x_1 - x_2|^{-2}$. An estimate for the constant c with $\alpha = 0$ based on the discussion below is about 2.35.

We begin the proof of the theorem by recalling that the eigenfunctions of $-\Delta$ are of the form

$$\Psi_{-k^2}(x) = \frac{1}{4\pi}\sum_{i=1}^{N} \frac{\exp(-k|x-x_i|)}{|x-x_i|} q_i$$

with eigenvalue $-k^2$; here $q=(q_1,...,q_N)$ is a solution to $A(k)q = 0$ with k such that this equation has a homogeneous solution. Thus, at least for $k>0$, $N(k)$ is the number of values $k' \geqslant k$, times multiplicities, for which the self-adjoint matrix function $A(k')$ has eigenvalue zero.

Rather than working with $A(k)$ directly, we consider the matrix function

$$B(k) \equiv C^{-1/2}(A(k)+C)C^{-1/2}, \tag{2.3}$$

where C is a diagonal matrix with strictly positive entries c_i, $i=1,2,...,N$ as yet unspecified. Note that $A(k)$ has eigenvalue zero if and only if $B(k)$ has eigenvalue one. Further, the multiplicities are the same (the respective eigenvectors are related by $q \rightarrow C^{1/2}q$). Consequently $N(k)$ is the number of values $k' \geqslant k$ for which $B(k')$ has eigenvalue one.

The derivative of $B(k)$ with respect to k can be written

$$\frac{d}{dk}B_{ij}(k) = -(c_i)^{-1/2}\exp(-k|x_i-x_j|)(c_j)^{-1/2} \tag{2.4}$$

which is a *negative definite* matrix. It follows from first order perturbation theory that if $\lambda(k)$ is an eigenvalue of $B(k)$, then as a function of k it is strictly decreasing in k. Since $\lambda(k)$ goes to $-\infty$, $k \rightarrow \infty$, we have that $N(k)$ is the number of eigenvalues of $B(k)$ greater than or equal to one. But this is estimated by

$$N(k) \leqslant \mathrm{tr}B^2(k) = \sum_{\substack{i,j\\ i\neq j}}^{N} \frac{\exp(-2k|x_i-x_j|)}{c_i|x_i-x_j|^2 c_j} + \sum_{i=1}^{N}\left(\frac{k+\alpha-c_i}{c_i}\right)^2. \tag{2.5}$$

Now if α is positive, we can replace c_i by $c_i+\alpha$ and use the inequality $(c_i+\alpha)^{-1} \leqslant (c_i)^{-1}$ in (2.5). If α is negative we simply make the variable substitution $k' = k+\alpha$. In either case we can write

$$N(k+\kappa) \leqslant \sum_{\substack{i,j\\ i\neq j}}^{N} \frac{\exp(-2(k+\kappa)|x_i-x_j|)}{c_i|x_i-x_j|^2 c_j} + \sum_{i=1}^{N}\left(\frac{k-c_i}{c_i}\right)^2. \tag{2.6}$$

Estimate (2.6) holds for arbitrary c_i's >0 and by continuity holds for any of the c_i's infinite; in particular they can be taken as functions of k. We attempt to minimize the rhs by differentiating with respect to the c_i's and setting the derivatives to zero. This leads to the matrix equation with $k>0$,

$$c_i^{-1} + \sum_{j}^{N'} L_{ij}(k)c_j^{-1} = \frac{1}{k}, \quad i=1,2,...,N' \leqslant N, \tag{2.7}$$

where

$$L_{ij}(k) = \begin{cases} \dfrac{\exp(-2(k+\kappa)|x_i-x_j|)}{k^2|x_i-x_j|^2}, & i\neq j, \\ 0, & i=j, \end{cases} \tag{2.8}$$

and where for later purposes, we consider the equations $N' \leqslant N$ dimensional.

Lemma 1: *Equations* (2.7) *are solvable with* c_i*'s* >0 *provided*

$$\sup_i \sum_{j\neq i}^{N'} L_{ij}(k) < 1.$$

In this case, each $c_i(k)$ *satisfies*

$$c_i(k)/k > \sup\{1, L_{ij}(k)\}, \quad i,j \in 1,2,...,N'. \tag{2.9}$$

Proof of Lemma: Equation (2.7) can be written $(1+L)\chi = \chi^0$ with χ^0,χ the vectors with entries $\chi_i^0 = 1/k$, $\chi_i = c_i^{-1}$. The hypothesis of the lemma says that the Shur–Holmgren norm of the self-adjoint matrix L is <1 so that the operator norm of L is also <1. The hypothesis also implies that $(1-L)\chi^0$ has positive entries. Thus χ is given by the convergent Neumann expansion $\chi = (1+L^2+L^4+\cdots)\times(1-L)\chi^0$ which has strictly positive entries and so $c_i>0$ for each i. Finally, we have

$$\chi_i^0 = \chi_i + \sum_{j}^{N} L_{ij}\chi_j > L_{ij}\chi_j$$

and $\chi_i^0 - \chi_i > 0$ by the matrix equation which combine to give (2.9). ■

Returning to the proof of the theorem we *define* the $c_i(k)$'s as follows: For $k_{l-1} \leqslant k < k_l$, $l=1,2,...,N$ set $c_{l+1}(k) = c_{l+2}(k) = \cdots = c_N(k) = \infty$ and let $c_1(k),\cdots,c_l(k)$ be the solution to Eq. (2.7) with $N'=l$. The solution is assured by the lemma and the definition of the k_l's. Substituting these values for the c_i's into the estimate (2.6) and using the inequality

$$\frac{k}{c_i(k)}L_{ij}(k) < \frac{2L_{ij}(k)}{1+L_{ij}(k)} \tag{2.10}$$

which follows readily from the estimate (2.9) of the lemma, we obtain

$$\begin{aligned} N(k+\kappa) &\leqslant \sum_{i,j}^{l} \frac{kL_{ij}(k)}{c_i(k)} + (N-l) \\ &\leqslant 2\sum_{i,j}^{l} \frac{L_{ij}(k)}{1+L_{ij}(k)} + (N-l), \quad k_{l-1} \leqslant k < k_l \end{aligned} \tag{2.11}$$

which is estimate (2.2) of the theorem.

From estimate (2.11) we have

$$\begin{aligned} \sum_{e_i < -k_2} |e_i| &= -\int_{\kappa+0}^{\infty} k^2\, dN(k) \\ &= \kappa^2 N(\kappa+0) + 2\int_{\kappa+0}^{\infty} kN(k)dk \\ &\leqslant \kappa^2(N-1) + 4\sum_{i,j}^{N}\int_0^{\infty} \frac{(k+\kappa)L_{ij}(k)}{1+L_{ij}(k)}dk \\ &\quad + 2\sum_{l=1}^{N-1}(N-l)\int_{k_{l-1}}^{k_l}(k+\kappa)\,dk, \end{aligned} \tag{2.12}$$

where we have estimated the sum on the rhs of (2.11) by a sum up to N.

A typical term from the middle sum on the rhs of (2.12) can be estimated

$$\int_0^\infty \frac{(k+\kappa)L_{ij}(k)}{1+L_{ij}(k)}\,dk = \frac{1}{|x_i-x_j|^2}\int_0^\infty \frac{(p+\kappa|x_i-x_j|)\exp(-2p-\kappa|x_i-x_j|)}{p^2+\exp(-2p-2\kappa|x_i-x_j|)}\,dp$$

$$\leqslant \frac{1}{|x_i-x_j|^2}\left(\int_0^\infty \frac{pe^{-2p}\,dp}{p^2+e^{-2p}} + \kappa|x_i-x_j|\int_0^\infty \frac{dp}{p^2+\exp(-2\kappa|x_i-x_j|)}\right)$$

$$\leqslant \frac{1}{|x_i-x_j|^2}\left(\int_0^\infty \frac{pe^{-2p}\,dp}{p^2+e^{-2p}} + \pi/2\kappa|x_i-x_j|\exp(-\kappa|x_i-x_j|)\right) \leqslant \text{const}\,|x_i-x_j|^{-2}. \qquad (2.13)$$

Here we have made the substitution $p = k\,|x_i - x_j|$ and used the facts that the function $u(1+u)^{-1}$ is increasing, $u \geqslant 0$, and ue^{-u} is bounded, $u \geqslant 0$.

It remains to estimate the last sum in (2.12). We have

$$2\sum_{l=1}^{N-1}(N-l)\int_{k_{l+1}}^{k_l}(k+\kappa)\,dk = \sum_{l=1}^{N-1}k_l^2 + 2\kappa\sum_{l=1}^{N-1}k_l \leqslant 2\sum_{l=1}^{N-1}k_l^2 + \kappa^2(N-1) = 2\sum_{i<j}^{N}|x_i-x_j|^{-2} + \kappa^2(N-1) \qquad (2.14)$$

since $2\kappa k_l \leqslant \kappa^2 + k_l^2$. Combining (2.13), (2.14), with (2.12), we obtain (2.1), which completes the proof of the theorem.

We conclude with some remarks concerning the case when the boundary condition parameter α is pin dependent, i.e.,

$$\lim_{r_i\downarrow 0}\left(\frac{\partial}{\partial r_i}r_i - \alpha_i r_i\right)\Psi = 0, \quad i = 1,2,\ldots,N. \qquad (2.15)$$

In this situation [cf. Eq. (1.3)] the diagonal elements of $A(k)$ are $A_{ii}(k) = -(1/4\pi)(k+\alpha_i)$ and the off-diagonal elements are as before. The argument expressing $N(k)$ as the number of eigenvalues of $B(k)$ greater than one still holds. To indicate the dependence of $N(k)$ on the α_i's we write $N(k,\alpha_1,\alpha_2,\ldots,\alpha_N)$ and set $N_\alpha(k) = N(k,\alpha,\alpha,\ldots,\alpha)$.

Lemma 2: *$N(k,\alpha_1,\alpha_2,\ldots,\alpha_N)$ is monotone decreasing in each of the α_i's. If $\alpha = \inf\alpha_i$, $\beta = \sup\alpha_i$, then*

$$N_\alpha(k) \geqslant N(k,\alpha_1,\alpha_2,\ldots,\alpha_N) \geqslant N_\beta(k). \qquad (2.16)$$

Proof: By first order perturbation theory, the eigenvalues of $B(k)$ are decreasing functions of the α_i's. ■

[1]G. Flamand, "Mathematical theory of non-relativistic two- and three-particle systems with point interactions," in *Cargèse Lectures in Theoretical Physics*, edited by F. Lurcat (Gordon and Breach, New York, 1965), 247–87.

[2]M. Birman, The spectrum of singular boundary problems, Math. Sb. **55**, 124–74 (1961) [Amer. Math. Soc. Trans. **53**, 23–80 (1966)].

[3]J. Schwinger, "On the bound states of a given potential," Proc. Nat. Acad. Sci., U.S.A. **47**, 122–29 (1961).

[4]E. Lieb and W. Thirring, "Bound for the kinetic energy of fermions which proves the stability of matter," Phys. Rev. Lett. **35**, 687–89 (1975).

On rotating plane-fronted waves and their Poincaré-invariant differential geometry

H. Urbantke

Institut für Theoretische Physik, Universität Wien, Austria
(Received 9 October 1978)

After reviewing the rotating plane-fronted wave type solutions of the scalar wave equation and Maxwell's vacuum equations in flat space (we point out that there are Yang–Mill analogs as well), we study their Poincaré-invariant geometry by discussing their characteristic differential invariants and a noninertial curvilinear coordinate system canonically associated with them. In one of the appendices we treat the shearfree and the nondiverging null hypersurfaces in *complex* Minkowski space, in another one we derive the Yang–Mills version of Robinson's theorem on null electromagnetic fields.

1. INTRODUCTION

Rotating plane and plane-fronted waves as solutions to the scalar wave equation and to Maxwell's and Einstein's vacuum equations have been discussed in the literature. For the former, we may in particular refer to Ref. 1, where they appear under the title "functionally invariant" solutions. We review them here (Sec. 2A) together with a derivation in Appendix A which we also extend to the complex case. For the latter, we refer to Ref. 2, where a particular coordinate system is used to simplify the field equations. One gets, in fact, explicit solutions in terms of these coordinates which are in complete analogy to the case of nonrotating plane-fronted waves, and the whole gain in generality is hidden in the metric coefficients or in the relations between these coordinates and the Cartesian ones (Sec. 2B). It is easy to see that these solutions have exact Yang–Mills analogs, an observation which parallels Coleman's[3] for the nonrotating situation (Sec. 2C). Some general features of Yang–Mills null fields, in particular the analog of Robinson's theorem,[4] are given in Appendix B. Section 2D comments on the rotating plane-fronted gravitational waves of Ref. 2 and mainly explains why the present flat space considerations of the following sections are irrelevant for this case. In Sec. 3 we treat the differential invariants of a rotating one-parameter family of null hyperplanes: In general, there is a distinguished parameter and two differential invariants which, when regarded as functions of that parameter, describe the family in the sense of "natural geometry." The relation between the Lorentz group and the Moebius group in two dimensions has allowed us to borrow most of the development from existing theory of conformal differential geometry of plane curves.[5] In Sec. 4 we discuss the various envelopes which are associated with a rotating one-parameter family of null hyperplanes. The above-mentioned coordinate system turns out to be intimately related to these envelopes, which explains why it is possible for it to be "canonical" without there being any a priori symmetries in the arrangement of the hyperplanes. This relation is obtained in Sec. 5, where we derive the relation between the canonical and Cartesian inertial coordinates.

Notations:

Metric signature
$ds^2 = \eta_{\mu\nu}\, dx^\mu\, dx^\nu, \eta_{\mu\nu} = \text{diag}\ (1,1,1,-1)$.

Greek indices range and sum over 1,2,3,4.

$\epsilon_{\mu\nu\alpha\beta}$ is the permutation symbol, $\epsilon_{1234} = +1$.

Comma: partial derivative.

Electromagnetism: $F_{\mu\nu} = A_{\nu,\mu} - A_{\mu,\nu}$.

Yang–Mills theory: $\vec{A}_\mu = (A^a_\mu) = (A^1_\mu,\ldots,A^r_\mu)$, where Latin indices range and sum over $1,\ldots,r$ = dimension of gauge group.

$\vec{A} \times \vec{B}$ is defined by $(\vec{A}\times\vec{B})^c = C^c_{ab} A^a B^b$, where C^c_{ab} are the structure constants of the gauge group. The field tensor is then

$$\vec{F}_{\mu\nu} = \vec{A}_{\nu,\mu} - \vec{A}_{\mu,\nu} + \vec{A}_\mu \times \vec{A}_\nu.$$

2. ROTATING PLANE-FRONTED WAVES

A. Scalar rotating plane waves

A scalar field Φ in Minkowski space is called a plane wave if the hypersurface $\Phi =$ const. are lightlike, or null, hyperplanes. This means that the family of hypersurfaces $\Phi(x^\mu) = \tau$ must also be expressible in the form $k_\mu(\tau)x^\mu = p(\tau)$, where $k_\mu(\tau)k^\mu(\tau) \equiv 0$. If the direction of $k_\mu(\tau)$ is constant, we have the usual plane waves with parallel rays. Here we want to consider the *rotating* ray case where $k'(\tau)$ is *not* proportional to $k(\tau)$($' := d/d\tau$). From $k^2 = 0$ we have $kk' = 0$, and thus in the rotating case k' is spacelike, $k'^2 > 0$.

Conversely, given $p(\tau)$ and $k(\tau)$ with $k^2 = 0$, we can associate with the family of null hyperplanes $k(\tau)x = p(\tau)$ a scalar field $\Phi(x)$, given by the implicit equation

$$k(\Phi)x = p(\Phi). \tag{2.1}$$

By implicit differentiation, we find

$$\Phi_{,\mu} = \frac{k_\mu}{p' - k'x},$$

$$\Phi_{,\mu\nu} = -\frac{p'' - k''x}{p' - k'x}\Phi_{,\mu}\Phi_{,\nu} + \frac{k'_\mu\Phi_{,\nu} + k'_\nu\Phi_{,\mu}}{p' - k'x} \tag{2.2}$$
$$\equiv + \frac{p'' - k''x}{p' - k'x}\Phi_{,\mu}\Phi_{,\nu} - (\ln(p' - k'x))_{,\mu}\Phi_{,\nu} - (\ln(p' - k'x))_{,\nu}\Phi_{,\mu}.$$

From $k^2 = 0$, $kk' = 0$ we see that Φ satisfies the wave equation $\Box\Phi = 0$; its gradient field $\nabla\Phi$ is null, geodesic and divergence-free, which (flat space!) also implies the absence of

 0022-2488/79/091851-10$01.00 © 1979 American Institute of Physics

shear.[6] It is well-known that all null, geodesic, hypersurface-orthogonal, shear and divergence-free vector fields can be obtained this way. (For a method of showing this, cf. Appendix A.)

Our scalar fields may also be characterized by another property: They are what has been termed "*functionally invariant*" solutions of the wave equation. This means, if $\Phi(x)$ is a solution of $\Box\Phi = 0$ and $F(\cdot)$ is an arbitrary function of one variable, then $\Phi_1(x) := F(\Phi(x))$ again satisfies the wave equation. In fact, this condition is immediately seen to be equivalent to $\Box\Phi = 0, (\nabla\Phi)^2 = 0$, whose simultaneous solutions are—as we pointed out—given by (2.1), which in this context has been called the "Smirnov–Sobolev formula.[1]" The following side remark may be interesting: Because of the linearity of the wave equation, it makes sense also physically to ask for *complex-valued* functionally invariant solutions. In this case *not* all solutions are given by (a complex version of) (2.1), except for space dimensions lower than 3, as found by Erugin.[7] Because of the inaccessibility of this reference, we rederive his results in Appendix A, following the geometric ideas of Friedlander,[8] thereby achieving considerable simplification and unification.

B. Electromagnetic rotating plane-fronted waves

An electromagnetic field $F_{\mu\nu}(x)$ is called a null field if its invariants $F_{\mu\nu}F^{\mu\nu}$ and $F_{\mu\nu}{}^*F^{\mu\nu}$ both vanish. One can deduce that $F_{\mu\nu}$ is then of the form $F_{\mu\nu} = \alpha_\mu k_\nu - \alpha_\nu k_\mu$ where $k^2 = 0 = \alpha_\mu k^\mu$. When the source-free Maxwell equations hold, k must be geodesic and shearfree; conversely, for a given null geodesic and shearfree k one can find α_μ such that $\alpha_\mu k^\mu = 0$ and $F_{\mu\nu} = \alpha_\mu k_\nu - \alpha_\nu k_\mu$ satisfies the source-free Maxwell equations (Robinson's theorem, cf. Refs. 4 and 6). When the vector field k is twistfree, i.e., hypersurface orthogonal, one can take $k_\mu = u_{,\mu}$ and then find a gauge for the vector potential so that $A_\mu = au_{,\mu}$, where a is some function. The field then becomes $F_{\mu\nu} = a_{,\mu}u_{,\nu} - a_{,\nu}u_{,\mu}$. Fields of this kind have a "functional invariance" property in the sense that if $F_{\mu\nu}$ satisfies Maxwell's equations, then $F_{1\mu\nu} := F_{\mu\nu}F(u)$ with arbitrary F also satisfies them, and conversely, if some $F_{\mu\nu}$ has this property, it must be of the above form.

We will now call a null field $F_{\mu\nu}$ a plane-fronted wave, if the ray vector field k is twist- and divergence-free, i.e., $k = \nabla u$, where $(\nabla u)^2 = 0 = \Box u$. Thus k is the gradient of a scalar plane wave, and the term "plane-fronted" is introduced only because the amplitude $a_{,\mu}$ may in general not be constant over the wave surfaces. We shall be interested here in the rotating ray case. Maxwell's equations $F_{\mu\nu}{}^{,\nu} = 0$ then become

$$2a^{,\nu}u_{,\mu\nu} + u_{,\mu}\Box a = 0, \tag{2.3}$$

using $\Box u = 0$, $\nabla u\nabla a = 0$. According to Ref. 2, this may be simplified further by observing that $u_{,\mu\nu}$ has the form [cf. (2.2)]

$$u_{,\mu\nu} = \beta u_{,\mu}u_{,\nu} + \gamma_{,\mu}u_{,\nu} + \gamma_{,\nu}u_{,\mu}. \tag{2.4}$$

Here γ is determined only up to addition of an arbitrary function of u, with a corresponding change in β, and may be checked to satisfy

$$\gamma_{,\mu}u^{,\mu} = 0, \quad e^{-\gamma}\Box e^\gamma \equiv \gamma_{,\mu}{}^{\mu} + \gamma_{,\mu}\gamma^{,\mu} = 0. \tag{2.5}$$

The conditions to be satisfied by a may then be written

$$\nabla u\cdot\nabla a = 0, \quad \Box(e^\gamma a) = 0. \tag{2.6}$$

Explicit solutions for a may then be obtained using a curvilinear noninertial system of coordinates u,v,ξ,η which is particularly adapted to the rotating ray system and in which the Minkowski line element takes the following form,

$$ds^2 = d\xi^2 + d\eta^2 - \frac{4v}{\xi}d\xi du + 2du\,dv + \left(\frac{v^2}{\xi^2} + \xi A(u,\xi,\eta)\right)du^2, \tag{2.7}$$

where $A := h(u) + j(u)(\xi + i\eta) + j^*(u)(\xi - i\eta)$ with some real function h and some complex function j of u. These functions characterize the particular rotating ray system under consideration, as will be discussed in the subsequent sections. The existence of this coordinate system can be deduced from Kundt's canonical form of the rotating plane-fronted gravitational waves[2] upon recognition[9] of their Kerr–Schild form. The relation to Cartesian inertial coordinates will be given later; for the moment it may suffice to state that the flatness of the metric ds^2 can be checked by calculating its Riemann tensor and that the properties $(\nabla u)^2 = 0 = \Box u$ can be checked directly in terms of the new coordinates. The inverse metric is

$$\left(\frac{\partial}{\partial s}\right)^2 = \partial_\xi^2 + \partial_\eta^2 + \frac{4v}{\xi}\partial_v\partial_\xi + 2\partial_u\partial_v + \left(\xi A - \frac{3v^2}{\xi^2}\right)\partial_v^2, \tag{2.8}$$

where symmetrized tensor products are understood, as in ds^2. Using the appropriate Christoffel symbols, we find

$$u_{;\mu\nu} = \frac{v}{\xi^2}u_{,\mu}u_{,\nu} - (\ln\xi)_{,\mu}u_{,\nu} - (\ln\xi)_{,\nu}u_{,\mu}, \tag{2.9}$$

so that $p' - k'x$ [cf. (2.2)], $e^{-\gamma}$, and ξ may differ from each other only by u-dependent factors. Equations (2.6) now become in the new coordinates

$$\frac{\partial a}{\partial v} = 0, \quad (\partial_\xi^2 + \partial_\eta^2)a = 0, \tag{2.10}$$

so that

$$a = \mathrm{Re}\,\mathscr{F}(u,\xi + i\eta), \tag{2.11}$$

where $\mathscr{F}$ is analytic in the complex variable $\xi + i\eta$.

In the nonrotating case ($u_{;\mu\nu} \equiv 0$) the same formula for a holds with ξ,η replaced by the original Cartesian inertial coordinates. Thus the gain in richness of the solutions considered here is hidden in the functions $j(u)$, $h(u)$ that appear in (2.7). In fact, the new coordinates are quite canonical in the sense that (2.7) is form invariant only under

(a) $u = f(u_1)$,

$v = \frac{v_1}{f'} + \frac{f''}{f'^2}\xi^2$ (f monotonic, three times differentiable),

$$\begin{aligned} h &\to h_1(u_1) = h(f(u_1))f'^2(u_1), \\ j &\to j_1(u_1) = j(f(u_1))f'^2(u_1) + I(f(u_1)), \end{aligned} \tag{2.12}$$

where

$$I(f(\cdot)) := \frac{f'''}{f'} - \frac{3}{2}\left(\frac{f''}{f'}\right)^2 \tag{2.13}$$

is "Schwarz' derivative," to be discussed later (Sec. 3);

$$\text{(b)} \quad \eta = \eta_1 + c \quad (c \text{ a real constant}), \qquad h \to h_1(u) = h(u) - c\,\mathrm{Im}j(u). \tag{2.14}$$

We see that under (a) $h(u)\,du^2$ and $\mathrm{Im}j(u)du^2$ are invariant differential forms, and therefore $h/\mathrm{Im}j$ is invariant under (a), but changed by a constant under (b). We must now distinguish two cases:

I. $\mathrm{Im}j(u) \not\equiv 0$. Here

$$\epsilon := \mathrm{sgn}\,\mathrm{Im}j(u) = \pm 1$$

is an invariant, and

$$\sigma(u) := \int |\mathrm{Im}j(u)|^{1/2} du = \int |\mathrm{Im}j_1(u_1)|^{1/2} du_1 \tag{2.15}$$

is an invariant coordinate, i.e., a distinguished choice for u. From the formal properties of $I(f(\cdot))$, to be listed in Sec. 3, we also find that

$$\epsilon \mathrm{Re}j(\sigma) = \frac{\mathrm{Re}j(u) - I(\sigma(u))}{\mathrm{Im}j(u)} \tag{2.16}$$

is an invariant. Finally, by our remarks above,

$$\frac{d}{d\sigma}\frac{h}{\mathrm{Im}j} = |\mathrm{Im}j(u)|^{-1/2}\frac{d}{du}\frac{h(u)}{\mathrm{Im}j(u)} = \epsilon\frac{dh(\sigma)}{d\sigma} \tag{2.17}$$

is an invariant.

II. $\mathrm{Im}j(u) \equiv 0$. In this case, $h(u)$ is not affected by (b), and by (a) we can reach $\mathrm{Re}j(u) \equiv 0$, i.e., $j(u) \equiv 0$, which determines the coordinate u up to "projective" transformations $u \to (\alpha u + \beta)/(\gamma u + \delta)$ with real $\alpha,\beta,\gamma,\delta$ satisfying $\alpha\delta - \beta\gamma \neq 0$ [see Sec. 3 for the relevant properties of $I(f(\cdot))$]. Under these transformations $h(u)$ remains a "relative" invariant, so that $h(u) \equiv 0$ represents an invariant subcase.

Thus, in order to understand our solutions, it will be important to understand these invariants and the canonical coordinates geometrically. In particular, one has to realize that these coordinates will be valid only in a part of Minkowski space, and will eventually develop singularities. This is connected to the existence of various kinds of envelopes associated with the rotating ray system. We will discuss these topics in the subsequent sections.

C. Yang–Mills fields

Coleman[3] has pointed out the existence of (nonrotating) plane wave solutions in non-Abelian gauge theory. Here we proceed to construct rotating plane-fronted Yang–Mills (YM) waves.

A *null* YM field $\vec{F}_{\mu\nu}$ is defined to be of the form $\vec{F}_{\mu\nu} = \vec{a}_\mu k_\nu - \vec{a}_\nu k_\mu$, $k^2 = 0 = \vec{a}_\mu k^\mu$. We show in Appendix B that Robinson's theorem holds in the non-Abelian case. We also show there that if k_μ is twistfree and thus can be taken as a gradient $u_{,\mu}$ then there is a gauge where the potentials are of the form $\vec{A}_\mu = \vec{a} u_{,\mu}$, where $(\nabla u)^2 = 0$. In this gauge the nonlinear terms in the expression of the field in terms of the potential drop out: $\vec{A}_\mu \times \vec{A}_\nu = 0$, due to the antisymmetry of the structure constants. Thus we may take $\vec{a}_\mu = \vec{a}_{,\mu}$ and must have $\nabla\vec{a}\cdot\nabla u = 0$. It follows that the nonlinear terms in the YM field equations also vanish: $\vec{A}^\nu \times \vec{F}_{\mu\nu} = 0$. This means that we can copy everything about rotating plane-fronted electromagnetic waves from Sec. 2B, putting arrows over the a.

D. Remark on gravitational waves

Rotating plane-fronted gravitational waves were given by Kundt.[2] They are of the same form as in (2.7), but $A(u,\xi,\eta)$ is now of the more general form $\mathrm{Re}\mathscr{F}(u,\xi + i\eta)$, where $\mathscr{F}$ is analytic in $\xi + i\eta$, but not linear. They can be written in an infinity of ways in the "Kerr–Schild form" $ds^2 = ds_0^2 + H du^2$, i.e., as a perturbation of a flat background ds_0^2, which is given by (2.7).[10] This is because we can split off arbitrary portions linear in $\xi + i\eta$ from H. Since basically only the total metric ds^2 is essential, such a splitting could only be useful physically if it were unique. Thus in *general* relativity the invariants considered in Sec. 2B and to be considered later lose their significance.

We also remark here that, because of the form (B8) of the energy–momentum tensor, it is easy to construct combined gravitational and Yang–Mills rotating plane fronted waves, using Kundt's formulas. But we stress that we want to consider only flat space in this article and therefore want to leave out gravitation further on.

3. DIFFERENTIAL INVARIANTS OF THE RAY SYSTEM

For fixed $u = $ const, the rays generate the null hyperplane $k(\tau)x = p(\tau)$. In writing this equation, we introduce a twofold redundancy, as we can reparametrize τ by letting $\tau = f(\tau_1)$ and we can recalibrate k by letting $k = \lambda k_1, p = \lambda p_1$ (a null vector has no preferred Lorentz invariant normalization). We now study the question whether it is possible to find quantities which are Poincaré invariant as well as invariant under reparametrization and recalibration and which characterize our one-parameter family of null hyperplanes $k(\tau)x = p(\tau)$. Note that p has a nontrivial translational behavior. For visualization we imagine some arbitrary inertial observer who observes the three-dimensional ray direction as a point ("searchlight spot") on his "celestial" sphere. This point will then trace a curve on the sphere which we may also study via its stereographic projection. It is well known that a Lorentz transformed observer will see a conformally transformed curved (on his sphere as well as on the stereographic plane); the conformal transformation sends circles into circles.[11] We shall now find a conformally invariant parameter on this curve and a conformally invariant kind of "curvature" that will characterize the curve up to conformal transformations, in the sense of "natural geometry." Finally we will find an invariant characterizing the kinematical aspects of how the point traces over the sphere.

A. The Pick–Liebmann parameter[5]

Consider a curve of null directions: $k = k(\tau), k^2(\tau) = 0$, where $(dk/d\tau)^2 > 0$; then $(' := d/d\tau)$

$$\sigma = \int \frac{|\epsilon_{\mu\nu\alpha\beta} k^{\mu} k'^{\nu} k''^{\alpha} k'''^{\beta}|^{1/2}}{k'^2} d\tau \tag{3.1}$$

is a natural parameter for the curve (in the general case; we shall see the significance of $d\sigma/d\tau \equiv 0$ in a moment), in the sense that it is

(1) manifestly Lorentz invariant

(2) invariant against recalibration $k_\mu(\tau) \to s(\tau) k_\mu(\tau)$

(3) invariant against reparametrization $\tau \to f(\tau)$.

When we introduce the complex coordinate $Y(\tau)$ for the representative point of the null direction in the stereographic plane of an inertial observer as

$$Y := \frac{k_x + ik_y}{k_z + k_t}, \tag{3.2}$$

then under Lorentz transformations, $Y \to (\alpha Y + \beta)/(\gamma Y + \delta)$ with complex constants $\alpha,...,\delta$ satisfying $\alpha\delta - \beta\gamma \neq 0$.

An important quantity in the sequel will be "Schwarz' derivative"

$$I(Y(\tau)) := \frac{Y'''}{Y'} - \frac{3}{2}\left(\frac{Y''}{Y'}\right)^2 \equiv (\ln Y')'' - \tfrac{1}{2}(\ln Y')'^2$$

$$= -2\sqrt{Y'}\left(\frac{1}{(Y')^{1/2}}\right)''. \tag{3.3}$$

Some easily verified properties of it are:

(1') It is invariant under $Y \to (\alpha Y + \beta)/(\gamma Y + \delta)$.

(2') Under reparametrization $\tau_1 = \tau_1(\tau)$, $Y(\tau) = Y_1(\tau_1)$, it behaves as

$$I(Y(\tau)) = I(Y_1(\tau_1))\tau_1'^2 + I(\tau_1(\tau)) \tag{3.4}$$

(3') $I(Y(\tau)) = 0$ iff $Y(\tau) = (\alpha\tau + \beta)/(\gamma\tau + \delta)$ for some complex constants $\alpha,...,\delta$ satisfying $\alpha\delta - \beta\gamma \neq 0$.

(1') and (2') imply that $\int |\mathrm{Im} I(Y(\tau))|^{1/2} d\tau$ has the same invariances as the parameter σ above; in fact, from the stereographic formulas

$$k_x = \sqrt{2}\,\mathrm{Re} Y, \quad k_y = \sqrt{2}\,\mathrm{Im} Y,$$

$$k_z = \frac{1 - |Y|^2}{(2)^{1/2}}, \quad k_t = \frac{1 + |Y|^2}{(2)^{1/2}}, \tag{3.5}$$

one may verify by straightforward calculation that

$$\frac{\epsilon_{\mu\nu\alpha\beta} k^{\mu} k'^{\nu} k''^{\alpha} k'''^{\beta}}{(k'^2)^2} = \mathrm{Im} I(Y(\tau)). \tag{3.6}$$

If this is not equal to 0, its sign ϵ is also invariant. If it is equal to 0, we see that by (3.4) we can reach $\mathrm{Re} I(Y_1(\tau_1)) \equiv 0$ by a suitable choice of τ_1, and then from property (3') of $I(Y(\tau))$ that we can find an observer for whom $Y(\tau) \equiv \tau$: the stereographic point traces the real axis, the celestial point traces a great circle, and thus traces some circle for any other observer as well. Hence σ may be used as a parameter for noncircular curves. (There is no invariant parameter in the circular case, because the Lorentz group acts multiply transitively on circles.)

B. The Thomsen–Takasu invariant[5]

In the noncircular case one can find a curvaturelike invariant, using the preferred parameter σ. When Y is referred to σ, we have $\mathrm{Im} Y(\sigma) = :\epsilon = \pm 1$ and see at once that $\mathrm{Re} I(Y(\sigma))$ is an invariant. In terms of the arbitrary parameter τ, this equals

$$\frac{\mathrm{Re} I(Y(\tau)) - I(\sigma(\tau))}{\mathrm{Im} I(Y(\tau))}, \tag{3.7}$$

and is therefore of the fifth differential order. Specifying it as a function of σ will fix the curve on the celestial sphere up to changes of inertial observer. For instance, it is a constant iff the curve is a (generalized) loxodrome, i.e., an isogonal trajectory of a family of circles that have two points in common (the stereographic image is conformally related to a logarithmic spiral). In this case, its value is related to the angle of intersection.

To express this invariant in a manifestly Lorentz covariant form, $k(\tau)$ is equipped with a preferred calibration; namely, one defines

$$v(\tau) := \frac{|\epsilon_{\mu\nu\alpha\beta} k^{\mu} k'^{\nu} k''^{\alpha} k'''^{\beta}|^{1/2}}{k'^2} \frac{k}{(k'^2)^{1/2}}, \tag{3.8}$$

which is independent of the original calibration and parametrization of k. When referred to σ, we have $v'^2 = 1$, and v''^2 is the lowest order differential invariant one can form. Again a tedious but straightforward verification shows that

$$K := v''^2 = 2\mathrm{Re} I(Y(\sigma)). \tag{3.9}$$

All other differential invariants depend functionally on K and its derivatives. This is because $\epsilon = \det(vv'v''v''')$ $= \pm 1 \neq 0$ says that v, v', v'', v''' form a basis, and all scalar products between them follow once K is known (cf. Table C5) in Appendix C)—hence all expansion coefficients of higher derivatives $v^{iv},\cdots$ may be calculated in terms of K. We shall need the result

$$v^{iv}(\sigma) = (1 - \tfrac{1}{2}K'')v - \tfrac{3}{2}K'v' - Kv''. \tag{3.10}$$

In Appendix C we show that K is indeed minus the invariant defined by Thomsen, who gives a geometric interpretation in terms of certain circles associated with the curve and angles between them.

C. The kinematical invariant

We now want to extract from the function $p(\tau)$ an invariant that characterizes the temporal behavior of the rays. We can use our preferred parameter σ and preferred calibration: $k = v(\sigma)$, i.e., put $p = p(\sigma)$. Since p is a Lorentz scalar, we have only to overcome the difficulty of its nontrivial translational behavior: $p \to p + ka$ under $x \to x + a$. However, this is easy with the help of (3.10): Multiplying it with a, we see that ka is annihilated by the invariant linear operator

$$(\tfrac{1}{2}K'' - 1) + \tfrac{3}{2}K'\frac{d}{d\sigma} + K\frac{d^2}{d\sigma^2} + \frac{d^4}{d\sigma^4},$$

which therefore, when applied to p, produces the desired invariant,

$$p^{iv} + Kp'' + \tfrac{3}{2}K'p' + (\tfrac{1}{2}K'' - 1)p$$

$$\equiv \frac{d}{d\sigma}\Big(p''' + Kp' + \tfrac{1}{2}K'p - \int p d\sigma\Big). \qquad (3.11)$$

An interpretation of it will be given in the next section. In the circular case, where we canot use σ, we have a linear dependence among k, k', k'', k''' which can replace (3.10). Using the calibration (3.5) and a parametrization plus Lorentz frame where $Y(\tau)\equiv\tau$, this relation is simply $k'''\equiv 0$.

4. ENVELOPES ASSOCIATED WITH A ONE-PARAMETER FAMILY OF NULL HYPERSURFACES

In general, our family $k(\tau)x = p(\tau)$ of null hyperplane will have no simple symmetry properties. Nevertheless, the canonical coordinates are remarkably well adapted to it. This is because the ray system has certain singularities in the form of enveloping curves, surfaces and a hypersurface, and the coordinates are tied intimately to these objects. This will be shown in the following section; here we just describe the various envelopes.

Let $k(\tau)x = p(\tau)$, where $k^2(\tau) = 0$, be the family to be considered; if we assume $k\wedge k'\neq 0$ ($\wedge$ indicates exterior multiplication) as before, we have from $kk' = 0$ that $k'^2 > 0$, i.e., k' is spacelike.

(a) Intersecting the hyperplane τ with its neighbor $\tau + \delta\tau$ ($\delta\tau\to 0$), we get a *characteristic* 2*-plane* within each hyperplane, defined by the pair of equations

$$k(\tau)x = p(\tau), \quad k'x = p'. \qquad (4.1)$$

(Note the assumption $k\wedge k'\neq 0$.) Because of $(k\wedge k')^2 = k^2k'^2 - (kk')^2 = 0$, the characteristic 2-planes are null (i.e., touch the light cone at each of their points). As τ varies, they trace a hypersurface whose equation is obtained by eliminating τ between Eqs. (4.1). For those points where $k''x\neq p''$, we may solve $k'x = p'$ for τ as $\tau = \phi(x)$; then on our hypersurface $F(x) := k(\phi(x))x - p(\phi(x)) = 0$, hence $\nabla F = k(\phi(x))$, i.e., the hypersurface touches the hyperplanes along the characteristic 2-planes, thus forming their *enveloping hypersurface* and therefore being null. (It may be characterized by the relation $\theta^2 = |\sigma|^2$ between its expansion and shear.)

[One could think of an exceptional case where the characteristic 2-planes all coincide and do not trace a hypersurface. However, for this to happen one must have $(k\wedge k')'\propto k\wedge k'$, whence $k\wedge k'\wedge k'' = 0$; but $(k\wedge k'\wedge k'')^2 = -(k'^2)^3 < 0$. Thus the characteristic 2-planes cannot even be parallel.]

(b) Intersecting the hyperplane τ with two neighbors gives, in the coincidence limit, a *characteristic line* within each characteristic 2-plane, satisfying

$$k(\tau)x = p(\tau), \quad k'x = p', \quad k''x = p''. \qquad (4.2)$$

Vectors along this line are orthogonal to k, k', k'' and thus parallel to $l := {}^*k\wedge k'\wedge k''$ (where $*$ means dual). One calculates $l^2 = (k'^2)^3 > 0$, thus these lines always exist and are spacelike. On varying τ, these lines generate a 2-surface with normal bivectors $k(\tau(x))\wedge k'(\tau(x))$ at points where $k'''x\neq p'''$. Hence this 2-surface is null and is the envelope of the characteristic 2-planes, touching them along their characteristic lines.

There may be an exceptional case where all characteristic lines coincide and do not generate a surface. Necessary for this is $l'\propto l$ or $(k\wedge k'\wedge k'')'\propto k\wedge k'\wedge k''$, which generally characterizes parallel characteristic lines. This condition is equivalent to $k\wedge k'\wedge k''\wedge k'''\equiv 0$ and will be treated later.

(c) Continuing, we get a *characteristic point* on each characteristic line by the simultaneous solution of

$$k(\tau)x = p(\tau), \quad k'x = p', \quad k''x = p'', \quad k'''x = p'''. \qquad (4.3)$$

Varying τ, this point will in general ($k\wedge k'\wedge k''\wedge k'''\neq 0$) trace a curve $x = x(\tau)$ which makes (4.3) into identities. Differentiating these with respect to τ and using them once more, we find the conditions $kx' = 0, k'x' = 0, k''x' = 0$ and $k^{iv}x + k'''x' = p^{iv}$ for the tangent to this curve. From the first three of them $x' = \lambda\, {}^*k\wedge k'\wedge k'' = \lambda l$, which shows that the curve touches the characteristic lines and is their (spacelike) envelope. λ may be calculated from the last condition, since k^{iv} may be expressed linearly in terms of $k,\ldots,k'''$. Using the invariant parameter σ and the preferred calibration introduced in Sec. 3, we get $x'^2 = \lambda^2$, and from (3.10) that $\lambda\epsilon = p^{iv} + Kp'' + \frac{3}{2}K'p' + (\frac{1}{2}K'' - 1)p$, which is just the kinematical invariant of Sec. 3C. Thus its integral is nothing but the arc length on the envelope of the characteristic lines. The invariants of this curve are expressible in terms of K and λ. In particular, our curve turns out to have vanishing (first) curvature (i.e., null first normals) and characterizes the whole family $k(\tau)x = p(\tau)$ in the following sense: Given a spacelike curve which is not contained in a hyperplane and has vanishing (first) curvature, its tangents sweep over a null 2-surface. The tangent planes of the latter sweep over a null hypersurface whose tangent hyperplanes form a one-parameter family of null hyperplanes.

We now turn to the exceptional cases. One of them occurs when the characteristic point does not trace a curve, but remains fixed, $x'(\tau)\equiv 0$. Then $\lambda = 0$, i.e., the kinematic invariant vanishes. All hyperplanes of the family pass through this point, and shifting it to the origin makes $p(\tau)\equiv 0$. We may refer to this case as the "conical case."

The other exceptional cases arise when the characteristic points do not exist or remain undetermined on the characteristic lines. The condition for this is $k\wedge k'\wedge k''\wedge k'''\equiv 0$, or equivalently, $\det(kk'k''k''') = 0$, i.e., k''' depends linearly on k, k', k''. In this case we have already seen that the characteristic lines are parallel, or coincide if there is the same relation among $p(\tau),\ldots,p'''(\tau)$ as is among $k(\tau),\ldots,k'''(\tau)$ [so that by a translation we can reach $p(\tau)\equiv 0$]. Since we know that there exist inertial observers for which the spatial part $\vec{k}(\tau)$ traces a great circle on their celestial spheres, this case may be called "cylindrical," or "axial" in the subcase.

It should be remarked that our analysis is nothing but the standard treatment of the envelopes associated with a one-parameter family of planes in Euclidean R^3, adapted to four-dimensional Minkowski spacetime and the possibility of null hyperplanes, null 2-planes,...,therein.

5. RELATION BETWEEN CANONICAL AND INERTIAL COORDINATES

We now want to establish the relation between the canonical coordinates (2.7) and Cartesian inertial coordinates. In principle, there is a systematic procedure for this: If ds^2 is flat as checked by calculating its Riemann tensor, one can use the fact that the gradient of a Cartesian inertial coordinate is a geodetically parallel vector field. Thus one can parallelly transport some vierbein from a given point to all others, no matter along what curve, to obtain four independent gradient fields, whose "potentials" are Cartesian inertial coordinates. Using a spinor dyad and the associated null tetrad, this procedure can be reduced to the solution of one single (complex) Riccati equation and some quadratures.[10] In the present case this Riccati equation cannot be solved explicitly, except for some very special choices of $j(u)$, $h(u)$. However, it is possible to find an expression of the canonical coordinates in terms of Cartesian ones which is explicit, except for the fact that $u(x)$ has to be imagined as being a solution $\tau = u(x)$ of $k(\tau)x = p(\tau)$. This is achieved using the fact that there is also a systematic procedure for constructing the canonical coordinates. The most elegant version of it uses Cartan's structure equations of flat space in terms of null tetrads together with some modest amount of exterior form technology and will be presented first. (We are specializing from unpublished work of Plebański.)

A. Construction of the canonical coordinates

We choose a null cotetrad, i.e., four one-forms e^1, e^2, e^3, e^4 such that the metric becomes

$$ds^2 = 2(e^1e^2 + e^3e^4), \quad e^3, e^4 \text{ real}, \quad e^1 = e^{2*} \text{ complex}, \tag{5.1}$$

and orient it such that e^3 is along the ray direction $k_\mu dx^\mu$. The freedom left consists of boosts along $\vec{k}$, together with rotations around it, which may be written

$$e^{1\prime} = e^{i\phi}e^1, \quad e^{2\prime} = e^{-i\phi}e^2, \quad e^{3\prime} = e^{\omega}e^3, \quad e^{4\prime} = e^{-\omega}e^4, \tag{5.2}$$

and of "null rotations"

$$e^{1\prime} = e^1 + C^*e^3, \quad e^{2\prime} = e^2 + Ce^3, \quad e^{3\prime} = e^3,$$
$$e^{4\prime} = e^4 - Ce^1 - C^*e^2 - CC^*e^3. \tag{5.3}$$

Using the notation of Ref. 12, the Cartan structure equations of Minkowski space may be written

$$de^1 = -e^1\wedge\Gamma_{12} - e^3\wedge\Gamma^*_{31} - e^4\wedge\Gamma_{42}, \tag{5.4a}$$
$$de^2 = +e^2\wedge\Gamma_{12} - e^3\wedge\Gamma_{31} - e^4\wedge\Gamma^*_{42}, \tag{5.4b}$$
$$de^3 = e^1\wedge\Gamma^*_{42} + e^2\wedge\Gamma_{42} - e^3\wedge\Gamma_{34}, \tag{5.4c}$$
$$de^4 = e^1\wedge\Gamma_{31} + e^2\wedge\Gamma^*_{31} + e^4\wedge\Gamma_{34}, \tag{5.4d}$$
$$d\Gamma_{42} + \Gamma_{42}\wedge(\Gamma_{12} + \Gamma_{34}) = 0, \tag{5.5a}$$
$$d(\Gamma_{12} + \Gamma_{34}) + 2\Gamma_{42}\wedge\Gamma_{31} = 0 \quad \text{and c.c.} \tag{5.5b}$$
$$d\Gamma_{31} + (\Gamma_{12} + \Gamma_{34})\wedge\Gamma_{31} = 0, \tag{5.5c}$$

where Γ_{ab} are the connection forms: Γ_{34} is real, Γ_{12} is pure imaginary, Γ_{31}, and Γ_{42} are complex. Under (5.2) they transform as

$$\Gamma'_{42} = e^{\omega+i\phi}\Gamma_{42}, \quad \Gamma'_{12} + \Gamma'_{34} = \Gamma_{12} + \Gamma_{34} + d(\omega + i\phi),$$
$$\Gamma'_{31} = e^{-(\omega+i\phi)}\Gamma_{31}, \tag{5.6}$$

while under (5.3) they undergo

$$\Gamma'_{42} = \Gamma_{42}, \quad \Gamma'_{12} + \Gamma'_{34} = \Gamma_{12} + \Gamma_{34} + 2C\Gamma_{42},$$
$$\Gamma'_{31} = \Gamma_{31} + C(\Gamma_{12} + \Gamma_{34}) + C^2\Gamma_{42} + dC. \tag{5.7}$$

The fact that the ray system considered here should be geodesic, twist- shear-, and divergence-free but rotating may be expressed[12] as

$$\Gamma_{42} = \Gamma_{423}e^3, \quad \Gamma_{423} \neq 0. \tag{5.8}$$

From (5.5a) we have $\Gamma_{42}\wedge d\Gamma_{42} = 0$, implying the existence of two complex functions A and B such that $\Gamma_{42} = -AdB$. By (5.8) Γ^*_{42} is proportional to Γ_{42}, hence there is a complex function K such that $dB^* = KdB = KK^*dB^*$, i.e., a real ψ such that $dB^* = e^{2i\psi}dB$, where $d\psi\wedge dB = 0$. Therefore, $e^{i\psi}dB = e^{-i\psi}dB^*$ is real and closed, so locally there is a real function r with $dr = e^{i\psi}dB$. From $\Gamma_{42} = -Ae^{-i\psi}dr$ we can reach by (5.2),

$$\Gamma'_{42} = -dr, \quad e^3 \propto dr. \tag{5.9}$$

In the new frame $d\Gamma'_{42} = 0$, thus (5.5a) gives $\Gamma'_{12} + \Gamma'_{34} \propto \Gamma'_{42}$, and by (5.3) we can reach $\Gamma''_{12} + \Gamma''_{34} = 0$, i.e., $\Gamma''_{12} = 0$, $\Gamma''_{34} = 0$. (5.5b) now gives $\Gamma''_{42}\wedge\Gamma''_{31} = 0$, i.e., $\Gamma''_{31} = -\frac{1}{2}j\,dr$ with some function j, for which we deduce from (5.5c) $0 = d\Gamma''_{31} = -\frac{1}{2}dj\wedge dr$, i.e., $j = j(r)$. By

$$\Gamma''_{42} = -dr, d\Gamma''_{12} = 0 = \Gamma''_{34}, \quad \Gamma''_{31} = -\tfrac{1}{2}j(r)\,dr, \tag{5.10}$$

we have now satisfied (5.5), obtaining simple expressions for the Γ's.

Inserting (5.10) into (5.4), we obtain (observe $e^{3''} \propto dr$)

$$de^{1''} = e^{4''}\wedge dr = de^{2''} \quad \text{and c.c.,} \tag{5.11a,b}$$
$$de^{3''} = -(e^{1''} + e^{2''})\wedge dr, \tag{5.11c}$$
$$de^{4''} = -\tfrac{1}{2}(je^{1''} + j^*e^{2''})\wedge dr. \tag{5.11d}$$

At this point the following easily proved lemma is useful:

Lemma: If r is a function and α is a one-form, then $d\alpha\wedge dr = 0$ is necessary and (locally) sufficient for α to be expressible as $\alpha = da + bdr$ for some functions a, b. Applying this to $e^{1''}$, $e^{2''}$, $e^{4''}$, we may write

$$e^{1''} = d\zeta + (w + f^*)dr \quad \text{and c.c,} \quad e^{4''} = dw - Adr,$$

where ζ, f are complex, w and A real. Inserting this back into (5.11), we get from (5.11c), putting $e^{3''} = gdr$

$$0 = de^{3''} + (d\zeta + d\zeta^*)\wedge dr = d(e^{3''} + (\zeta + \zeta^*)dr),$$

i.e., $(g + \zeta + \zeta^*)dr$ is (locally) a perfect differential $dF(r)$, where F is real. Introducing $\zeta_1 := \zeta - \frac{1}{2}F'$, we get $g = -(\zeta_1 + \zeta^*_1)$, thus

$$e^{3''} = -(\zeta_1 + \zeta^*_1)dr,$$
$$e^{1''} = d\zeta_1 + (\tfrac{1}{2}F'' + w + f^*)dr = d\zeta_1 + (w + f^*_1)dr,$$

where $f_1 = f + \frac{1}{2}F''$. Now (5.11c) is satisfied, and we turn to (5.11a) which gives

$$(dw + df^*_1)\wedge dr = dw\wedge dr, \quad \text{i.e.,} f_1 = f_1(r).$$

We can write

$$e^{1''} = d\zeta_1 + (w + \mathrm{Re}f_1 - i\mathrm{Im}f_1)dr$$
$$= d\left(\zeta_1 - i\int \mathrm{Im}f_1 dr\right) + (w + \mathrm{Re}f_1)dr$$
$$= d\zeta_2 + w_2 dr,$$
$$e^{3''} = -(\zeta_2 + \zeta_2^*)dr, \quad e^{4''} = dw_2 - \tfrac{1}{2}A_2 dr,$$

where

$$\zeta_2 := \zeta_1 - i\mathrm{Im}f_1 dr, \quad w_2 := w + \mathrm{Re}f_1, \quad \tfrac{1}{2}A_2 := A + \mathrm{Re}f_1',$$

and have now satisfied (5.11a,b). Finally, (5.11d) requires

$$0 = (-dA_2 + j(r)d\zeta_2 + j^*(r)d\zeta_2^*)\wedge dr$$
$$= d(-A_2 + j(r)\zeta_2 + j^*(r)\zeta_2^*)\wedge dr,$$

which means that the last bracket equals some function $-\frac{1}{2}h(r)$ of r. This determines the form of A_2, and (5.11d) is now satisfied.

Omitting the indices 2, we conclude that we have rotated the original cotetrad $\{e^1,\cdots\}$ to a new one $\{e^{1''},\cdots\}$ such that there are functions r, w, ζ in terms of which

$$e^{1''} = d\zeta + w dr, \quad e^{2''} = d\zeta^* + w dr,$$
$$e^{3''} = -(\zeta + \zeta^*)dr, \tag{5.12}$$
$$e^{4''} = dw - \tfrac{1}{2}A dr, \quad A := h(r) + j(r)\zeta + j^*(r)\zeta^*.$$

Calculating $e^{1''}\wedge e^{2''}\wedge e^{3''}\wedge e^{4''} = -(\zeta + \zeta^*)d\zeta\wedge d\zeta^*\wedge dr\wedge dw$, we find that r, w, $\mathrm{Re}\zeta$, $\mathrm{Im}\zeta$ may be used locally as coordinates where $\zeta + \zeta^* \neq 0$. If we finally write

$$\zeta = \frac{\xi + i\eta}{(2)^{1/2}}, \quad r = \frac{u}{(2)^{1/2}}, \quad w = -\frac{v}{\xi}, \tag{5.13}$$

$ds^2 = 2(e^{1''}e^{2''} + e^{3''}e^{4''})$ goes over into (2.7). We shall use the coordinates ζ, w, r in the following, however.

B. Relation to Cartesian inertial coordinates

After having seen how to construct canonical coordinates "out of thin air," i.e., out of any given null cotetrad with $e^3 \propto k_\mu dx^\mu$, we can establish the relation to inertial coordinates by making a suitable choice for the original $\{e^a\}$ in terms of inertial coordinates. Following Ref. 12, we write the ray vectors k as in (3.5), and introduce Cartesian null coordinates.

$$Z = \frac{x + iy}{(2)^{1/2}}, \quad Z^*, \quad U = \frac{z + t}{(2)^{1/2}}, \quad V = \frac{z - t}{(2)^{1/2}}. \tag{5.14}$$

Then

$$k_\mu dx^\mu = dU - |Y|^2 dV + Y^* dZ + Y dZ^* \tag{5.15}$$

and

$$ds^2 = dx^2 + dy^2 + dz^2 - dt^2 \equiv 2(dZdZ^* + dUdV)$$
$$\equiv 2((dZ - YdV)(dZ^* - Y^*dV) + dV(dU - |Y|^2dV + Y^*dZ + YdZ^*)). \tag{5.16}$$

Thus we may take as the initial cotetrad

$$e^1 = dZ - Y dV, \quad e^2 = \text{c.c.}, \quad e^3 = kdx, \quad e^4 = dV. \tag{5.17}$$

One finds for the connection forms (cf. Ref. 12)

$$\Gamma_{42} = -dY, \quad \Gamma_{12} + \Gamma_{34} = 0, \quad \Gamma_{31} = 0. \tag{5.18}$$

Our rotating family of null hyperplanes $k(\tau)x = p(\tau)$ is given by specifying $p(\tau)$ and $Y(\tau)$. If τ is used as coordinate, $\tau = \tau(x)$ has to be taken from this implicit equation. Thus $Y = Y'd\tau$, and $\Gamma_{42} = -Y'd\tau$ may be transformed into $\Gamma'_{42} = -d\tau$ by (5.2) and (5.6) with $\omega + i\phi = -\ln Y'$, so that we may take $r = \tau(x)$, as expected. We shall write

$$Y' = Re^{-i\psi}, \quad R > 0, \quad \psi \text{ real}. \tag{5.19}$$

In the new cotetrad, $\Gamma'_{12} + \Gamma'_{34} = -(\ln Y')'d\tau, \Gamma'_{31} = 0$, and we remove $\Gamma'_{12} + \Gamma'_{34}$ by (5.3) and (5.7) with $C = -\frac{1}{2}Y''/Y'$. This gives

$$e^{1''} = e^{i\psi}e^1 - \frac{1}{2}\frac{Y^{*''}}{Y^{*'}}\frac{e^3}{R}, \quad e^{3''} = \frac{e^3}{R}, \tag{5.20}$$
$$e^{4''} = \mathrm{Re}^4 + \frac{1}{2}\frac{Y''}{Y'}e^{i\psi}e^1 + \frac{1}{2}\frac{Y^{*''}}{Y^{*'}}e^{-i\psi}e^2 - \frac{1}{4}\left|\frac{Y''}{Y'}\right|^2\frac{e^3}{R},$$

and the connection forms

$$\Gamma''_{42} = -d\tau, \quad \Gamma''_{12} + \Gamma''_{34} = 0, \quad \Gamma''_{31} = -\tfrac{1}{2}I(Y(\tau))d\tau, \tag{5.21}$$

where again Schwarz' derivative appears. Now we can read off the function j and see immediately that one of the invariants encountered in Sec. 2B agrees with the Thomsen–Takasu invariant of Sec. 3. To see the agreement between the other ones, we must fullow through identification of the coordinates, comparing (5.20) with (5.12). From $-(\zeta + \zeta^*)dr = e^{3''} = kdx/R$ we read off, using $kx = p$, $kdx = (p' - k'x)d\tau$

$$\zeta + \zeta^* = -\frac{p' - k'x}{R} = e^{i\psi}(Z - YV) + \text{c.c.} - \frac{p'}{R}. \tag{5.22}$$

Next observe that $e^{1''} - e^{2''}$ should be a perfect differential $d(\zeta - \zeta^*)$,

$$e^{1''} - e^{2''} = e^{i\psi}(dZ - YdV) - \text{c.c.}$$
$$- i\psi'\left(\frac{p'}{R} - 2\mathrm{Re}e^{i\psi}(Z - YV)\right)d\tau$$
$$= d\left[e^{i\psi}(Z - YV) - \text{c.c.} - i\int\frac{p'\psi'}{R}d\tau\right],$$

where we have expressed Y''/Y' in terms of (5.19). Thus

$$\zeta = e^{i\psi}(Z - YV) - \frac{p'}{2R} - \frac{i}{2}\int\frac{\psi' p'}{R}d\tau \tag{5.23}$$

up to an imaginary constant of integration.

We now calculate $w = (e^{1''} - d\zeta)/d\tau$ to find

$$w = RV - \frac{p'R'}{R^2} + \frac{p''}{2R} + \frac{R'}{R}\mathrm{Re}[e^{i\psi}(Z - YV)]$$
$$+ \psi'\mathrm{Im}[e^{i\psi}(Z - YV)]$$
$$= \frac{p'' - k''x}{2R} - \frac{R'}{R^2}(p' - k'x). \tag{5.24}$$

This already ends the determination of the coordinate transform, and the comparison of both expressions for $e^{4''}$ only serves to express $h(\tau)$ in terms of k, i.e., of Y. After some calculation, expressing everything in terms of R, ψ, one finds

$$h(\tau)=R\left(\frac{1}{R}\left(\frac{p'}{R}\right)'\right)'+\frac{p'}{R}\psi'^2+R\left(\frac{\psi'}{R}\right)'\int\frac{\psi'p'}{R}d\tau. \tag{5.25}$$

To see that this leads indeed to the kinematical invariant (3.11), we assume that τ has been chosen to be the distinguished parameter σ (in the noncircular case). In terms of R, ψ, this means

$$\frac{R'\psi'}{R}-\psi''=-R\left(\frac{\psi'}{R}\right)'=\operatorname{Im}I(Y(\sigma))=\epsilon=\pm 1, \tag{5.26}$$

while the Thomsen–Takasu invariant is

$$K=2\operatorname{Re}I(Y(\sigma))=2\left(\frac{R''}{R}-\frac{3}{2}\left(\frac{R'}{R}\right)^2\right)+\psi'^2.$$

When we observe that p in (3.11) corresponds to k calibrated such that $k'^2=1$ we see that we have to replace p with Rp in (5.25) before comparing with (3.11). Now a partial integration and the use of (5.26) at several places leads, after some algebra, to the coincidence of (5.25) with the bracketed expression in (3.11). [In the circular case we can take $Y(\tau)\equiv\tau$, $\psi\equiv 0$, which makes $k'''=0$, corresponding to $h=p'''$.]

We can now discuss the significance of the canonical coordinates. By construction, $\tau=r$ describes the various null hyperplanes $kx=p$. $\operatorname{Re}\zeta=0$ gives their envelope, whereas $r=$ const, $\operatorname{Re}\zeta=$ const are 2-planes parallel to the characteristic 2-plane within each null hyperplane. [Another interpretation of $\operatorname{Re}\zeta$ follows from (5.8) which gives $\Gamma''_{423}=(2\operatorname{Re}\zeta)^{-1}$, since in Ref. 6 it is shown generally that $|\Gamma_{423}|$ is the angular velocity of the spatial ray direction with respect to a geodesic observer; this can, of course, be checked directly for our ray system.] $w=$ const cuts out from each of these parallel planes of family of lines parallel to the characteristic line. $\operatorname{Im}\zeta$ is the arc length on each of them. These results follow from (4.1), (4.2), (5.23), (5.24), and (2.7). It is therfore geometrically clear that such a coordinate system must have singularities.

APPENDIX A: SHEARFREE AND NONDIVERGING NULL HYPERSURFACES IN COMPLEX MINKOWSKI SPACE, AND COMPLEX FUNCTIONALLY INVARIANT SCALAR WAVES

Erugin[7] has obtained the complex functionally invariant solutions of the scalar wave equation and has, in particular, pointed out the existence of solutions not given by the complex version (2.1). We rederive his results here by a geometric method borrowed from Ref. 8, thereby achieving considerable simplification, avoiding to discriminate between several cases. Since this method continues to consider $\Phi(x)=c$ as the equation of a hypersurface even for complex-valued Φ, we must assume $\Phi(x)$ to permit a differentiable extension to complexified space–time, i.e., we must restrict ourselves to analytic solutions. (It seems to us that Erugin has made this assumption somewhere in his paper as well without mentioning it.)

To study the simultaneous solutions of $(\nabla\Phi)^2=0$, $\Box\Phi=0$, we consider the two-dimensional wave surfaces cut out from $\Phi=$ const by hypersurfaces $t=$ const ($t\cdots$ inertial time in the real case, nonnull Cartesian coordinate in the complex case) which carry ordinary Euclidean geometry. It is known (although we are aware only of Refs. 10 and 13 and three-dimensional expressions in Refs. 14 and 15 from which it is easily deduced), that vanishing shear for the hypersurfaces $\Phi=c$ means $H^2=K$, and vanishing divergence means $H^2=K=0$, where H and K are the mean and Gaussian curvature, resp., of the two-dimensional wave surfaces. Standard texts on surfaces in *real* Euclidean space now inform us that $H^2=K\neq 0$ implies that the wave surfaces are spheres, while $H^2=K=0$ requires plane wave surfaces. The corresponding shearfree hypersurfaces are therefore light cones in the diverging, and null hyperplanes in the nondiverging case. Some texts, e.g., Ref. 16, also contain the results for complex Euclidean space: $H^2=K\neq 0$ characterizes Monge surfaces, i.e., metrically nondegenerate surfaces ruled by complex, null, nonparallel straight lines, while $H^2=K=0$ characterizes "isotropic cylinders," cylinders with complex null generators. Spheres and planes are special cases. [One can obtain the results in the complex case by referring the surface to null coordinates: then the Gauss equation(s) associated with the zero(s) in the matrix of the second fundamental form following from $H^2=K(=0)$ yield straightness (and constant direction) for the null coordinate lines of one of the two families.]

To obtain the null hypersurface whose $t=0$ section is a given isotropic cylinder we proceed as follows. Let l be a 4-vector parallel to the generators of the cylinder. Then l is a null vector also in the four-dimensional sense. If $y(\alpha)$ describes any curve (different from a generator) on the cylinder, then the normals k of the hypersurface are null and also tangent and have to be chosen orthogonal to $dy/d\alpha$ and to l. Since it is generated by its null tangents, the hypersurface will have a parametric representation of the form

$$x(\alpha,\beta,\gamma)=y(\alpha)+\beta l+\gamma k(\alpha), \tag{A1}$$

where the first two terms on the right describe the cylinder. To pass from a parametric representation to an equation, i.e., to eliminate α,β,γ, we appeal to the fact that all null vectors orthogonal to the given fixed null vector l are contained in two fixed 2-planes (one self-dual, the other anti-self-dual) each of which we imagine as spanned by l and another fixed vector m satisfying $m^2=0=ml$. Hence $k(\alpha)$ has the form $k(\alpha)=\lambda(\alpha)l+\mu(\alpha)m$, and since it varies differentiably, it must stay in one of the planes. Therefore, we may easily eliminate β,γ from (A1) by forming

$$lx=ly(\alpha),\quad mx=my(\alpha), \tag{A2}$$

and may now eliminate α in the form $f(lx,mx)=0$.

A family of null hypersurfaces of this kind is obtained by considering a family of isotropic cylinders, i.e., taking a family of curves $y(\alpha,\tau)$ and a family of null vectors $l(\tau)$. Then, of course, we must also have $m=m(\tau)$, satisying $m^2(\tau)=m(\tau)l(\tau)=l^2(\tau)=0$. The above elimination procedure leads to $f(l(\tau)x,m(\tau)x,\tau)=0$, and if this is to be the same hypersurface as $\Phi(x)=\tau$ for all τ, then $\Phi(x)$ is given by the implicit equation

$$f(l(\Phi)x,m(\Phi)x,\Phi)=0. \quad \text{(A3)}$$

One may check by implicit differentiation that Φ satisfies $(\nabla\Phi)^2=0, \Box\Phi=0$. In the real case, l and m must be proportional, which gets us back to (2.1). If, in the complex case, l and m are constant, one gets Erugin's explicit solutions $\Phi=\phi(lx,mx)$. Erugin's implicit solutions, not given in Ref. 1 because of their complicated form, correspond to the general case (A3).

APPENDIX B: YM NULL FIELDS AND ROBINSON'S THEOREM

We call a YM field null iff all isospin components of the field are null with the same degenerate principal null direction. The field can then be written $\vec{F}_{\mu\nu}=\vec{a}_\mu k_\nu-\vec{a}_\nu k_\mu$, where $k^2=0, \vec{a}_\mu k^\mu=0$. From this, one derives that the symmetric spinor $\vec{\Phi}_{MN}$ corresponding to $\vec{F}_{\mu\nu}$ (cf. Ref. 17 for spinors) must be of the form

$$\vec{\Phi}_{MN}=\vec{f}\kappa_M\kappa_N, \quad \text{(B1)}$$

where κ is the spinor corresponding to the null vector k. It has been pointed out[18] that the source-free YM equations imply the geodesic-shearfree condition of the null congruence tangent to k. Indeed, these equations have the spinor form $D^{MM'}\vec{\Phi}_{MN}=0$, where $D^{MM'}$ is the gauge covariant (in curved space, the combined Riemannian and gauge covariant) derivative in spinor form. Inserting (B1), we get ($\nabla\cdots$ Riemannian covariant derivative)

$$0=D^{MM'}(\vec{f}\kappa_M\kappa_N)$$
$$=\kappa_M\kappa_N D^{MM'}\vec{f}+\vec{f}(\kappa_M\nabla^{MM'}\kappa_N+\kappa_N\nabla^{MM'}\kappa_M). \quad \text{(B2)}$$

Transvecting with κ^N, we arrive at the geodesic-shearfree condition

$$\kappa^N\kappa_M\nabla^{MM'}\kappa_N=0 \quad \text{(B3)}$$

(Refs. 6 and 18.) Thus the non-Abelian character of the YM theory plays no role in the argument, since the unknown potential enters only the $D^{MM'}\vec{f}$ term in (B2) which drops out.

One would like to know whether the converse statement (and thus *Robinson's theorem*) will hold in the non-Abelian case as well, namely that from a geodesic-shearfree k one could always construct a sourceless null YM field by a suitable choice of $\vec{a}_\mu$ (or $\vec{f}$). We shall show now that this is indeed the case. (B3) implies the existence of some $g^{M'}$ such that

$$\kappa_M\nabla^{MM'}\kappa_N=g^{M'}\kappa_N. \quad \text{(B4)}$$

We want to show that the system

$$\kappa_M D^{MM'}\vec{f}+\vec{f}(g^{M'}+\nabla^{MM'}\kappa_M)=0 \quad \text{(B5)}$$

has a solution $\vec{f}$, because multiplying (B5) with κ_N and using (B4) then gets us back to (B2). The existence of solutions to (B5) depends on the question whether the integrability condition to it is satisfied identically in $\vec{f}$. Spinorially, this is obtained by applying $\kappa^N D_{NM'}$ to (B5). Using (B5) and (B4), it can be written as

$$\kappa^M\kappa^N D_N{}^{M'}D_{MM'}\vec{f}+\vec{f}\{g_{M'}\nabla^{MM'}\kappa_M$$
$$+\kappa^N\nabla_{NM'}g^{M'}+\kappa^N\nabla_{NM'}\nabla^{MM'}\kappa_M\}=0 \quad \text{(B6)}$$

Now the curly brackets vanishes on the basis of (B4)—nothing of the non-Abelianess enters here— as one can see by some straightforward spinor algebra (and, in the curved space case, using a spinorial Ricci identity). To treat the first term, we note that $\vec{f}$ is a space–time scalar, such that the spinor form of the relevant isospin Ricci identity is just [(...) means symmetrization]

$$D_{(N}{}^{M'}D_{M)M'}\vec{f}\equiv\vec{\Phi}_{MN}\times\vec{f}, \quad \text{(B7)}$$

which vanishes by (B1). Thus also the first term in (B6) is zero identically in $\vec{f}$, Q.E.D. {Vectorially, (B5) is equivalent to a system of form $D_k\vec{f}=K\vec{f},\ D_m\vec{f}=M\vec{f}$, where m_μ is the vector corresponding to $\kappa_M\bar{\mu}^*_{M'}$, μ_M being a spinor satisying $\kappa_M\mu^M=1$; D_k,D_m are covariant directional derivatives. Integrability is checked using the commutator $[D_k,D_M]\vec{f}$, which by the relevant isospin Ricci identity equals $D_{[k,m]}\vec{f}+k^\mu m^\nu\vec{F}_{\mu\nu}\times\vec{f}$, where $[k,m]$ is the Lie bracket. By the definition of the null field, $\vec{F}_{\mu\nu}k^\mu=0$, which reduces the problem to the Abelian case.}

It has been pointed out[18] that for null fields with a twist-free (i.e., hypersurface orthogonal) ray congruence k a particularly simple gauge for the potential may be found. Indeed, in this case k_μ is (proportional to) a gradient $u_{,\mu}$, such that the YM curvature form $\vec{F}:=\frac{1}{2}\vec{F}_{\mu\nu}dx^\mu\wedge dx^\nu$ may be written as $\vec{a}\wedge du$, vanishing upon restriction to the hypersurfaces $u=$ const. Thus the potential form $\vec{A}_\mu dx^\mu=:\vec{A}$ may be gauged away within each of these, i.e., may be gauged to be $\vec{A}=\vec{a}du$, whence $\vec{F}=d\vec{a}\wedge du$.

For null fields, the energy–momentum tensor $T_{\mu\nu}$ becomes

$$T_{\mu\nu}=-\vec{a}_\lambda\vec{a}^\lambda k_\mu k_\nu, \quad \text{(B8)}$$

where a scalar product in the sense of an invariant group metric is understood. For any observer with 4-velocity v^ν, the Poynting 4-vector $S_\mu=T_{\mu\nu}v^\nu$ of the null field is lightlike. It has been remarked[18] that the converse is also true (if the invariant group metric used is definite).

APPENDIX C: THOMSEN'S INVARIANT

Thomsen defines his conformal invariant of a plane (or spherical) curve as follows. (We give only the formal description without geometrical motivation.) Start from $v(\sigma)$ as defined in (3.8), satisfying $v'^2=1$. Let the 4-vector $y(\sigma)$ be a solution of the conditions

$$yv=yv'=yv''=0,\quad y^2=1. \quad \text{(C1)}$$

Define

$$x:=v+y. \quad \text{(C2)}$$

Let $\bar{v}$ be the solutin of the conditions

$$\bar{v}^2=\bar{v}x=\bar{v}x'=0,\quad \bar{v}v=1. \quad \text{(C3)}$$

Then the invariant is

$$b:=-2(vx'')(\bar{v}x''). \quad \text{(C4)}$$

We wish to verify that $b=-v''^2=-K$.

First we assemble the following table of scalar products between v, v', v'', v''' by differentiating the relations $v^2 = 0$, $v'^2 = 1$:

	v	v'	v''	v'''
v	0	0	-1	0
v'	0	1	0	$-K$
v''	-1	0	K	$\frac{1}{2}K'$
v'''	0	$-K$	$\frac{1}{2}K'$	K^2+1

(C5)

The entry $v'''^2 = K^2 + 1$ follows from $\det(vv'v''v''') = \epsilon = \pm 1$ using $1 = (\det)^2 = -$ determinant of table $= v'''^2 - K^2$. Differentiating the determinant, we find $\det(vv'v''v^{iv}) = 0$, i.e., v^{iv} is a combinatin of v, v', v'' with coefficients that are found from the scalar products vv^{iv}, $v'v^{iv}$,$v''v^{iv}$, which in turn are obtained by differentiating the last line of the table. This way we obtain (3.10).

Next we need y of (C1), which obviously must be a multiple of ${}^*v \wedge v' \wedge v''$, whose square is just (minus the first upper 3×3 minor of the table) $= 1$. Thus

$$y = {}^*v \wedge v' \wedge v'', \tag{C6}$$

$$y' = {}^*v \wedge v' \wedge v''', \tag{C7}$$

$$\begin{aligned} y'' &= {}^*(v \wedge v'' \wedge v''' + v \wedge v' \wedge v^{iv}) \\ &= {}^*v \wedge v'' \wedge v''' - Ky, \end{aligned} \tag{C8}$$

where in the last step we have used (3.10). y' satisfies

$$y'v' = 0, \quad y'^2 = 0 \tag{C9}$$

(the second equation comes from another minor of the table). Thus when x is defined as in (C2), it satisfies [by (C5), (C1), (C9)]

$$x^2 = 1, \quad x'^2 = 1 \tag{C10}$$

as required in Thomsen's procedure. Also, from (C5) and (C8),

$$vx'' = -1. \tag{C11}$$

Making an ansatz

$$\bar{v} = \alpha v + \beta v' + \gamma v'' + \delta v''', \tag{C12}$$

the conditions (C3) in the order $\bar{v}v = 1$, $\bar{v}x = 0$, $\bar{v}x' = 0$, $\bar{v}^2 = 0$ determine the coefficients as (use $\bar{v}y = \delta v'''y = -\delta\epsilon$, $\bar{v}y' = \gamma v''y' = \gamma\epsilon$)

$$\begin{aligned} &\gamma = -1, \quad \delta = \epsilon (= \pm 1), \\ &\beta = \epsilon(K+1), \quad 2\alpha = \epsilon K' - K - 2. \end{aligned} \tag{C13}$$

Inserting (C12), (C13), (C11), into (C4) yields the desired result

$$b = -K. \tag{C14}$$

[1] N.S. Koshlyakov, M.M. Smirnov, and E.B. Gliner, *Differential Equations of Mathematical Physics* (North-Holland, Amsterdam, 1964).
[2] W. Kundt, Z. Phys. **163**, 77 (1961).
[3] S. Coleman, Phys. Lett. B **70**, 59 (1977).
[4] I. Robinson, J. Math. Phys. **2**, 290 (1961).
[5] W. Blaschke and G. Thomsen, *Differentialgeometrie* (Springer-Verlag, Berlin, 1929), Vol. III.
[6] P. Jordan, J. Ehlers, and R.K. Sachs, Abh. Math.-Naturwiss. Kl. Akad. Wiss. Lit. Mainz. **1961d**, 1.
[7] N.P. Erguin, Uchenye Zapiski, Leningrad State University, Math. Sci. Ser., No. 15 (1948) (In Russian; we are indebted to E.B. Gliner for supplying a copy of this reference.)
[8] F.G. Friedlander, Proc. Cambridge, Philos. Soc. Math. Phys. Sci. **43**, 360 (1947).
[9] H.K. Urbantke, Acta Phys. Austriaca **35**, 396 (1972).
[10] H.K. Urbantke, Acta Phys. Austriaca **41**, 1 (1975).
[11] R. Penrose, Proc. Cambridge Philoc. Soc. Math. Phys. Sci. **55**, 137 (1959);
[12] .W. Misner, K.S. Thorne, and J.A. Wheeler, *Gravitation* (Freeman, San Francisco, 1973), Chap. 41.
[12] G.C. Debney, R.P. Kerr, and A. Schild, J. Math. Phys. **10**, 1842 (1969).
[13] R. Kantowski, J. Math. Phys. **9**, 336 (1968); R. Geroch, A. Held, and R. Penrose, *ibid.* **14**, 874 (1973).
[14] A. Duschek, *Höhere Mathematik* (Springer-Verlag, Vienna, 1958), Vol. II, 2nd ed., p. 361.
[15] *Die Hütte* (W. Ernst&Sohn, Berlin, 1948), Vol. **I**, 27th ed., p. 161.
[16] K. Strubecker, *Differentialgeometrie* (Sammlung Göschen, de Gruytes, Berlin, 1959), Vol. III, Nr. 1180/1180a.
[17] F.A.E. Pirani, in *Brandeis Summer Institute* 1964, edited by S. Deser and K. Ford (Prentice-Hall, Englewood Cliffs, N.J.,1965), Vol. **I**.
[18] R. Güven, Middle East Technical University, Ankara, Turkey, preprint (June 1978) (unpublished).

The role of the Onsager–Machlup Lagrangian in the theory of stationary diffusion process

Kunio Yasue
Département de Physique Théorique, Université de Genève, CH-1211 Genève, Swiss

(Received 9 February 1979)

The Onsager–Machlup Lagrangian is shown to have a direct relevance to a cost function for a stochastic control problem. It is found that any stationary diffusion process can be regarded as a solution to the stochastic control problem, that is, it is controlled optimally by the Onsager–Machlup Lagrangian. A deterministic limit of the stationary diffusion process is also obtained as a solution to an ordinary (nonrandom) control problem which is equivalent to the usual variational problem with respect to the Onsager–Machlup Lagrangian.

I. INTRODUCTION

In recent nonequilibrium statistical mechanics one encounters the problem of finding deterministic paths (or most probable paths) for dynamical systems described by stochastic differential equations of Itô type (or equivalently by Fokker–Planck equations). I shall call it the Onsager–Machlup (OM) problem hereafter.

Beginning with Onsager and Machlup's monumental work,[1] many authors have approached the OM problem with different methods. They seem to be classified mainly into three groups, that is, those who utilized the path integral formalism,[2–5] those who worked with the canonical operator formalism,[6–8] and those who preferred the probabilistic formalism.[9,10] I have also proposed a fundamental approach to the OM problem in the light of Onsager and Machlup's original spirit.[11]

In the present paper, reformulating the OM problem as a stochastic control problem, I will show a basic role of the OM Lagrangian in the theory of stationary diffusion process.

The main sources for the stochastic calculus fully utilized in this paper were Nelson's lecture note[12] and Itô's papers.[13,14]

II. STATIONARY DIFFUSION PROCESS

By the notion of an n-dimensional stationary diffusion process, I denote an $\mathbb{R}^n$-valued Markov process $X(t)$, $t\in[0,\infty)$, described by a stochastic differential equation of Itô type,

$$dX = a(X)dt + dB, \quad X(0) = x_0, \tag{1}$$

where the drift vector field a is assumed to be bounded and of class C^2, $B(t)$, $t\in[0,\infty)$, denotes an n-dimensional standard Brownian motion (i.e., an $\mathbb{R}^n$-valued Wiener process with diffusion constant equal to unity), and $x_0\in\mathbb{R}^n$. It is a Markov process with invariant measure $\mu(d^n x)$ (i.e., stationary probability distribution). A probability distribution of $X(t)$, $p^t(d^n x) = \text{Prob}\{X(t)\in d^n x\}$, solves the Fokker–Planck equation,

$$\frac{\partial}{\partial t}p^t = -\operatorname{div}(ap^t) + \operatorname{div}\operatorname{grad} p^t, \tag{2}$$

weakly, and it has an asymptote $\lim_{t\to\infty}p^t = \mu$ in $L_\infty(\mathbb{R}^n)$.

Let $\{\mathscr{F}_t\}_{t\in[0,\infty)}$ be an increasing family of σ-algebras of measurable events such that $X(t)$ is $\mathscr{F}_t$-measurable for each $t\in[0,\infty)$. Then one can introduce the notion of mean derivative of $X(t)$ by

$$DX(t) = \lim_{h\downarrow 0} E\left[\frac{X(t+h)-X(t)}{h}\,\middle|\,\mathscr{F}_t\right], \tag{3}$$

where $E[\cdot|\mathscr{F}_t]$ means to take a conditional expectation with respect to the σ-algebra $\mathscr{F}_t$. Equations (1) and (3) give us

$$DX(t) = \lim_{h\downarrow 0} E\,[dX((t,t+h])/h\,|\mathscr{F}_t] = a(X(t)). \tag{4}$$

The mean derivative of $f(X(t))$, where f is a function of class C^2, is also defined

$$Df(X(t)) = \lim_{h\downarrow 0} E\left[\frac{f(X(t+h)) - f(X(t))}{h}\,\middle|\,\mathscr{F}_t\right], \tag{5}$$

obtaining

$$\begin{aligned} Df(X(t)) &= \lim_{h\downarrow 0} E\,[df(X)((t,t+h])/h\,|\mathscr{F}_t] \\ &= (a\cdot\operatorname{grad} f + \operatorname{div}\operatorname{grad} f)(X(t)). \end{aligned} \tag{6}$$

Here I have made use of the chain rule in stochastic calculus[13] and of Eq. (1).

A considerable number of stationary diffusion processes with which physicists have encountered are of the gradient type, i.e., the drift vector field is written

$$a = -\operatorname{grad} A, \tag{7}$$

where A is a positive bounded $\mathbb{R}$-valued function of class C^3. I shall restrict myself to such a case as Eq. (7).

Now, following Nelson,[12] I shall define the mean velocity and the mean acceleration of the stationary diffusion process $X(t)$, $t\in[0,\infty)$, by

$$DX(t), \quad t\in[0,\infty) \tag{8}$$

and

$$D^2X(t), \quad t\in[0,\infty), \tag{9}$$

respectively. A straightforward calculation gives us

$$\begin{aligned} D^2X(t) &= DDX(t) \\ &= (a\cdot\operatorname{grad} a + \operatorname{div}\operatorname{grad} a)(X(t)) \end{aligned}$$

 © 1979 American Institute of Physics

$$= \text{grad}(\tfrac{1}{2}|a|^2 + \text{div}a)(X(t)), \quad (10)$$

where the gradient property of the drift vector is essential in deriving the last expression in Eq. (10). Then a question arises: What does Eq. (10) tell us about? This is the very starting point of the present approach to the OM problem which will be developed in the following sections.

III. STOCHASTIC CONTROL PROBLEM

Prior to proceeding with the OM problem it seems worthwhile to make here a remark on the role of Newton's equation of motion

$$(\text{mass})\times(\text{acceleration}) = (\text{force}). \quad (11)$$

It has complementary characters; on the one hand, when (mass) and (force) are known a trajectory of the particle should be obtained by integrating (acceleration), on the other hand, when (mass) and (acceleration) are known (force) acting on the particle should be found (mass) times (acceleration). The latter property of the equation of motion gives us the following reformulation of the OM problem.

Let us introduce an $\mathbb{R}$-valued function

$$V_a = -\tfrac{1}{2}|a|^2 - \text{div}a, \quad (12)$$

call it OM potential. Regarding the boundedness of a, I can make V_a negative by adding a constant. So V_a is assumed to be negative hereafter. Then one may interpret from Eq. (10) that the stationary diffusion process $X(t)$, $t\in[0,\infty)$, is controlled as if it obeys the equation of motion

$$D^2X(t) = -\text{grad}V_a(X(t)). \quad (13)$$

By the notion of a reduced OM Lagrangian, I denote an $\mathbb{R}$-valued function

$$L'_{\text{OM}}(z(t),\dot z(t)) = \tfrac{1}{2}|\dot z(t)|^2 - V_a(z(t)), \quad (14)$$

where $z(t)$, $t\in[0,\infty)$, is a path of class C^1 in $\mathbb{R}^n$ and $\dot z(t) = dz(t)/dt$. The theorem below allows us to reformulate the OM problem as a stochastic control problem.

Theorem 1: *The stationary diffusion process* $X(t)$, $t\in[0,\infty)$, *described by* Eq. (1) *is a solution to the stochastic control problem,*

$$dZ = b(Z)dt + dB, \quad Z(0) = x_0, \quad (15)$$

where the drift vector b of class L (Lipschitz functions) is controlled to minimize a cost function

$$\text{Cost}(b) = E\left[\overline{\lim_{T\to\infty}}\frac{1}{T}\int_0^T L'_{\text{OM}}(Z(t),DZ(t))dt\right], \quad (16)$$

and $E[\cdot]$ *means to take an expectation.*

It is worthwhile to notice that the theorem is essentially the same statement as that of Holland,[15] and here I follow his proof.

Proof: I have an identity

$$\text{div grad}A - \tfrac{1}{2}|\text{grad}A|^2 - V_a = 0, \quad (17)$$

which can be written

$$\text{div grad}A + \min_b(b\cdot\text{grad}A + \tfrac{1}{2}|b|^2 - V_a) = 0, \quad (18)$$

where the minimum is taken over all $\mathbb{R}^n$-valued functions of class L. Then for any b, I have an inequality

$$\text{div grad}A + b\cdot\text{grad}A + \tfrac{1}{2}|b|^2 - V_a \geqslant 0. \quad (19)$$

Substituting $Z(t)$ in the inequality, integrating over $[0,T]$ and dividing it by T, I obtain

$$\frac{1}{T}\int_0^T(b\cdot\text{grad}A + \text{div grad}A)(Z(t))dt + \frac{1}{T}\int_0^T(\tfrac{1}{2}|b|^2 - V_a)(Z(t))dt \geqslant 0. \quad (20)$$

The chain rule in stochastic calculus gives us

$$\int_0^T(b\cdot\text{grad}A + \text{div grad}A)(Z(t))dt = \int_0^T dA(Z(t)) = A(Z(T)) - A(x_0), \quad (21)$$

which yields

$$\frac{A(Z(T)) - A(x_0)}{T} + \frac{1}{T}\int_0^T(\tfrac{1}{2}|b|^2 - V_a)(Z(t))dt \geqslant 0. \quad (22)$$

Taking the expectation I find

$$\frac{E[A(Z(T))] - A(x_0)}{T} + E\left[\frac{1}{T}\int_0^T(\tfrac{1}{2}|b|^2 - V_a)(Z(t))dt\right] \geqslant 0. \quad (23)$$

Now what is left for us is to pass to the limit $T\to\infty$, obtaining

$$E\left[\overline{\lim_{T\to\infty}}\frac{1}{T}\int_0^T(\tfrac{1}{2}|b|^2 - V_a)(Z(t))dt\right] \geqslant 0. \quad (24)$$

This claims $\text{Cost}(b)\geqslant 0$. To complete the proof I claim

$$DZ(t) = b(Z(t)), \quad (25)$$

and that the cost function is minimized by a, that is,

$$\text{Cost}(a) = 0. \quad (26)$$

This is because the equality in Eq. (20) is achieved by setting $b = -\text{grad}A = a$. QED.

I have found a basic role of the reduced OM Lagrangian (14) in the theory of stationary diffusion process. Notice that Theorem 1 gives us a global characterization of the original stationary diffusion process $X(t)$, $t\in[0,\infty)$, because the cost function contains the long-time average $\lim_{T\to\infty}(1/T)\int_0^T\cdots dt$. However it is also possible to obtain a local version of Theorem 1 in the sense that for each $t\in[0,\infty)$ the original process $X(s)$, $s\in[0,t]$, minimizes a t dependent cost function. This will be shown in the next section.

IV. ONSAGER–MACHLUP LAGRANGIAN AS A COST FUNCTION

In the previous paper[11] I found that the probability distribution of the stationary diffusion process $X(t)$, $t\in[0,\infty)$, is given approximately in the form

$$p^t \propto \exp\left[-\int_0^t L_{\text{OM}}(z(s),\dot z(s))ds\right]_{\max}, \quad (27)$$

where

$$L_{\text{OM}}(z(t),\dot z(t)) = \tfrac{1}{2}|\dot z(t) - a(z(t))|^2 + \text{div}a(z(t)) \quad (28)$$

is the OM Lagrangian and $[\cdot]_{\max}$ means to take a maximum value over continuous path $z(t)$'s connecting x_0 and $x\in d^n x$. The expression (28) allows us to conclude that a variational problem

$$\int_0^t L_{\mathrm{OM}}(z(s),\dot z(s))ds = \text{minimum}, \tag{29}$$

determines a most probable path. Namely the most probable path, say $\bar z(t)$, satisfies the Euler–Lagrange equation

$$\ddot{\bar z}(t) - \tfrac12 \operatorname{grad}|a(\bar z(t))|^2 - \operatorname{grad}\operatorname{div}a(\bar z(t)) = 0, \tag{30}$$

$$\bar z(0) = x_0. \tag{31}$$

Let us introduce a notion of OM action form

$$\begin{aligned} d\,I_{\mathrm{OM}}(z(t),\dot z(t),dt,dz(t)) &= L_{\mathrm{OM}}(z(t),\dot z(t))dt \\ &= L'_{\mathrm{OM}}(z(t),\dot z(t))dt - a(z(t))dz(t). \end{aligned} \tag{32}$$

Then I obtain

Theorem 2: The statement of Theorem 1 is also valid with respect to the local cost function

$$\begin{aligned} \mathrm{Cost}'(b) &= E\Big[\int_0^t dI_{\mathrm{OM}}(Z(s),DZ(s),ds,dZ(s))\Big] \\ &= E\Big[\int_0^t L'_{\mathrm{OM}}(Z(s),DZ(s))ds \\ &\quad - \int_0^t a(Z(s))\circ dZ(s)\Big], \end{aligned} \tag{33}$$

where $\int\cdots\circ dZ(s)$ denotes the stochastic integral of the Fisk–Stratonovich type.[13]

Proof: A straightforward stochastic calculus gives

$$\begin{aligned} dI_{\mathrm{OM}} &= (\tfrac12|DZ|^2 + \tfrac12|a|^2 + \operatorname{div}a)dt - a\circ dZ \\ &= (\tfrac12|b|^2 + \tfrac12|a|^2 + \operatorname{div}a)dt - a\cdot dZ - \tfrac12 da\cdot dZ \\ &= (\tfrac12|b|^2 + \tfrac12|a|^2 + \operatorname{div}a)dt \\ &\quad - a\cdot b\,dt - a\cdot dB - \operatorname{div}a\,dt \\ &= \tfrac12|b-a|^2dt - a\cdot dB. \end{aligned} \tag{34}$$

This yields by integration

$$\int_0^t dI_{\mathrm{OM}} = \int_0^t \tfrac12|b(Z(s)) - a(Z(s))|^2ds - \int_0^t a(z(s))\cdot dB(s), \tag{35}$$

and one finds by taking the expectation

$$\begin{aligned} \mathrm{Cost}'(b) &= E\Big[\int_0^t dI_{\mathrm{OM}}\Big] \\ &= E\Big[\int_0^t \tfrac12|b(Z(s)) - a(Z(s))|^2ds\Big] \\ &\geqslant 0, \end{aligned} \tag{36}$$

where the martingale property of the stochastic integral of Itô type is used. The local cost function is minimized by setting $b=a$. Q.E.D.

Theorem 2 claims that the original stationary diffusion process $X(t)$, $t\in[0,\infty)$, is completely characterized by the OM Lagrangian (28) or by the OM action form (32). It is optimally controlled to keep the mean OM action integral minimum.

V. DETERMINISTIC LIMIT

In the recent nonequilibrium statistical mechanics we have been much interested in the asymptotic behavior of a stationary diffusion process $X_\lambda(t)$, $t\in[0,\infty)$, of the type

$$dX_\lambda = a(X_\lambda)dt + \lambda^{-1}dB, \quad X_\lambda(0) = x_0, \tag{37}$$

in the limit $\lambda\to\infty$, where $\lambda\in\mathbb{R}$ is a system size parameter. The limit $\lambda\to\infty$ corresponds to an infinite volume limit.

In the limit $\lambda\to\infty$, randomness of the dynamical system (37) vanishes and $X_\lambda(t)$, $t\in[0,\infty)$, might be expected to converge to a deterministic path in some sense. To investigate the deterministic limit is the final goal of the OM problem. I shall approach the goal by making use of the results obtained in the previous sections.

Let us start with the stochastic control problem

$$dZ_\lambda = b(Z_\lambda)dt + \lambda^{-1}dB, \quad Z_\lambda(0) = x_0, \tag{38}$$

with respect to the cost function

$$\begin{aligned} \mathrm{Cost}'(b) &= E\Big[\int_0^t dI_{\mathrm{OM}}(Z_\lambda,DZ_\lambda,ds,dZ_\lambda)\Big] \\ &= E\Big\{\int_0^t L'_{\mathrm{OM}}(Z_\lambda,DZ_\lambda)ds - \int_0^t a(Z_\lambda)\circ dZ_\lambda\Big\}, \end{aligned} \tag{39}$$

which recovers the original stationary diffusion process $X_\lambda(t)$, $t\in[0,\infty)$, uniquely. From Eqs. (38) and (39) it is easy to observe that in the limit $\lambda\to\infty$ the stochastic control problem tends to an ordinary (nonrandom) control problem

$$\frac{dz(t)}{dt} = b(z(t)), \quad z(0) = x_0, \tag{40}$$

under the control

$$\begin{aligned} \mathrm{Cost}'(b) &= \int_0^t dI_{\mathrm{OM}}(z(s),\dot z(s),ds,dz(s)) \\ &= \int_0^t L'_{\mathrm{OM}}(z(s),\dot z(s))ds - \int_0^t a(z(s))\cdot dz(s) \\ &= \int_0^t L_{\mathrm{OM}}(z(s),\dot z(s))ds \\ &= \text{minimum}. \end{aligned} \tag{41}$$

Namely, a deterministic limit of $X_\lambda(t)$ is obtained as an optimally controlled differentiable path $\bar z(t)$, $t\in[0,\infty)$, with respect to the cost function (41). Condition (41) can be achieved by a solution to the Euler–Lagrange equation

$$\frac{d}{dt}\frac{\partial L_{\mathrm{OM}}}{\partial \dot z(t)} - \frac{\partial L_{\mathrm{OM}}}{\partial z(t)} = 0. \tag{42}$$

The deterministic limit of the stationary diffusion process thus obtained completely agrees with that obtained in the previous paper,[11] i.e., Eq. (30).

I have found in the realm of the stochastic control theory that the variational problem with respect to the OM Lagrangian gives us the deterministic path of the original stationary diffusion process. It is worthwhile to notice that the deterministic path is also a solution to the ordinary (nonrandom) control problem (40) under the control

$$\mathrm{Cost}(b) = \overline{\lim_{T\to\infty}}\frac1T\int_0^T L'_{\mathrm{OM}}(z(t),\dot z(t))dt = \text{minimum}. \tag{43}$$

This is guaranteed by the fact that the reduced OM Lagrangian differs from the OM Lagrangian only in a total differential.

VI. CONCLUDING REMARKS

I have investigated the OM problem from a stochastic control theoretical point of view. The OM Lagrangian was found to play a role of the cost function by which the original stationary diffusion process is controlled optimally. In this section I will make some remarks on the OM problem.

Firstly I want to mention that from the present viewpoint, unlike the path-integral approach [2–5] and the probabilistic one, [9,10] the canonical operator formalism [6–8] seems to have a possibility of generating a profound approach to the OM problem. This can be seen easily by rewriting the stochastic control problem (15) and (16) in terms of a white noise,

$$\dot{X}(t) = b(X(t)) + \dot{B}(t), \quad X(0) = x_0, \tag{44}$$

where $\dot{B}(t)$ is a white noise [16] defined on $\mathscr{S}'(\mathbb{R}^n)$ (tempered distribution space). The canonical operator formalism can be mostly visualized in the realm of Hida calculus [17] (a stochastic calculus of generalized Brownian functionals) in such a way that

$$\dot{B}(t) = \left(\frac{\partial}{\partial \dot{B}(t)}\right) + \left(\frac{\partial}{\partial \dot{B}(t)}\right)^*, \tag{45}$$

$$\left[\left(\frac{\partial}{\partial \dot{B}(t)}\right), \left(\frac{\partial}{\partial \dot{B}(s)}\right)^*\right] = \delta(t-s), \tag{46}$$

where ( )* denotes an adjoint operation and [,] a commutator. [17–20] By substituting Eq. (45) into Eq. (44), one finds that the Wick expansion, frequently used in the canonical operator formalism, is nothing but the Wiener–Itô decomposition. Therefore, the present approach in the realm of the stochastic control theory can be put on the line of the canonical operator formalism with the help of Hida calculus. The details of this will be given in a forthcoming paper. [21]

Secondly I want to point out that the present formulation of the OM problem from the stochastic control theoretical point of view gives us a profound way of understanding the quantum vacuum structure of the non-Abelian gauge field and the instanton effect. This is because the whole instanton effect can be calculated by means of a stationary diffusion process of the same type as treated here. [22–24] Details will be published elsewhere.

ACKNOWLEDGMENT

The author would like to express his sincerest thanks to Professor C.P. Enz for valuable discussions, encouragement, and support. He is also grateful to Professor T. Toyoda and to Professor T. Takabayasi for their continuous encouragements. He enjoyed helpful discussions with T. Nakagomi in the early stage of this work.

[1] L. Onsager and S. Machlup, Phys. Rev. **91**, 1505 (1953).
[2] H. Haken, *Synergetics* (Springer-Verlag, Berlin, 1977).
[3] R. Graham, Phys. Rev. Lett. **38**, 51 (1977).
[4] W. Horsthemke and A. Bach, Z. Physik B **22**, 189 (1975).
[5] H. Dekker, Phys. Lett. A **69**, 241 (1978).
[6] C.P. Enz, Physica A **89**, 1 (1977).
[7] C.P. Enz and L. Garrido, Phys. Rev. A **14**, 1258 (1976).
[8] T. Gotô, Prog. Theor. Phys. **60**, 1298 (1978).
[9] R.L. Stratonovich, Sel. Trans. Math. Stat. Prob. **10**, 273 (1971).
[10] D. Dürr and A. Bach, Commun. Math. Phys. **60**, 153 (1978).
[11] K. Yasue, J. Math. Phys. **19**, 1671 (1978).
[12] E. Nelson, *Dynamical Theories of Brownian Motion* (Princeton U.P., Princeton, 1967).
[13] K. Itô, Appl. Math., Optim. **1**, 374 (1975).
[14] K. Itô, "Stochastic Calculus," in *Lecture Notes in Physics* (Springer-Verlag, Berlin, 1975), Vol. 39.
[15] C.J. Holland, Commun. Pure Appl. Math. **30**, 755 (1977).
[16] T. Hida, *Stationary Stochastic Processes* (Princeton U.P., Princeton, 1970).
[17] T. Hida, "Causal Analysis in terms of White Noise," talk given at the conference "Bielefeld Encounters in Physics and Mathematics II," December 1–7, 1978.
[18] T. Hida, *Analysis of Brownian Functionals* (Carleton U.P., Ottawa, 1975).
[19] T. Hida, Appl. Math. Optim. **2**, 82 (1975).
[20] T. Hida, Math. Program. Study **5**, 53 (1976).
[21] K. Yasue, in preparation.
[22] K. Yasue, Phys. Rev. Lett. **40**, 665 (1978).
[23] K. Yasue, Phys. Rev. D **18**, 532 (1978).
[24] G. Jona-Lasinio, "Semi-Classical Limit and Stochastic Dynamics," talk given at the conference "Bielefeld Encounters in Physics and Mathematics II," December 1–7, 1978.

A model for the stochastic origins of Schrödinger's equation

Mark Davidson
1322 La Loma Ave., Berkeley, California 14708

(Received 2 March 1979; accepted for publication 23 April 1979)

A model for the motion of a charged particle in the vacuum is presented which, although purely classical in concept, yields Schrödinger's equation as a solution. It suggests that the origins of the peculiar and nonclassical features of quantum mechanics are actually inherent in a statistical description of the radiative reactive force.

I. INTRODUCTION

Stochastic models of quantum mechanics attempt to reconcile the postulates of quantum theory with modern probability theory, and to provide a space–time picture of quantum phenomena. The traditional inspiration for this effort is rooted in the extensive debates of the 1920's and 30's over the interpretation of quantum mechanics. A standoff developed, which persists to this day, between the Bohr complementarity school and the statistical school usually associated with Einstein.[1–3]

Today the Bohr interpretation is much more widely accepted. It asserts, in a nutshell, that given a physical state, then there is a state vector of some Hilbert space which describes this state completely, but only statistical properties about the physical system can be deduced from this presumed complete description. A number of forceful (and unresolved) completeness arguments againsts the Bohr view have been made,[4–7] and a number of the founders of modern quantum theory did not accept this view, including Einstein,[2] Schrodinger,[5] and De Broglie.[8]

There are several reasons why the Bohr view is dominant. Rigorous no-go theorems make stochastic or hidden variable models difficult to construct.[9–12] Despite these, there are statistical theories which do reproduce all of the statistical assertions of quantum mechanics, such as the differential space theory of Wiener and Siegel.[13] Any such theory must have some nonlocal features to avoid conflict with Bell's theorem,[12] and this can present conceptual problems. The Bohr view provides a justification for ignoring the puzzling questions of the origins of quantum mechanics, and for concentrating on applications of the theory. The accomplishments of the last half century have validated this point of view.

Most stochastic or hidden variable models have some nonclassical or difficult to understand features to them. For example, Bohm's early hidden variable theory[14] required the existence of a nonclassical quantum mechanical potential to be consistent with Schrodinger's equation. The Fenyes–Nelson stochastic model[15–17] also has a nonclassical quality about it. The dynamical assumption of Nelson,[16] for example, is not derived from first principles, and implies the existence of nonclassical forces acting on the particle. In most statistical models of quantum mechanics there is a gap in the derivation of quantum mechanical laws from classical laws, usually in the form of postulating a quantum mechanical potential or its equivalent. These gaps make the models unconvincing. An exception is the derivation of Schrodinger's equation from stochastic electrodynamics (SED),[18] where all quantum behavior is derived from a classical Langevin equation. The mathematics of this derivation are quite complicated, however, and there are several points of nonrigor owing to the singular nature of the random force in this model. Moreover, the SED model yields a Moyal type of phase space picture,[19] whereas the Markov model of Fenyes and Nelson seems better adapted to describing quantum mechanics.

This paper presents a simple model, within the Fenyes–Nelson scheme, which provides an explanation of the origin of the quantum mechanical potential, and of the steady state Schrodinger's equation. This model describes the diffusion of charged particles, and it includes the radiative reactive force. Neutral particles are not considered, but all known finite mass neutral particles are believed to be bound states of charged particles, so the results derived are not limited by this. The vacuum in which these charged particles move is assumed to have a finite temperature, but this temperature may be taken to zero. Inherent in the derivation is the concept of a vacuum alive with fluctuations and randomness. This concept of a nonempty vacuum has been slowly creeping back into physics with the work of Wheeler.[20] Boyer,[21] the models of Bohm and Vigier,[22] and De Broglie,[23] and more subtly in the whole quantum field effort with its infinite vacuum fluctuations.

The model presented is not a complete treatment of the problem. It relies on two reasonable postulates: The charged particles are described by a continuous Markov process in configuration space, and they are assumed to satisfy Gibbs' classical distribution, where the radiative reactive force is included. In the limit of zero temperature, these postulates imply the Schrodinger equation and the existence of a quantum mechanical potential, provided the diffusion constant of the theory has a certain value.

II. THE MODEL

Consider the Schrodinger equation for a single particle in a potential V:

$$\left[-\frac{\hbar^2}{2m}\Delta + V\right]\psi = i\hbar\frac{\partial\psi}{\partial t}, \quad \psi = e^{R+iS}. \tag{1}$$

 0022-2488/79/091865-05$01.00 © 1979 American Institute of Physics

It is equivalent to the following two equations:

$$\frac{\partial \rho}{\partial t} = -\nabla \cdot \frac{\hbar}{m} \nabla S \rho, \quad \rho = \psi^* \psi \tag{2}$$

and

$$\frac{\hbar^2}{2m} (\nabla S)^2 + V - \frac{\hbar^2}{2m} \frac{\Delta \rho^{1/2}}{\rho^{1/2}} = -\hbar \frac{\partial S}{\partial t}, \tag{3}$$

where R and S are chosen to be real. Equation (2) simply reflects the conservation of probability, and Eq. (3) is the Hamilton–Jacobi equation, but with an extra quantum mechanical potential:

$$V_{\mathrm{QM}} = -\frac{\hbar^2}{2m} \frac{\Delta \rho^{1/2}}{\rho^{1/2}}. \tag{4}$$

Were it not for this potential term, Schrodinger's equation could be interpreted as the diffusion of Newtonian particles whose initial conditions were not completely specified. This potential term is required in most classical models of quantum mechanics. For example, Madelung's hydrodynamic model requires it,[24] Bohm's early hidden variable model requires it,[14] and De Broglie's theory of the double solution requires it.[8] This term is also implicit in the dynamical assumption of Nelson,[16] where Eqs. (2) and (3) are interpreted as diffusion equations for a continuous Markov process. It is the possible origin of this extra term which shall be examined in this paper.

The quantum mechanical potential implies an unusual force, which acts on the particle, but which depends on the statistical properties of an ensemble of particle trajectories. This kind of behavior is difficult to understand in classical statistical mechanics. Indeed, it is this extra potential term which leads to quantum interference effects, and the difficulty of describing quantum interference in terms of classical statistical theories has been forcefully stated by Feynman.[25] Despite this, it appears that the model presented does give a possible explanantion of this extra potential in a classical statistical theory. The reason is that the radiative reactive force playes a large role in the theory about to be presented. Preacceleration associated with this radiative reactive force was not considered by Feynman in his arguments.

Consider a charged particle in motion in the physical vacuum. Let this particle be described by classical mechanics, and let its motion be nonrelativistic. Then it satisfies the equation

$$m_0 \mathbf{a}(t) = \int_0^\infty ds\, e^{-s} \mathbf{F}(\mathbf{x}(t + \tau s), t), \tag{5}$$

where

$$\tau = \frac{2}{3} \frac{q^2}{m_0 c^3}, \tag{6}$$

and where q is the charge of the particle, m_0 its mass, and τ has units of time. For an electron, $\tau \approx 10^{-22} s$, if q and m_0 are taken to be the observed charge and mass of the electron. For most practical calculations, such a brief preacceleration can be ignored. It has played little role in Newtonian physics. As is shown by Rohrlich,[26] Eq. (5) is the unique nonrelativistic limit of a perfectly well-defined relativistic theory.

There are two ways that the preacceleration effect can become amplified in the model to be presented. First of all, if q and m_0 are not the observed charge and mass, but rather are bare quantities, then τ can be much larger. If the diffusion constant of the Markov theory, which will be used to describe the particle is large, then preacceleration also becomes more important.

Suppose that the vacuum is alive with random field fluctuations, and suppose that it has a small temperature T. A more precise definition of this concept will not be attempted. It will only be assumed that the classical Gibbs distribution is satisfied. If the radiative force were ignored, then the particle would reach a state of equilibrium at temperature T, and its spatial density would be given by the classical Gibbs distribution,

$$\rho(x) = e^{-V(x)/kT}, \tag{7}$$

up to a normalization constant, where k is Boltzman's constant. This equation may be written

$$kT \nabla \ln(\rho) = -\nabla V = \mathbf{F}_{\mathrm{ext}}. \tag{8}$$

Equation (8) would not be satisfied by a charged particle which experiences a significant radiative force. The statistical distribution in this case is simply not known. Two assumptions shall be made to generalize Eq. (8) to include radiative forces in the simplest possible way.

The first assumption is that the charged particle, in thermal equilibrium with the vacuum, is described by a continuous Markov process on configuration space. Using Nelson's notation[16] $\mathbf{x}$ is assumed to satisfy the stochastic differential equation

$$d\mathbf{x}(t) = \mathbf{b}(x(t))dt + d\,\mathbf{W}(t), \tag{9}$$

where $\mathbf{W}$ is a three-dimensional Wiener process with

$$E\,(dW_i(t) dW_j(t)) = 2\nu\, \delta_{i,j}\, dt \tag{10}$$

and where ν is called the diffusion constant. This type of process was studied by Nelson,[16,17] and he showed that Schrodinger's equation could be derived, with a dynamical assumption, provided $\nu = \hbar/2m$. In fact, this result can be generalized,[27] and any value of ν greater than zero can be used to develop a model of Schrodinger's equation. The solutions to (9) are Markov processes on configuration space, and in general, velocities are not well defined. This Markov description must be viewed as an approximation to the actual motion of the particle, valid so long as dt is not too small in Eq. (9). If Eq. (9) were taken to be true for arbitrarily small dt, then the particle would be relativistic, and the nonrelativistic approximation would be inaccurate.

Imagine that the charged particle, in interaction with the finite temperature vacuum and subject to an external potential V, has reached a stationary state of thermal equilibrium described by a probability density $\rho(x)$. Consider the following conditional expectation:

$$\mathbf{F}_E(x) = -E\left(\int_0^\infty ds\, e^{-s} \nabla V(x(t + \tau s)) \;\Big|\; \mathbf{x}(t) = \mathbf{x} \right). \tag{11}$$

From Eq. (5), it is seen that this expresses the expected value of the total force on the particle, including preacceleration, given that at time t the particle's trajectory passed through the point $\mathbf{x}$. This equation represents the best estimate that can be made of the instantaneous force acting on the particle at position $\mathbf{x}$ and time t.

By analogy with the classical Gibbs distribution [Eq. (8)], the following equation for the charged particle is postulated:

$$kT\,\nabla \ln(\rho) = \mathbf{F}_E(x). \qquad (12)$$

This constitutes the second postulate. All it says is that the classical Gibbs distribution is satisfied for the total force given by Eq. (11), and including radiative effects. Implicit in Eq. (12) is the assumption that $\mathbf{F}_E$ has vanishing curl. This will prove to be consistent.

$\mathbf{F}_E$, as expressed in Eq. (11), will depend on $\mathbf{b}$ in Eq. (9), and therefore Eq. (12) will be a differential equation for ρ. To derive this equation, the Markov transition function is used:

$$P_{t-u}(y,x) = \lim_{d^3y\to 0} \frac{1}{d^3y} P(\mathbf{x}(t)\in d^3y \mid \mathbf{x}(u) = \mathbf{x}), \quad t>u \qquad (13)$$

which satisfies the forward and backward equations of Kolmogorov[28]:

$$\frac{\partial}{\partial t} P_{t-u}(y,x) + \nabla_y \cdot \mathbf{b}(y) P_{t-u}(y,x) - \nu \Delta_y P_{t-u}(y,x) = 0, \quad t>u, \qquad (14)$$

$$\frac{\partial}{\partial u} P_{t-u}(y,x) + \mathbf{b}(x)\cdot\nabla_x P_{t-u}(y,x) + \nu\Delta_x P_{t-u}(y,x) = 0, \quad t>u. \qquad (15)$$

$\mathbf{F}_E$ may be written as

$$\mathbf{F}_E(x) = -\int_0^\infty ds\, e^{-s} \int d^3y\, P_{\tau s}(y,x)\nabla V(y). \qquad (16)$$

P must satisfy two limiting conditions: The first is a statement of continuity, and the second is a statement of ergodicity:

$$P_0(y,x) = \delta^3(y-x), \qquad (17)$$

$$P_\infty(y,x) = \rho(y). \qquad (18)$$

Equation (18) requires some qualifications. If the density ρ vanishes at some point, then Eq. (18) is not quite valid, as has been shown by Albeverio and Hoegh-Krohn.[29] In this case, space is divided up into disjoint regions bounded by surfaces $\rho(x) = 0$, and the Markov transition function vanishes unless x and y are in the same region. Equation (18) is true if x and y are in the same region in this case, and this is sufficient for the results below.

From Eq. (18) and Eq. (14) it follows that, taking the limit $t\to\infty$ in (14),

$$\mathbf{b} = \nu\nabla \ln(\rho). \qquad (19)$$

Now, using the backward equation [Eq. (15)] together with the expression for $\mathbf{F}_E$ [Eq. (16)] one obtains

$$(\mathbf{b}\cdot\nabla + \nu\Delta)\mathbf{F}_E(x) = -\int_0^\infty ds\, e^{-s}\int d^3y \frac{\partial}{\partial \tau s} P_{\tau s}(y,x)\nabla V(y), \qquad (20)$$

where it has been assumed that the order of differentiation and integration can be freely interchanged. Integrating (20) by parts, and using (17) then yields

$$[1-\tau(\mathbf{b}\cdot\nabla + \nu\Delta)]\mathbf{F}_E(x) = -\nabla V(x). \qquad (21)$$

At this point, the Gibbs distribution [Eq. (12)] is used to substitute for $\mathbf{F}_E$ in Eq. (21). One finds

$$[1-\tau(\mathbf{b}\cdot\nabla + \nu\Delta)]\nabla \ln[\rho(x)] = -\frac{1}{kT}\nabla V(x). \qquad (22)$$

Defining R by

$$R = \tfrac{1}{2}\ln(\rho), \qquad (23)$$

and using (19) and (22), one finds:

$$\nabla[R - \tau\nu((\nabla R)^2 + \Delta R)] = -\frac{1}{2kT}\nabla V. \qquad (24)$$

Integrating this expression, and rewirting it, one obtains

$$[-2\tau\nu kT\Delta + V + 2kTR]e^R = \lambda e^R, \quad \lambda = \text{const}. \qquad (25)$$

This can also be written as

$$\rho(x) = \exp\left[-\frac{1}{kT}\left(V(x) - 2\tau\nu kT\frac{\Delta\rho^{1/2}}{\rho^{1/2}} - \lambda\right)\right]. \qquad (26)$$

This last expression clearly displays the existence of an extra, and unusual, potential given by

$$-2\tau\nu kT\Delta\rho^{1/2}/\rho^{1/2}. \qquad (27)$$

This extra potential term is due to the radiative reactive force, and it has exactly the same form (including the right sign) as the quantum mechanical potential [Eq. (4)]. Equation (25) bears a remarkable similarity in form to the Schrodinger steady state equation.

The strength of the radiative preacceleration effects depend on the magnitude of the gradient of (27) relative to the gradient of V. This depends on the factor

$$\gamma = 2\tau\nu kT = \frac{4}{3}\frac{q^2\nu kT}{m_0c^3}. \qquad (28)$$

This factor γ determines the magnitude of radiative effects. It is interesting that one cannot distinguish between different values of q^2, ν, and m_0, but only different values of γ. For small T, γ can be large if the ratio $q^2\nu/m_0$ is large. Since ν is a free parameter in this model, a large radiative correction is possible for large ν, regardless of the size of the other factors.

Suppose that

$$\gamma = \frac{\hbar^2}{2m} = \frac{4}{3}\frac{q^2\nu kT}{m_0c^3}, \qquad (29)$$

where m is the physical mass, and the possibility that it is different from the bare mass m_0 has been allowed. Then Eq. (25) becomes

$$\left(-\frac{\hbar^2}{2m}\Delta + V + 2kTR\right)e^R = \lambda e^R \qquad (30)$$

and (26) becomes

$$\rho(x) = \exp\left[-\frac{1}{kT}\left(V(x) - \frac{\hbar^2}{2m}\frac{\Delta\rho^{1/2}}{\rho^{1/2}} - \lambda\right)\right]. \qquad (31)$$

Equation (30) has the same form as Schrodinger's equation, except for the extra term in the potential, $2kTR$. This extra term can be interpreted as representing the diffusion force. It prevents the occurence of zeroes in $\psi = e^R$. Equation (31) is the analog of the Gibbs distribution for neutral particles [Eq.

(7)], with the quantum mechanical potential included, but due to radiative forces. If T is very small, then (30) becomes

$$[(-\hbar^2/2m)\Delta + V]\psi \simeq \lambda\psi \tag{32}$$

which is just Schrodinger's stationary state equation.

Equation (30) is a classical model for steady state quantum mechanics with one free parameter, the temperature of the vacuum T. It is nonlinear, and in general difficult to solve. In the limit $T \to 0$, Eq. (32) becomes exact. It is an experimental question what T is, assuming that the model is taken seriously.

Although the possibility that m and m_0 are different has been allowed, it is interesting to note that if $m = m_0$ or if m and m_0 are proportional with a fixed factor, then both sides of Eq. (29) have the same mass dependence. This means that νq^2 may be chosen to be mass independent. If q is taken to be the electronic charge, then ν could be mass independent. This is consistent with the generalization of the Fenyes–Nelson model,[27] where any value of ν can be used to construct a model of quantum mechanics. If ν is mass independent, then the underlying thermal agitation could be gravitational in nature. This could be consistent with Wheeler's concepts of superspace.[20]

If Eq. (32) is a good approximation, that is if T is small, then energy levels are quantized, provided the usual Hamiltonian operator is taken as the energy operator. Quantization of the energy levels of harmonic oscillators leads, through fairly well known arguments,[30] to a derivation of the Planck radiation law. The present theory, if correct, could influence the equilibrium of radiation at finite temperature. This could provide a way out of the Rayleigh–Jeans spectrum.

The question that remains is what could determine T, and how could a more complete model be constructed. If T is nonzero, then it is reasonable to expect to see black-body electromagnetic radiation at this temperature in the vacuum. The spectrum of radiation in the vacuum is not exactly black-body, but in the microwave region, a Planck spectrum has been observed at a temperature of 2.76° K.[31] The problem with this is that the radiation may not yet have reached thermal equilibrium. It is possible that T equals this radiation temperature, and this deserves some consideration, but this does not appear to be a necessity, and this possibility will not be considered here.

The results of this section sould be compared with the SED Langevin approach.[18] In that model, Schrodinger's equation is derived for the diffusion of an electron in interaction with zero point background radiation. A number of approximations are made to derive general results, and the radiative reactive force playes a crucial role. It is hoped that the present model complements and perhaps sheds some light on the SED calculation. Although less complete, the present model is much simpler than the SED model, and it is felt that this simplicity helps to isolate the essential ingredients in the relationship between quantum mechanics and stochastic theories with radiative reactive forces.

III. CONCLUSION

Charged particles in interaction with a low temperature vacuum can be expected to satisfy a Schrodinger type equation. This result offers an explanation of the quantum mechanical potential as essentially due to radiative reactive forces in a stochastic theory. It also suggests that an extra term may be present and possibly observable in Schrodinger's equation if the vacuum temperature is not zero.

The main limitation of the model presented is that it makes no attempt to account in a detailed way for the Markov motion of the particles from, say, a Langevin approach in terms of random forces. However, by using only simple postulates, independent of the details of the vacuum's structure, it is felt that the derivation of Schrodinger's equation is less model dependent and more straightforward than, say, the SED calculation,[18] although both calculations are similar in spirit. Moreover, the SED approach may not contain all of the relevant vacuum fluctuations. It does not include gravitational fluctuations or fluctuations in the vector fields which mediate the weak interactions, both of which could be important for the electron. The model presented here does not really care what fields are involved, so long as the generalized Gibbs distribution is satisfied and the motion is described by a Markov process. In this sense it may be more general than the SED approach.

The future of this model will hinge on the ability to generalize it to the time dependent case, and to make it relativistic. These are major problems at the present. The importance of the preacceleration in the model helps to explain the nonlocal character of hidden variable models of quantum mechanics. In the classical theory, the acceleration at a particular time depends on the force for all future times. Treating this type of dynamical system statistically, one is forced to conclude that the most likely value for the force which will be experienced by a particle at a given position and time will depend on the properties of the ensemble, that is, it will depend on ρ. Any measurement made on the system will change ρ, and this will change the expected force on the particle instantaneously. This peculiar property is understood in terms of the preacceleration of charged particles, and should not be considered unphysical, unless preacceleration is also considered unphysical.

It is believed that the results presented can be generalized to many-particle systems. The possibilities that T is the temperature of the cosmic background radiation, or that the thermal agitation of the vacuum is gravitational in nature, with ν independent of mass, are intriguing and should provide fertile areas for exploration.

[1] M. Jammer, *The Philosophy of Quantum Mechanics* (Wiley, New York, 1974).

[2] P.A. Schilpp, Ed., *Albert Einstein: Philosoper-Scientist* (Harper and Row, New York, 1959).

[3] S. Rozental, Ed. *Neils Bohr—His Life and Work as seen by his Friends and Colleagues* (Wiley, New York, 1967).

[4] A. Einstein, B. Podolsky, and N. Rosen, Phys. Rev. **47**, 777–80 (1935).

[5] E. Schrödinger, Proc. Camb. Phil. Soc. **31**, 555–62 (1935); Proc. Cambridge Philos. Soc. **32**, 446–52 (1936); Die Naturwissenschaften **23**, 807–12, 824–28, 848–49 (1935).

[6] E.P. Wigner, "Remarks on the Mind-Body Question," in *The Scientist Speculates*, edited by I.J. Good (Heinemann, London, 1961).

[7]A. Landé, *Foundations of Quantum Theory* (Yale U.P. New Haven, 1955); *From Dualism to Unity in Quantum Physics* (Cambridge U. P., London, 1960). See Also: Nucl. Phys. **1**, 133–34 (1956) for a review of the first book, and Nucl. Phys. **3**, 132–34 (1957) for "An Anti-Review."
[8]L. De Broglie, *Non-Linear Wave Mechanics* (Elsevier, Amsterdam, 1960).
[9]J. Von Neumann, *Mathematical Foundations of Quantum Mechanics* (Princeton U. P., Princeton, 1955).
[10]J.M. Jauch and C. Piron, Helvetica Physica Acta **36**, 827–37 (1963).
[11]S. Kochen and E.P. Specker, Nuovo Cimento B **10**, 518–22 (1972).
[12]J.S. Bell, Rev. Mod. Phys. **38**, 447–52 (1966).
[13]N. Wiener, A. Siegel, B. Rankin, W.T. Martin, *Differential Space, Quantum Systems, and Prediction* (M.I.T. Press, Cambridge, 1966).
[14]D. Bohm, Phys. Rev. **85**, 166–79 (1952); Phys. Rev. **85**, 180–93 (1952).
[15]I. Fényes, Zeitschrift für Physik **132**, 81–106 (1952).
[16]E. Nelson, *Dynamical Theories of Brownian Motion* (Princeton U. P., Princeton, 1967).
[17]E. Nelson, Phys. Rev. **150**, 1079–85 (1966).
[18]L. de la Pena-Auerbach and A.M. Cetto, J. Math. Phys. **18**, 1612–22 (1977).
[19]J.E. Moyal, Proc. Cambridge Philos. Soc. **45**, 99–124 (1949).
[20]C.W. Misner, K.S. Thorne, and J.A. Wheeler, *Gravitation* (Freeman, San Francisco, 1973).
[21]T.H. Boyer, Phys. Rev. D **11**, 790–808, 809–30 (1975).
[22]D. Bohm and J.P. Vigier, Phys. Rev. **96**, 208 (1954).
[23]L. De Broglie, *La Thermodynamique de la particule Isolée* (Gauthier-Villars, Paris, 1964).
[24]E. Madelung, Z. Phys. **40**, 322–26 (1926).
[25]R.P. Feynman, "The Concept of Probability in Quantum Mechanics," in *The Second Berkeley Symposium on Mathematical Statistics and Probability* (University of Cal. P., Berkeley, 1951).
[26]F. Rohrlich, *Classical Charged Particles* (Addison–Wesley, Reading, Massachusetts, 1965).
[27]M. Davidson, "A Generalization of the Fényes–Nelson Stochastic model of Quantum Mechanics," to appear in Lett. Math. Phys.; "A Dynamical Theory of Markovian Diffusion," to appear in Physica A.
[28]J.L. Doob, *Stochastic Processes* (Springer, Berlin, 1965).
[29]S. Albeverio and R. Hoegh-Krohn, J. Math. Phys. **15**, 1745–47 (1974).
[30]T.S. Kuhn, *Black-Body Theory and the Quantum Discontinuity* (Clarendon, Oxford, 1978).
[31]M.S. Longair, Rep. Prog. Phys. **34**, 1125–248 (1971).

Exact solution of a time-dependent quantal harmonic oscillator with damping and a perturbative force

D. C. Khandekar and S. V. Lawande

Bhabha Atomic Research Centre, Theoretical Reactor Physics Section, 5th floor, Central Complex, Trombay, Bombay, India 400 085
(Received 21 August 1978)

The problem of a quantal harmonic oscillator with damping and a time-dependent frequency acted on by a time-dependent perturbative force is exactly solved. The wavefunctions are found in Schrödinger representation using the theory of explicitly time-dependent invariants and also by an expansion of the Feynman propagator. The propagator is obtained in exactly closed form by an explicit path integration of the classical Lagrangian. It is found that the wavefunctions and the propagator depend only on the solution of classical damped oscillator through a single function $\rho(t)$. The function $\rho(t)$ itself may be obtained as a solution of a second order nonlinear differential equation under the appropriate set of initial conditions.

1. INTRODUCTION

Exact solutions of the Schrödinger equation with explicitly time-dependent Hamiltonians are available only in few cases. Such problems are mostly solved by using approximation methods such as perturbation theory. An exactly solvable system which has received considerable attention in literature is that of a harmonic oscillator with variable frequency.[1-8] More recently Lewis and Reisenfeld[9] have developed a general theory of explicitly time-dependent invariants for quantum systems characterized by explicitly time-dependent Hamiltonians. They have derived a simple relation between eigenstates of such an invariant and solutions of the corresponding Schrödinger equation and have applied it to the case of a harmonic oscillator with time dependent frequency. Leach[10] has obtained generalized invariants for quadratic Hamiltonians. Exact closed form for the wavefunctions of a time-dependent linear oscillator perturbed by an inversely quadratic potential has been obtained by Camiz *et al.*[11] using Schrödinger formalism and a generating function. The same problem has also been solved in I[12] using Feynman path integrals and in II[13] based on a simple relation between the classical action S and the generating function of a canonical transformation involving an explicitly time dependent invariant.

The present paper discusses an exact quantum theory of a classical forced oscillator with a time dependent frequency and a velocity dependent damping term described by the equation of motion:

$$\ddot{x} + \gamma\dot{x} + \omega^2(t)x = f(x), \tag{1}$$

where the frequency $\omega(t)$ is assumed to be a regular function of time, $f(t)$ is the time dependent external perturbative force, and γ is the constant damping coefficient. As shown by Havas,[14] Eq. (1) may be obtained from the Lagrangian

$$L = e^{\gamma t}\left[\frac{\dot{x}^2}{2} - \frac{\omega^2(t)}{2}x^2 + f(x)\,x\right], \tag{2}$$

from which the Hamiltonian H is readily obtained as

$$H = e^{-\gamma t}\frac{p^2}{2} + e^{\gamma t}\left[\frac{w^2(t)x^2}{2} + f(t)x\right]. \tag{2'}$$

The aim of this work is to show than an exact closed form for the solution of the Schrödinger equation

$$i\frac{\partial\psi}{\partial t} = \widehat{H}(t)\psi, \tag{3}$$

where $\widehat{H}$ is the quantum Hamiltonian operator corresponding to H of Eq. (2a) exists. At time $t = t_0$ this solution goes over into the corresponding solution of the free damped oscillator with a constant frequency obtained by Bopp.[15] It is first shown that a time dependent Hermitian invariant operator $\hat{I}(t)$ exists for the problem. $\hat{I}(t)$ is then expressed in terms of raising and lowering operators and its eigenvalues and eigenfunctions are constructed in the manner shown by Dirac.[16] These are then used to obtain the solutions of Schrödinger equation using the theory of Ref. 9.

An alternative manner of quantizing the system is through the Feynman propagator $K(x'',t'';x',t')$ defined as the path integral[1]

$$K(x'',t'';x',t') = \int \exp\left(i\int_{t'}^{t''} L\,dt\right)\mathscr{D}(x(t)), \tag{4}$$

where L is the Lagrangian and integrations are over all paths starting at $x' = x(t')$ and terminating at $x'' = x('')$. Since the Lagrangian of Eq. (2) is quadratic, the propagator may be evaluated either by using the Van Vleck–Pauli formula[17-19] or by Feynman's theorem both of which involve essentially the computation of the classical action $S_{cl}(x'',t'';x',t')$ and differ only in the manner in which the normalization factor is obtained. In fact, for a quadratic Lagrangian the propagator takes the form[20]

$$K(x'',t'';x',t') = F(t'',t')\exp\left[\frac{i}{\hbar}S_{cl}(x'',t'';x',t')\right], \tag{5}$$

where $F(t'',t')$ is an entirely time dependent function. In the Van Vleck–Pauli formula, $F(t'',t')$ is given by

$$\left[\frac{i}{2\pi}\frac{\partial^2}{\partial x''\partial x'}S_{cl}(x'',t'';x',t')\right]^{1/2}$$

while Feynman and Hibbs[20] describe $F(t'',t')$ as a conditional path integral. Papadopoulos[21] has recently evaluated this conditional path integral for a general quadratic Lagran-

 0022-2488/79/091870-08$01.00 © 1979 American Institute of Physics

gian. In Ref. 22 Goovaerts has evaluated $F(t'',t')$ by taking recourse to the Schrödinger equation. However, in the present paper $K(x'',t'';x',t')$ is obtained directly from Eq. (3) by interpreting it as a limit of multiple Riemann integrals. Such explicit path integration has the virtue of yielding both the normalizing factor $F(t'',t')$ and the contribution due to the classical path simultaneously. An exact closed form of the propagator has been obtained which depends only on the solutions of the corresponding classical problem. Further, it is shown that the propagator admits an expansion

$$K(x'',t'';x',t') = \sum_n \psi_n^*(x'',t'')\psi_n(x',t') \quad (6)$$

in a natural manner leading to the time-dependent wavefunctions $\psi_n(x,t)$ of the Schrödinger equation (3). The expansion of the propagator in terms of the eigenfunctions of the Hermitian inariant operator $\hat{I}(t)$ also follows from Eq. (6).

For the sake of completeness an outline of the theory of explicitly time dependent invariants and their relation to the Schrödinger equation and the Feynman propagator is included in Sec. 2. Derivation of the invariant operator $\hat{I}(t)$, its eigenfunction, and solution of the Schrödinger equation has been discussed in Sec. 3. Sections 4 and 5 refer the evaluation of the propagator and its eigenfunctions in terms of the wavefunctions of the Schrödinger equation. Throughout this paper the units $\hbar = m = 1$ have been used.

2. TIME DEPENDENT INVARIANT, SCHRÖDINGER EQUATION AND THE PROPAGATOR

Consider a system characterized by a Hamiltonian $\hat{H}(t)$ which is an explicit function of time. Following Lewis,[9] we assume the existence of an explicitly time-dependent Hermitian invariant operator $\hat{I}(t)$. It is clear that $\hat{I}(t)$ satisfies the equation

$$\frac{d\hat{I}}{dt} = \frac{\partial \hat{I}}{dt} + \frac{1}{i}[\hat{I},\hat{H}] = 0 \quad (7)$$

and $\hat{I}^\dagger = \hat{I}$. Further, the eigenfunctions $\phi_n(x,t)$ of $\hat{I}(t)$ are assumed to form a complete orthonormal set corresponding to the eigenvalue λ. For simplicity it is assumed that the eigenstates of $\hat{I}(t)$ are nondegenerate so that the eigenvalue λ is the only quantum number required to describe the system. This certainly applies for the case studied in this paper. Thus

$$\hat{I}\phi_\lambda(x,t) = \lambda\phi_\lambda(x,t), \quad (8)$$

$$(\phi'_\lambda,\phi_\lambda) = \delta_{\lambda'\lambda}. \quad (9)$$

The solutions $\psi_n(x,t)$ of the time-dependent Schrödinger equation are related to $\phi_\lambda(x,t)$ by a relation[9]

$$\psi_\lambda(x,t) = e^{i\alpha_\lambda(t)}\phi_\lambda, \quad (10)$$

where α_λ obeys an equation

$$\frac{d\alpha_\lambda}{dt} = \left\langle \phi_\lambda \,\middle|\, i\frac{\partial}{\partial t} - \hat{H} \,\middle|\, \phi_\lambda \right\rangle. \quad (11)$$

Since each of ψ_λ satisfies the Schrödinger equation, the general solution may be written as

$$\psi(x,t) = \sum_\lambda C_\lambda e^{i\alpha_\lambda(t)}\phi_\lambda(x,t), \quad (12)$$

C_λ being the time-dependent coefficients given by

$$C_\lambda = e^{i\alpha_\lambda(t)}\langle \phi_\lambda|\psi\rangle. \quad (13)$$

In order to obtain a relation between the propagator and ϕ_λ we see that Eq. (12) for $t'' > t'$ can be expressed as

$$\begin{aligned}\psi(x'',t') &= \sum_\lambda C_\lambda e^{i\alpha_\lambda(t'')}\phi_\lambda(x'',t'') \\ &= \sum e^{i\alpha_\lambda(t')}\int \phi_\lambda^*(x',t')\,dx' e^{i\alpha_\lambda(t'')}\phi_\lambda(x'',t'') \\ &= \int\Big\{\sum_\lambda e^{i[\alpha_\lambda(t'')-\alpha_\lambda(t')]}\phi_\lambda(x'',t'')\phi_\lambda^*(x',t')\Big\} \\ &\quad\times \psi(x',t')\,dx',\end{aligned} \quad (14)$$

where in the second step on the left Eq. (13) has been used. Comparing Eq. (14) with the definition of the propagator, viz.,

$$\psi(x'',t'') = \int K(x'',t'';x',t')\psi(x',t')\,dx', \quad t'' > t', \quad (15)$$

it follows that

$$K(x'',t'';x',t') = \sum_\lambda e^{i\{\alpha_\lambda(t'')-\alpha_\lambda(t')\}}\phi_\lambda^*(x',t')\phi_\lambda(x'',t''), \quad (16)$$

which is a generalization of the usual expansion formula for time-independent Hamiltonians given by Feynman and Hibbs.[20]

3. DERIVATION OF THE INVARIANT OPERATOR AND WAVEFUNCTIONS OF THE SCHRÖDINGER EQUATION

The Hermitian invariant operator $\hat{I}(t)$ for the system characterized by the Hamiltonian (2') is assumed to have the form

$$\begin{aligned}\hat{I}(t) = \tfrac{1}{2}[a(t)x^2 + b(t)\{x,p\}_+ + c(t)p^2 \\ + d(t)x + h(t)p + k(t)],\end{aligned} \quad (17)$$

where a, b, c, d, h, and k are real functions of time, $\{x,p\}_+$ is the conventional anticommutator, and the numerical multiplicative factor has been chosen for convenience. From Eqs. (7), (17), and the usual commutation relation $[x,p] = i$ one obtains the following set of equations:

$$\dot{a} - 2b\omega^2 e^{\gamma t} = 0, \quad (18)$$

$$\dot{b} + ae^{-\gamma t} - c\omega^2 e^{\gamma t} = 0, \quad (19)$$

$$\dot{c} + 2be^{-\gamma t} = 0, \quad (20)$$

$$\dot{d} + (2fb - h\omega^2)e^{\gamma t} = 0, \quad (21)$$

$$\dot{h} + de^{-\gamma t} + 2cfe^{\gamma t} = 0, \quad (22)$$

$$\dot{k} + hfe^{\gamma t} = 0. \quad (23)$$

The set of first order differential equations (18)–(23) can be solved readily to yield the functions a, b, c, d, h, and k using which the invariant $\hat{I}(t)$ can be written as

$$\begin{aligned}\hat{I}(t) = \frac{1}{2}\Big[\Big(\frac{x}{\rho}e^{\gamma t/2} + V\Big)^2 \\ + e^{-\gamma t/2}\rho p - e^{\gamma t/2}\Big(\dot{\rho} - \frac{\gamma\rho}{2}\Big)x - U\Big],\end{aligned} \quad (24)$$

where ρ obeys the equation

$$\rho + \omega^2\rho - \rho^{-3} = 0 \quad (25)$$

and Ω, U, and V are defined as

$$\Omega^2(t)=\omega^2(t)-\gamma^2/4, \quad (26)$$

$$U(t)=\int_{t_0}^{t} G(\tau)\cos\phi(\tau,t)\,d\tau, \quad (27)$$

$$V(t)=\int_{t_0}^{t} G(\tau)\sin\phi(\tau,t)\,d\tau. \quad (28)$$

The functions $G(\tau)$ and $\phi(\tau,t)$ are defined as

$$G(\tau)=\rho(\tau)f(\tau)\,e^{\gamma\tau/2}, \quad (29)$$

$$\phi(\tau,t)=\mu(\tau)-\mu(t), \quad (30)$$

where $\mu(t)$ is related to ρ by a relation

$$\rho^2\dot{\mu}=1. \quad (31)$$

Any particular solution of Eq. (25) may be used to construct $\hat{I}(t)$. However, in this paper we assume that at time $t=t_0$ the wavefunction must reduce to the one corresponding to the free damped oscillator with constant frequency ω_0. Consequently, Eq. (25) must be solved with the initial conditions

$$\rho(t_0)=1/\sqrt{\Omega_0},\quad \dot{\rho}(t_0)=0, \quad (32)$$

with

$$\Omega_0^2=\omega_0^2-\gamma^2/4.$$

Next, we define a pair of time dependent canonical lowering and raising operators a and $a^\dagger$ defined by

$$\hat{a}=\frac{1}{\sqrt{2}}(P+iQ),\quad \hat{a}^\dagger=\frac{1}{\sqrt{2}}(P-iQ), \quad (33)$$

where

$$P=\frac{e^{\gamma t/2}}{\rho}+V,$$

$$Q=e^{-\gamma t/2}\rho p+\left(\dot{\rho}+\frac{\gamma\rho}{2}\right)e^{\gamma t/2}x-U. \quad (34)$$

It may be easily verified that $\hat{a}$ and $\hat{a}^\dagger$ satisfy the commutation relation

$$[\hat{a},\hat{a}^\dagger]=\hat{a}\hat{a}^\dagger-\hat{a}^\dagger\hat{a}=1, \quad (35)$$

which implies that the operator $\hat{a}^\dagger\hat{a}$ is a number operator with nonnegative integer eigenvalues. The invariant operator $\hat{I}(t)$ of Eq. (24) can now be written as

$$\hat{I}=(\hat{a}^\dagger\hat{a}+\tfrac{1}{2}). \quad (36)$$

Equation (36) implies that the normalized eigenstates ϕ_λ of $\hat{I}$ are same as normalized eigenstates ϕ_n of $\hat{a}^\dagger\hat{a}$ defined by

$$\hat{a}^\dagger\hat{a}\phi_n=n\phi_n,\quad n=0,1,2,\cdots. \quad (37)$$

The relative phases of the eigenstates ϕ_n may be fixed by demanding that the usual relations

$$\hat{a}\phi_n=\sqrt{n}\phi_{n-1},\quad \hat{a}^\dagger\phi_n=\sqrt{n+1}\phi_{n+1} \quad (38)$$

are satisfied. The eigenvalue spectrum of $\hat{I}$ is clearly given by

$$\lambda_n=(n+\tfrac{1}{2}),\quad n=0,1,2,\cdots. \quad (39)$$

In order to obtain ϕ_n explicitly, one first obtain $\phi_0(x,t)$ from the relation

$$\hat{a}\phi_0=0. \quad (40)$$

Upon substituting the explicit value of $\hat{a}$ from Eq. (33), one obtains a first order differential equation for ϕ which when solved yields

$$\phi_0(x,t)=\frac{e^{\gamma t/4}}{\pi^{1/4}\rho^{1/2}}\exp(-\tfrac{1}{2}P^2)\exp\left(\frac{i}{2}R\right), \quad (41)$$

where

$$R=\left[x^2e^{\gamma t/2}\left(\frac{\dot{\rho}}{\rho}-\frac{\gamma}{2}\right)+\frac{U}{\rho}e^{\gamma t/2}x\right]. \quad (42)$$

The eigenstate ϕ_n can be obtained by applying the operator $\hat{a}^\dagger$ successively to ϕ_0. Thus

$$\phi_n(x,t)=\frac{(\hat{a}^\dagger)^n}{\sqrt{n!}}\phi_0(x,t), \quad (43)$$

which finally gives

$$\phi_n(x,t)=\frac{e^{\gamma t/4}}{[(\pi\rho)^{1/2}\cdot 2^n n!]^{1/2}}\exp\left\{-\frac{1}{2}P^2+\frac{i}{2}R\right\}\times H_n\left(\frac{x}{\rho}e^{\gamma t/2}+V\right), \quad (44)$$

where $H_n(x)$ is the usual Hermite polynomial of order n. When $\gamma=0$ and $f(t)=0$, these reduce to the eigenstates of the invariant operator corresponding to the free time dependent oscillator discussed in I.

In order to obtain the solutions $\psi_n(x,t)$ of Schrödinger equation, one has to compute the phase $\alpha_n(t)$ from the equation

$$\frac{d\alpha_n}{dt}=\left\langle\phi_n\middle|i\frac{\partial}{\partial t}-\hat{H}\middle|\phi_n\right\rangle. \quad (45)$$

After an explicit evaluation of the matrix element on the right of Eq. (45) we obtain

$$\frac{d\alpha_n}{dt}=-\frac{(n+\frac{1}{2})}{\rho^2}-\frac{(U^2-V^2)}{2\rho^2}. \quad (46)$$

Equation (46) may now be integrated to yield

$$\alpha_n(t)=-(n+\tfrac{1}{2})\mu(t)-F(t), \quad (47)$$

where the phase at $t=t_0$ has been fixed as

$$\alpha_n(t_0)=-(n+\tfrac{1}{2})\mu(t_0) \quad (48)$$

and $F(t)$ is defined as

$$F(t)=\int_{t_0}^{t}\frac{[U^2(\tau)-V^2(\tau)]}{2\rho^2(\tau)}d\tau. \quad (49)$$

Finally the normalized wavefunctions $\psi_{n(x,t)}$ of the Schrödinger equation are given by

$$\psi_n(x,t)=e^{i\alpha_n(t)}\phi_n(x,t)$$
$$=e^{i\alpha_n(t)}\cdot\frac{e^{\gamma t/4}}{[2^n n!(\pi\rho)^{1/2}]^{1/2}}\exp\left(-\frac{1}{2}P^2+\frac{i}{2}R\right)\times H_n\left(\frac{x}{\rho}e^{\gamma t/2}+V\right), \quad (50)$$

where P and R are defined in Eqs. (34) and (42), respectively. Several limiting cases follow from this general expression. First of all note that when $t=t_0$, this wavefunction reduces to the form given by Bopp.[15] Secondly, when the dissipative term is absent ($\gamma=0$), Eq. (50) yields the wavefunction derived by Goovaerts[22] by a path integral approach. Finally, when $\gamma=0$ and $f(t)=0$, Eq. (50) reduces to the wavefunction for the free oscillator with time-dependent frequency obtained in Ref. I.

It may be remarked here that such closed form of the oscillator wavefunction may be of practical interest in the semiclassical treatment of molecular scattering[23] and in the phenomenological theory of lasers.[24] For practical applications, one has to obtain the unknown function $\rho(t)$. This can be obtained by a numerical integration of the nonlinear differential equation (25) for a given frequency function $\omega(t)$ and the damping coefficient γ with appropriate initial conditions. Knowing $\rho(t)$, integration of Eq. (31) provides the other unknown function $\mu(t)$ while the function $U(t)$ and $V(t)$ may be obtained from their defining Eqs. (27) and (28) once the perturbative force $f(t)$ is specified.

It may also be added here that the above wavefunction has essentially the form

$$\psi_n(x,t) = A_n(t)\exp\{-\tfrac{1}{2}[xB(t) - C(t)]^2\} \times H_n(x\,D(t) + E(t)) \tag{51}$$

and once this form is assumed, the unknown functions A_n, B, C, D, and E may be determined by substituting Eq. (51) in the Schrödinger equation (3) and subsequently solving the five coupled differential equations. Alternatively, the method of generating function described in Ref. 11, which again presupposes the form of the generating function, may be used to arrive at the wavefunction $\psi_n(x,t)$. However, the method of explicitly time-dependent invariants described in the preceding sections and the path integral approach described in the subsequent section do not require any ad hoc assumptions about the nature of the wavefunction to be obtained.

4. FEYNMAN PROPAGATOR

The propagator defined by the path integral of Eq. (4) may be expressed in the form of a multiple Riemann integral

$$K(x'',t'';x',t') = \lim_{N\to\infty} A_n \times \int\cdots\int \exp\left\{i\sum_{k=1}^{N} S_k(x_k,x_{k-1},\epsilon)\right\}\prod_{k=1}^{N-i} dx_k, \tag{52}$$

where $x_k = x(t_k)$ and A_N is the normalization in the Nth approximation. The action $S_k(x_k,x_{k-1},\epsilon)$ over an infinitesimally small time interval $t_k - t_{k-1} = \epsilon$ may be approximated by

$$S_k = \int_{t_{k-1}}^{t_k} L\,dt \simeq \epsilon L = e^{\gamma t_k}\left[\frac{1}{2\epsilon}(x_k - x_{k-1})^2 - \frac{\epsilon}{2}\omega_k^2 x_k^2 - \epsilon f_k x_k\right]. \tag{53}$$

The normalization factor A_N essentially corresponds to the free particle normalization including the dissipative factor $e^{\gamma t}$ and is given by

$$A_N = \prod_{k=1}^{N}\left(\frac{e^{\gamma t_k}}{2\pi i\epsilon}\right)^{1/2}. \tag{54}$$

The integrations involved in Eq. (52) can be performed by using the formula

$$\int_{-\infty}^{\infty}\exp[i(\alpha x^2 - (a+b+c)x]\,dx = \left(\frac{i\pi}{\alpha}\right)^{1/2}\exp\left[-i\frac{(a^2+b^2+c^2)}{4\alpha}\right] \times\exp\left[\frac{-i}{2\alpha}(ab+bc+ca)\right] \tag{55}$$

recursively. Using Eq. (53) and the definitions

$$\beta_k = e^{\gamma t_k}/\epsilon, \quad g_k = \epsilon^2\beta_k f_k,$$
$$\alpha_k = \tfrac{1}{2}[(\beta_k + \beta_{k+1}) - \beta_k\omega_k^2\epsilon^2], \tag{56}$$

the final result of these $(n-1)$ integrations can be expressed as

$$K(x'',t'';x',t') = \lim_{N\to\infty}(a_N/2\pi)^{1/2} \times\exp(ip_N x'^2 + iq_N x''^2 - ia_N x'x'') \times\exp[i(b_N x' + c_N x'' - r_N)], \tag{57}$$

where

$$a_N = \beta_N\xi_{N-1}, \tag{58}$$

$$p_N = \frac{\beta_1}{2}\left(1 - \sum_{k=1}^{N-1}\frac{\beta_1\beta_k\xi_{k-1}}{\beta_{k+1}}\right), \tag{59}$$

$$q_N = \frac{\beta_N}{2}\left(1 - \frac{\xi_{N-1}}{\xi_{N-2}}\right), \tag{60}$$

$$b_N = \sum_{k=1}^{N-1}\frac{\eta_k\xi_k\beta_1}{\beta_{k+1}}, \tag{61}$$

$$c_N = \frac{\eta_{N-1}\xi_{N-1}}{\xi_{N-2}}, \tag{62}$$

$$r_N = \sum_{k=1}^{N-1}\frac{\eta_k^2\xi_k}{2\beta_{k+1}\xi_{k-1}}. \tag{63}$$

The quantities (ϵ_k,η_k) are further defined by

$$\epsilon_k = \prod_{j=1}^{k}\left(\frac{\beta_{j+1}}{2\theta_j}\right), \quad k\geqslant 1, \tag{64}$$

$$\eta_k = g_k + \beta_k\eta_{k-1}/2\theta_{k-1}, \quad \eta_1 = g_1, \tag{64'}$$

$$\theta_k = \alpha_k - \beta_k^2/4\theta_{k-1}, \quad k\geqslant 2, \quad \theta_1 = \alpha_1. \tag{65}$$

The evaluation of the coefficients defined through Eqs. (58)–(63) in the limit $N\to\infty$ is the main issue in obtaining the analytical form of the propagator. Defining two new variables q_k and Λ_k by the relation

$$\frac{q_{k+1}}{q_k} = \Lambda_k = \frac{2\theta_k}{\beta_{k+1}}, \tag{66}$$

Eq. (65) is cast in the form

$$q_{k+1} = (1 - \omega_k^2\epsilon^2 + \epsilon^{-\gamma\epsilon})q_k - e^{-\gamma\epsilon}q_{k-1}, \tag{67}$$

which in the limit $\epsilon\to 0$ reduces to the differential equation

$$\ddot{q} + \omega^2(t)q + \gamma\dot{q} = 0. \tag{68}$$

Since from Eq. (66)

$$\Lambda_1 = \frac{q_2}{q_1} = \frac{2\theta_1}{\beta_2} = (1 - \omega_1^2\epsilon^2 + e^{-\gamma\epsilon}), \tag{69}$$

it follows from Eq. (67) (by setting k equal to 1) that

$$q_0 = q(t') = 0, \tag{70}$$

which provides one of the initial condition for the solution of Eq. (68). A solution of Eq. (68) satisfying the condition (70) is given by

$$q(t) = s(t)\, e^{-\gamma t/2} \sin[\nu(t) - \nu(t')], \quad (71)$$

where the functions obey the equations

$$\ddot{s} + \Omega^2(t) s - s^{-3} = 0, \quad (72)$$

$$\dot{\nu} s^2 = 1, \quad (73)$$

with

$$\Omega^2 = \omega^2 - \gamma^2/4.$$

Comparing these equations with Eqs. (25) and (34) the functions $s(t)$ and $\nu(t)$ may be identified with $\rho(t)$ and $\mu(t)$. Moreover, Eq. (71) provides a physical interpretation for $\rho(t)$ and $\mu(t)$ as quantities related to the amplitude and the phase of a classical damped oscillator with a real time-dependent frequency. We shall therefore use the symbols ρ and μ in place of s and ν respectively hereafter.

The limits of all coefficients (58)–(63) have been obtained in the Appendix A. It is shown there that

$$\lim_{N\to\infty} a_N = \frac{1}{\sigma'\sigma''} \csc\phi(t',t''), \quad (74)$$

$$\lim_{N\to\infty} p_N = \frac{1}{2}\left[\frac{\cot\phi(t'',t')}{\sigma'^2} - \frac{\dot\sigma'}{\sigma'} e^{\gamma t'}\right], \quad (75)$$

$$\lim_{N\to\infty} q_N = \frac{1}{2}\left[\frac{\cot\phi(t'',t')}{\sigma''^2} - \frac{\dot\sigma'' e^{\gamma t''}}{\sigma''}\right], \quad (76)$$

$$\lim_{N\to\infty} b_N = \frac{\csc\phi(t'',t')}{\sigma'} e^{\gamma t'/2} \int_{t'}^{t''} G(t) \sin\phi(t'',t)\, dt, \quad (77)$$

$$\lim_{N\to\infty} c_N = \frac{\csc\phi(t'',t')}{\sigma''} e^{\gamma t''/2} \int_{t'}^{t''} G(t) \sin\phi(t,t')\, dt, \quad (78)$$

$$\lim_{N\to\infty} r_N = \csc\phi(t'',t') \times \int_{t'}^{t''}\int_{t'}^{t} G(t) G(s) \sin\phi(t'',t) \sin\phi(s,t')\, ds\, dt, \quad (79)$$

where the function $G(t)$ is the same defined in Eq. (29). Substituting these limits in the Eq. (57) for the propagator, one obtains the following closed form:

$$\begin{aligned} K(x'',t''; x',t') &= \frac{1}{[2\pi i \sigma'\sigma'' \sin\phi(t'',t')]^{1/2}} \\ &\times \exp[\tfrac{1}{2} i(\sigma''\dot\sigma'' e^{\gamma t''} y''^2 - \sigma'\dot\sigma' e^{\gamma t'} y'^2)] \\ &\times \exp\{\tfrac{1}{2} i \csc\phi(t'',t')([(y''^2 + y'^2)\cos\phi(t'',t') - 2y'y''] \\ &+ y'' e^{\gamma t''/2} \int_{t'}^{t''} G(t) \sin\phi(t,t')\, dt \\ &+ y' e^{\gamma t'/2} \int_{t'}^{t''} G(t) \sin\phi(t'',t)\, dt \\ &- 2\int_{t}^{t''}\int_{t'}^{t''} G(t) G(s) \sin\phi(s,t') \sin\phi(t'',t)\, ds\, dt)\} \end{aligned} \quad (80)$$

with

$$\sigma = \rho e^{-\gamma t/2}, \quad y = x/\sigma. \quad (80')$$

It may be noted that when $\gamma = 0$ and $\omega(t)$ is a real positive constant ω_0 the solution of Eq. (25) yields $\rho(t) = 1(\omega_0)$ and by Eq. (34) $\mu(t) = \omega_0 t$. In this case, the propagator of Eq. (80) reduces to the expression given by Feynman and Hibbs[20] for the forced oscillator. When the perturbative force is absent $[f(t) \equiv 0]$ and $\omega(t)$ is a constant ω_0, Eq. (80) yields the propagator for a free damped oscillator evaluated by Papadopoulos.[25] Finally, the expression of Ref. 12 is also reproduced when γ and $f(t)$ are set equal to zero.

5. EXPANSION OF THE PROPAGATOR

It will now be shown that the propagator of Eq. (59) admits expansion in the form given by Eq. (6) or Eq. (16). First, the propagator is cast in an appropriate form using the following results derived in Appendix B:

$$\frac{1}{\sin\phi(t'',t')}\int_{t'}^{t''} G(t) \sin\phi(t,t')\, dt = U'' + \{\cot\phi(t'',t')\} V'' - \{\csc\phi(t'',t')\} V', \quad (81)$$

$$\frac{1}{\sin\phi(t'',t')}\int_{t'}^{t''} G(t) \sin\phi(t'',t)\, dt = -U' + V' \cot\phi(t'',t') - V'' \csc\phi(t'',t'), \quad (82)$$

$$\begin{aligned} &\frac{1}{\sin\phi(t'',t')}\int_{t'}^{t''}\int_{t'}^{t''} G(t) G(s) \sin\phi(s,t') \sin\phi(t'',t)\, ds\, dt \\ &= -\tfrac{1}{2}(V'^2 + V''^2)\cot\phi(t'',t') + V'V'' \csc\phi(t'',t') + (F'' - F'), \end{aligned} \quad (83)$$

where the functions U, V, and F have been defined by Eqs. (27), (28) and (49) and the prime and double prime denote the quantities evaluated at time t' and t'' respectively. With the help of these results and after suitable rearrangement of terms, the propagator of Eq. (80) may be rewritten as

$$\begin{aligned} K(x'',t''; x',t') &= \frac{1}{(\pi\sigma'\sigma'')^{1/2}} \\ &\times \exp[\tfrac{1}{2} i(\sigma''\dot\sigma'' e^{\gamma t''} y''^2 - \sigma'\dot\sigma' e^{\gamma t'} y'^2)] \\ &\times \exp[\tfrac{1}{2} i(U''y'' - U'y')] \frac{e^{-i\phi}}{(1 - e^{-2i\phi})^{1/2}} \\ &\times \exp\left[\frac{-1}{1 - e^{2i\phi}}(X''^2 + X'^2 - 2X'X'' e^{-i\phi})\right] \\ &\times \exp\{\tfrac{1}{2}(X''^2 + X'^2)\} \exp[-i(F'' - F')], \end{aligned} \quad (84)$$

where $\phi = \phi(t'',t')$ and

$$X(t) = y + V(t). \quad (85)$$

Using Mehler's formula,[26]

$$\exp[-(x^2 + y^2 - 2xyz)/(1 - z^2)]/(1 - z^2)^{1/2} = \exp[-(x^2 + y^2)] \sum_{n=0}^{\infty} \frac{z^n}{2^n n!} H_n(x) H_n(y), \quad (86)$$

and, letting $z = e^{-i\phi}$, $x = x''$, $y = x'$ and noting that $\phi = \mu(t'') - \mu(t')$, it is easy to see that the expansion of Eq. (6) or that of Eq. (16) follows with

$$\psi_n(x,t) = e^{i\alpha_n(t)} \phi_n(x,t), \quad (87)$$

where

$$\alpha_n(x,t) = -(n + \tfrac{1}{2})\mu(t) - F(t) \quad (88)$$

and

$$\phi_n(x,t) = \frac{1}{\pi^{1/4}\sqrt{\sigma}} \exp[\tfrac{1}{2} i(e^{\gamma t}\sigma\dot\sigma y^2 + Uy)] \times \exp(-\tfrac{1}{2}X^2) H_N(X). \quad (89)$$

With the help of the defining equation for X it is easily verified that $\phi_n(x,t)$ of Eq. (89) are the same as the eigenfunctions of $\hat{I}(t)$ obtained in Eq. (44) and that $\alpha_n(t)$ of Eq. (88) coincide with the phases obtained earlier in Eq. (47) of Sec. 2. This completes the derivation of the wavefunction from the Feynman propagator.

6. CONCLUSIONS

Two alternative approaches were considered to solve the problem of a quantal oscillator with damping and time-dependent frequency acted on by a time-dependent perturbative force. In the first approach, an explicitly time-dependent invariant Hermitian operator and its eigenvalues and eigenfunctions were obtained. These were used along with appropriate phase factors to construct the wavefunctions of the Schrödinger equation in exact closed form. The second approach involved a path integration of the classical Lagrangian yielding an exact Feynman propagator. The wavefunctions were then obtained by an expansion of the propagator. It was found that the wavefunctions and the propagator depend only on a single function $\rho(t)$ related to the amplitude of the classical damped oscillator with a time dependent frequency. The function $\rho(t)$ in turn may, in practical situations, be obtained from a numerical integration of a second order nonlinear differential equation under appropriate set of initial conditions.

APPENDIX A

In this appendix limits of Eqs. (58)–(63) as $N\to\infty$ have been evaluated.

In order to obtain the desired limits, all the coefficients must be expressed in terms of the ratio q_j/q_k. This may be done by noting (as $q_0 = 0$)

$$\xi_1 = \prod_{j=1}^{k}(1/A_j) = q_1/q_{k+1} = \frac{\epsilon\dot{q}_0}{Qq_k + \epsilon\dot{q}_k} + O(\epsilon^2)$$
$$= \frac{\epsilon\dot{q}_0}{q_k\ \eta\epsilon^2) \simeq \epsilon\dot{q}_0(t')/q_k} \tag{A1}$$

and

$$\eta_k = g_k + \frac{\eta_{k-1}q_{k-1}}{q_k} = \sum_{j=1}^{k} g_j\, q_j/q_k$$
$$\simeq \frac{1}{\epsilon}\int_{t'+\epsilon}^{t_k}\frac{g(\tau)q(\tau)}{q_k}\,d\tau. \tag{A2}$$

Substitution of these values in Eqs. (58)–(63) gives us the desired result. It is then easy to evaluate limits. First

$$\lim_{N\to\infty} a_n = \lim_{N\to\infty}\frac{\beta_1 q_1}{q_N} = \lim_{\epsilon\to 0} e^{\gamma(t'+\epsilon)}\epsilon\dot{q}(t')/\epsilon q(t'')$$
$$= \frac{e^{\gamma(t'+t'')}}{\rho'\rho''}\csc\phi(t'',t') = \frac{\csc\phi(t'',t')}{\sigma'\sigma''}, \tag{A3}$$

where

$$\phi(t'',t') = \mu(t'') - \mu(t')$$

and

$$\sigma = \rho e^{-\gamma t/2},\quad \sigma' = \sigma(t'),\quad \sigma'' = \sigma(t'').$$

Next,

$$\lim_{N\to\infty} p_N = \lim_{N\to\infty}\frac{\beta_1}{2}\left(1 - \sum_{k=1}^{N}\frac{\beta_1}{\beta_{k+1}}\frac{q_1^2}{q_kq_{k+1}}\right)$$
$$= \lim_{\epsilon\to 0}\frac{e^{\gamma(t'+\epsilon)}}{2\epsilon}\left(1 - \epsilon\right.$$
$$\left.\times\int_{t'+\epsilon}^{t''} e^{-\gamma(t-t')}\dot{q}^2(t')/q^2(t)\,dt\right). \tag{A4}$$

Substituting the value of q and $\dot{q}$ explicitly from Eq. (66) and noting that $\dot{\mu}\rho^2 = 1$, the integration involved in the evaluation of the limit can be easily performed to yield

$$\lim_{\epsilon\to 0}\frac{e^{\gamma t'}}{2\epsilon}(1 + \gamma\epsilon + \cdots)$$
$$\times\{1 + \epsilon[\dot{\mu}'\cot\phi(t'',t') - \dot{\mu}'\cot(t'+\epsilon,t')]\}$$
$$= \lim_{\epsilon\to 0}\frac{\gamma e^{\gamma t'}}{2} + \frac{e^{\gamma t'}}{2}\left[\dot{\mu}'\cot\phi(t'',t')\right.$$
$$\left. + \frac{\sin\phi(t'+\epsilon,t') - \epsilon\dot{\mu}'\cos\phi(t'+\epsilon,t')}{\epsilon\sin\phi(t'+\epsilon,t')}\right]. \tag{A5}$$

The latter limit exists and is equal to $\ddot{\mu}'/2\dot{\mu}' = -\dot{\rho}'/\rho'$ as can be verified by expanding $\phi(t'+\epsilon,t')$ up to second order in ϵ. Thus we obtain

$$\lim_{N\to\infty} p_N = \frac{\gamma e^{\gamma t'}}{2} + \frac{e^{\gamma t'}}{2}\left(\dot{\mu}'\cot\phi(t'',t') - \frac{\dot{\rho}'}{\rho'}\right)$$
$$= \frac{e^{\gamma t'}}{2}\left[\frac{1}{\rho'^2}\cot\phi(t'',t') - \left(\frac{\dot{\rho}'}{\rho'} - \frac{\gamma}{2}\right)\right]$$
$$= \frac{1}{2}\left[\frac{\cot\phi(t'',t')}{\sigma'^2} - \frac{e^{\gamma t'}\dot{\sigma}'}{\sigma'}\right]. \tag{A6}$$

Next,

$$\lim_{N\to\infty} q_N = \lim_{N\to\infty}\frac{\beta_N}{2}\left(1 - \frac{q_{N-1}}{q_N}\right)$$
$$= \frac{\beta_N}{2}\left(1 - \frac{q_N - \epsilon\dot{q}_N}{q_N}\right) = \frac{e^{\gamma t''}}{2}\frac{\dot{q}(t'')}{q(t'')}$$
$$= \frac{e^{\gamma t''}}{2}\left[\frac{\cot\phi(t'',t')}{\rho''} + \left(\frac{\dot{\rho}''}{\rho''} - \frac{\gamma}{2}\right)\right]$$
$$= \frac{1}{2}\left[\frac{\cot\phi(t'',t')}{\sigma''^2} + \frac{e^{\gamma t''}\dot{\sigma}''}{\sigma''}\right].$$

Now

$$\lim_{N\to\infty} b_N = \beta_1\sum\frac{\eta_kq_1}{\beta_{k+1}q_k}$$
$$= \lim_{\epsilon\to 0}\int_{t'+\epsilon}^{t''} d\tau\,\frac{1}{\epsilon^2}\,\frac{e^{\gamma(\tau-t')}\epsilon\dot{q}(t')}{q(\tau)}$$
$$\times\int_{t'+\epsilon}^{\tau}\frac{g(s)\,q(s)}{q(\tau)}\,ds. \tag{A7}$$

Substituting explicitly the values of $g(s)$ from Eq. (51), $q(t)$ from Eq. (66) and then carrying out a partial integration w.r.t. τ one obtains by noting

$$\int\dot{\mu}(\tau)\csc^2\phi(\tau,t')\,d\tau = -\cot\phi(\tau,t'),$$

$$\lim_{N\to\infty} b_N = \frac{\csc\phi(t'',t')e^{\gamma t'/2}}{\sigma'}$$
$$\times\int_{t'}^{t''} G(t)\,\sin\phi(t'',t)\,dt. \tag{A8}$$

The next limit to be evaluated is

$$\lim_{N\to\infty} c_N = \lim_{N\to\infty} \eta_{N-1} q_{N-1}/q_N$$

$$= \lim_{\substack{N\to\infty\\ \epsilon\to 0}} \frac{1}{\epsilon}\int_{t'+\epsilon}^{t_{N-1}} \frac{g(s)q(s)q_{N-1}}{q_{N-1}q_N}\, dS$$

$$= \frac{e^{\gamma t''/2}\csc\phi\,(t'',t')}{\sigma''}$$

$$\times \int_{t'}^{t''} G(t)\ \sin\phi\,(t'',t)\, dt \qquad \text{(A9)}$$

and lastly

$$\lim_{N\to\infty} r_N = \lim_{N\to\infty} \sum_{k=1}^{N-1} \frac{\eta_k^2\, q_k}{2\beta_{k+1}\, q_{k+1}}$$

$$= \lim_{\epsilon\to 0}\frac{1}{\epsilon^2}\int_{t'+\epsilon}^{t''-\epsilon} \frac{\epsilon q(\tau)\, e^{-\gamma\tau}}{q(\tau+\epsilon)}$$

$$\times \int_{t'}^{\tau}\int_{t'}^{\tau} \frac{g(s)g(s')q(s)q(s')}{q^2(\tau)}\, ds\, ds'.$$

Substituting explicitly the values of q, q', and g from Eqs. (66) and (51), we get

$$\lim_{N\to\infty} r_N = \int_{t'}^{t''} \frac{\dot\mu(\tau)\, d\tau}{\sin^2\phi\,(\tau,t)}$$

$$\times \int_{t'}^{\tau}\int_{t'}^{\tau} e^{\gamma s} f(s) e^{\gamma s'} f(s')\, q(s)\, q(s')\, ds\, ds'.$$

Partially integrating once w.r.t. τ and rearranging the terms, we get

$$\lim_{N\to\infty} r_N = \csc\phi\,(t'',t')$$

$$\times \int_{t'}^{t''}\int_{t'}^{t} G(s)G(t)\sin\phi\,(t'',t)\, ds\, dt, \qquad \text{(A10)}$$

which is the result of Eq. (63).

APPENDIX B

In this appendix Eqs. (76)–(78) have been derived. Consider the integral

$$\int_s^t G(\tau)\ \sin\phi\,(\tau,t)\, d\tau$$

$$= -\int_{t_0}^{s} G(\tau)\ \sin\phi\,(\tau,t)\, d\tau + \int_{t_0}^{t} G(\tau)\ \sin\phi\,(\tau,t)\, d\tau. \qquad \text{(B1)}$$

Now, since from the definition of

$$\phi\,(\tau,t) = \phi\,(\tau,s) + \phi\,(s,t) \qquad \text{(B2)}$$

it follows that

$$\sin\phi\,(\tau,t) = \sin\phi\,(\tau,s)\cos\phi\,(s,t) + \cos\phi\,(\tau,s)\sin\phi\,(s,t). \qquad \text{(B3)}$$

Using Eq. (B3) in the first term on the right-hand side of Eq. (B1) and dividing throughout by $\sin\phi\,(t,s)$, one obtains

$$\frac{1}{\sin\phi\,(t,s)}\int_s^t G(\tau)\sin\phi\,(\tau,t)\, d\tau$$

$$= -V(s)\cot\phi\,(t,s) + V(t)\csc\phi\,(t,s) + U(s). \qquad \text{(B4)}$$

Thus Eq. (76) is obtained if one sets $t = t'$, $s = t''$ while Eq. (77) results when $t = t''$, $s = t'$ is set in Eq. (B4). Next consider the double integral

$$J = \int_{t'}^{t''}\int_{t'}^{t} I(t,t'',s,t')\, ds\, dt$$

with

$$I = G(t)\, G(s)\sin\phi\,(s,t')\sin\phi\,(t'',t). \qquad \text{(B5)}$$

After breaking the range of integration at $t = t_0$, one may write

$$J = J_1 + J_2 + J_3 + J_4, \qquad \text{(B6)}$$

where

$$J_1 = \int_{t_0}^{t''}\int_{t_0}^{t} I\, ds\, dt,\quad J_2 = -\int_{t_0}^{t''}\int_{t_0}^{t'} I\, ds\, dt,$$

$$J_3 = \int_{t_0}^{t'}\int_{t_0}^{t} I\, ds\, dt,\quad J_4 = -\int_{t_0}^{t'}\int_{t_0}^{t'} I\, ds\, dt, \qquad \text{(B7)}$$

Replacing $\sin\phi\,(s,t')$ by

$$\sin\phi\,(s,t') = \sin\phi\,(s,t'')\cos\phi\,(t'',t') + \cos\phi\,(s,t'')\sin\phi\,(t'',t'),$$

it is easy to obtain

$$J_1 = -\tfrac{1}{2}V''^2\cos\phi\,(t'',t') + \sin\phi\,(t'',t')\tilde{J}_1(t''), \qquad \text{(B8)}$$

where

$$\tilde{J}_1(\tau) = \int_{t_0}^{\tau}\int_{t_0}^{t} G(t)\, G(s)\cos\phi\,(s,\tau)\sin\phi\,(\tau,t)\, ds\, dt. \qquad \text{(B9)}$$

To simplify $\tilde{J}_1(\tau)$, we further differentiate both sides of Eq. (B9) w.r.t. τ to obtain

$$\frac{d\tilde{J}_1(\tau)}{d\tau} = \frac{U^2(\tau) - V^2(\tau)}{2\rho^2(\tau)}. \qquad \text{(B10)}$$

Since by definition $\tilde{J}_1(t_0) = 0$, Eq. (B10) on integration yields

$$\tilde{J}_1(t'') = \int_{t_0}^{t''}\left(\frac{U^2 - V^2}{2\rho^2}\right) d\tau = F(t''). \qquad \text{(B11)}$$

Substituting Eq. (B11) in Eq. (B8), one gets

$$J_1 = -\tfrac{1}{2}V''^2\cos\phi\,(t'',t') + F''\sin\phi\,(t'',t'). \qquad \text{(B12)}$$

Next, it is straightforward to show that

$$J_2 = V'V'' \qquad \text{(B13)}$$

while J_3 and J_4 may be simplified after substituting

$$\sin\phi\,(t'',t) = \sin\phi\,(t'',t)\cos\phi\,(t'',t) + \cos\phi\,(t'',t')\sin\phi\,(t',t) \qquad \text{(B14)}$$

in the definition of J_3 and J_4 to obtain

$$J_4 = V'U'\sin\phi\,(t'',t') - V'^2\cos\phi\,(t'',t') \qquad \text{(B15)}$$

while

$$J_3 = \frac{V'^2}{2}\cos\phi\,(t'',t') - \tilde{J}_3(t')\sin\phi\,(t'',t'), \qquad \text{(B16)}$$

where

$$\tilde{J}_3(\tau) = \int_0^{\tau}\int_{t_0}^{\tau} G(t)\, G(s)\sin\phi\,(s,\tau)\cos\phi\,(\tau,t)\, ds\, dt. \qquad \text{(B17)}$$

Differentiating $\tilde{J}_3(\tau)$ w.r.t. τ, we get

$$\frac{d\widetilde{J}_3}{d\tau} = G(\tau)\,V(\tau) + \frac{1}{2\rho^2}(U^2 - V^2). \quad \text{(B18)}$$

Using easily verified results

$$\dot{U} = G + V/\rho^2, \quad \dot{V} = -U/\rho^2, \quad \text{(B19)}$$

Eq. (B18) is cast in the form

$$\frac{d\widetilde{J}_3}{d\tau} = \frac{d}{dt}(UV) + \frac{1}{2\rho^2(U^2 - V^2)}, \quad \text{(B20)}$$

which on integration over (t_0, t') yields

$$\widetilde{J}_3(t') = U'V' + F'. \quad \text{(B21)}$$

Substituting values J_1, J_2, J_3, and J_4 from Eqs. (B12), (B13), (B15), (B16), and Eq. (B21) in Eq. (B6) and then dividing it throughout by $\sin\phi\,(t'',t)$, we obtain the result of Eq. (78).

[1]R.P. Feynman, Rev. Mod. Phys. **20**, 367 (1948); Phys. Rev. **84**, 108 (1951).
[2]J. Schwinger, Phys. Rev. **91**, 728 (1953).
[3]K. Husimi, Progr. Theoret. Phys. (Kyoto) **9**, 381 (1953).
[4]A.M. Dykhne, Zh. Eksp. Teor. Fiz. **38**, 570 (1960) [Sov. Phys. JETP **11**, 411 (1960)].
[5]M. Kolsrud, Phys. Rev. **104**, 1186 (1956).
[6]R.M. Kolsrud, Phys. Rev. **106**, 205 (1957).
[7]F. Hetwek and A. Schluter, Z. Naturforschr. **12a**, 844 (1957).
[8]C. Backus, A. Lenard, and R. Zulsud, Z. Naturforschr. **15a**, 1458 (1969).
[9]H.R. Lewis and W.B. Reisenfeld, J. Math. Phys. **10**, 1458 (1969).
[10]P.G.L. Leach, J. Math. Phys. **18**, 1608, 1902 (1977).
[11]P. Camiz, A. Gerardi, C. Marchioro, E. Presutti, and E. Scacciatelli, J. Math Phys. **12**, 2040 (1971).
[12]D.C. Khandekar and S.V. Lawande, J. Math. Phys. **16**, 1458 (1975), hereafter referred as I.
[13]D.C. Khandekar and S.V. Lawande, to be published in Phys. Lett. A, hereafter referred as II.
[14]P. Havas, Nuovo Cimento Suppl. **5**, 363 (1957).
[15]F. Bopp, Z. Angew. Phys. **14**, 699 (1962).
[16]P.A.M. Dirac, *Principles of Quantum Mechanics* (Clarendon, Oxford, 1958), 4th ed.
[17]J.H. Van Vleck, Proc. Nat. Acad. Sci. USA **14**, 178 (1928).
[18]W. Pauli, Ausgewalte Kapitel der Feldquantisierung (Lecture notes) (ETH Zurich, 1952), pp. 139–52.
[19]A.V. Jones and G.J. Papadopoulos, J. Phys A **4**, L86 (1971).
[20]R.P. Feynman and A.R. Hibbs, *Quantum Mechanics and Path Integrals* (McGraw-Hill, New York, 1965), pp. 58–64.
[21]G.J. Papadopoulos, Phys. Rev. D **11**, 2870 (1975).
[22]M.J. Goovaerts, Physica **77**, 379 (1974).
[23]X. Chapuisat and J. Manz, Molecular Physics **26**, 1577 (1973).
[24]I.R. Senitzky, Phys. Rev. B **95**, 1115 (1954).
[25]G.J. Papadopoulos, J. Physics A **7**, 209 (1974).
[26]A. Erdelyi (Ed.), *Higher Transcendental Functions* (McGraw-Hill, New York, 1953), Vol. 2, p. 194.

Explicit solution of the wave equation for arbitrary power potentials with application to charmonium spectroscopy

H. J. W. Müller-Kirsten, G. E. Hite, and S. K. Bose[a)]
Department of Physics, University of Kaiserslautern 6750 Kaiserslautern, Germany

(Received 1 December 1978)

We present an explicit and almost complete series solution of the Schrödinger equation for an arbitrary quark-confining power potential with or without a weak Coulomb component or other corrections. In particular, we derive two pairs of high-energy asymptotic expansions of the bound-state eigenfunctions together with a corresponding expansion of the eigenvalue determined by the secular equation. We also obtain a pair of uniformly convergent expansions and discuss other types of solutions. Various properties of the solutions and eigenvalues are examined including the scattering problem of the cutoff potential and the behavior of Regge trajectories. Finally, the relevance of these investigations to the spectroscopy of heavy quark composites is discussed. In particular, we derive approximate expressions for leptonic decay rates. Examples are given to demonstrate the usefulness of these results for theoretical discussion and as alternatives for numerical integration techniques. A subsequent paper will deal with the normalization of the bound-state wavefunctions and the corresponding derivation of explicit series expressions for certain decay rates.

1. INTRODUCTION

Over the last few years numerous attempts have been undertaken in order to understand the level spacing and the decay rates of heavy mesonic states in the ψ and Υ regions (see, for example, Refs. 1–5). Most of these attempts start from a nonrelativistic consideration of the bound-state problem for a vector (or scalar) interaction and employ numerical techniques for the integration of the relevant wave equation. Effects of spin–orbit coupling, tensor forces, and hyperfine structure are considered as additional corrections. Although the gross features of the spectra are recovered in most of these attempts (a well known difficulty is the problem of pseudoscalar states), a detailed comparison[6] of these attempts shows that one or the other of the parameters involved can differ by as much as 50% in different models. such a difference in the values of parameters which contribute effects of second or higher order may not be highly significant at the present stage of knowledge, but it indicates a certain range of uncertainty in their values. Although numerical methods seem indispensable in this type of work, it would be reassuring to have at one's disposal theoretical formulas for extracting at least good approximations for the level spacings, leptonic decay rates, etc., since these would also permit a better understanding of the relative magnitudes of various contributing effects. This is one reason why we undertake in the following a detailed perturbation theoretical solution of the wave equation for the superposition of a quark confining power potential and a short-range gluon exchange Coulomb component.

There are also other motivations for undertaking this type of work (quite apart from its interest in quantum mechanics and mathematical physics). The discovery[7] of the Υ states at Fermilab in the mass spectrum of muon pairs produced in the bombardment of nuclei by high-energy protons has raised the question whether the linear potential is indeed the proper or most convenient phenomenological ansatz for the quark confining interaction. In fact, it is difficult to fit all the essential data in both the ψ and Υ regions with one and the same linear potential. Quigg and Rosner[8] therefore began an investigation of the spectroscopy resulting from a logarithmic potential or even an arbitrary power potential.[9] These investigations are useful not only for singling out the approximate form of the potential (which, of course, will eventually have to be derived from field theory), but can also be a valuable aid in exploring the spectroscopy of numerous other states which may be discovered by the new generation of accelerators. In the following, however, we consider applications only in the field of charmonium spectroscopy.

We begin by considering in detail the eigenvalue problem defined by the Schrödinger equation for a superposition of Coulomb and arbitrary quark confining potentials. We then derive various types of solutions for the wave functions as well as explicit expansions for the energy eigenvalues and Regge trajectories. Our methods of solution are very general and parallel the methods used for solving the wave equation for a logarithmic potential[10] or the methods of solution of more complicated standard differential equations such as the Mathieu[11] or spheroidal wave equation.[12] In Secs. 2 and 3 we derive two pairs of high energy asymptotic expansions for the discrete eigenfunctions together with the corresponding asymptotic expansion for the eigenvalues. In Sec. 4 we derive a pair of uniformly convergent expansions for the solutions and discuss their relevance for the scattering problem of the cutoff confinement potential. In Sec. 5, then, we investigate the physical implications of the asymptotic expansions of the energy eigenvalues and Regge trajectories. In particular, we extend our considerations by incorporating spin-dependent corrections (spin–orbit, tensor, spin–spin) and calculate the splitting of the ^3P_J levels of charmonium. Finally, we calcu-

[a)]A.v. Humboldt foundation fellow on leave from the University of Gorakhpur, Gorakhpur, India.

 0022-2488/79/091878-13$01.00 © 1979 American Institute of Physics

late the S-wave bound-state wave function at the origin and leptonic decay rates. In Sec. 6 we make some concluding remarks.

2. A FIRST PAIR OF ASYMPTOTIC EIGENSOLUTIONS

We consider the Schrödinger equation for an unscreened power potential gr^{λ} which is modified by the addition of a Coulomb component, i.e.,

$$V(r) = g_1 r^{\lambda} - g_0/r + V_0, \tag{1}$$

where $\lambda \geqslant 1$ and $g_1 > 0$, and V_0 is a constant. Separating off the motion of the center of mass in the usual way, we obtain the radial wave equation for the relative motion of the two particles of masses m_1, m_2, i.e.,

$$\frac{d^2\psi}{dr^2} + \frac{2\mu}{\hbar^2}\left[E - \frac{l(l+1)\hbar^2}{2\mu r^2} - V(r)\right]\psi = 0, \tag{2a}$$

where, as usual, $\Psi = (1/r)\psi(r)P_l^m(\cos\theta)e^{im\varphi}$, $\mu = m_1 m_2/(m_1 + m_2)$ is the reduced mass of the two particles and r is their separation.

Inserting the potential, we have the equation

$$\frac{d^2\psi}{dr^2} + \left(\alpha - \beta r^{\lambda} + \frac{\delta}{r} - \frac{\gamma}{r^2}\right)\psi = 0, \tag{2b}$$

where we have set

$$\alpha = 2\mu(E - V_0)/\hbar^2, \quad \beta = 2\mu g_1/\hbar^2,$$
$$\delta = 2\mu g_0/\hbar^2, \quad \gamma = l(l+1) \equiv L^2 - \tfrac{1}{4}. \tag{3}$$

Next we set

$$r = e^z \quad (-\infty < z < \infty). \tag{4}$$

Setting also

$$\psi = e^{z/2}\phi, \tag{5}$$

we obtain our basic equation

$$\frac{d^2\phi}{dz^2} + [-L^2 + v(z)]\phi = 0, \tag{6}$$

where

$$v(z) = \alpha\, e^{2z} - \beta\, e^{(2+\gamma)z} + \delta\, e^{z}. \tag{7}$$

Our next step is to find *that* value of z, say z_0, for which $v(z)$ becomes maximal. In the vicinity of this maximum, $v(z) - L^2$ can become positive and the solutions therefore oscillatory as required for the existence of eigenvalues. Thus, setting $(dv/dz)_{z=z_0} = 0$ and soving for z_0, we find

$$z_0 = \ln\rho + \ln\left(1 + \frac{\delta}{2\alpha\rho\lambda} - \frac{(\lambda+1)\delta^2}{2(2\alpha\rho\lambda)^2} + \frac{(\lambda+1)(\lambda+2)\delta^3}{3(2\alpha\rho\lambda)^3} - \cdots\right), \tag{8}$$

where

$$\rho = \left[\frac{2\alpha}{(2+\lambda)\beta}\right]^{1/\lambda} \tag{9}$$

for $\alpha > 0, \beta > 0$.

Expanding $v(z)$ in the neighborhood of the maximum at z_0, we obtain

$$v(z) = v(z_0) + \sum_{i=2}^{\infty} \frac{(z - z_0)^i}{i!} v^{(i)}(z_0), \tag{10}$$

where, for $i = 0,1,2,\cdots$,

$$v^{(i)}(z_0) = 2\alpha\, e^{2z_0}\left[2^{i-1} - (2+\lambda)^{i-1}\cdot\left(1 + \frac{\delta}{2\alpha\rho\lambda} - \frac{(\lambda+1)\delta^2}{2(2\alpha\rho\lambda)^2} + \frac{(\lambda+1)(\lambda+2)\delta^3}{3(2\alpha\rho\lambda)^3} - \cdots\right)^{\lambda}\right] + \delta e^{z_0}. \tag{11}$$

For $i = 0$ this expression is positive, for $i = 1$ it is zero, and for $i > 1$ it is negative [as required for a maximum of $v(z)$ at $z = z_0$ for $\alpha > 0$]. We now set

$$h = [-2v^{(2)}(z_0)]^{1/4},$$

i.e.,

$$h^4 = 4\alpha\lambda\rho^2\left[1 + (\lambda+3)\frac{\delta}{2\alpha\rho\lambda} + \left(\frac{\delta}{2\alpha\rho\lambda}\right)^2 + \cdots\right], \tag{12}$$

$$h^2 = 2(\alpha\lambda)^{1/2}\rho\left[1 + (\lambda+3)\frac{\delta}{4\alpha\rho\lambda} - \frac{(\lambda+1)(\lambda+5)}{2}\left(\frac{\delta}{4\alpha\rho\lambda}\right)^2 + \cdots\right], \tag{13}$$

and change the independent variable in Eq. (6) to

$$\omega = h(z - z_0).$$

The equation then becomes

$$\frac{d^2\phi}{d\omega^2} + \left(\frac{-L^2 + v(z_0)}{h^2} - \frac{\omega^2}{4}\right)\phi = \sum_{i=3}^{\infty}\left(\frac{v^{(i)}(z_0)}{2v^{(2)}(z_0)}\right)\frac{\omega^i}{i!h^{i-2}}\phi, \tag{14}$$

where

$$\frac{v^{(i)}(z_0)}{v^{(2)}(z_0)} = \frac{1}{\lambda}[(2+\lambda)^{i-1} - 2^{i-1}] - \frac{\delta}{\lambda(2\alpha\rho\lambda)} \times [(2+\lambda)^{i-1} + \lambda - 2^{i-1}(\lambda+1)] + \frac{\delta^2(\lambda+2)}{\lambda(2\alpha\rho\lambda)^2} \times [(2+\lambda)^{i-1} + \lambda - 2^{i-1}(\lambda+1)] + \cdots. \tag{15}$$

In particular, we have, for $i = 3,4$,

$$\frac{v^{(3)}(z_0)}{v^{(2)}(z_0)} = (\lambda+4) - \frac{\delta(\lambda+1)}{2\alpha\rho\lambda}, + \frac{\delta^2(\lambda+1)(\lambda+2)}{(2\alpha\rho\lambda)^2} + \cdots$$

and

$$\frac{v^{(4)}(z_0)}{v^{(2)}(z_0)} = (\lambda^2 + 6\lambda + 12) - \frac{\delta(\lambda+1)(\lambda+5)}{2\alpha\rho\lambda} + \frac{\delta^2(\lambda+1)(\lambda+2)(\lambda+5)}{(2\alpha\rho\lambda)^2} + \cdots.$$

In the following (except for Sec. 4) we assume that the parameter h is large (i.e., $h^2 \gg 1$), so that $1/h$ is the perturbation parameter of our expansions; the latter will therefore be arranged in descending powers of h. It will be seen, moreover, that our expansions are asymptotic in h and are therefore ideally suited for large but finite values of h, so that the least term of the expansion comes rather late and is correspondingly small. It is well known that asymptotic expansions are useful even for moderately large values of the parameter which is assumed to be large, and using modern methods for handling such expansions, their usefulness can be extended into regions of even smaller values of it. As was pointed out previously,[10] the dominant terms of our expansions correspond to those provided by the ususal WKB approximation. For the parameters used below, h^2 is larger than 4.

For large values of h, the right-hand side of Eq. (14) can—to a first approximation—be neglected. The corresponding behavior of the "eigenvalues"

$$[-L^2+v(z_0)]/h^2$$

can then be determined by comparing the equation with the equation of parabolic cylinder functions. The solutions are square integrable only if

$$[-L^2+v(z_0)]/h^2=\tfrac{1}{2}q,$$

where q is an odd integer, i.e., $2n+1, n=0,1,2,\cdots$ (provided the wave function is required to vanish at infinity; otherwise it is only approximately an odd integer, as we explain later in Sec. 4). For the complete solution we set

$$[-L^2+v(z_0)]/h^2=\tfrac{1}{2}q+\Delta/h. \tag{16}$$

The quantity Δ in Eq. (16) remains to be determined. We proceed as follows: Substituting Eq. (16) into (14), we have an equation which can be written

$$\mathscr{D}_q\phi=\frac{2\Delta}{h}\phi-\sum_{i=3}^{\infty}\left(\frac{v^{(i)}(z_0)}{v^{(2)}(z_0)}\right)\frac{\omega^i}{i!h^{i-2}}\phi, \tag{17}$$

where

$$\mathscr{D}_q\equiv-2\frac{d^2}{d\omega^2}-q+\frac{1}{2}\omega^2. \tag{18}$$

Equation (17) is now in a form suitable for the application of our perturbation method. To a first approximation, $\phi=\phi^{(0)}$ is simply a parabolic cylinder function $D_{(q-1)/2}(\omega)$, i.e.,

$$\phi^{(0)}=\phi_q=D_{(q-1)/2}(\omega),\quad \mathscr{D}_q\phi_q=0. \tag{19}$$

We have

$$D_{(q-1)/2}(\omega)=2^{(q-3)/4}e^{-\omega^2/4}\Psi\left(\frac{3-q}{4},\frac{3}{2};\frac{\omega^2}{2}\right),$$

where Ψ is a confluent hypergeometric function. The function ϕ_q is well known to obey the recurrence formula

$$\omega\phi_q=(q,q+2)\phi_{q+2}+(q,q-2)\phi_{q-2}, \tag{20}$$

where

$$(q,q+2)=1,\quad (q,q-2)=\tfrac{1}{4}(q-1). \tag{21}$$

For higher powers we have

$$\omega^i\phi_q=\sum_{j=2i}^{-2i}S_i(q,j)\phi_{q+j} \tag{22}$$

and a recurrence relation can be written down for the coefficients S_i. The first approximation $\phi=\phi^{(0)}$ then leaves uncompensated terms amounting to

$$\mathscr{R}_q^{(0)}=\left[\frac{2\Delta}{h}-\sum_{i=3}^{\infty}\left(\frac{v^{(i)}(z_0)}{v^{(2)}(z_0)}\right)\frac{\omega^i}{i!h^{i-2}}\right]\phi_q(\omega)$$

$$=\frac{2\Delta}{h}\phi_q-\sum_{i=3}^{\infty}\frac{1}{h^{i-2}}\sum_{j=2i}^{-2i}\widetilde{S}_i(q,j)\phi_{q+j}(\omega), \tag{23}$$

where we have set

$$\widetilde{S}_i(q,j)=\frac{v^{(i)}(z_0)}{v^{(2)}(z_0)}\cdot\frac{1}{i!}S_i(q,j). \tag{24}$$

We rewrite Eq. (23) in the form

$$\mathscr{R}_q^{(0)}=\sum_{i=3}^{\infty}\frac{1}{h^{i-2}}\sum_{j=2i}^{-2i}[q,q+j]_i\phi_{q+j}(\omega), \tag{25}$$

where

$$[q,q]_3=2\Delta-\widetilde{S}_3(q,0),$$

and for $j\neq0$

$$[q,q+j]_3=-\widetilde{S}_3(q,j), \tag{26}$$

and for $i>3,\ -2i\leqslant j\leqslant 2i$,

$$[q,q+j]_i=-\widetilde{S}_i(q,j).$$

Since $\mathscr{D}_{q+j}=\mathscr{D}_q-j$, $\mathscr{D}_q\phi_{q+j}=j\phi_{q+j}$, a term $\mu\phi_{q+j}$ in $\mathscr{R}_q^{(0)}$ can be removed by adding to $\phi^{(0)}$ the contribution $\mu\phi_{q+j}/j$ except, of course, when $j=0$. Thus, the next order contribution of ϕ becomes

$$\phi^{(1)}=\sum_{i=3}^{\infty}\frac{1}{h^{i-2}}\sum_{\substack{j=2i\\ j\neq0}}^{-2i}\frac{[q,q+j]_i}{j}\phi_{q+j}(\omega). \tag{27}$$

In its turn this contribution leaves uncompensated

$$\mathscr{R}_q^{(1)}=\sum_{i=3}^{\infty}\frac{1}{h^{i-2}}\sum_{\substack{j=2i\\ j\neq0}}^{-2i}\frac{[q,q+j]_i}{j}\mathscr{R}_{q+j}^{(0)} \tag{28}$$

and yields the next contribution of ϕ:

$$\phi^{(2)}=\sum_{i=3}^{\infty}\frac{1}{h^{i-2}}\sum_{\substack{j=2i\\ j\neq0}}^{-2i}\frac{[q,q+j]_i}{j}\sum_{i'=3}^{\infty}\frac{1}{h^{i'-2}}\times\sum_{\substack{j'=2i'\\ j+j'\neq0}}^{-2i'}\frac{[q+j,q+j+j']_{i'}}{j+j'}\phi_{q+j+j'}. \tag{29}$$

Proceeding in this we obtain the solution

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

which is an asymptotic expansion in descending powers of h, valid for

$$z-z_0=O(1/h), \tag{30}$$

i.e., around $z=z_0$. However, the sum of contributions $\phi^{(0)}+\phi^{(1)}\cdots$ is a solution only if the sum of the terms in ϕ_q in $\mathscr{R}_q^{(0)},\mathscr{R}_q^{(1)},\cdots$ (left unaccounted for so far) is set equal to zero. Thus,

$$0=\sum_{i=3}^{\infty}\frac{1}{h^{i-2}}[q,q]_i+\sum_{i=3}^{\infty}\frac{1}{h^{i-2}}\times\sum_{\substack{j=2i\\ j\neq0}}^{-2i}\frac{[q,q+j]_i}{j}\sum_{i'=3}^{\infty}\frac{1}{h^{i'-2}}[q+j,q]_{i'}+\cdots$$

or

$$0=\frac{1}{h}[q,q]_3+\frac{1}{h^2}\left\{[q,q]_4+\sum_{\substack{j=6\\ j\neq0}}^{-6}\frac{[q,q+j]_3}{j}[q+j,q]_3\right\}$$

$$+O\left(\frac{1}{h^3}\right). \quad (31)$$

This is the equation from which Δ and hence the eigenvalues are determined. The expansion on the right-hand side is much simpler than may appear on first sight because many terms are zero, e.g., $\widetilde{S}_3(q,0)$, $\widetilde{S}_3(q,\pm 1)$.

Thus,

$$2h\Delta = \Big[\widetilde{S}_4(q,0) - \tfrac{1}{6}\widetilde{S}_3(q,6)\widetilde{S}_3(q+6,-6) + \tfrac{1}{6}\widetilde{S}_3(q,-6)\widetilde{S}_3(q-6,6) - \tfrac{1}{2}\widetilde{S}_3(q,2)\widetilde{S}_3(q+2,-2) + \tfrac{1}{2}\widetilde{S}_3(q,-2)\widetilde{S}_3(q-2,2)\Big] + O(1/h^2)$$

$$= \left[\frac{(q^2+1)}{2^4}\cdot\frac{v^{(4)}(z_0)}{v^{(2)}(z_0)} - \frac{(15q^2+7)}{2^4\cdot 3^2}\left(\frac{v^{(3)}(z_0)}{v^{(2)}(z_0)}\right)^2\right] + O(1/h^2)$$

$$= \frac{(q^2+1)}{2^4}\left((\lambda^2+6\lambda+12) - \frac{\delta(\lambda+1)(\lambda+5)}{2\alpha\rho\lambda} + \frac{\delta^2(\lambda+1)(\lambda+2)(\lambda+5)}{(2\alpha\rho\lambda)^2} + \cdots\right) - \frac{(15q^2+7)}{2^4 3^2}\times\left((\lambda+4) - \frac{\delta(\lambda+1)}{2\alpha\rho\lambda} + \frac{\delta^2(\lambda+1)(\lambda+2)}{(2\alpha\rho\lambda)^2} + \cdots\right)^2 + O\left(\frac{1}{h^2}\right).$$

Using Eqs. (3), (15), and (16), this gives us

$$(l+\tfrac{1}{2})^2 = v(z_0) - \tfrac{1}{2}qh^2 - \Delta h$$
$$= \left(\frac{\lambda\alpha\rho^2}{(\lambda+2)} + \frac{(\lambda^2+3\lambda-1)}{\lambda(\lambda+2)}\delta\rho + \frac{(\lambda^2+5\lambda-3)\delta^2}{4(\lambda+2)\lambda^2\cdot\alpha} + \cdots\right) - \tfrac{1}{2}qh^2$$
$$- \frac{(q^2+1)}{2^5}\left((\lambda^2+6\lambda+12) - \frac{\delta(\lambda+1)(\lambda+5)}{2\alpha\rho\lambda} + \frac{\delta^2(\lambda+1)(\lambda+2)(\lambda+5)}{(2\alpha\rho\lambda)^2} + \cdots\right)$$
$$+ \frac{(15q^2+7)}{2^5\cdot 3^2}\left((\lambda+4) - \frac{\delta(\lambda+1)}{2\alpha\rho\lambda} + \frac{\delta^2(\lambda+1)(\lambda+2)}{(2\alpha\rho\lambda)^2} + \cdots\right)^2 + \frac{c_q(\delta)}{h^2} + \cdots. \quad (32)$$

In the case of the pure power potential ($\delta = 0$), one more term of this expansion is known yielding [the coefficient of the term of $O(1/h^2)$ is the contribution of Hite and Bose]

$$(l+\tfrac{1}{2})^2 = \frac{h^4}{4(2+\lambda)} - \tfrac{1}{2}qh^2 - \Delta h$$
$$= \frac{h^4}{4(2+\lambda)} - \tfrac{1}{2}qh^2 - \frac{1}{2^5\cdot 3^2}[9(q^2+1)\times(\lambda^2+6\lambda+12) - (15q^2+7)(\lambda+4)^2]$$
$$+ \frac{q}{2^8h^2}[(-320c_6 + 2240c_5c_3 + 544c_4^2 - 7200c_4c_3^2 + 5640c_3^4)q^2 + 8(-200c_6 + 760c_5c_3 + 268c_4^2 - 1836c_4c_3^2 + 1155c_3^4)) - \cdots, \quad (33a)$$

where

$$c_i = \frac{1}{\lambda(i!)}[(2+\lambda)^{i-1} - 2^{i-1}].$$

The quality of the expansion (33a) is (perhaps) best illustrated by the expansion for $\lambda = 1$ (the linear potential):

$$(l+\tfrac{1}{2})^2 = \frac{h^4}{12} - \tfrac{1}{2}qh^2 + \frac{51q^2+1}{72} + \frac{q}{256h^2}\times(-19.851862q^2 + 0.29640) + O\left(\frac{1}{h^3}\right). \quad (33b)$$

It should be noted that this expansion is invariant under the joint interchange $q \to -q$ and $h^2 \to -h^2$ which converts one solution into another.

It is clear that the approximation provided by the first four terms of this expansion is best for large values of h and small q. In Sec. 5 we will work with this approximation. Of course, more terms can be estimated without explicit calculation (we make use of this in the calculation of Regge trajectories). We observe finally that consecutive terms alternate in sign, thus indicating the Borel summability of the expansion.[13]

We have thus obtained a large-h asymptotic expansion of the eigenfunctions of the Schrödinger equation for an arbitrary power potential with a weak Coulomb-like component. The expansion is valid in the region $z = z_0$ or $\ln r = z_0$, with z_0 given by Eq. (8). A second, linearly independent solution in the same domain is obtained by changing the signs of ω and h throughout as we observe by looking at Eqs. (17) and (18). It can also be seen from these equations that a further pair of solutions is obtained by the interchanges

$$\omega \to i\omega, \quad q \to -q, \quad h \to ih$$

and

$$\omega \to -i\omega, \quad q \to -q, \quad h \to -ih.$$

The physical implications of these solutions and eigenvalues are discussed in Secs. 4 and 5.

Finally, we observe that we can define a creation operator a^+ and an annihilation operator a by

$$a^+ = -i\frac{d}{d\omega} + \frac{i}{2}\omega,$$
$$a = -i\frac{d}{d\omega} - \frac{i}{2}\omega,$$

respectively so that $[a,a^+] = 1$. The operator $\mathscr{D}_q$ can then be written

$$\mathscr{D}_q = 2a^+a - q + 1.$$

A vacuum state $|0\rangle$ or rather the ground state wave function $\langle\omega|0\rangle$ is defined accordingly by

$$a|0\rangle = 0 \quad \text{or} \quad a(\omega)\langle\omega|0\rangle = 0.$$

The perturbation method can then be carried through in terms of the operators a and a^+ as a method for calculating the quantum fluctuations around a local minimum of the potential.

3. A SECOND PAIR OF ASYMPTOTIC EIGENSOLUTIONS

We now derive a second pair of large-h asymptotic expansions for the eigenfunctions of the wave equation for our potential. This pair is valid in regions of large $|z|$ where the expansions obtained above are no longer applicable. Of course, the corresponding eigenvalue expansion will be identical with Eqs. (32) and (33) above. For reasons of simplicity we ignore the Coulomb part of the potential in this section although it is not difficult to include this as well.

Our starting point is Eq. (6) in which we insert for L^2 the expression (16) in terms of the quantity Δ which, again, is to be determined by iteration. We then have the equation

$$\frac{d^2\phi}{dz^2} + \left(v(z) - \frac{h^4}{4(2+\lambda)} + \tfrac{1}{2}qh^2 + \Delta h\right)\phi = 0. \qquad (34)$$

It is convenient to make the substitution

$$z = y - c. \qquad (35)$$

Then, choosing c such that

$$e^{-c} = \left(\frac{2\alpha}{\beta(2+\lambda)}\right)^{1/\lambda} = \rho, \qquad (36)$$

i.e., $c = -z_0$, Eq. (34) can be written

$$\frac{d^2\phi(y)}{dy^2} - \frac{h^4}{4\lambda}\left(\frac{\lambda}{\lambda+2} + \frac{2}{\lambda+2}e^{(2+\lambda)y} - e^{2y}\right)\phi(y), + (\tfrac{1}{2}qh^2 + \Delta h)\phi(y) = 0 \qquad (37)$$

Next we set[14]

$$\phi(y) = x(y)\exp\left(\pm\frac{h^2}{2\lambda^{1/2}}\int^y [v(y)]^{1/2}dy\right), \qquad (38)$$

where

$$v(y) = \frac{\lambda}{\lambda+2} + \frac{2}{\lambda+2}e^{(2+\lambda)y} - e^{2y} \qquad (39)$$

and $x(y)$ satisfies

$$\frac{d^2x}{dy^2} \pm \frac{h^2}{\lambda^{1/2}}v^{1/2}(y)\frac{dx}{dy} \pm \frac{h^2}{4\lambda^{1/2}}\cdot\frac{v'(y)}{v^{1/2}(y)}x + (\tfrac{1}{2}qh^2 + \Delta h)x = 0. \qquad (40)$$

From now on we consider only the equation for the upper signs. The equation for the lower signs leads to another solution which can be obtained from the solution we shall derive by changing the signs of h^2 and q throughout. Thus, choosing the upper signs in Eq. (40), we can rewrite the equation in the form

$$\mathscr{D}_q x = \frac{2}{h^2}\left(\frac{d^2x}{dy^2} + \Delta hx\right), \qquad (41)$$

where

$$\mathscr{D}_q \equiv -\frac{2}{\lambda^{1/2}}v^{1/2}\frac{d}{dy} - \frac{1}{2\lambda^{1/2}}\frac{v'}{v^{1/2}} - q. \qquad (42)$$

By construction, Δh is at most of $O(0)$ in h^2 for $h^2 \to \infty$. Hence, to a first approximation we can neglect the terms on the right-hand side of Eq. (41) and write for the solution to that order

$$x^{(0)} = x_q, \qquad (43)$$

where x_q is the solution of

$$\mathscr{D}_q x_q = 0, \qquad (44)$$

i.e.,

$$x_q(y) = \frac{C}{v^{1/4}}\exp\left[-\frac{q\lambda^{1/2}}{2}\int^y\frac{dy}{v^{1/2}(y)}\right], \qquad (45)$$

where C is an overall multiplicative constant which we ignore in the following except in the context of normalization.

Proceeding as in the derivation of our first solution, we evaluate d^2x_q/dy^2 and obtain

$$\frac{d^2x_q}{dy^2} + \Delta hx_q = \left(\Delta h + \frac{5}{16}\cdot\frac{v'^2}{v^2} + \frac{q\lambda^{1/2}v'}{2v^{3/2}} + \frac{q^2\lambda}{4}\cdot\frac{1}{v} - \frac{v''}{4v}\right)x_q. \qquad (46)$$

Looking at the solution (45), we observe the following relations:

$$\frac{x_{q+j}}{x_q} = \left(\frac{x_{q+1}}{x_q}\right)^j, \quad \frac{x_{q+j}}{x_q} = \frac{x_q}{x_{q-j}}. \qquad (47)$$

Further, since

$$\mathscr{D}_{q+j} = \mathscr{D}_q - j$$

and

$$\mathscr{D}_q x_{q+j} = jx_{q+j}, \qquad (48)$$

it is desirable to reexpress Eq. (46) as a sum over various x_{q+j} because then the perturbation procedure becomes particularly simple. This type of expansion is simply an expansion in terms of eigenfunctions, such as, for example, a Fourier expansion. In order to derive this expansion we have to use Eq. (45) and express y in terms of x_q. It is not difficult to convince oneself that the series reversion which this step implies is possible only if $v(y)$ is expanded around a point $y = y_0$ for which both

$$v(y_0) = 0 \quad \text{and} \quad v'(y_0) = 0.$$

Looking at Eq. (39), we see immediately that $y_0 = 0$. Then,

$$v(y) = \sum_{i=2}^{\infty}\frac{y^i}{i!}v^{(i)}(0), \qquad (49)$$

where for $i = 2,3,...$

$$v^{(i)}(0) = 2(\lambda+2)^{i-1} - 2^i. \qquad (50)$$

We then have (apart from an additive constant)

$$\frac{\lambda^{1/2}}{2}\int^y\frac{dy}{v^{1/2}(y)} = \frac{1}{2}\ln y + \sum_{i=1}^{\infty}\gamma_i y^i, \qquad (51)$$

where

$$\gamma_1 = -\frac{1}{12}(\lambda+4), \quad \gamma_2 = \frac{1}{2^4\cdot 3}(\lambda+2),$$
$$\gamma_3 = \frac{(\lambda+1)(\lambda-2)(\lambda+4)}{2^4\cdot 3^4\cdot 5},$$

etc. Expression (51) can now be substituted into the relation

$$\frac{x_{q-1}}{x_q} = \exp\left[\frac{\lambda^{1/2}}{2}\int^y\frac{dy}{v^{1/2}(y)}\right] \qquad (52)$$

and the resulting expansion in powers of y can be reversed. We then have

$$y^{1/2} = \sum_{i=0}^{\infty}d_{2i+1}\frac{x_{q-(2i+1)}}{x_q}, \qquad (53)$$

where

$$d_1 = 1, \quad d_3 = (1/12)(\lambda + 4),$$
$$d_5 = (1/2^5 \cdot 3^2)(5\lambda^2 + 34\lambda + 68),$$
$$d_7 = (1/2^7 \cdot 3^3 \cdot 5)(\lambda + 4)(79\lambda^2 + 446\lambda + 892), \text{ etc.}$$

Inserting Eq. (53) into (49) and inverting the series, we obtain

$$\frac{1}{v(y)} = \frac{1}{\lambda} \sum_{i=2,1,0}^{-\infty} \delta_{2i} \frac{x_{q+2i}}{x_q}, \tag{54}$$

where

$$\delta_4 = 1, \delta_2 = -\tfrac{2}{3}(\lambda + 4), \delta_0 = (1/2^2 \cdot 3)(\lambda^2 + 11\lambda + 22),$$
$$\delta_{-2} = -\frac{(\lambda + 4)}{2^5 \cdot 3^3 \cdot 5}(727\lambda^2 + 4718\lambda + 19\,232), \ldots.$$

In a similar way we find

$$\frac{v'^2}{v^2} = \sum_{i=2,1,0}^{-\infty} \tau_{2i} \frac{x_{q+2i}}{x_q}, \tag{55}$$

with coefficients

$$\tau_4 = 4, \tau_2 = 0, \tau_0 = (1/9)(\lambda^2 - \lambda - 2),$$
$$\tau_{-2} = -\frac{(\lambda + 4)}{2^3 \cdot 3^3 \cdot 5}(1495\lambda^2 + 9590\lambda + 38\,772), \ldots;$$

and similarly

$$\frac{v''}{v} = \sum_{i=2,1,0}^{-\infty} \epsilon_{2i} \frac{x_{q+2i}}{x_q}, \tag{56}$$

with coefficients

$$\epsilon_4 = 2, \ \epsilon_2 = \tfrac{2}{3}(\lambda + 4), \ \epsilon_0 = \tfrac{1}{6}(\lambda^2 - \lambda - 2),$$
$$\epsilon_{-2} = -\frac{(\lambda + 4)}{2^4 \cdot 3^3 \cdot 5}(667\lambda^2 + 4778\lambda + 19\,352), \cdots;$$

and

$$\frac{v'}{v^{3/2}} = \frac{1}{\lambda^{1/2}} \sum_{i=2,1,0}^{-\infty} \kappa_{2i} \frac{\chi_{q+2i}}{\chi_q}, \tag{57}$$

with coefficients

$$\kappa_4 = 2, \kappa_2 = -\tfrac{2}{3}(\lambda + 4), \kappa_0 = 0,$$
$$\kappa_{-2} = -\frac{(\lambda + 4)}{2^4 \cdot 3^3 \cdot 5}(1151\lambda^2 + 7114\lambda + 28\,922), \ldots.$$

These expansions can now be substituted in Eq. (46). Then

$$\frac{d^2 x_q}{dy^2} + \Delta h x_q = \sum_{j=2,1,0}^{-\infty} (q, q + 2j) x_{q+2j}, \tag{58}$$

where for $i \neq 0$

$$(q, q + 2i) = \frac{5}{16}\tau_{2i} + \frac{q}{2}\kappa_{2i} + \frac{q^2}{4}\delta_{2i} - \frac{1}{4}\epsilon_{2i}, \tag{59}$$

in particular

$$(q, q + 4) = \tfrac{1}{4}(q + 1)(q + 3),$$
$$(q, q + 2) = -\tfrac{1}{6}(\lambda + 4)(q + 1)^2,$$

and for $i = 0$

$$(q,q) = \Delta h + \frac{5}{16}\tau_0 + \frac{q}{2}\kappa_0 + \frac{q^2}{4}\delta_0 - \frac{1}{4}\epsilon_0$$
$$= \Delta h - \frac{1}{2^5 \cdot 3^2}[9(q^2 + 1)(\lambda^2 + 6\lambda + 12) - (15q^2 + 7)(\lambda + 4)^2]. \tag{60}$$

Thus, the first approximation $x^{(0)} = x_q$ leaves uncompensated on the right-hand side of Eq. (41) a sum of terms amounting to

$$\mathscr{R}_q^{(0)} = \frac{2}{h^2} \sum_{j=2,1,0}^{-\infty} (q, q + 2j) x_{q+2j}. \tag{61}$$

Using Eq. (48), we see that these terms can be taken care of by adding to $x^{(0)}$ the next order contribution

$$x^{(1)} = \frac{2}{h^2} \sum_{\substack{j=2,1,\ldots \\ j \neq 0}}^{-\infty} \frac{(q, q + 2j)}{2j} x_{q+2j}, \tag{62}$$

excluding, of course, the term in x_q. The coefficient of x_q in Eq. (61) set equal to zero, i.e.,

$$(q,q) = 0,$$

yields an expression for Δ (to the same order of approximation) which is identical with the expression obtained previously for our other type of eigenvalue expansion and thus verfies our previous results.

The complete solution is obtained in our standard fashion[10–12] leading to the sum

$$x = x^{(0)} + x^{(1)} + x^{(2)} + \ldots$$

in descending powers of h^2. The corresponding equation for Δ and thus the eigenvalues is

$$0 = (q,q) + \frac{2}{h^2} \sum_{\substack{j=2,1,\ldots \\ j \neq 0}}^{-\infty} \frac{(q, q + 2j)}{2j}(q + 2j, q) + \left(\frac{2}{h^2}\right)^2 \times \sum_{\substack{j=2,1,\ldots \\ j \neq 0}}^{-\infty} \sum_{\substack{j'=2,1,\ldots \\ j+j' \neq 0}}^{-\infty} \frac{(q, q + 2j)(q + 2j, q + 2j + 2j')}{2j(2j + 2j')} \times (q + 2j + 2j', q) + \cdots. \tag{63}$$

Successive contributions $x^{(0)}, x^{(1)}, \ldots$ of x form a rapidly decreasing sequence provided that

$$\frac{2}{h^2} \cdot \frac{x_{q \pm 2}}{x_q} < 1,$$

i.e.,

$$\exp\left[\mp \lambda^{1/2} \int^y \frac{dy}{v^{1/2}(y)}\right] < \tfrac{1}{2} h^2 \tag{64}$$

This relation allows arbitrarily large values of y (since $h^2 \to \infty$) but excludes the region around $y = 0$ of $z = -c$ in view of the logarithmic term in Eq. (51). Since $-c = z_0$ of Sec. 2, the latter region is precisely the region in which our previous expansion is valid.

Investigating the solution (38) in the limit $z \to -\infty$ (i.e., $r \to 0$), one can show that

$$\phi(y) \sim r^{\pm [h^2/2(\lambda + 2)^{1/2}] - [(\lambda + 2)^{1/2}/2]q + O(1/h^2)} = r^{l + 1/2}$$

on using the square root of expansion (33a). This solution can therefore be identified as the regular solution (a similar case has been discussed previously in the last paper of Ref. 12).

Finally, we observe that operators a_{q+j} defined by

$$a_{q+j} = \exp\left\{\frac{j\lambda^{1/2}}{2}\int\frac{dy}{[v(y)]^{1/2}}\right\}\cdot\mathscr{D}_q$$

obey the following relations

$$a_{q+j}x_q = 0, \quad a_{q+j}x_{q+j} = jx_q,$$
$$a_{q-j}x_{q+j} = jx_{q+2j}$$

together with

$$[\mathscr{D}_q, a_{q+j}] = -ja_{q+j}.$$

We suspect that an investigation of these relations in conjunction with the better-known operators defined in Sec. 2 will reveal a deeper insight into the structure of our perturbation solutions.

Thus, we now have two pairs of large-h asymptotic expansions of the eigenfunctions of the wave equation for a generalized power potential together with a corresponding expansion for the eigenvalues. These expansions cover (presumably) the entire range of the independent variable. We could proceed to demonstrate that the two types of eigenfunctions we have derived are proportional to each other in their common region of validity. Such a verification would proceed along the lines of Ref. 11 and 12.

4. UNIFORMLY CONVERGENT SOLUTIONS AND THE SCATTERING PROBLEM OF THE CUTOFF POTENTIAL

For the extension of our analysis to the cutoff potential which also permits scattering, it is useful and desirable to have yet another type of solution. For this reason we now derive an expansion which is uniformly convergent for finite values of h. This type of solution is similar to a corresponding solution of the wave equation for the logarithmic potential.[10,15] We therefore skip the proof of its convergence,[15] but investigate in more detail its relevance for the scattering problem.

Our starting point is Eq. (6) which we write as

$$\frac{d^2\phi}{dz^2} - L^2\phi = (\beta e^{(2+\lambda)z} - \alpha e^{2z})\phi. \quad (65)$$

For simplicity we again ignore the Coulomb part of the potential (i.e., we take $\delta = 0$). Setting

$$\phi = e^{\pm L^2 z}f(z), \quad (66)$$

we find that f satisfies the following equation,

$$\frac{d^2f}{dz^2} \pm 2L\frac{df}{dz} = \alpha e^{2z}\left(\frac{\beta}{\alpha}e^{\lambda z} - 1\right)f. \quad (67)$$

From now on we consider only the upper of these equations. The solution of the lower equation then follows by changing the sign of L. We solve Eq. (67) by iteration. Thus, if f_L is the solution of

$$\frac{d^2f_L}{dz^2} + 2L\frac{df_L}{dz} = 0,$$

we have

$$f_L = \text{const.} \quad \text{or} \quad f_L \propto e^{-2Lz}.$$

We take the first of these alternatives, because the second form leads to $\phi = e^{-Lz}f$. Ignoring an overall constant, we take $f_L = 1$, and set

$$f = 1 + \sum_{i=1}^{\infty}\alpha^i f_i, \quad (68)$$

where f_i is the solution of

$$\frac{d^2}{dz^2}f_i + 2L\frac{d}{dz}f_i = e^{2z}\left(\frac{\beta}{\alpha}e^{\lambda z} - 1\right)f_{i-1}. \quad (69)$$

Solving for f_1, we set $f_1 = e^{2z}g_1(z)$ and then

$$g_1(z) = e^{\lambda z}h_1 - \frac{1}{4(L+1)}.$$

Solving the resulting equation for h_1, we find

$$h_1 = \frac{\beta}{\alpha}\cdot\frac{1}{(\lambda+2)(\lambda+2L+2)}.$$

Proceeding in this way, we obtain the solution

$$\begin{aligned}
f = 1 &+ \left(\frac{\beta e^{(\lambda+2)z}}{(\lambda+2)(\lambda+2L+2)} - \frac{\alpha e^{2z}}{4(L+1)}\right) \\
&+ \left[\frac{\beta^2 e^{(2\lambda+4)z}}{4(\lambda+2)^2(\lambda+2L+2)(\lambda+L+2)}\right. \\
&- \frac{\alpha\beta e^{(\lambda+4)z}}{(\lambda+4)(\lambda+2L+4)}\left(\frac{1}{4(L+1)}\right. \\
&+ \left.\left.\frac{1}{(\lambda+2)(\lambda+2L+2)}\right) + \frac{\alpha^2 e^{4z}}{32(L+1)(L+2)}\right] \\
&+ \left[\frac{\beta^3 e^{(3\lambda+6)z}}{3\cdot 4(\lambda+2)^3(\lambda+L+2)(\lambda+2L+2)(2L+3\lambda+6)}\right. \\
&- \frac{\alpha\beta^2 e^{(2\lambda+6)z}}{(2\lambda+6)(2\lambda+6+2L)} \\
&\times\left(\frac{[1/4(L+1)] + [1/(\lambda+2)(\lambda+2L+2)]}{(\lambda+4)(\lambda+2L+4)}\right. \\
&+ \left.\frac{1}{4(\lambda+2)^2(\lambda+L+2)(\lambda+2L+2)}\right) \\
&+ \frac{\alpha^2\beta e^{(\lambda+6)z}}{(\lambda+2L+6)(\lambda+6)}\left(\frac{1}{32(L+1)(L+2)}\right. \\
&+ \left.\frac{1/4(L+1) + 1/(\lambda+2)(\lambda+2L+2)}{(\lambda+4)(\lambda+2L+4)}\right) \\
&- \left.\frac{\alpha^3 e^{6z}}{32\cdot 12(L+1)(L+2)(L+3)}\right] \\
&+ O(\alpha^4, \alpha^3\beta, \alpha^2\beta^2, \alpha\beta^3, \beta^4). \quad (70)
\end{aligned}$$

As pointed out earlier, a second solution is obtained by reversing the sign of L or by replacing l by $-l-1$. The solution $\phi = e^{Lz}f(z)$ is obviously the so-called regular solution.

We now look at the expansion (70) more closely and sum the terms containing leading powers of z, i.e., terms in $\alpha, \alpha^2, \alpha^3, \cdots$ or terms in $\beta, \beta^2, \beta^3, \cdots$. Summing the terms in $\alpha, \alpha^2, \alpha^3, \cdots$, we obtain

$$f \to f^{(\alpha)} = \frac{L!J_L(\alpha^{1/2}e^z)}{(\frac{1}{2}\alpha^{1/2}e^z)^L} \quad (71)$$

and summing the terms in $\beta, \beta^2, \beta^3, \cdots$, we obtain

$$f \to f^{(\beta)} = \left(\frac{2L}{\lambda+2}\right)! I_{2L/(\lambda+2)}\left(\pm\frac{2\beta^{1/2}}{\lambda+2}e^{[(\lambda+2)/2]z}\right)$$
$$\left(\pm\frac{\beta^{1/2}}{\lambda+2}e^{[(\lambda+2)/2]z}\right)^{2L/(\lambda+2)} \quad (72)$$

where J and I are Bessel functions of real and imaginary arguments, respectively. In each case an overall multiplicative constant has been ignored. The expression (71) is independent of the potential (i.e., β) and so tells us simply that in this case the solution can be expressed as a Bessel function. The expression (72), on the other hand, gives us the behavior of the wave function for $z\to\infty$ and so for $r\to\infty$. The function I_n has the following asymptotic behavior:

$$I_n(x) = \frac{e^x}{(2\pi x)^{1/2}}\left[1 + O\left(\frac{1}{x}\right)\right].$$

Considering only the dominant term of this expansion and the leading terms summed by Eq. (72), we have (inserting a constant N)

$$\psi(r)\simeq N\left(\frac{2L}{\lambda+2}\right)!\exp\left(\pm\frac{2\beta^{1/2}}{\lambda+2}r^{(\lambda+2)/2}\right)\Big/\left(\frac{\pm\beta^{1/2}}{\lambda+2}\right)^{2L/(\lambda+2)}\left(\frac{4\pi\beta^{1/2}}{\lambda+2}\right)^{1/2}r^{\lambda/4}. \tag{73}$$

The expression with the minus sign gives the asymptotic behavior of the bound state wave function. This expression is independent of the energy parameter α and thus indicates that the set of discrete states is complete, i.e., there are no scattering solutions. However, if the potential is cut off at some finite distance R_0, the asymptotic behavior of ψ is no longer given by Eq. (73) and scattering becomes possible. In fact, requiring the wave function to be continuous at R_0, the S matrix is

$$S = e^{-ikR_0}\cdot\frac{d\psi_R(R_0)/dr + ik\psi_R(R_0)}{d\psi_R(R_0)/dr - ik\psi_R(R_0)}, \tag{74}$$

where $\alpha = k^2$ and the subscript R indicates that the solution used is the regular solution discussed above. The Jost function $f_-(k) = f_-(e^{-i\pi}k)$ is given by

$$f_+(k) = e^{ikR_0}\left(\frac{d\psi_R(R_0)}{dr} - ik\psi_R(R_0)\right). \tag{75}$$

Its zeros determine the eigenvalues or—as was shown in the context of the linear power potential[16]—an expansion

$$q = q_0 + O(\exp[-R_0^{(\lambda+2)/2}]), \tag{76}$$

where q_0 is exactly an odd integer, and q only approximately. For $R_0\to\infty$ the cutoff of the potential is removed and we regain the case of only discrete states. Of course, the order of the limits $r\to\infty$, $R_0\to\infty$ is not reversible (in one case we have a scattering problem; in the other we do not).

The asymptotic form (72) suggests that solutions can be found in terms of cylindrical functions. This is in fact the case. Changing the independent variable of Eq. (65) to

$$\omega = (\beta e^{(2+\lambda)z} - \alpha e^{2z})^{1/2}/(1+\lambda/2),$$

we obtain the equation

$$\frac{d^2\phi}{d\omega^2} + \frac{1}{\omega}\left[1 - \frac{\lambda^2}{2}\cdot\frac{\alpha}{\beta}e^{-\lambda z}\Big/\left(1+\frac{\lambda}{2}-\frac{\alpha}{\beta}e^{-\lambda z}\right)^2\right]\frac{d\phi}{d\omega} - \left\{\left[L^2+\omega^2\left[1+\frac{\lambda}{2}\right]^2\right]\left(1-\frac{\alpha}{\beta}e^{-\lambda z}\right)^2\Big/\omega^2\left(1+\frac{\lambda}{2}-\frac{\alpha}{\beta}e^{-\lambda z}\right)^2\right\}\phi = 0.$$

For $z\to\infty$ this equation becomes

$$\frac{d^2\phi}{d\omega^2} + \frac{1}{\omega}\cdot\frac{d\phi}{d\omega} - \left(1 + \frac{L^2}{(1+\lambda/2)^2}\cdot\frac{1}{\omega^2}\right)\phi = 0.$$

A solution of this equation is

$$\phi^{(0)} = I_{2L/(\lambda+2)}(\omega), \tag{77}$$

which agrees with Eq. (72). The asymptotic form (77) gives us a further important hint at the solutions of the wave equation for a power potential. It demonstrates the very close similarity between the form (65) of this equation and equations of higher transcendental functions outside the circle of hypergeometric and allied functions such as the modified Mathieu equation. In fact, any solution of Eq. (65) parallels a corresponding (admittedly somewhat simpler) solution of the modified Mathieu equation,[11] and hints at the solutions of Eq. (65) can be found by consulting the literature on this equation.

5. THE POWER POTENTIAL AND PARTICLE SPECTROSCOPY

We now investigate some of the implications of our formulas for power potentials. Of course, for the linear potential, many physical implications have been investigated in great detail over the last few years—so in this case our main results are explicit series expansions and thus asymptotic formulas for quantities which have previously been evaluated only numerically. Formulas of this type are a useful aid for numerical calculations and theoretical discussion.

Setting $\lambda = 1$ in Eq. (32) we obtain

$$(l+\tfrac{1}{2})^2 = \left(\tfrac{1}{3}\alpha\rho^2 + \delta\rho + \delta^2/4\alpha + \cdots\right) - \tfrac{1}{2}qh^2$$

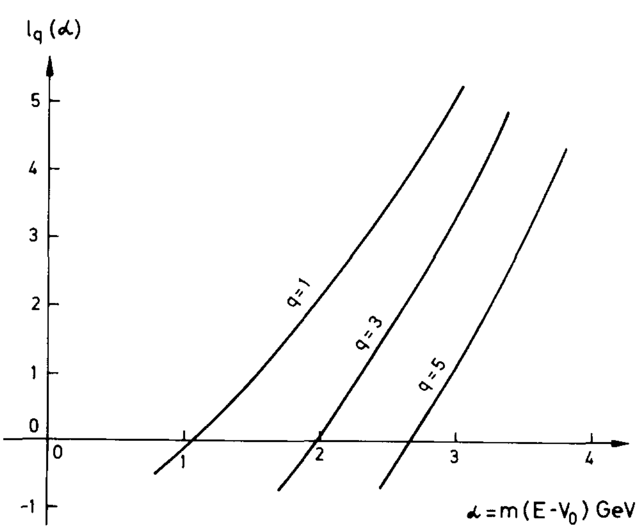


FIG. 1. The first three Regge trajectories for the linear potential and $m = 1.65$ GeV, $\beta = 0.3101$ GeV^3, and $\delta = 0$.,

$$-\frac{(q^2+1)}{2^5}\left(19-\frac{6\delta}{\alpha\rho}+\frac{9\delta^2}{\alpha^2\rho^2}+\cdots\right)$$
$$+\frac{5(15q^2+7)}{2^5\cdot 3^2}\left(5-\frac{2\delta}{\alpha\rho}+\frac{16\delta^2}{5\alpha^2\rho^2}+\cdots\right)$$
$$+\frac{c_q(\delta)}{h^2}+\cdots, \quad (78)$$

where $q=2n+1$, $n=0,1,2,\ldots$, and

$$\rho=2\alpha/3\beta \quad (\text{for } \lambda=1 \text{ only})$$

and (79)

$$h^2=2\rho\alpha^{1/2}\left(1+\frac{\delta}{\alpha\rho}-\frac{3\delta^2}{8\alpha^2\rho^2}+\cdots\right) \quad (\text{for}$$

$\lambda=1$ only).

Taking the square root of Eq. (78), we obtain the Regge trajectories $l_q\equiv\alpha_q(E)$. In Fig. 1 we show the behavior of these trajectories for $\delta=0$, $\beta=0.3101$ GeV^3, and quark masses $m=1.65$ GeV$/c^2$. The distortion of these trajectories for values of δ which are sensible in the present context is very small; so we do not show this separately. We observe that the trajectories are almost linear over the range of immediate interest. Since

$$l+\tfrac{1}{2}\sim(2/\beta)\left(\frac{\alpha}{3}\right)^{3/2}$$

we see that the slope is proportional to $1/\beta$.

Next we solve the expansion (32) for α in order to obtain E. We use the following abreviations:

$$N_q^2=(\lambda+2)(l+\tfrac{1}{2})^2-(2^4\cdot 3^2)^{-1}(3q^2-1)$$
$$\times(\lambda+2)(\lambda-2)(\lambda+1),$$
$$T_q=N_q+\tfrac{1}{2}(\lambda+2)q,$$
$$\frac{1}{B}=\lambda^{1/2}[2/(2+\lambda)\beta]^{1/\lambda},$$
$$R_q=-(2^4\cdot 3^2)^{-1}(15q^2+7)(\lambda+2)(\lambda+1)(\lambda+4)$$
$$+(2^5)^{-1}(q^2+1)(\lambda+2)(\lambda+1)(\lambda+5). \quad (80)$$

We observe that the linear potential differs from other confining potentials of power $\lambda\neq 2$ in that the q-dependent part of the dominant quantity N_q^2 is positive [this follows from the factor $(\lambda-2)$ in N_q^2]. Then

$$\alpha^{(1/2)+(1/\lambda)}/B$$
$$\simeq\frac{1}{2}(\lambda+2)q+\left(N_q^2-\frac{c_q\cdot(\lambda+2)}{2T_q}\right.$$
$$-\frac{4N_q(\lambda^2+3\lambda-1)+(\lambda+2)q(\lambda^2+3\lambda-2)}{4\lambda^{3/2}(BT_q)^{\lambda/(\lambda+2)}}\delta$$
$$\left.-\frac{R_q}{2\lambda^{1/2}T_q(BT_q)^{\lambda/(\lambda+2)}}\delta\right)^{1/2}, \quad (81)$$

where $c_q(\delta)$ is the coefficient of the term of $O(1/h^2)$ in Eq. (32). $c_q(\delta=0)$ is given in Eq. (33). In the case of the linear plus Coulomb potential this expression becomes

$$\frac{2}{3\beta}\alpha^{3/2}=\frac{3}{2}q+\left(N_q^2-\frac{3c_q}{2T_q}-\frac{3(2N_q+q)\delta}{2(BT_q)^{1/3}}-\frac{R_q}{2T_q(BT_q)^{1/3}}\delta\right)^{1/2}. \quad (82)$$

For $\lambda=1$, $l=0$ this is

$$\frac{2}{3\beta}\alpha^{3/2}=\frac{3}{2}q+\left(f(q)+\frac{\delta}{\beta^{1/3}}g(q)\right)^{1/2}, \quad (83)$$

where

$$f(q)=\frac{3q^2+17}{3\cdot 2^3}-\frac{3q(19.85186q^2+0.29640)}{2^9\{\frac{3}{2}q+[(3\cdot 2^3)^{-1}(3q^2+17)]^{1/2}\}} \quad (84a)$$

and

$$g(q)=-\frac{3q+6[(3\cdot 2^3)^{-1}(3q^2+17)]^{1/2}}{2(\frac{3}{2})^{1/3}\{\frac{3}{2}q+[(3\cdot 2^3)^{-1}(3q^2+17)]^{1/2}\}^{1/3}}+\frac{(6q^2+1)}{18^{2/3}\{\frac{3}{2}q+[(3\cdot 2^3)^{-1}(3q^2+17)]^{1/2}\}^{4/3}}. \quad (84b)$$

These expressions neglect the term of $O(\delta)/h^2$.

In Table I we give the values of $f(q)$ and $g(q)$ for $q=1,3,5,7$ corresponding to $n=1,2,3,4$. Table II gives the lowest S states obtained from Eq. (83) and the values of the

TABLE I. Particular values of quantities appearing in Eq. (96). The last column indicates the rate at which $df/dq/[f(q)]^{1/2}$ becomes independent of q.

q	$f(g)$	$-g(q)$	$-df/dq$	$-dg/dq$	$\frac{-df/dq}{[f(q)]^{1/2}}$
1	0.78441	2.44575	0.04965	0.36523	0.05606
3	1.29595	3.39112	0.18617	0.52027	0.16354
5	2.29508	4.42526	0.31420	0.50615	0.20740
7	3.79006	5.40592	0.44073	0.47439	0.22638

TABLE II. The lowest S and P states of charmonium obtained from Eq. (83) using as input $\delta/\beta^{1/3}=0.25$ GeV (see text) and $m=1.65$ GeV. β was determined by requiring the difference of the masses of the states $1\,^3S_1$ and $2\,^3S_1$ to be 0.588 GeV. Then, $\beta=0.3660$ GeV^3 and $\delta=0.1788$ GeV. The P levels do not allow a comparison with data because Eq. (83) does not take into account spin-dependent forces.

State	Mass (GeV) (calculated)	Mass (GeV) (observed)
$1\,^3S_1$	3.096	3.096
$1\,^3P$	(3.469)	
$2\,^3S_1$	3.684	3.684
$2\,^3P$	(3.952)	
$3\,^3S_1$	4.173	4.16 (?)
$4\,^3S_1$	4.606	4.6 (?)

parameters involved (δ,β,m). The values predicted for the higher S states agree with those stated by other authors (e.g., Refs. 6 and 17). The P states, of course, have to be corrected for tensor forces and spin–orbit and spin–spin coupling and are therefore not given in Table II.

Spin-dependent corrections have been considered in many previous investigations[4,5,18–20] and are generally taken over from the corresponding work on positronium. To leading order in $(v/c)^2$, where v is the relative velocity of the quark and antiquark (each of mass m), the correction to be added to the potential $V(r)=g_1 r-(g_0/r)$ is

$$V_c(r)=\frac{3}{2m^2}\cdot\frac{1}{r}\cdot\frac{dV(r)}{dr}\mathbf{L}\cdot\mathbf{S}+\frac{1}{6m^2}\nabla^2 V(r)\boldsymbol{\sigma}_1\cdot\boldsymbol{\sigma}_2-\frac{1}{12m^2}\left[\frac{d^2V(r)}{dr^2}-\frac{1}{r}\cdot\frac{dV(r)}{dr}\right]S_{12}. \quad (85)$$

Here, $\mathbf{L}$ is the orbital angular momentum operator, $\mathbf{S}=\frac{1}{2}(\boldsymbol{\sigma}_1+\boldsymbol{\sigma}_2)$, and S_{12} is the standard tensor operator, i.e.,

$$S_{12}=3(\boldsymbol{\sigma}_1\cdot\hat{r})(\boldsymbol{\sigma}_2\cdot\hat{r})-\boldsymbol{\sigma}_1\cdot\boldsymbol{\sigma}_2.$$

In the usual way we have, for $S=1$ and, respectively, $J=L-1$, L, and $L+1$,

$$\mathbf{L}\cdot\mathbf{S}=-(L+1),-1,L \quad (86)$$

Similarly we have, for $S=1$ and, respectively,[21] $L=J-1,J,J+1$,

$$S_{12}=-\frac{2(J-1)}{2J+1},2,-\frac{2(J+2)}{2J+1}. \quad (87)$$

Substituting V into V_c, we obtain terms which are singular at $r=0$ (i.e., more divergent than $1/r^2$). Since there are no acceptable bound-state solutions for such singular potentials, we have to regularize the singularity by the introduction of a cutoff parameter a into the potential. We choose this parameter by using the following replacement in the singular terms[22]:

$$\frac{1}{r}\frac{d}{dr}\left(-\frac{g_0}{r}\right)\rightarrow\frac{1}{r}\cdot\frac{g_0}{r^2+a^2}. \quad (88)$$

Substituting V into V_c, we then have

$$V_c(r)=\frac{3}{2m^2}\left[\frac{g_1}{r}+\frac{g_0}{r(r^2+a^2)}\right]\mathbf{L}\cdot\mathbf{S}+\frac{1}{3m^2}\left[\frac{g_1}{r}+2\pi g_0\delta^3(\mathbf{r})\right]\boldsymbol{\sigma}_1\cdot\boldsymbol{\sigma}_2+\frac{1}{12m^2}\left[\frac{g_1}{r}+\frac{2g_0 r}{(r^2+a^2)^2}+\frac{g_0}{r(r^2+a^2)}\right]S_{12}. \quad (89)$$

TABLE III. Effective Coulomb and confinement potential couplings after addition of spin–orbit, tensor, and spin–spin contributions applied to charmonium states $1\ ^{2S+1}P_J$ (note: the values uncorrected for these contributions are $\delta=0.1788$ GeV, $\beta=0.3660$ GeV^2). The cutoff a is taken to be 2.25 GeV$^{-1}\simeq$0.45f, and the quark mass $m=1.65$ GeV.

State	δ (GeV)	β (GeV^2)
$1\ ^3P_2$	-0.0822	0.3620
$1\ ^3P_1$	$+0.3305$	0.3702
$1\ ^3P_0$	$+0.6253$	0.3728

TABLE IV. The charmonium states $1\ ^{2S+1}P_J$ assuming $m=1.65$ GeV and $a=2.25$ GeV^{-1}.

State	Calculated (GeV)	Observed (GeV)[a]
$1\ ^3P_2$	3.530	$3.561\pm7\times10^{-3}$
$1\ ^3P_1$	3.435	$3.511\pm7\times10^{-3}$
$1\ ^3P_0$	3.337	$3.413\pm9\times10^{-3}$

[a]See Refs. 5 and 17.

The contact term in $\delta^3(\mathbf{r})$ is best treated at the very end as an additional perturbation. For simplicity we consider the expectation value of this term as a further constant contribution to E which we write as $\delta'\,\boldsymbol{\sigma}_1\cdot\boldsymbol{\sigma}_2$, $\delta'=$ const.

The values of our parameters, i.e., δ and β, were fixed such that α is determined predominantly by the leading contributions of Eq. (81) so that our expansion retains its validity as an asymptotic expansion in the sense that the first two terms provide an approximation of the quantity to be evaluated. In particular we required the contributions of δ-dependent terms to be small enough so that the square root in Eq. (83) does not become imaginary. Under these conditions the particle masses seem to be well described by our expressions.

Next we assume that on the average the separation of the quark and antiquark in the meson is such that $0\ll r/a<1$. In this case we can expand the regularized denominators in rising powers of r. Since the Coulomb coupling is in any case supposed to be small, i.e., $g_0/m^2a^4\ll g_1$, we ignore all powers of r higher than the first since these would have to be treated as additional perturbations of the linear confinement potential. Under these conditions we have to add to $V(r)$ the contribution

$$\widetilde{V}_c(r)=-\frac{d_0}{r}+d+d_1 r, \quad (90)$$

where

$$d_0=-\frac{3(\mathbf{L}\cdot\mathbf{S})}{2m^2}\left(g_1+\frac{g_0}{a_2}\right)-\frac{g_1(\boldsymbol{\sigma}_1\cdot\boldsymbol{\sigma}_2)}{3m^2}-\frac{S_{12}}{12m^2}\left(g_1+\frac{g_0}{a^2}\right),$$
$$d=\delta'(\boldsymbol{\sigma}_1\cdot\boldsymbol{\sigma}_2), \quad (91)$$
$$d_1=-\frac{3g_0(\mathbf{L}\cdot\mathbf{S})}{2m^2a^4}+\frac{g_0S_{12}}{12m^2}.$$

In Table III we give the values of the couplings of the Coulomb and confinement potentials corrected for these contributions for the well-known state $1\ ^{2S+1}P_J$ of charmonium [i.e., $\delta=m(g_0+d_0)$, $\beta=m(g_1+d_1)$]. The cutoff a is given the value of $0.45f$. We see from Table III that β changes only slightly due to the constraint that a be large, whereas δ changes substantially. In Table IV we give the mass differences between these states calculated from Eq. (84) after correction for $l\neq0$ [i.e., adding $3\,l(l+1)$ to $(3q^2+17)/24$ throughout].

In view of the simplicity (and perhaps crudity) of our arguments, the agreement with the experimentally observed differences[17] can be considered to be very good. This algebraic rather than numerical way of deriving these results has the advantage of demonstrating more convincingly the importance of the effective Coulomb part of the interaction. Moreover, looking at d_0 given in Eq. (91) and at the values of **L·S** and S_{12} given by Eqs. (86) and (87), we see that the **L·S** contribution dominates over the S_{12} contribution and thus gives rise to the ordering ${}^3P_0$, ${}^3P_1$, ${}^3P_2$ instead of ${}^3P_0$, ${}^3P_2$, ${}^3P_1$ (the latter is generally obtained for baryon–antibaryon states[21]).

Since the quark mass is twice the reduced mass μ, the mass M_q of a bound quark–antiquark pair in state q is given by

$$M_q = 4\mu + E.$$

We observe that, since

$$\alpha \simeq (BT_q)^{2\lambda/(\lambda+2)},$$

the level spacing has the following dependence on the reduced mass of the quarks

$$M_q - M_{q'} \propto \mu^{-\lambda/(\lambda+2)}. \tag{92}$$

Thus, the level spacing decreases with increasing power of the potential.

Decay widths play an important role in exploring the origin of a newly found hadronic state. The leptonic and hadronic decay widths of a vector quark–antiquark bound state such as ψ and Υ can be expressed in terms of the S-wave bound state wave function at the origin. Thus,[18]

$$\Gamma(\psi \to l\bar{l}) = \frac{16\pi\alpha^2 e_Q^2}{m_\psi^2}|\Psi(0)|^2 \tag{93}$$

and

$$\Gamma(\psi \to \text{hadrons}) = \frac{160}{81}(\pi^2 - 9)\frac{\alpha_s^3}{m_\psi^2}|\Psi(0)|^2. \tag{94}$$

Here, α is the fine structure constant [not to be confused with α of Eq. (3)], α_s the strong coupling constant ($\frac{4}{3}\alpha_s = g_0$), and e_Q is the charge of the constituent quark of ψ. We can now use the following semiclassical formula (derived by Quigg and Rosner[23]) for the value of the S-wave wave function at the origin, in order to derive an explicit expression for the leptonic widths:

$$|\Psi_n(0)|^2 = \frac{(2\mu)^{3/2}}{4\pi^2}E_n^{1/2}\frac{dE_n}{dn}$$

$$= \frac{\beta\hbar^3}{2\pi^2}\frac{d}{dq}\left(\frac{2}{3\beta}\alpha^{3/2}\right). \tag{95}$$

The differentiation involved in this formula assumes continuity in q, which is approximately true and, of course, best when q is large. Expression (95) can now be evaluated explicitly with the help of Eq. (83). For $l = 0, \lambda = 1$ we obtain

$$|\Psi_n(0)|^2 = \frac{\beta\hbar^3}{2\pi^2}\left[\frac{3}{2} + \left(\frac{df}{dq} - \frac{\delta}{\beta^{1/3}}\cdot\frac{dg}{dq}\right)\bigg/ \left(f(q) - \frac{\delta}{\beta^{1/3}}g(q)\right)^{1/2}\right]_{q=2n+1}. \tag{96}$$

In Table I we give the q-dependent quantities in this expression for $q = 1,3,5,7$. The last column of Table I shows the rate at which the δ-independent wave function approaches a constant value for increasing q. This is what one expects on the basis of Eqs. (84a) and (96). With the help of Eqs. (81) and (95) it can be seen that, for a confinement potential of power λ,

$$|\Psi_n(0)|^2 \propto n^{2(\lambda-1)/(\lambda+2)},$$

whereas for the logarithmic potential[10]

$$|\Psi_n(0)|^2 \propto 1/(2n+1)$$

and for the Coulomb potential[24]

$$|\Psi_n(0)|^2 \propto 1/n^3.$$

In Table V we give the leptonic decay widths of ψ(3096) and its radial excitations. The available experimental data are taken from Ref. 25. The agreement between calculated and observed values is reasonable and perhaps even good, if we remember that Eq. (96) is strictly valid only for large values of q. The Coulomb potential seems to produce a slight change of these decay widths. In view of the importance of a proper quantitative understanding of these widths, it seems essential to derive the exact form of $|\Psi_n(0)|^2$, i.e., as an asymptotic expansion in descending powers of h^2. The derivation of such a form requires the normalization of our wave functions. This can be done[26] but is beyond the scope of the present investigation (asymptotic expansions of this type have been derived for the normalization constants of eigensolutions of several periodic equations—see, for example, Refs. 11 and 12). Nevertheless, the numbers in Table V strongly support the idea of treating the Coulomb potential as a perturbation of the confinement potential. It would clearly be wrong to argue that the wave function at the origin ($r = 0$) is dominated by the Coulomb potential, since in that

TABLE V. Leptonic decay widths calculated from Eqs. (93) and (96); all parameters are as in Table II.

q	M_q(GeV)	$\Gamma(\psi\to l\bar{l})$ (keV) for no Coulomb modification	$\Gamma(\psi\to l\bar{l})$ (keV) with Coulomb modification	$\Gamma(\psi\to l\bar{l})$ (keV) by experiment [a]
1	3.096	3.33	3.54	4.80 ± 0.60
3	3.684	2.35	2.38	2.10 ± 0.30
5	4.16	1.84	1.79	
7	4.60	1.51	1.42	

[a] See Ref. 25.

case the decay widths of successive radial excitations would decrease much more rapidly.

6. CONCLUDING REMARKS

In the above we have demonstrated that it is by no means essential to use numerical integration techniques in order to obtain the meson spectrum and its decay widths from the nonrelativistic wave equation. Perturbation theory can be used to handle almost any problem of this type (e.g., also the baryonium spectroscopy which we have currently under investigation). The solutions that we derived above for an arbitrary confinement potential with or without a weak Coulomb component are of two types: asymptotic eigensolutions or (presumably uniformly convergent) standard series solutions.

In the case of the latter we have shown that solutions of different forms exist, e.g., expansions in terms of functions appearing in the potential or in terms of Bessel functions. In fact, comparing Eq. (6) with the modified Mathieu equation,[11,27] we see that both equations are of the same general type. The solutions of Eq. (6) therefore parallel the solutions of the modified Mathieu equation. This correspondence applies also to the asymptotic eigensolutions.[11] Thus, considerable information on all aspects of the problem defined by the wave equation for a confinement potential can be obtained from the literature on the Mathieu equation.

In the above we exploited particularly the asymptotic eigenexpansions and eigenvalues. On the basis of earlier investigations of asymptotic expansions of Mathieu functions and their eigenvalues,[11,28] we expect these expansions to be such that successive terms alternate in sign, thus indicating the Borel summability of the expansion.[28] Our application of these results for physical predictions could be improved by the computation of further terms or even by the application of Dingle's converging factors.[13] A further aspect of considerable interest is, of course, the calculation of the normalization constants (again in the form of asymptotic expansions), since these will allow a more precise determination of leptonic decay widths.[26]

The considerations of this paper were based on the assumption that the Coulomb potential is sufficiently weak to allow it to be treated as a perturbation of the confinement potential. This procedure is suggested by the slow decrease of the leptonic decay widths of successive radial excitations of ψ (3.096) and the good agreement with experimental data achieved for the level shift of the $1\,^{2S+1}P_J$ states. One might ask the following:How is it possible that the behavior of the wave function at the origin is determined predominantly by the confinement potential and not by the Coulomb potential? Our answer is that the perturbation procedure we use is carried out in the region of finite z (i.e., $r\neq 0$) for sufficiently small values of δ. $\Psi(0)$ is therefore the value obtained by finally extrapolating the wave function to the origin. An improvement of our procedure which we consider worth examining is the replacement of the Coulomb potential by the improved form

$$-\frac{g_0}{r\ln(r/r_0)}$$

as suggested by asymptotic freedom arguments.[29] The alternative method of treating the power potential as a perturbation of the Coulomb potential (this has been investigated in Ref. 30) does not apply here, since then the confinement would be lost; a relevant area of application of this case is baryonium spectroscopy.

[1]E. Eichten, K. Gottfried, T. Kinoshita, K.D. Lane, and T.-M. Yan, Phys. Rev. Lett. **34**, 369 (1975); and **36**, 500 (1976).

[2]J.F. Gunion and R.S. Willey, Phys. Rev. D **12**, 174 (1975).

[3]W. Celmaster, H. Georgi, and M. Machacek, Harvard University preprint 77/A051, 77/A060 (unpublished).

[4]R. Barbieri, R. Kögerler, Z. Kunszt, and R. Gatto, Nucl. Phys. B **105**, 125 (1976).

[5]D. Pignon and C.A. Piketty, University of Paris preprints LPTENS 77/21, 78/2 (unpublished).

[6]T. Appelquist, R.M. Barnett, and K. Lane, SLAC-PUB-2100 (1978), Ann. Rev. Nucl. Particle Sci. **28** (1979).

[7]L.M. Lederman, Proceedings of the 1977 International Symposium on Lepton and Photon Interactions at High Energies, edited by F. Gutbrod, DESY, Hamburg, 1977, p. 567.

[8]C. Quigg and J.L. Rosner, Fermilab preprint 77/82-THY (1977) (unpublished).

[9]C. Quigg and J.L. Rosner, Fermilab preprint 77/90-THY (1977) (unpublished); The arbitrary power potential has also been investigated by S.K. Bose, A. Jabs, and H.J.W. Müller-Kirsten, Phys. Rev. D **13**, 1489 (1976); and H.J.W. Müller-Kirsten, Phys. Rev. D **12**, 1103 (1975).

[10]H.J.W. Müller-Kirsten and S.K. Bose, J. Math. Phys. to be published.

[11]R.B. Dingle and H.J.W. Müller, J. Reine Angew. Math. **211**, 11 (1962); H.H. Aly, H.J.W. Müller-Kirsten, and N. Vahedi-Faridi, J. Math. Phys. **16**, 961 (1975).

[12]H.J.W. Müller, J. Reine Angew. Math. **211**, 33 (1962); **212**, 26 (1963); J. Math. Phys. **11**, 355 (1970).

[13]R.B. Dingle, *Asymptotic Expansions: Their Derivation and Interpretation* (Academic, London, 1973).

[14]For a discussion of this method of solution, see R.B. Dingle, Ref. 13, Chap. 13.

[15]The proof of uniform convergence proceeds along the lines of F. Gesztesy and L. Pittner, J. Phys. A **11**, 687 (1978).

[16]H.J.W. Müller-Kirsten and N. Vahedi, Phys. Rev. D **14**, 3079 (1976).

[17]H.F.W. Sadrozinski, in Proceedings of the 1977 International Symposium on Lepton and Photon Interactions at High Energies, edited by F. Gutbrod, Hamburg, DESY, 1967, p. 47.

[18]J.D. Jackson, Lectures on the New Particles, Lawrence Berkeley report LBL-5500 (1976), also published in the Proceedings of the 1976 Summer Institute on Particle Physics, edited by M.C. Zipf, SLAC, Stanford, California, 1976.

[19]H.J. Schnitzer, Phys. Rev. D **13**, 74 (1975).

[20]H.J. Schnitzer, Phys. Rev. Lett. **35**, 1540 (1975).

[21]See, for instance, C.B. Dover, S.H. Kahana, and T.L. Trueman, Phys. Rev. D **16**. 799 (1977); and W.W. Buck, C.B. Dover, and J.M. Richard, Brookhaven National Laboratory preprint BNL-23315 (1977) (unpublished).

[22]A similar procedure was used by J.G. Wills, D.B. Lichtenberg, and J.T. Kiehl, Phys. Rev. D **15**, 3358 (1977).

[23]C. Quigg and J.L. Rosner, Fermilab preprint 77/106-THY (1977) (unpublished).

[24]L.D. Landau and E.M. Lifshitz, *Relativistic Quantum Theory* (Pergamon, New York, 1971), p. 288.

[25]P.A. Rapidis *et al.*, Phys. Rev. Lett. **39**, 526 (1977).

[26]H.J.W. Müller-Kirsten and R.S. Kaushal, Explicit normalization of bound-state wave functions and the calculation of decay widths, preprint, University Kaiserslautern, 1978 (to be published in J. Math. Phys.).

[27]See for example, F.M. Arscott, *Periodic Differential Equations* (Pergamon, Oxford, 1964); J. Meixner and F. Schäfke, *Mathieusche Funktionen und Sphäroidfunktionen* (Springer, Berlin, 1954).

[28]R.B. Dingle and H. J.W. Müller, J. Reine Angew. Math. **216**, 123 (1964). In this paper the behavior of the coefficients of the late terms of the asymptotic expansion of the eigenvalues is derived. The same type of behavior

was later found for the coefficients of the late terms of the asymptotic expansion of the eigenvalues in the related case of the anharmonic oscilator—see, for example, C. Bender and T.T. Wu, Phys. Rev. **184**, 1231 (1969); C. Bender, J. Math. Phys. **11**, 796 (1970); J.J. Loeffel, A. Martin, B. Simon, and A.S. Wightman, Phys. Lett. B. **30**, 656 (1969).
[29]F.L. Feinberg, Phys. Rev. Lett. **39**, 316 (1977); T. Appelquist, M. Dine, and I. Muzinich, Phys. Lett. B **69**, 231 (1977); W. Fischler, CERN-TH 2321, 1977 (unpublished); E.C. Poggio, Phys. Lett. B **71**, 381 (1977); E. Eichten, K. Gottfried, T. Kinoshita, K.D. Lane, and T.-M. Yan, Cornell University preprint CLNS-375, 1978 (unpublished).
[30]H.J.W. Müller, Ann. Phys. (Leipzig) **15**, 395 (1965); **16**, 255 (1965); H.J.W. Müller and K. Schilcher, J. Math. Phys. **9**, 255 (1968); G.J. Jafrate and L.B. Mendelsohn, Phys. Rev. A **2**, 561 (1970); H. J.W. Müller-Kirsten and N. Vahedi-Faridi, J. Math. Phys. **14**, 1291 (1973); J. McEnnan, L.D. Kissel, and R.H. Pratt, University of Pittsburgh preprint PITT-148 (1975), Phys. Rev. (to be published).

Solution of a quantum mechanical eigenvalue problem with long range potentials

Frank H. Stillinger
Bell Telephone Laboratories, Murray Hill, New Jersey, 07974

(Received 13 February 1979; accepted for publication 27 April 1979)

Wavefunctions and eigenvalues for the Schrödinger equation on the half-line $x \geq 0$ are examined in the presence of a potential $v_0 x^{-2} + v_1 x^{-1} + v_2 x^{-1/2}$. With a special choice for the constant v_0 the wave equation can be solved in terms of parabolic cylinder functions. In this case the spectrum is determined by an implicit equation that arises from the boundary condition that must be imposed at $x = 0$. Depending on v_1 and v_2, the spectrum can contain and infinite number of discrete values, a finite number, or none. It is pointed out that continuous variations in v_1 or v_2 can convert negative energy bound states into positive energy responances, or *vice versa*, and the threshold behavior has been investigated.

I. INTRODUCTION

It is often enlightening to achieve exact solutions to model problems in quantum mechanics, even though the models may not directly represent physically realizable systems. The example treated in this paper largely fits that description. It entails the exact determination of wavefunctions and energy levels for a single particle subject to a central potential of the form

$$V(r) = v_0 r^{-2} + v_1 r^{-1} + v_2 r^{-1/2}. \tag{1.1}$$

The method of solution requires that v_0 have a definite value; however v_1 and v_2 are arbitrary, and the resulting flexibility in V generates an interesting richness of behavior.

The following Sec. II provides some preliminary transformations which facilitate the solution of selected quantum-mechanical problems in terms of parabolic cylinder functions. Section III shows how the spectrum specifically for potential (1.1) must be determined in principle, while Sec. IV carries that determination to essential completion. Continuum solutions are examined in Sec. V.

One of the primary reasons for interest in the model potential (1.1) is that it can produce sharp resonance behavior. If $v_1 < 0$ and $v_2 > 0$, $V(r)$ will display a wide barrier around an attractive core region. Presuming that v_1 is sufficiently negative to produce bound states, variation in v_2 can move their energies up or down, and in particular can move a bound state to zero energy (the continuum edge). This special circumstance which separates bound-state character from resonance character is studied to some extent here with emphasis on threshold behavior (Secs. IV and V). However, we intend the present work to serve as the foundation for a later, more complete analysis of these continuum edge encounters.

II. PRELIMINARY TRANSFORMATIONS

Our objective is construction of a class of solutions over $x \geqslant 0$ for the one-dimensional Schrödinger equation

$$\Phi''(x) + B(x)\Phi(x) = 0 \tag{2.1}$$

subject to suitable boundary conditions. Upon introduction of appropriate reduced units one has

$$B(x) = 2E - 2V(x), \tag{2.2}$$

where E is the total energy and $V(x)$ is the potential energy function. Equation (2.1) is also relevant to the radial motion with a central potential $V(r)$ in a space of D dimensions.[1] For that case the radial wavefunction $R(r)$ may be written

$$R(r) = r^{(1-D)/2}\Phi(r), \tag{2.3}$$

where Φ is a solution to Eq. (2.1) with

$$B(x) = 2E - 2V(x) - C(D, \Lambda)x^{-2}, \tag{2.4}$$

and

$$C(D, \Lambda) = \Lambda(\Lambda + D - 2) + \tfrac{1}{4}(D-1)(D-3). \tag{2.5}$$

In this last expression $\Lambda = 0,1,2,\cdots$ is the quantum number for angular momentum.

If Φ can be expressed in the form

$$\Phi(x) = f(x)\phi[g(x)], \tag{2.6}$$

then direct substitution shows that the differential equation (2.1) will be satisfied provided ϕ is a solution to

$$\phi'' + \left[\frac{2f'}{fg'} + \frac{g''}{(g')^2}\right]\phi' + \left[\frac{f''}{f(g')^2} + \frac{B}{(g')^2}\right]\phi = 0. \tag{2.7}$$

It will be advantageous to eliminate the ϕ' term. This will occur by requiring f to be determined by g in the manner:

$$f = C_0(g')^{-1/2}, \tag{2.8}$$

where C_0 is any nonzero constant. Thereupon the differential equation for ϕ adopts the following form:

$$\phi''(g) + \left[\frac{B}{(g')^2} - \frac{g'''}{2(g')^3} + \frac{3(g'')^2}{4(g')^4}\right]\phi(g) = 0. \tag{2.9}$$

We will now demand that the coefficient of ϕ in Eq. (2.9) be quadratic in g,

$$\phi''(g) + (ag^2 + bg + c)\phi(g) = 0, \tag{2.10}$$

where a, b, and c are constants. In other words, we demand that ϕ obey the general differential equation for parabolic cylinder functions.[2] Therefore, we will have

$$B(x) = \frac{g'''}{2g'} - \frac{3(g'')^2}{4(g')^2} + (g')^2(ag^2 + bg + c). \tag{2.11}$$

 0022-2488/79/091891-05$01.00 © 1979 American Institute of Physics

Next it is necessary to identify functions $g(x)$ which upon substitution in Eq. (2.11) will confer upon $B(x)$ the requisite form (2.4) or (2.2). In doing so we must ensure that the x independent term in $B(x)$ can sweep through all possible values for $2E$ to avoid missing any eigenvalues. Several simple examples can be listed.

(1) By choosing

$$g(x) = x, \tag{2.12}$$

one obtains

$$B(x) = ax^2 + bx + c. \tag{2.13}$$

This is the form appropriate for the one-dimensional harmonic oscillator with

$$V(x) = -\frac{a}{2}\left(x + \frac{b}{2a}\right)^2, \quad E = \frac{c}{2} - \frac{b^2}{8a}, \tag{2.14}$$

provided $a < 0$. After selecting a and b to represent appropriately the curvature and position of the quadratic potential, independent variation of c will yield any E value, including the well-known spectrum of equally spaced eigenvalues.

(2) With the choice

$$g(x) = x^{2/3}, \tag{2.15}$$

Eq. (2.11) yields

$$B(x) = (5/36x^2) + (4/9)(ax^{2/3} + b + cx^{-2/3}). \tag{2.16}$$

It is known on general grounds[3] that an x^{-2} term in $B(x)$ would prevent occurrence of any eigenstates provided its coefficient exceeded 1/4. However, that does not happen here. One could therefore proceed to determine energies and eigenfunctions for the case ($a \leqslant 0$):

$$V(x) = -\frac{5}{72x^2} - \frac{2ax^{2/3}}{9} - \frac{2c}{9x^{2/3}}, \quad E = \frac{2b}{9}. \tag{2.17}$$

(3) The case of primary interest in this paper corresponds to

$$g(x) = x^{1/2}, \tag{2.18}$$

for which

$$B(x) = (3/16x^2) + (1/4)(a + b/x^{1/2} + c/x). \tag{2.19}$$

Once again an inverse square term arises with a magnitude consistent with the existence of eigenstates. The natural separation of B is obviously the following:

$$V(x) = -\frac{3}{32x^2} - \frac{b}{8x^{1/2}} - \frac{c}{8x}, \quad E = \frac{a}{8}. \tag{2.20}$$

We shall see in detail how the signs and magnitudes of b and c determine whether the number of eigenstates is infinite, finite and positive, or zero.

III. SPECTRUM DETERMINATION

Case 3 above leads to the following generic wavefunction:

$$\Phi(x) = 2^{1/2}C_0 x^{1/4}\phi(x^{1/2}), \tag{3.1}$$

where ϕ is a solution to Eq. (2.10) over the positive real axis. In this section we shall be concerned with bound states, i.e., $a < 0$. Setting

$$g = \frac{z}{2^{1/2}|a|^{1/4}} + \frac{b}{2|a|}, \quad \phi(g) = w(z), \tag{3.2}$$

Eq. (2.10) transforms to one of the standard forms for parabolic cylinder functions[2]

$$w''(z) - (\tfrac{1}{4}z^2 + A)w(z) = 0, \tag{3.3}$$

where

$$A = -\frac{b^2}{8|a|^{3/2}} - \frac{c}{2|a|^{1/2}}. \tag{3.4}$$

The independent solutions to differential Eq. (3.3) are conventionally denoted by $U(A,z)$ and $V(A,z)$. The latter diverges as z approaches infinity; only the former is an acceptable solution. Thus,

$$\Phi(x) = 2^{1/2}C_0 x^{1/4} U\left(A, 2^{1/2}|a|^{1/4}x^{1/2} - \frac{b}{2^{1/2}|a|^{3/4}}\right) \tag{3.5}$$

encompasses all eignefunctions.

It is unacceptable on general grounds for $\Phi(x)$ to behave as $x^{1/4}$ at the origin; instead the leading order must be $x^{3/4}$ at this point.[3] Consequently $x = 0$ must be a zero of U in Eq. (3.5):

$$U\left(-\frac{b^2}{8|a|^{3/2}} - \frac{c}{2|a|^{1/2}}, -\frac{b}{2^{1/2}|a|^{3/4}}\right) = 0. \tag{3.6}$$

This condition will only be satisfied for a discrete set of negative a values which, through the second of Eqs. (2.20), determines the energy spectrum.

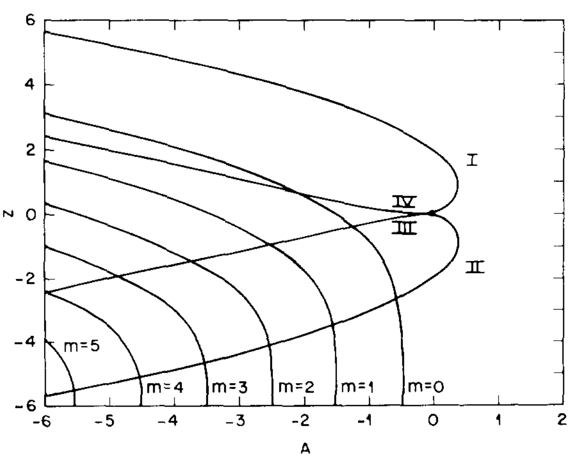


FIG. 1. Graphical determination of eigenvalues according to implicit Eq. (3.6). The curves labeled $m = 0,1,2,\cdots$ are loci of zeros for $U(A,z)$. The four curves emanating from the origin represent special cases of Eq. (4.1): I: $b = -1, c = -1$; II: $b = 1, c = -1$; III: $b = 1, c = 4$; IV: $b = -1$, $c = 4$.

IV. EIGENVALUE RESULTS

The task of deducing eigenvalues and their behavior is simplified considerably by using a graphical analysis in the real A, z plane. The roots of $U(A,z)$ correspond to curves in that plane, and their intersections with another curve determined by the form of Eq. (3.6) yield the energy spectrum. This last curve can be constructed by eliminating $|a|$ between the specific forms shown in Eq. (3.6) for the variables A and z, with the result

$$A = -\frac{z^2}{4} - \frac{c|z|^{2/3}}{2^{2/3}|b|^{2/3}}; \tag{4.1}$$

since

$$z = -\frac{b}{2^{1/2}|a|^{3/4}} \tag{4.2}$$

along this curve the sign of z is opposite to that of b.

Figure 1 shows the A, z plane with loci of zeros for $U(A,z)$, and selected examples of the curve (4.1) for each of the four sign assignments for b and c.

The following properties should be noted for the zeros of $U(A,z)$.

(1) The zeros occur along an infinite set of curves confined to the left half plane.

(2) Each curve intersects the negative A axis once and only once at a point of the form $-(2m+3/2)$, $m = 0,1,2,\cdots$. The integer m is a convenient index for the curves.

(3) All of the curves lie below the upper branch of the parabola $A = -z^2/4$.

(4) For each curve $A(z)$ is a single-valued function with unique inverse and the property

$$\lim_{z\to-\infty} A(z) = -(m+1/2). \tag{4.3}$$

We now discuss each of the four sign assignments separately.

(i) $b<0$, $c<0$. By referring to Eq. (2.20) we see that this case makes both the $x^{-1/2}$ and the x^{-1} parts of the interaction repulsive. Since the attractive x^{-2} term alone is incapable of producing bound states[3] it is obvious that addition of these extra repulsions will not change that situation. The graphical manifestation is clear in Fig. 1, for curve (4.1) lies above the upper branch of the parabola $A = -z^2/4$ and hence cannot intersect any zeros of U.

(ii) $b>0$, $c<0$. The $x^{-1/2}$ term in $V(x)$ is now negative, but the x^{-1} term remains positive. Since the former will dominate V at large x and since it has such extreme range, it is clear that an infinite number of bound states should exist. This is also clear from Fig. 1, since the corresponding position of the curve for Eq. (4.1) lies below the lower branch of $A = -z^2/4$ and intersects each U-zero branch in turn. If the magnitude of the negative quantity c is very large, Eq. (4.3) may be used to derive the following limiting distribution of eigenvalues:

$$E \approx -\frac{b^2}{32|c|} + \frac{b^3(m+1/2)}{32|c|^{5/2}}. \tag{4.4}$$

Such equally spaced levels are characteristic of harmonic oscillators, and indeed that is what this limit has generated. The interplay between attractive $x^{-1/2}$ and strongly repulsive x^{-1} potential terms gives $V(x)$ a negative broad minimum at

$$x_{\min} \approx 4c^2/b^2, \tag{4.5}$$

in the neighborhood of which quadratic behavior obtains. The low-lying states in this case are sufficiently localized around $x_{\min}$ to resemble harmonic oscillator states, and Eq. (4.4) reflects that fact.

(iii) $b>0$, $c>0$. The potential is negative everywhere and of course long-ranged. The bound states are infinite in number, since curve (4.1) must cross each U-zero curve (see Fig. 1). As b shrinks to zero, the curve for Eq. (4.1) approaches the negative A axis. In this limit $V(x)$ contains no $x^{-1/2}$ part. Remark 2 above is relevant for locating intersections, and one obtains the Rydberg series

$$E(b=0) = -\frac{c^2}{128(m+3/4)^2}. \tag{4.6}$$

For small b it should be a relatively straightforward matter to calculate successive orders of perturbation to this last result.

(iv) $b<0$, $c>0$. This is the most interesting of the four cases in that the number of bound states depends crucially upon the magnitudes of b and c. It is also the case which permits resonances to exist since $V(x)$ will have a positive barrier.

As b changes continuously from positive to negative, the preceding case (iii) transforms to the present case with Rydberg series (4.6) marking the boundary between the two. Any given eigenvalue increases continuously as b decreases, since V is being made less attractive. As we shall now see, it is possible to decrease b to a point at which any given eigenvalue increases to zero and ceases to belong to the bound-state spectrum. This last phenomenon is associated with intersections in the second quadrant of Fig. 1 that recede to infinity. In order to identify conditions that produce such behavior, it suffices to have an asymptotic expansion, valid for large $|A|$, for the U-zero curve with index m in Fig. 1[4]:

$$z_m \sim 2|A|^{1/2} - \frac{(-a_{m+1})}{|A|^{1/6}} - \frac{(-a_{m+1})^2}{20|A|^{5/6}} + O(|A|^{-3/2}),$$
$$m = 0,1,2,\cdots, \tag{4.7}$$

where a_n is the nth (negative) zero of the Airy function $\mathrm{Ai}(t)$. The first few a_n have the following values[5]:

$$\begin{aligned} a_1 &= -2.3381\ 0741,\\ a_2 &= -4.0879\ 4944,\\ a_3 &= -5.5205\ 5983,\\ a_4 &= -6.7867\ 0809. \end{aligned} \tag{4.8}$$

Equation (4.1) can be used for the intersections of interest to eliminate A from Eq. (4.7), and then z can be expressed in terms of a and b according to Eq. (3.6). Taking into account the second of Eqs. (2.20) we find (for $\Delta_m > 0$),

$$E_m = -[15|b|^{4/3}/64(-a_{m+1})^2]\Delta_m + O(\Delta_m^2), \tag{4.9}$$

where

$$\Delta_m = a_{m+1} + c/|b|^{2/3}. \tag{4.10}$$

This shows that the mth eigenvalue vanishes when Δ_m vanishes, i.e., when

$$c/|b|^{2/3} = -a_{m+1}. \quad (4.11)$$

For slightly larger values of $c/|b|^{2/3}$, E_m varies linearly with Δ_m. A more detailed analysis shows that higher order corrections in Eq. (4.9) involve only positive integer powers of the quantity Δ_m. It is easy to show that at threshold each bound state exhibits the following asymptotic exponential order:

$$\Phi(x) = \exp[-(2/3)|b|^{1/2}x^{3/4} + o(x^{3/4})]. \quad (4.12)$$

Consequently the bound-state property of square integrability is retained at threshold.

V. CONTINUUM SOLUTIONS

All positive energy states are unbound, and they form a dense continuum. The second of Eqs. (2.20) establishes that $a > 0$ for these unbound states. In place of Eq. (3.2) for bound states, we now set

$$g = \frac{z}{2^{1/2}a^{1/4}} - \frac{b}{2a},$$
$$\phi(g) = w(z). \quad (5.1)$$

Consequently w must satisfy

$$w''(z) + (\tfrac{1}{4}z^2 - A)w(z) = 0, \quad (5.2)$$

$$A = \frac{b^2}{8a^{3/2}} - \frac{c}{2a^{1/2}}. \quad (5.3)$$

The real solutions to parabolic cylinder Eq. (5.2) conventionally[2] are denoted by $W(A,z)$ and $W(A,-z)$. Consequently, aside from a normalization factor the general solution has the form

$$w(z) = \cos\theta W(A,z) + \sin\theta W(A,-z), \quad (5.4)$$

where θ must be chosen to satisfy boundary conditions.

Just as was the case with bound states, continuum wavefunctions $\Phi(x)$ must behave as $x^{3/4}$ near the origin.[3] This in turn requires that $w(z)$ vanish when $x = 0$, that is when

$$z = \frac{b}{2^{1/2}a^{3/4}}. \quad (5.5)$$

In order for this to occur θ must obey the following relation:

$$\tan\theta(a) = -\frac{W((b^2/8a^{3/2}) - (c/2a^{1/2}),(b/2^{1/2}a^{3/4}))}{W((b^2/8a^{3/2}) - (c/2a^{1/2}), -(b/2^{1/2}a^{3/4}))}. \quad (5.6)$$

This completes the determination of positive energy states, at least in principle.

These continuum solutions are useful in locating resonances, i.e., complex energy states with pure diverging current boundary conditions at $x = +\infty$:

$$\ln\Phi_r \sim i(2E)^{1/2}x. \quad (5.7)$$

Here E will have a negative imaginary part whose magnitude determines the resonance width in the usual way. Along with such a resonance there will also exist a time-reversed "antiresonance" state with pure converging-current boundary conditions:

$$\ln\Phi_{ar} \sim -i(2E^*)^{1/2}x. \quad (5.8)$$

By using the large z asymptotic forms[2] for $W(A,z)$ and $W(A,-z)$ one can show that the resonance condition (5.7) leads to the following implicit equation:

$$U\left[i\left(\frac{b^2}{8a^{3/2}} - \frac{c}{2a^{1/2}}\right), \frac{b\exp(-\pi i/4)}{2^{1/2}a^{3/4}}\right] = 0. \quad (5.9)$$

The complex resonance energies are again given by the second of Eqs. (2.20) now in terms of complex a values which satisfy Eq. (5.9). The analog of the last equation for the time-reversed states is

$$U\left[-i\left(\frac{b^2}{8a^{3/2}} - \frac{c}{2a^{1/2}}\right), \frac{b\exp(\pi i/4)}{2^{1/2}a^{3/4}}\right] = 0. \quad (5.10)$$

For any solution a of Eq. (5.9), a^* solves (5.10).

Analytic connections can be established between bound states just below threshold and the corresponding resonance and antiresonance energy pair above threshold. This is accomplished by giving the originally real parameter b an infinitesimal imaginary part as it passes a threshold value. If the imaginary part is negative, Eq. (3.6) for bound states converts automatically to Eq. (5.9) for resonances. If the imaginary part is positive, Eq. (3.6) converts to Eq. (5.10) for antiresonances. Therefore, the bound state energy in the complex b plane displays a cut at threshold, with the resonance and antiresonance pair at corresponding positions along the cut.

It is significant that Eqs. (5.9) and (5.10) lead to *identically* the same small Δ_m series that was indicated in Eq. (4.9) for bound states at threshold, but now with negative Δ_m. This series, which is evidently asymptotic rather than convergent, yields a real result above threshold whereas we know that the resonance state has an imaginary part. Clearly this imaginary part has a zero asymptotic series in positive powers of Δ_m.

VI. DISCUSSION

Although it is proper to regard the potential $V(r)$ in Eq. (1.1) as artificial, it is worth noting that an electrostatic charge density $\rho(r)$ exists which would cause a unit charge to experience just that potential. By employing Poisson's equation, one finds the appropriate density in three dimensions to be

$$\rho(r) = v_0 \lim_{\epsilon\to 0} f(\epsilon,r) + v_1\delta(r) + v_2/16\pi r^{5/2}, \quad (6.1)$$

where

$$f(\epsilon,r) = 3/2\pi\epsilon^4, \quad (0 \leqslant r < \epsilon),$$
$$= -1/2\pi r^4, \quad (\epsilon \leqslant r). \quad (6.2)$$

Similar results can easily be obtained for other values of D, the space dimension.

There may be some interest eventually in comparing exact eigenvalues for the present model with those approximate semiclassical eigenvalues that follow from quantization of the classical action.[6-8] In this regard, we mention in passing that the general radial potential containing terms of arbitrary strength proportional to r^{-2}, $r^{-3/2}$, r^{-1}, and $r^{-1/2}$ is a case for which the classical equations of motion can be integrated in closed form using Legendre elliptic integrals of the first and third kinds.[9] The present model is just a special case of this more general potential.

The available theory of the asymptotic properties of parabolic cylinder functions[2,4] is not sufficient to provide the imaginary part of our resonance energies near threshold, as discussed in Sec. V above. In principle the necessary information could be achieved through direct numerical study of Eq. (5.9), though this is likely to be a cumbersome procedure. On this account it is useful to turn to the quasiclassical WKB method. That approximation indicates that the inverse lifetime just above threshold will be dominated by a factor of the type ($K > 0$),

$$\exp(-K/|\Delta_m|^{3/2}). \tag{6.3}$$

The imaginary part of the resonance energy is proportional to the inverse lifetime, and thus exhibits the same factor. Since (6.3) has a zero asymptotic series in positive powers of Δ_m we see that the WKB approximation is consistent with the failure of imaginary terms to appear in Eq. (4.9) for $\Delta_m < 0$.

The present model may provide a convenient testing ground for the method of complex coordinate rotation[10,11] that seems to be computationally useful for locating resonances.[12,13] In particular, the one-dimensional nature of our model should permit extensive studies to be carried out on the effect of various basis set choices in complex-coordinate-rotation calculations. The extent of agreement between the computed thresholds and the behavior of the real part of the energy there, in comparison with the exact result in Eq. (4.9), provides a natural measure of numerical accuracy. At the same time it would be instructive to see how resonance lifetimes predicted by such calculations compare with the WKB result.

The WKB approximation predicts that the energy in the complex b plane possesses an essential singularity at threshold. A natural sequel to the present study would therefore involve generating the first few terms in an exact b power series for eigenvalues. Standard methods of power series analysis could then be employed to test for a singularity of the WKB type at the known thresholds.

[1]F.H., Stillinger, J. Math. Phys. **18**, 1224 (1977).
[2]M. Abramowitz and I.A. Stegun, *Handbook of Mathematical Functions*, NBS Applied Mathematics Series, No. 55 (U.S. Government Printing Office, Washington, D.C., 1968), Chapter 19.
[3]L.D. Landau and E.M. Lifshitz, *Quantum Mechanics*, translated by J.B. Sykes and J.S. Bell (Addison–Wesley, Reading, Massachussetts, 1958), pp. 118–21.
[4]F.W.J. Olver, J. Res. Nat. Bur. Stand B **63**, 131 (1959).
[5]Ref. 2, p. 478.
[6]A. Einstein, Verhandl. deut. physik. Ges. **19**, 82 (1917).
[7]M.L. Brillouin, J. Phys. Paris **7**, 353 (1926).
[8]J.B. Keller, Ann. Phys. (N.Y.) **4**, 180 (1958).
[9]E.T. Whittaker and G.N. Watson, *A Course of Modern Analysis* (Cambridge U.P., 1952), p. 515.
[10]E. Balslev and J.M. Combes, Commun. Math. Phys. **22**, 280 (1971).
[11]B. Simon, Commun. Math. Phys. **27**, 1 (1972).
[12]T.N. Rescigno and W.P. Reinhardt, Phys. Rev. A **10**, 2240 (1974).
[13]G. Doolen, J. Nuttall, and R. Stagat, Phys. Rev. A **10**, 1612 (1974).

A scale limit of φ^2 in φ^4 Euclidean field theory. I

Z. Haba[a)]

Department of Theoretical Physics, University of Bielefeld, Bielefeld, Germany

(Received 28 June 1978)

It is shown that a scale limit (appropriately defined) of the Wick square of the free Euclidean field in $d<4$ dimensions with mass m_0 exists and is a random field with values independent at every point. When $m_0 \to 0$ a stable distribution is obtained. The same limit is then calculated in φ^4 cutoff field theory. After taking the scale limit the regularization can be removed. The limit field is again independent at every point but with different density and mean different from zero. The interpretation of the results in lattice approximation is given. The problem of restoring the correlations between different points is considered as a perturbation around an independent-value field.

I. INTRODUCTION

Field theoretic models depend always on some parameters (mass, coupling constants), which describe the correlation length or the interaction strength. It is interesting whether a limit of the model exists when the parameters go to some limit values. Because the mass (and other dimensional parameters) fixes the length scale, the problem can be investigated by means of scale transformations. From the mathematical point of view the problem is to find a limit distribution under a scale transformation which is closely related to classical limit theorems.[1]

Let φ_{m_0} be the Euclidean free scalar field in d-dimensional space with mass m_0 [Gaussian random field with covariance $(-\Delta + m_0^2)^{-1}$]. We define the following scale transformations between different field theories

$$R(\lambda): \varphi_{m_0}(x) \to \varphi_{\lambda m_0}(x), \tag{1.1}$$

$$T_\gamma(\lambda): \varphi_{m_0}(x) \to \lambda^{\gamma} \varphi_{m_0}(x), \tag{1.2}$$

$$S(\lambda): \varphi_{m_0}(x) \to \varphi_{m_0}(x/\lambda). \tag{1.3}$$

There is a relation between them in the sense that the operator $U(\lambda) = T_{1-d/2}(\lambda) S(\lambda) R(\lambda)$ is a unitary operator in the Fock space $\oplus_n L_n$ (see Refs. 2 and 3). Therefore, any of the scale transformation (1.1)–(1.3) can be interpreted in terms of the remaining two.

We are interested mainly in the limit when the fields become uncorrelated, i.e., the correlation length $1/m$ goes to zero. If this limit exists then due to some general theorems[4] the characteristic function of the random field should be given by (we exclude derivatives)

$$L(f) = \exp\Big\{ -\tfrac{1}{2}a^2 \int f^2(x) d^dx + im \int f(x) d^dx + \int d^dx\, d\sigma(s) [e^{isf(x)} - 1 - isf(x)] \Big\}, \tag{1.4}$$

where $d\sigma$ is a positive measure.

Let us take the simplest example—the free field. Then

$$L_{m_0}[f] = E[\exp(i\varphi_{m_0}(f))] = \exp\Big[-\tfrac{1}{2} \int f(x) G_{m_0}(x-y) f(y) d^dx\, d^dy \Big], \tag{1.5}$$

where

$$G_{m_0}(x-y) = \frac{1}{(2\pi)^d} \int \frac{e^{ip(x-y)}}{p^2 + m_0^2} d^dp. \tag{1.6}$$

We can see that the limit of $R(\lambda)\varphi_{m_0} = \varphi_{\lambda m_0}$ when $\lambda \to \infty$ is trivial ($L[f] = 1$). However, $(L_{m_0})^{\lambda^2}$ is also a characteristic function because L_{m_0} is infinitely divisible[1] and we may consider

$$\lim_{\lambda \to \infty} L^{\lambda^2}_{\lambda m_0}[f] = \exp\Big[-\tfrac{1}{2}(1/m_0^2) \int f^2(x) d^dx \Big]. \tag{1.7}$$

So, this limit exists and is of the form (1.4).

The limit of infinite correlation length has been extensively investigated (see the probabilistic formulation in Ref. 5 and 6) due to its relevance to phase transitions. A limit of infinite bare mass and coupling constants in superrenormalizable theories is interesting as a tool for construction of new models of field theory, which from the conventional point of view can be nonrenormalizable (see Refs. 7 and 8). We are not able to do this limit in a renormalized cutoff free theory. Therefore, our physical mass goes to infinity together with the bare mass. That is the essential difference between our limit and Glimm and Jaffe infinite scale limit.[3,7]

In this paper we obtain a scale limit of $:\varphi^2:$, which is a random field with values independent at every point[4] (Sec. II). Using this result we consider the scale limit of φ^2 in φ^4 field theory (Sec. III). Then in Sec. IV an interpretation of these results in lattice approximation is considered. Finally, in Sec. V we discuss the possibility to restore the correlations between different points by a perturbation around the independent value field.

II. A SCALE LIMIT OF THE WICK SQUARE

The Wick square is, after the free field, the second simplest model of field theory. Its characteristic function can be computed explicitly as the integral is Gaussian

[a)] On leave of absence from Institute of Theoretical Physics, University of Wroclaw, Poland.

 0022-2488/79/091896-09$01.00 © 1979 American Institute of Physics

$$L[f] = \int d\mu_{G_{m_0}}(\varphi)\exp\left(\frac{i}{2}:\varphi^2:(f)\right)$$

$$= \exp(-\tfrac{1}{2}\operatorname{Tr}\{\ln[1 - if(-\Delta + m_0^2)^{-1}]$$

$$+ if(-\Delta + m_0^2)^{-1}\}). \tag{2.1}$$

The logarithm can be represented by means of an integral over a parameter[9] and we get

$$L[f] = \exp\left(\frac{1}{2}\int_0^\infty \frac{ds}{s} e^{-m_0^2 s} \times \operatorname{Tr}[e^{-s(-\Delta - if)} - (1 + isf)e^{s\Delta}]\right). \tag{2.2}$$

We will next represent the trace by means of the Feynman–Kac formula (we assume $d < 4$)

$$L[f] = \exp\left(\frac{1}{2}\int_0^\infty \frac{ds}{s} e^{-m_0^2 s}\int d^dx\, dW^s_{(x,x)}(x(\cdot)) \times\left\{\exp\left[i\int_0^s f(x(\tau))d\tau\right] - 1 - i\int_0^s f(x(\tau))d\tau\right\}\right) \tag{2.3}$$

where $dW^s_{(x,x)}(x(\cdot))$ is the Wiener measure (see Ref. 10) over closed paths, which start in x at $\tau = 0$ and end at x at $\tau = s$.

From the representation (2.3) it can be seen that the Wick square is infinitely divisible. In fact Eq. (2.3) is the Levy–Khinchine representation of an infinitely divisible random variable in infinite dimensional spaces[11] (for another argument see Ref. 9). Therefore,

$$L^{\epsilon^d}[f] = \exp\left(\frac{\epsilon^d}{2}\int_0^\infty \frac{ds}{s} e^{-m_0^2 s}\int d^dx\, dW^s_{(x,x)} \times\left\{\exp\left[i\int_0^s f(x(\tau))d\tau\right] - 1 - i\int_0^s f(x(\tau))d\tau\right\}\right)$$

$$= \int d\mu_\epsilon(\varphi_\epsilon)\exp[i/2\varphi_\epsilon(f)] \tag{2.4}$$

is also a characteristic function of a field φ_ϵ. It is not simple to obtain the measure $d\mu_\epsilon$ corresponding to its Fourier transform (2.4). We will consider this problem in Sec. IV. However, the correlation functions of φ_ϵ are different from these of $:\varphi^2:$ only by multiplicative factors. φ_ϵ fulfills all Osterwalder–Schrader axioms as $:\varphi^2:$ does. So, φ_ϵ is not "essentially" different from $:\varphi^2:$. If $1/\epsilon^d$ is an integer we can get $:\varphi^2:$ from φ_ϵ as a sum

$$:\varphi^2: = \sum_{i=1}^{1/\epsilon^d}\varphi_\epsilon^{(i)} \tag{2.5}$$

of independent random fields with the characteristic function (2.4). Let us now rescale φ_ϵ and define

$$\Phi_\epsilon(x) = \epsilon^{-d}\varphi_\epsilon(x/\epsilon) = T_{-d}(\epsilon)S(\epsilon)\varphi_\epsilon(x). \tag{2.6}$$

According to the unitary equivalence mentioned in Sec. I, Φ_ϵ is related to the Wick square with the mass m_0/ϵ. We put $f_\epsilon(x) = f(\epsilon x)$, then we let $\epsilon\to 0$ and prove

Theorem 1: Let $f(x)$ be a continuous function with compact support in R^d. Then the weak limit $\epsilon\to 0$ of $\Phi_\epsilon(f)$ exists (if $d < 4$). Namely

$$\lim_{\epsilon\to 0} E[e^{i/2\Phi_\epsilon(f)}]$$

$$= \lim_{\epsilon\to 0}\int d\mu_\epsilon(\varphi_\epsilon)e^{i/2(f_\epsilon,\varphi_\epsilon)}$$

$$= \lim_{\epsilon\to 0} L^{\epsilon^d}[f_\epsilon]$$

$$= \exp\left\{\frac{1}{2(2\pi)^{d/2}}\int_0^\infty ds\, e^{-m_0^2 s}s^{-1-d/2} \times\int d^dx[e^{isf(x)} - 1 - isf(x)]\right\}$$

$$= \int d\mu_0(\Phi_0)e^{i/2\Phi_0(f)}. \tag{2.7}$$

The limiting field Φ_0 is a random field with values independent at every point.

Proof: The Wiener measure is translation invariant. So, we can write $x(t) = x + \omega(t)$ and integrate over paths pinned up at 0 [we denote this by $dW^s_0(\omega(\cdot))$]. We will then separate the ω-independnet part in Eq. (2.4). In this way we can rewrite the formulae (2.4) and (2.7) in the form

$$L^{\epsilon^d}[f_\epsilon] = \exp\left(\frac{\epsilon^d}{2}\int\frac{ds}{s}e^{-m_0^2 s}\int dW^s_0(\omega(\cdot))\int d^dx[e^{isf(\epsilon x)} - 1 - isf(\epsilon x)]\right)\exp\left(\frac{\epsilon^d}{2}\int\frac{ds}{s}\exp(-m_0^2 s)\int dW^s_0(\omega(\cdot))\right.$$

$$\times\int d^dx\, e^{isf(\epsilon x)}\left\{\exp\left[i\int_0^s (f(\epsilon x + \epsilon\omega(\tau)) - f(\epsilon x))d\tau\right] - 1 - i\int_0^s (f(\epsilon x + \epsilon\omega(\tau)) - f(x))d\tau\right.$$

$$\left.\left. - i(e^{-isf(\epsilon x)} - 1)\int_0^s (f(\epsilon x + \epsilon\omega(\tau)) - f(x))d\tau\right\}\right)$$

$$= \exp\left(\frac{1}{2(2\pi)^{d/2}}\int\frac{ds}{s^{1+d/2}}\exp(-m_0^2 s)\int d^dx[e^{isf(x)} - 1 - isf(x)]\right)$$

$$\times\exp\Big(\frac{1}{2}\int\frac{ds}{s}\exp(-m_0^2 s)\int dW_0^s(\omega(\cdot))\int d^dx\, e^{isf(x)}\Big\{\exp\Big[i\int_0^s(f(x+\epsilon\omega(\tau))-f(x))d\tau\Big]-1$$

$$-i\int_0^s(f(x+\epsilon\omega(\tau))-f(x))d\tau-i(e^{-isf(x)}-1)\int_0^s(f(x+\epsilon\omega(\tau))-f(x))d\tau\Big\}\Big), \tag{2.8}$$

where we have taken the advantage of

$$\int dW_0^s(\omega(\cdot))=1/(2\pi s)^{d/2}. \tag{2.9}$$

In order to prove the theorem we need only to show that the term in the exponential (we denote it by M_ϵ) in the second factor vanishes as $\epsilon\rightarrow 0$. This will be the case if we are allowed to take the limit $\epsilon\rightarrow 0$ in M_ϵ under the integral sign. We will show that the integrand is dominated by an integrable function. Then the result follows from Lebesque dominated convergence theorem. We can use the inequalities

$$|e^{i\alpha}-1|\leqslant|\alpha|,\quad |e^{i\alpha}-1-i\alpha|\leqslant|\alpha|^2/2$$

to get the estimate

$$m_\epsilon=\Big|e^{isf(x)}\Big\{\exp\Big[i\int_0^s(f(x+\epsilon\omega(\tau))-f(x))d\tau\Big]-1-i\int_0^s(f(x+\epsilon\omega(\tau))-f(x))d\tau$$

$$-i(e^{-isf(x)}-1)\int_0^s(f(x+\epsilon\omega(\tau))-f(x))d\tau\Big\}\Big|$$

$$\leqslant\tfrac{1}{2}\Big(\int_0^s|f(x+\epsilon\omega(\tau))-f(x)|d\tau\Big)^2+s|f(x)|\int_0^s|f(x+\epsilon\omega(\tau))-f(x)|d\tau$$

$$\leqslant\tfrac{1}{2}s\cdot\int_0^s|f(x+\epsilon\omega(\tau))-f(x)|^2d\tau+s|f(x)|\int_0^s|f(x+\epsilon\omega(\tau))-f(x)|d\tau.$$

Clearly

$$|f(x+\epsilon\omega(\tau))-f(x)|<\sup_{y\in\Omega}|f(y)|(\chi_\Omega(x+\epsilon\omega(\tau))+\chi_\Omega(x)),$$

where $\chi_\Omega(x)$ is the characteristic function of the support Ω of f and we get

$$m_\epsilon\leqslant 2\Big(\sup_{y\in\Omega}|f(y)|\Big)^2 s\Big(s\chi_\Omega(x)+\int_0^s\chi_\Omega(x+\epsilon\omega(\tau))d\tau\Big).$$

Now, the function on the right side is integrable as

$$\int_0^\infty\frac{ds}{s}e^{-m_0^2 s}\int dW_0^s(\omega(\cdot))d^dx(\sup|f|)^2 s\Big(s\chi_\Omega(x)+\int_0^\infty\chi_\Omega(x+\epsilon\omega(\tau))d\tau\Big)$$

$$\leqslant\frac{2}{(2\pi)^{d/2}}\Big(\int ds\, e^{-m_0^2 s}s^{1-d/2}\Big)(\sup|f|)^2|\Omega|. \tag{2.10}$$

So, the limit $\epsilon\rightarrow 0$ in Eq. (2.8) can be taken pointwise. Q.E.D.

Remark: Fields with values independent at every point have been advocated some time ago by Klauder[12] as resulting from an omission of gradient terms in the functional measure. The s density $s^{-1}e^{-m_0^2 s}$ was suggested on the basis of some mass shift arguments (see also another derivation in Ref. 13). It seems that there are two ways of the omission of gradients, either to let the gradient term to tend to zero or to let it go to infinity and redefine the functional measure. Klauder's limit would correspond to the first method whereas our model as it follows from Sec. IV is a realization of the second possibility.

The question arises if the limit field Φ_0 still has anything in common with the initial Wick square $:\varphi^2:$. The common feature of Eqs. (2.7) and (2.3) is their infinite divisibility and therefore similar Levy–Khinchin representation. In fact, the limit $\epsilon\rightarrow 0$ shrinks only the normalized to 1 measure $(2\pi s)^{d/2}dW_{(x,x)}(x(\cdot))$ on closed paths passing through the point x to a δ-type measure concentrated only at this point. This leads to a similar combinatorial structure of the Green's functions in both theories and further to the same expression for the "pressure" defined by[14]

$$p(d)=\lim_{\Omega\rightarrow R^d}\frac{1}{|\Omega|}\ln\int\exp\Big[-\frac{\sigma}{2}:\varphi^2:(\chi_\Omega)\Big]d\mu_{G_{m_0}}(\varphi) \tag{2.11}$$

in φ^2 theory

and by

$$p(d) = \lim_{\Omega\to R^d} \frac{1}{|\Omega|} \ln \int \exp\left[-\frac{\sigma}{2}\Phi_0(\chi_\Omega)\right] d\mu_0(\Phi_0) \quad (2.12)$$

in Φ_0 theory (χ_Ω is the characteristic function of $\Omega \subset R^d$). Both limits can be calculated easily from Eqs. (2.3) and (2.7) corresp. and shown to be equal to

$$p(d) = \frac{1}{2(2\pi)^{d/2}} \int_0^\infty ds\, e^{-m_0^2 s} s^{-1-d/2}(e^{-s\sigma} - 1 + s\sigma). \quad (2.13)$$

This result was obtained earlier in Refs. 15 and 16 (Euclidean pressure coincides with relativistic vacuum energy) and Ref. 14 for $d = 1,2,3$. Equation (2.13) makes sense for any noninteger dimension less than four (see e.g., Ref. 17 for the relevance of noninteger dimensions). Assuming $d < 2$ the integral (2.13) can be computed (formula 3.434, Ref. 18) and then continued analytically for other d

$$p(d) = \frac{1}{(2\pi)^{d/2}} \frac{1}{d} \Gamma\left(1 - \frac{d}{2}\right) \times [m_0^d - (m_0^2 + \sigma)^{d/2} + \tfrac{1}{2}\sigma d m_0^{d-2}]. \quad (2.14)$$

It is singular for $d/2$ to be equal to a positive integer. For $d = 2$ we get a finite result through a limiting procedure $d\to 2$. A similar limiting procedure for $d\to 4$ gives the free theory result (i.e., corresp. to linear perturbation $\sigma\varphi$)

$$p(4) = A\sigma^2$$

but the constant A is now infinite.

The limit measure μ_0 (2.7) is still parameterized by the free field mass m_0. We could not perform the limit (2.7) in a massless theory because the estimate (2.10) fails. However, after the limit $\epsilon\to 0$ is taken the limit measure $d\mu_0$ (2.7) is defined also for $m_0 = 0$ (we shall denote this measure $d\mu_c$). In such a case we get the stable distributions[1] which play an exceptional role in probability as a limit of any distribution under renormalization group transformations (see also Ref. 6). Calculating the integral over s for $m_0 = 0$ in Eq. (2.7) we get for $d \neq 2$ (see Ref. 19)

$$L_c[f] = \exp\left[-\gamma_d \int d^dx |f(x)|^{d/2} \times\left(1 + i\beta \frac{f(x)}{|f(x)|} tg \frac{(\pi d)}{4}\right)\right], \quad (2.15)$$

where

$$\beta = \begin{cases} 1 & \text{for } 0 < d < 2, \\ -1 & \text{for } 2 < d < 4, \end{cases}$$

and for $d = 2$

$$L_c[f] = \exp\left[-\gamma_2 \int d^2x |f(x)| \times\left(1 - i \frac{f(x)}{|f(x)|} \frac{2}{\pi} \ln|f(x)|\right)\right], \quad (2.16)$$

where γ_d are constants which are finite for any $0 < d < 4$ (see Ref. 19 for explicit expressions for them).

The stable distributions are invariant under the following scale transformation

$$\Phi_c(x) \to \lambda^{2-d}\Phi_c(x/\lambda). \quad (2.17)$$

[There is an anomaly in two dimensions. Equation (2.16) is scale invariant if we restrict $f(x)$ to $\int f(x) d^2x = 0$ similarly as in relativistic theory (see Ref. 20).] This scale transformation coincides with that of the Wick square of a free massless Euclidean field. Let us notice finally that all the correlation functions of Φ_c are infinite. So, this massless limit cannot be obtained by taking a limit of the Green's functions with $m_0 \neq 0$.

III. PERTURBATION BY φ^4 INTERACTION

We are now going to compute the scale limit of $:\varphi^2:$ field in φ^4 field theory. [We are not interested in Green's functions of φ field itself because they vanish in the $\epsilon\to 0$ limit (see Sec. IV).] Again we replace $:\varphi^4:$ by the rescaled field Φ_ϵ (2.6) and φ^4 by Φ_ϵ^2. The free measure $d\mu_\epsilon$ (2.4) is then multiplied by the Gibbs factor $\exp[(-g_0^2/8)\int \Phi_\epsilon^2(x) d^dx]$. However, the interaction $\int \Phi_\epsilon^2 d^dx$ exists only if Φ_ϵ is properly regularized. So, let

$$\Phi_{\epsilon,\rho}(x) = \int \rho_x(y)\Phi_\epsilon(y) d^dy. \quad (3.1)$$

We are interested in computing the $\epsilon\to 0$ limit of

$$S_{\epsilon,\rho}[J] = N_{\epsilon,\rho}^{-1} \int d\mu_\epsilon(\varphi_\epsilon) \times \exp\left[-\frac{g_0^2}{8}\int \Phi_{\epsilon,\rho}^2(x) d^dx\right] \exp i(J,\Phi_\epsilon),$$

where

$$N_{\epsilon,\rho} = \int d\mu_\epsilon(\varphi_\epsilon) \exp\left[-\frac{g_0^2}{8}\int \Phi_{\epsilon,\rho}^2(x) d^dx\right] \quad (3.2)$$

with the regularization ρ fixed. Then, we will remove the regularization. Clearly, it would be much more interesting first to construct (cutoff free) φ^4 theory and then to perform the $\epsilon\to 0$ limit. This is quite a difficult problem. As indicated in Sec. I it requires a knowledge of the behavior of φ^4 when the bare mass $m_0\to\infty$. With cutoff fixed we get an independent value field because then the correlation length goes to zero. In Sec. V we will show how to recover the correlations among different points by a perturbation around the independent value model. We think that this is an alternative point of view at the $\epsilon\to 0$ limit in the φ^4 theory without cutoffs.

We can reduce the problem of calculating the limit of $S_{\epsilon,\rho}[J]$ to that in Sec. II by means of the following representation:

$$\exp\left[-\frac{g_0^2}{8}\int \Phi_{\epsilon,\rho}^2(x) d^dx\right] = \int d\mu_I(\psi) \exp\left[\frac{i}{2}(\psi,\Phi_{\epsilon,\rho})\right]$$

$$= \int d\mu_I(\psi) \exp\left[\frac{i}{2} g_0(\psi_\rho,\Phi_\epsilon)\right], \tag{3.3}$$

where $\psi_\rho(y) = \int\rho_x(y)\psi(x)d^dx$ and μ_I is the Gaussian measure with covariance I.

This integral exists with $\Phi_\epsilon(x)$ in the support of $d\mu_\epsilon$ if, e.g., $\rho_x(y)$ is a continuous function of both variables with compact support as will be further assumed.

Next, we insert Eq. (3.3) into Eq. (3.2) and apply Fubini's theorem in order to exchange Φ and ψ integrations. We get

$$S_{\epsilon,\rho}[J] = N_{\epsilon,\rho}^{-1} \int d\mu_I(\psi) \int d\mu_\epsilon(\varphi_\epsilon) \times \exp[(i/2)(\Phi_\epsilon,J + g_0\psi_\rho)]. \tag{3.4}$$

The integral over φ_ϵ has been already computed in Sec. II, Eq. (3.4). Using results of Sec. II we can prove

Theorem 2: Let $J(x), \rho_x(y)$ be continuous functions (on R and $R\times R$ corresp.) with compact supports. Then the limit $\epsilon\to 0$ of $S_{\epsilon,\rho}[J]$ exists and is equal to

$$S_{0\rho}[J] = N_{0,\rho}^{-1} \int d\mu_0(\Phi_0) \exp\left[-(g_0^2/8)\int \Phi_{0,\rho}^2(x)d^dx\right] \times \exp[(i/2)(\Phi_0,J)]. \tag{3.5}$$

Proof: If $\rho_x(y)$ is continuous with bounded support then also $\psi_\rho(x)$ has these properties for ψ belonging to the support of $d\mu_I$. We have shown already in Sec. II the existence of the $\epsilon\to 0$ limit of $\int d\mu_\epsilon(\varphi_\epsilon)\exp[(i/2)(\Phi_\epsilon,J + \psi_\rho)]$ for each ψ. In order to show the limit of $S_{\epsilon,\rho}[J]$ it is sufficient (due to the Lebesque dominated convergence theorem) to show that the absolute value of the integrand in Eq. (3.4) {equal to $L^{\epsilon^d}[J_\epsilon + g_0\psi_{\rho,\epsilon}]$, see (2.4)} is bounded. But we have

$$|L^{\epsilon^d}[J_\epsilon + g_0\psi_{\rho,\epsilon}]| = \exp\left\{\frac{1}{2}\int_0^\infty \frac{ds}{s} e^{-m_0^2 s} \times \int dW_0^s(\omega(\cdot)) \int d^dx\left[\cos\left(\int_0^s \times (J + g_0\psi_\rho)(x + \epsilon\omega(\tau))d\tau\right) - 1\right]\right\} \leqslant 1$$

as $\cos\alpha - 1 \leqslant 0$. So going with ϵ to zero we get, due to Theorem 1 of Sec. II,

$$S_{0,\rho}[J] = \int d\mu_I(\psi) \int d\mu_0(\Phi_0) \times \exp[(i/2)(J + g_0\psi_\rho,\Phi_0)].$$

We are again allowed to exchange Φ_0 and ψ integrations (Fubini theorem) getting the result (3.5). Q.E.D.

We are now able to remove the regularization ρ appearing in Eq. (3.5). The function $\rho_x(y)$ contains a volume cutoff and a smearing out over a finite space region. Let $x\to y$, then we face the problem of defining

$$\lim_{x\to y} \Phi_0(x)\Phi_0(y).$$

This problem has been solved for independent value fields by Hegerfeldt and Klauder[21] using the Wilson–Zimmermann method[22] of forming the renormalized powers. Due to Wilson and Zimmermann an expansion exists for $x\to y$

$$\Phi(x)\Phi(y) = \sum w_n(x,y)\Phi_R^{2(n)}(x) + R(x,y), \tag{3.6}$$

where w_n are functions singular at $x = y$ and $R(x,y)$ is regular when $x\to y$. $\Phi_R^{2(n)}$ can be considered as a renormalized square. For independent value fields there is only one sigular term in the expansion (3.6). The way to pick it up is given by Ref. 21.

Theorem 3: Let $\rho_x^2(y)\to c\delta(x-y)$ and let $f(x)$ be a continuous function with a compact support. Then the operator $\Phi_{0,\rho}^2(f) = \int\Phi_{0,\rho}^2(x)f(x)d^dx$ converges strongly to $\Phi_R^2(f)$.

The need for multiplicative renormalization and no subtraction are remarkable features of the construction of renormalized powers of independent value fields in comparison with φ^4 theory we have started from. This multiplicative renormalization is contained in the assumption that $\rho_x^2(y)\to c\delta(x-y)$ and not $\rho_x(y)$ itself. Here, c is a dimensional constant which has to have dimension $(\text{length})^{-d}$ if $g_0^2\Phi_R^2(f)$ is to be dimensionless. If we construct this constant from the constants m_0 and g_0 being at our disposal then the Fock space scale covariance (generated by $U(\lambda)$ Eqs. (1.1)–(1.3) and below) will be preserved. From dimensional reasons we should have $c = m_0^d u(\gamma^2)$, where u is an arbitrary function of the dimensionless coupling constant $\gamma^2 = g_0^2 m_0^{d-4}$. After the limit $\rho\to c\delta$ is taken it is possible to calculate the expression (see Ref. 21)

$$\int d\mu_0(\Phi_0)\exp\left[\frac{-g_0^2}{8}\Phi_R^2(f)\right]\exp\left[\frac{i}{2}\Phi_0(J)\right] = \exp\left(\frac{1}{2(2\pi)^{d/2}}\int ds\, e^{-m_0^2 s} s^{-1-d/2} \times \int d^dx\{\exp[isJ(x) - \tfrac{1}{2}g_0^2 m^d u s^2 f(x)] - 1 - isJ(x)\}\right). \tag{3.7}$$

Now, we set $f = \chi_\Omega$ and go with Ω to R^d. The normalization factor $N_{0,\rho}$ in Eq. (3.5) corresponds to $J = 0$ in Eq. (3.7). Dividing by $N_{0,\rho}$ we get in the limit $\Omega\to R^d$ a finite result for the characteristic function

$$S_{g_0}[J] = \exp\left\{\frac{1}{2(2\pi)^{d/2}}\int_0^\infty ds \exp(-m_0^2 s - \tfrac{1}{2}g_0^2 m_0^d u s^2) \times s^{-1-d/2}\int d^dx(e^{isJ(x)} - 1 - isJ(x))\right\}$$

$$\times \exp\left\{-\frac{i}{2(2\pi)^{d/2}}\int_0^\infty ds\, e^{-m_0^2 s} s^{-d/2}\right.$$

$$\left.\times(1-e^{-(1/2)g_0^2 m_0^d u s^2})\int d^d x J(x)\right\}$$

$$\stackrel{df}{=}\int d\mu_{g_0}(\Phi)\exp\left[\frac{i}{2}\Phi(J)\right]. \quad (3.8)$$

This is again a random field with values independent at every point but now the spectral function $d\sigma(s)$ [Eq. (1.4)] is modified by an interaction term in comparison with the free case. The Fock space scale covariance (mentioned in Sec. I) present in Eq. (3.2) is preserved in the scale limit (Klauder's independent value field[12] does not fulfill this scale covariance). The mean value of the field is different from zero and equal to

$$\langle\Phi(x)\rangle = -\frac{1}{(2\pi)^{d/2}}\int_0^\infty ds\, e^{-m_0^2 s} s^{-d/2}$$

$$\times[1-\exp(-\tfrac{1}{2}g_0^2 m_0^d u s^2)]. \quad (3.9)$$

This nonzero mean value of Φ should not be considered as surprising. The $:\varphi^2:$ field can have a negative mean value also in the constructive φ^4 field theory if a mass shift $+\sigma{:}\varphi^2{:}$ is present in the interaction (see Ref. 23). In our case $(:\varphi^2:)^2$ differs from $:\varphi^4:$ by an (infinite!) positive mass shift term.

In Sec. II the pressure in the limit theory has been calculated and shown to coincide with the pressure in the initial φ^2 field theory (mass shift model). Using Eq. (3.7) we can calculate the pressure for $\varphi^4\sim\Phi^2$ interaction with a mass shift $\sigma\Phi$

$$p_{g_0} = \lim_{\Omega\to R^d}\frac{1}{|\Omega|}\ln\int d\mu_0(\Phi_0)$$

$$\times\exp\left[-\frac{g_0^2}{8}\Phi_R^2(\chi_\Omega)-\frac{\sigma}{2}\Phi_0(\chi_\Omega)\right.$$

$$\left.+\frac{g_0^2}{8}\int d\mu_0(\Phi_0)\Phi_R^2(\chi_\Omega)\right]$$

$$=\frac{1}{2(2\pi)^{d/2}}\int ds\, e^{-m_0^2 s} s^{-1-d/2}$$

$$\times[\exp(-s\sigma-g_0^2 m^d u s^2/2)-1+\sigma s+\tfrac{1}{2}g_0^2 m^d u s^2]. \quad (3.10)$$

A comparison of this result with the initial φ^4 needs more detailed investigation and is postponed to a forthcoming paper.

IV. INTERPRETATION VIA THE LATTICE APPROXIMATION

We will explain in this section the significance of the rescaling of the Green's functions of $:\varphi^2:$ and the meaning of the measure $d\mu_\epsilon(\varphi_\epsilon)$ (2.4) in terms of the conventional Feynman–Kac integral. Let us notice that formulae (2.1) and (2.4) are just like characteristic functions of the gamma distribution[19] in an infinite dimensional space. In the finite n-dimensional case we can derive by means of an expansion into eigenvectors the formula

$$\exp[-\lambda\operatorname{Tr}\ln(A-iT)+\lambda\operatorname{Tr}\ln A]$$

$$=\int e^{(i/2)(y,Ty)}e^{(-1/2)(y,Ay)}\frac{(\det A)^\lambda}{2^{\lambda n}(\Gamma(\lambda))^n}$$

$$\times\prod_{l=1}^n\frac{dy_l}{|y_l|^{1-2\lambda}}. \quad (4.1)$$

For $\lambda=\frac{1}{2}$ there is an essential simplification in Eq. (4.1) due to the disappearance of the denominator in the integration measure. Such a denominator looks rather ugly in a functional measure. So, we will represent it in terms of an exponential by introduction of a new variable (such a representation is not unique, but is suggested by the requirement that the additional variable Θ be Gaussian.)

$$\prod_{l=1}^n\frac{1}{|y_l|^{1-2\lambda}}$$

$$=\frac{\delta^{nd/2}}{(2\pi)^{n/2}}\int\exp\left(-\tfrac{1}{2}\delta^d\sum_{l=1}^n|y_l|^{2-4\lambda}\Theta_l^2\right)\prod_{l=1}^n d\Theta_l \quad (4.2)$$

Then, setting $y_l=\varphi(l\delta)$, $\Theta_l=\Theta(l\delta)$ (where δ is the lattice spacing) and

$$(y,Ay)$$

$$=\sum\delta^{d-2}(\varphi(l\delta)-\varphi(l'\delta))^2+\sum_l m_0^2\varphi^2(l\delta)\delta^d$$

$$=\sum\delta^d\varphi(l\delta)(-\Delta+m_0^2)_{ll'}\varphi(l'\delta), \quad (4.3)$$

$$T_{ll'}=\delta^d f(l\delta)\delta_{ll'},$$

we can rewrite Eqs. (4.1) and (4.2) in the form

$$\exp[-\lambda\operatorname{Tr}\ln(-\Delta+m_0^2-if)+\lambda\operatorname{Tr}\ln(-\Delta+m_0^2)]$$

$$=\int[\det\tfrac{1}{2}(-\Delta+m_0^2)]^{\lambda-1/2}\prod_x\frac{\delta^{d/2}\pi^{1/2}}{\Gamma(\lambda)}d\Theta(x)[\det(-\Delta+m_0^2)]^{1/2}$$

$$\times\prod_x\frac{d\varphi(x)}{(2\pi)^{1/2}}\exp\left[-\tfrac{1}{2}\int\Theta^2(x)(\varphi^2(x))^{1-2\lambda}d^dx\right]$$

$$\times \exp\left[-\tfrac{1}{2} \int \varphi(x)(-\Delta + m_0^2)\varphi(x) d^d x \right] \exp\left[(i/2) \int \varphi^2(x) f(x) d^d x \right]. \tag{4.4}$$

This formula shows that passing from the field $:\varphi^2:$ to φ_ϵ in Sec. II. ($\lambda = \frac{1}{2}\epsilon^d$) is equivalent to a replacement of the original free Euclidean measure $d\mu_{G_{m_0}}$ by another one

$$d\mu_{G_{m_0}}(\varphi) \to d\mu_{G_{m_0}}(\varphi)[\det \tfrac{1}{2}(-\Delta + m_0^2)]^{-(1-\epsilon^d)/2} \prod_x \frac{\delta^{d/2}\pi^{1/2}}{\Gamma(\epsilon^d/2)} d\Theta(x) \exp\left[-\tfrac{1}{2} \int \Theta^2(x)(\varphi^2(x))^{1-\epsilon^d} d^d x \right]$$
$$= d\nu_\epsilon(\varphi,\Theta). \tag{4.5}$$

Then, the addition of interaction in Sec. III can be done in the usual way by means of the Gibbs factor

$$d\nu_\epsilon(\varphi,\Theta) \to \frac{\exp[-\frac{1}{8}g_0^2 \int \varphi^4(x) d^d x] d\nu_\epsilon(\varphi,\Theta)}{\int \exp[-\frac{1}{8}g_0^2 \int \varphi^4(x) d^d x] d\nu_\epsilon(\varphi,\Theta)}. \tag{4.6}$$

The measure 4.5 (corresponding to $\epsilon = 0$) was introduced previously by Klauder[13] ("augmented" field theory) as an alternative to the conventional quantization scheme. The modification (4.5) does not change equations of motion (because there are no gradients of the Θ field) but only the quasi-invariance property under translations $\varphi \to \varphi + h$. The field theory (4.5) or (4.6) is not canonical as the quasi-invariance of the measure is destroyed. Our derivation of the formula (4.4) shows that a precise meaning to the formal measure (4.5) can be attached through its Fourier transform (2.4).

Let us consider now the $\epsilon \to 0$ limit as formulated in Theorems 1 and 2. After rescaling (2.6)

$$\varphi^2(x) \to \epsilon^{-d} \varphi^2(x/\epsilon) \tag{4.7}$$

the measure changes as follows:

$$\exp\left[-\tfrac{1}{8}g_0^2 \int \varphi^4(x) d^d x \right] d\nu_\epsilon(\varphi,\Theta) \to \int \det[\tfrac{1}{2}(-\Delta + m_0^2)]^{-(1/2)(1-\epsilon^d)} \prod_x \frac{\delta^{d/2}\pi^{1/2}}{\Gamma(\epsilon^d/2)} d\Theta(x)$$
$$\times [\det(-\Delta + m_0^2)]^{1/2} \prod_x \frac{d\varphi(x)}{(2\pi)^{1/2}} \exp\left[-\tfrac{1}{2}(\epsilon^d)^{\epsilon^d} \int \Theta^2(x)(\varphi^2(x))^{1-\epsilon^d} d^d x \right]$$
$$\times \exp\left(-\frac{\epsilon^{-2}}{2} \int \varphi(x)(-\Delta)\varphi(x) d^d x \right) \exp\left(-\frac{m_0^2}{2} \int \varphi^2(x) d^d x \right) \exp\left(-\frac{\epsilon^{-d}}{8} g_0^2 \int \varphi^4(x) d^d x \right) \tag{4.8}$$

We can see that when $\epsilon \to 0$ the gradient term in the functional measure is not negligible at all. It *tends to infinity* and the exponential to zero. This suggests that the functional measure becomes concentrated on stepwise constant functions in the limit $\epsilon \to 0$. In fact, this can be shown starting from the characteristic function (1.4). If we add a source term $\int J(x)\varphi(x) d^d x$ generating the Green's functions of $\varphi(x)$, then after the scaling (4.7)

$$\int J(x)\varphi(x) d^d x \to \epsilon^{d/2} \int J(\epsilon x)\varphi(x) d^d x$$

and vanishes as $\epsilon \to 0$. Therefore, in Sec. II we did not take into consideration the φ field. Finally, Eq. (4.8) shows that the effective coupling constant $\epsilon^{-d} g_0^2$ goes to infinity as $\epsilon \to 0$.

In Ref. 13 Klauder deals with another limit of the measure (4.5), when he neglects the gradient terms. Klauder obtains a result different from ours (see Remark in Sec. II). We could eliminate the gradient terms in the formula (4.4) by a suitable scale transformation and reformulate the problem in a continuum theory as in Secs. II and III. However, we have gotten a divergent or trivial limit as a result.

V. RESTORING OF THE CORRELATIONS. AN EXPANSION AROUND INDEPENDENT VALUE FIELD

Fields with values independent at every point are rather unphysical. They fulfill Osterwalder–Schrader axioms[24] but have only trivial continuation to Minkowski space–time (see Ref. 25). We have shown that an independent value field can be a limit of a Euclidean field with nonzero correlation length. Now, the following problem arises. Assuming we guess a limit of a Euclidean model [we know the representation of this limit, Eq. (1.4)] when the correlation length goes to zero (physical mass to infinity). Can we get a theory with nonzero correlation length (this theory can be itself a limit of, e.g., a $P(\varphi)$ model with an infinite bare coupling)? A related problem for Klauder's model[13] has been investigated in Ref. 26 where a perturbation of the independent value measure by the interaction $\varphi(-\Delta)\varphi$ was considered. Unfortunately, only some trivial (tree) diagrams have been picked up.

We have shown in Sec. IV that our scale limit is not an omission of $\varphi(-\Delta)\varphi$ in the functional measure. Therefore, a different perturbation scheme is needed. For the φ^2 field theory (Sec. II) it is quite easy to get a perturbation expansion restoring the correlations. It can be obtained from Eq. (2.8). We have shown that the second factor in Eq. (2.8) tends to 1 as

$\epsilon \to 0$. Now, we will consider it as a perturbation. We assume that f is an analytic function

$$f(x+\epsilon\omega(\tau)) - f(x) = \sum_{i_1+\cdots i_d=k} \frac{\epsilon^k}{i_1!\cdots i_d!} f^{(k)}_{i_1\cdots i_d}(x)\omega_1^{i_1}(\tau)\cdots\omega_d^{i_d}(\tau), \tag{5.1}$$

where

$$f^{(k)}_{i_1\cdots i_d}(x) = \frac{\partial^k f}{\partial x_1^{i_1}\cdots \partial x_d^{i_d}}.$$

Then, we can expand the second factor in Eq. (2.8) as a power series in ϵ. Further we use Eq. (2.7) in order to represent the first factor as Fourier transform of the measure μ_0. A typical term in the expansion has the form

$$\prod_l f^{(k')}_{i_1\cdots i_d}(x_l)\exp[(i/2)(\Phi_0, f)] = \prod_l \left((-1)^{k'}(-i)\partial^{(k')}_{i_1\cdots i_d} 2\frac{\delta}{\delta\Phi_0(x_l)}\right)\exp[(i/2)\Phi_0(f)]. \tag{5.2}$$

Using Eq. (5.2) we can sum up the series again and write (2.8) in a compact form as a perturbation of the field with independent values at every point

$$\begin{aligned} L^{\epsilon''}[f_\epsilon] &= \int d\mu_0(\Phi_0)\exp\left(\frac{1}{2}\int_0^\infty \frac{ds}{s} e^{-m_0^2 s}\int dW_0^s(\omega(\cdot))\int d^dx\left\{\exp\left[2\int_0^s d\tau\, e^{-\epsilon\omega(\tau)\cdot\nabla}\frac{\delta}{\delta\Phi_0(x)}\right]\right.\right.\\ &\quad \left.\left. - e^{2[\delta/\delta\Phi_0(x)]} - 2i\int_0^s d\tau (e^{-\epsilon\omega(\tau)\nabla}-1)\frac{\delta}{\delta\Phi_0(x)}\right\}\right)\exp\left[\frac{i}{2}\Phi_0(f)\right]. \end{aligned} \tag{5.3}$$

In deriving Eq. (5.3) we replaced $f(x)$ in Eq. (2.8) by $-2i[\delta/\delta\Phi_0(x)]$ as follows from Eq. (5.2) and used the equation $e^{a\nabla}\times f(x) = f(x+a)$. Eqs. (2.8) and (5.3) are just identities but our aim was to show that an expansion around the independent value field makes sense and can completely restore the correlations.

The analogous problem with interaction is already less trivial. Due to Eq. (3.4) we can repeat all the steps leading to the formula (5.3) in order to get for the regularized generating functional

$$\begin{aligned} S_{\epsilon,\rho}[J] &= N^{-1}_{\epsilon,\rho}\int d\mu_I(\psi)d\mu_0(\Phi_0)\exp\left(\frac{1}{2}\int_0^\infty \frac{ds}{s} e^{-m_0^2 s}\int dW_0^s(\omega(\cdot))d^dx\left\{\exp\left[2\int_0^s d\tau\, e^{-\epsilon\omega(\tau)\cdot\nabla}\frac{\delta}{\delta\Phi_0(x)}\right]\right.\right.\\ &\quad \left.\left. - e^{2[\delta/\delta\Phi_0(x)]} - 2\int_0^s d\tau(e^{-\epsilon\omega(\tau)\nabla}-1)\frac{\delta}{\delta\Phi_0(x)}\right\}\right)\exp\left[\frac{i}{2}\Phi_0(J+g_0\psi_\rho)\right]. \end{aligned} \tag{5.4}$$

Now, we can integrate over ψ getting on the right-hand side the factor

$$\int d\mu_I(\psi)\exp[(i/2)\Phi_0(J+g_0\psi_\rho)] = \exp[(i/2)\Phi_0(J)]\exp\left[-\frac{g_0^2}{8}\int \Phi^2_{0,\rho}(x)d^dx\right] \tag{5.5}$$

which then should be differentiated in order to get $S_{\epsilon,\rho}[J]$. In the limit $\epsilon\to 0$ there is no differentiation and in such a case we were able to remove the regularization (Sec. III) by letting $\rho_x^2(y)\to\delta(x-y)$.

Such a simple solution is not possible for the series (5.4). Differentiation of the function (5.5) over Φ_0 generates power series in $\Phi_0(x)$ times the factor (5.5). We can again include $\exp[-(g_0^2/8)\int\Phi^2(x)d^dx]$ in the definition of the interaction measure as in (3.7)–(3.8). But now the limiting procedure $\rho_x^2(y)\to\delta(x-y)$ is not allowed, because with such a choice of $\rho_x(y)$

$$\frac{\partial}{\partial\Phi_0(x_l)}\exp\left(-\frac{g_0^2}{8}\int\Phi^2_{0,\rho}(x)d^dx\right)\to 0.$$

We have to let $\rho_x(y)\to\delta(x-y)$ if Φ_0 in the power series is not to disappear but then $g_0^2\Phi^2_{0,\rho}(x)\to g_0^2\delta(0)\Phi^2_R(x)$. So the coupling constant renormalization is needed. Then, after differentiation we get power series in $\Phi''(x)$, which should be renormalized according to the prescriptions given in Refs. 21 and 22. This leads in addition to the coupling constant renormalization and also to the wavefunction renormalization. After this is done, we get a renormalized power series in the field Φ [defined by Eq. (3.8) with g_0 replaced by renormalized coupling constant g]. Then, the integration over $d\mu_g$ (3.8) gives finite results in each order of perturbation expansion in ϵ. However, it is rather difficult to investigate the properties of the resulting theory due to the complexity of Eqs. (5.4)–(5.5). Nevertheless, it can be seen that such a theory does not coincide with the initial φ^4 field theory (3.4) we have started from. This follows from the quite different renormalization procedures applied to the conventional expansion in g_0 and the ϵ expansion around the independent value field. [If we performed the ψ integration in Eq. (3.4) and then renormalized the resulting φ_ϵ theory without taking the limit $\epsilon\to 0$ we would get the conventional φ^4 field theory with some inessential rescalings.]

VI. CONCLUDING REMARKS

It is expected that when correlation length goes to zero the values of a random field become independent at every point. The question is whether this limit field with values independent at every point is nontrivial (i.e., different from zero). If it is to be a nontrivial limit it should be carefully defined with an appropriate normalization chosen. So far much more attention was concentrated on the infinite correlation limit due to its relevance to phase transitions. It seems that both infinite correlation limit and zero correlation limit are analogs of classical limit theorems of probability (see Refs. 5 and 6).

We have shown that the scale limit of φ^2 exists and we have computed the corresponding limit in regularized φ^4 field theory. We think that such a limit can occur also in φ^4 without cutoff. Due to Glimm and Jaffe[3,7] there exists a scale limit of $(\varphi^4)_2$ with nonzero correlations indexed by the physical mass. Now, when the physical mass m in this model goes to infinity, the correlation length tends to zero. This is just an interchange of the limits $m_0 \to \infty$ in cutoff theory and $m \to \infty$ in the scale limit of φ^4 without cutoff. The field theories defined as limits of superrenormalizable models, when some parameters tend to infinity, are interesting as examples of more singular theories[7,8] than the well-understood superrenormalizable models. For a construction of such singular theories the independent-value field can be a good starting point. In Sec. V we have suggested an expansion around the independent value field. We have pointed out that the perturbation series can be renormalized in each order of the expansion parameter. However, a sum of an infinite number of terms is needed in order to draw conclusions about the correlations. Without summing the series we can calculate some expressions which do not depend on correlations, e.g., the pressure, or depend on them in an irrelevant way. In the forthcoming paper it will be shown that the scale limit may be considered as a low (Euclidean) momentum approximation of a large coupling limit of superrenormalizable theories and a small coupling limit for nonrenormalizable interactions.

ACKNOWLEDGMENTS

The author wishes to thank Professor L. Streit for hospitality at the Institute for Theoretical Physics of Bielefeld University and Gesellschaft Für Kernforschung for the financial support. He would like to thank Professor F. Jegerlehner for his reading of the manuscript and useful remarks and Professor G. Jona-Lasinio for an interesting discussion.

[1] I.A. Ibragimov and Yu. V. Linnik, *Independent and Stationary Sequences of Random Variables* (Wolters–Noordhoff, Armsterdam, The Netherlands, 1971).
[2] J. Glimm and A. Jaffe, Ann. Inst. Henri Poincaré **21**, 27 (1974).
[3] J. Glimm and A. Jaffe, "Critical Exponents and Renormalization in the φ^4 Scaling Limit," in *Quantum Dynamics: Models and Mathematics*, edited by L. Streit (Springer-Verlag, Berlin, 1976).
[4] I.M. Gel'fand and N.Y. Vilenkin, *Generalized Functions* (New York, 1964), Vol. 4.
[5] R.L. Dobrushin, Avtomodelnost i Renorm-grupa Obobshchenikh Sluchaynikh Poley (to be published).
[6] G. Joan-Lasinio, Nuovo Cimento B **26**, 99 (1975).
[7] J. Glimm and A. Jaffe, Phys. Rev. D **11**, 2816 (1975).
[8] G. Parisi, "On Non-Renormalizable Interactions," in *New Developments in QFT and Stat. Mechanics*, edited by M. Lévy and P. Mitter (Plenum, New York, 1977).
[9] K. Baumann and G. Hegerfeldt, J. Math. Phys. **18**, 171 (1977).
[10] E. Nelson, J. Math. Phys. **5**, 332 (1964).
[11] I.I. Gikhman and A.V. Skorohod, *The Theory of Stochastic Processes* (Nauka, Moscow, 1971) (in Russian), Vol. 1.
[12] J.R. Klauder, Acta Physica Austr. **41**, 237 (1975).
[13] J.R. Klauder, Phys. Rev. D **14**, 1952 (1976).
[14] F. Guerra, L. Rosen, and B. Simon, Ann. Inst. Henri Poincaré **25**, 16 (1976).
[15] L. Rosen, J. Math. Phys. **13**, 918 (1972).
[16] L. Pitner and L. Streit, Acta Phys. Austr. **38**, 361 (1973).
[17] K.G. Wilson and J.B. Kogut, Phys. Rep. **126**, 75 (1974).
[18] I.S. Gradsteyn and I.M. Rysik, *Table of Integrals, Series and Products* (Academic, New York, 1965).
[19] E. Lukacs, *Characteristic Functions* (Griffin, London, 1970).
[20] Z. Haba, Rep. Math. Phys. **9**, 187 (1976).
[21] G.C. Hegerfeldt and J.R. Klauder, Nuovo Cimento A **19**, 153 (1974).
[22] K. Wilson and W. Zimmermann, Commun. Math. Phys. **24**, 87 (1972).
[23] A. Klein and L.J. Landau, J. Math. Phys. **17**, 87 (1976).
[24] K. Osterwalder and R. Schrader, Commun. Math. Phys. **31**, 83 (1973).
[25] G.C. Hegerfeldt, Commun. Math. Phys. **35**, 155 (1974).
[26] S. Kövesi-Domokos, Nuovo Cimento A **33**, 769 (1976).

Comments on certain divergence-free tensor densities in a 3-space

S. J. Aldersley

Department of Applied Mathematics University of Waterloo, Waterloo, Ontario, Canada N2L 3G1 and National Auditing Department, Clarkson, Gordon & CO., Royal Trust Tower, P.O. Box 251, Toronto—Dominion Centre, Toronto, Canada, M5K 1J7

(Received 28 December 1977; accepted for publication 10 April 1979)

It is well known that a necessary and sufficient condition for the conformal flatness of a three-dimensional pseudo-Riemannian manifold can be expressed in terms of the vanishing of a third-order tensor density concomitant of the metric which has contravariant valence 2. This was first discovered by Cotton in 1899. It is shown that Cotton's tensor density is not the Euler–Lagrange expression corresponding to a scalar density built from one metric tensor. This tensor density is shown to be uniquely characterized by its conformal properties coupled with the demand that it be differentiable for arbitrary metrics.

Lovelock[1] has shown that in a 3-space, the only third order (2,0)-tensor density concomitants of a metric which are symmetric and divergence free are[2] $g^{1/2}g^{ij}$, $g^{1/2}G^{ij}$, and C^{ij} where

$$C^{ij}:=\epsilon^{iab}(R^j_a-\tfrac{1}{4}\delta^j_aR)_{|b}. \tag{1}$$

The tensor density C^{ij} was first introduced by Cotton[3] in connection with the conformal properties of three-dimensional Riemannian manifolds. He showed that the vanishing of C^{ij} is a necessary and sufficient condition for the conformal flatness of such manifolds. We shall call C^{ij} the Cotton tensor density.[4]

It is well known that

$$\sqrt{g}g^{ij}=E^{ij}(2\sqrt{g}),$$

and that

$$\sqrt{g}G^{ij}=E^{ij}(-\sqrt{g}R),$$

where E^{ij} is the Euler–Lagrange operator.[5] The question naturally arises as to whether there exists a scalar density $\mathscr{L}$ of the form

$$\mathscr{L}=\mathscr{L}(g_{ab};g_{ab,c};\dots;g_{ab,c_1\cdots c_n}) \tag{2}$$

for which

$$C^{ij}=E^{ij}(\mathscr{L}). \tag{3}$$

The obvious choice of $ag_{ij}C^{ij}$ (for some constant a) fails since $g_{ij}C^{ij}\equiv 0$ (cf. Ref. 4).

The function $\mathscr{L}$ in (2) can be treated as a differentiable function of the variables $g_{ab}, g_{ab,c},\cdots$. The derivatives d/dx^i in the operator E^{ij} become the operators

$$\frac{d}{dx^i}=g_{ab,i}\frac{\partial}{\partial g_{ab}}+g_{ab,ci}\frac{\partial}{\partial g_{ab,c}}+\cdots.$$

Equation (3) can then be interpreted as a partial differential equation for the function $\mathscr{L}$. Horndeski[6] has shown that the Cotton tensor density satisfies

$$E^{ij}(C^{ab})-\frac{\partial C^{ij}}{\partial g_{ab}}=0,$$

$$3\frac{d^2}{dx^rdx^s}\frac{\partial C^{ab}}{\partial g_{ij,krs}}-2\frac{d}{dx^r}\frac{\partial C^{ab}}{\partial g_{ij,kr}}+\frac{\partial C^{ab}}{\partial g_{ij,k}}+\frac{\partial C^{ij}}{\partial g_{ab,k}}=0,$$

$$3\frac{d}{dx^r}\frac{\partial C^{ab}}{\partial g_{ij,klr}}-\frac{\partial C^{ab}}{\partial g_{ij,kl}}+\frac{\partial C^{ij}}{\partial g_{ab,kl}}=0, \tag{4}$$

and

$$\frac{\partial C^{ab}}{\partial g_{ij,klm}}+\frac{\partial C^{ij}}{\partial g_{ab,klm}}=0.$$

Equations (4) are necessary conditions for the existence of a function $\mathscr{L}$ of the form (2) which satisfies (3).[6,7] They can be interpreted as "integrability conditions" for the partial differential equations (3). (Anderson[7] refers to them in this way.) In fact, Eq. (4) are sufficient for the existence of a scalar density $\mathscr{V}$ of the form

$$\mathscr{V}=\mathscr{V}(g_{ab};\dots;g_{ab,cde};h_{ab};\dots,h_{ab,cde}) \tag{5}$$

for which $C^{ij}=E^{ij}(\mathscr{V})$, where both g_{ab} and h_{ab} are positive or negative definite metric tensors.[6] Thus in the case of definite metrics, scalar density solutions to (3) can be obtained by enlarging the domain of definition assumed for $\mathscr{L}$. In this paper it is shown that no scalar density solutions to (3) of the form (2) exist (for metrics of arbitrary signature) provided certain differentiability requirements are imposed.[8]

Theorem: In a 3-space there does not exist a class $C^{\alpha+4}$ *scalar density* $\mathscr{L}$ *of the form* (2) *satisfying* (3).

Proof: For any tensorial concomitant of a metric tensor with the form

$$F:::=F:::(g_{ab};g_{ab,c};\dots;g_{ab,c_1\cdots c_\alpha}),$$

then the functions $F:::(t)$, for all $t\in\mathbb{R}$ given by

$$F:::(t):=F:::(g_{ab};tg_{ab,c};\dots;t^\alpha g_{ab,c_1\cdots c_\alpha}),$$

define a tensorial concomitant of the metric. Evidently (1) implies that

$$C^{ij}(t)=t^3C^{ij}. \tag{6}$$

Let $\mathscr{L}$ be a scalar density of the form (2) for which (3) holds. By direct computation we deduce that

$$E^{ij}(\mathscr{L}(t))=E^{ij}(\mathscr{L})(t)=C^{ij}(t). \tag{7}$$

From (6) and (7) we find that

$$t^3C^{ij}=E^{ij}(\mathscr{L}(t)). \tag{8}$$

It is evident that the d/dt derivative commutes with the E^{ij}

 0022-2488/79/091905-03$01.00 © 1979 American Institute of Physics

operator ($\mathscr{L}$ is differentiable). Thus differentiating three times with respect to t in (8) and then setting $t=0$ we obtain

$$C^{ij} = E^{ij}(\mathscr{L}'), \tag{9}$$

where

$$\mathscr{L}' := \frac{1}{3!}\left[\frac{d^3}{dt^3}(\mathscr{L}(t))\right]\bigg|_{t=0}. \tag{10}$$

To construct $\mathscr{L}'$ we apply the replacement theorem[9] to $\mathscr{L}(t)$ to deduce that

$$\mathscr{L}(t) = \mathscr{L}(g_{ab}; 0; t^2 g_{abcd}; \ldots; t^\alpha g_{abc_1\cdots c_\alpha}), \tag{11}$$

where $g_{abc_1\cdots c_\beta}$, $\beta = 2,\ldots,\alpha$, are the βth metric normal tensors.[9] Substituting from (11) into (10) and doing the indicated operations we find that

$$\mathscr{L}' = \eta^{abcde} g_{abcde}, \tag{12}$$

where η^{abcde} is (necessarily) a tensor density concomitant of g_{rs} only. Since η^{abcde} consists of a linear combination of terms of the form[10] $g^{rs}\epsilon^{tuv}$, each term in the sum for η^{abcde} will be totally antisymmetric in three of its five indices. Since g_{abcde} is symmetric in its first two indices and totally symmetric in its last three indices we must have [by (12)] that

$$\mathscr{L}' \equiv 0.$$

The above contradicts (9) and therefore contradicts the existence of a scalar density $\mathscr{L}$ of the form (2) satisfying (3). □

The Cotton tensor density thus possesses the unusual feature that it satisfies conditions (4) but cannot be obtained via (3) from a scalar density of the form (2). It is conceivable that by relaxing the demand that $\mathscr{L}$ be a scalar density one may be able to find a function $\mathscr{L}$ of the form (2) for which (3) holds.[11] The behavior exhibited by C^{ij} is not a common feature among expressions satisfying Eqs. (4) or their generalizations. In most cases the above behavior does not occur unless the relevant expressions are homogeneous of degree -1 in the field functions[11] as is the case with C^{ij}.

An interesting application of the proof of the theorem is the following related result:

Corollary: In a 3-space there does not exist a class C^3 scalar density concomitant of the metric and its derivatives which is conformally invariant.

Proof: Let $\mathscr{W}$ be such a concomitant. From the coordinate transformation $x^i = t\bar{x}^i$ for $t\in\mathbb{R}^+$ (arbitrary) it follows that

$$t^3\mathscr{W} = \mathscr{W}(t^2 g_{ab}; t^3 g_{ab,c}; \ldots, t^{\alpha+2} g_{ab,c_1\cdots c_\alpha}).$$

Since $\mathscr{W}$ is conformally invariant and hence homogeneous of degree zero in g_{ab} and its derivatives we find that at an arbitrary point P in the 3-space $t^3\mathscr{W}_P = \mathscr{W}_P(t)$ (cf. the notation in the proof of the theorem). Since P is arbitrary, $t^3\mathscr{W} = \mathscr{W}(t)$ and the steps in the proof of the theorem can now be followed starting at Eq. (10) (replacing $|_{t=0}$ by $\lim_{t\to 0}$). □

The differentiability condition in the previous corollary is necessary since the scalar density $\alpha_1 = (C^r_s C^s_r)^{1/2}$ and $\alpha_2 = (C^r_s C^s_t C^t_r)^{1/3}$ are both conformally invariant but neither is of class C^1 in any neighborhood of a conformally flat metric tensor. As York[12] has indicated, one can use scalar densities such as α_1, α_2, and $\alpha_1 f(\alpha_2/\alpha_1)$ (where f is a real valued differentiable function to obtain trace free, symmetric divergence free (2,0) tensor density concomitants T^{ij} of a metric tensor. Given any such scalar density ϕ one simply takes $T^{ij} = E^{ij}(\phi)$, provided one avoids those metrics for which $\alpha_1 = 0$ or $\alpha_2 = 0$. Any tensor density T^{ij} constructed in this manner has the additional property that $T^i_j := g_{jk}T^{ik}$ is conformally invariant.

From York's construction procedure it seems likely that any tensor density T^{ij} obtained in this way will not be differentiable at a conformally flat metric (nor will it even be defined there). This is due to the occurrence of terms involving negative powers of α_1 and α_2 in the resulting T^{ij}. By demanding differentiability for arbitrary metric tensors, we are led to the following characterization of the Cotton tensor density:

Proposition: In a 3-space let T^i_j be a class C^3 (1,1) tensor density concomitant of the metric and its derivatives (to some finite order) which is conformally invariant. Then $T^i_j = aC^i_j$ where a is an arbitrary constant.

Proof: Using the argument which begins the proof of the previous corollary we find that for all $t\in\mathbb{R}^+$,

$$t^3 T^i_j = T^i_j(t).$$

Following the same steps as in the proof of the above theorem starting at (10) (using $\lim_{t\to 0}$ instead of $|_{t=0}$), we obtain

$$T^i_j = \eta^{i\,abcde}_{j} g_{abcde},$$

where $\eta^{i\,abcde}_{j}$ is a tensor density concomitant of the metric. Thomas[9] (chapter 6) has shown that the third metric normal tensor is a linear combination of $R_{abcd|e}$ terms and thus, upon raising an index in the above equation, we find that

$$T^{ij} = \zeta^{ijabcde} R_{abcd|e},$$

where $\zeta^{\cdots}$ is a tensor density concomitant of the metric. Using the Bianchi identity along with Weyl[10] (to construct $\zeta^{\cdots}$) we obtain the nonvanishing terms

$$T^{ij} = b_1\epsilon^{iab}R^j{}_{a|b} + b_2\epsilon^{jab}R^i{}_{a|b} + b_3\epsilon^{ija}R_{|a}.$$

The dimensionally dependent identity $\delta^{ijab}_{rstu}\epsilon^{rst}R^u{}_{a|b} \equiv 0$ implies the T^{ij} is of the form

$$T^{ij} = c_1 C^{ij} + c_2\epsilon^{ija}R_{|a}.$$

The conformal invariance of T^i_j now implies that $c_2 = 0$ and hence establishes the proposition. □

It is remarkable that the two conditions, viz., conformal invariance and differentiability, are sufficient to obtain the Cotton tensor density in a 3-space from all possible (1,1)-tensor densities. Moreover, these two conditions also imply all the properties of the Cotton tensor density, i.e., $T^i_i = 0$, $T^i_{j|i} = 0$, and $T^{ij} = T^{ji}$! This unique characterization does not hold for spaces of dimension greater than three. For example, in a 4-space the conformally invariant tensor density $g^{1/2}C^{iabc}C_{jabc}$ is not divergence free nor is it trace free. (C^i_{abc} is the Weyl conformal curvature tensor.[13])

Finally, it is easily seen that the method of proof employed in the above proposition can be utilized (in somewhat modified form, depending upon the situation) to construct any *differentiable* (for metrics of arbitrary signature) tensorial concomitant of a metric tensor with conformal transformation properties. Of course, in general the procedure is more involved than the relatively simple proofs given for the cases considered here.

ACKNOWLEDGMENTS

The author would like to thank G.W. Horndeski, H.F. Ahner, D. Lovelock, and I.M. Anderson for discussions on the topic of this paper. This work was supported in part by the National Research Council of Canada.

[1] D. Lovelock, "Divergence-Free Third Order Concomitants of the Metric Tensor in Three Dimensions," in *Topics in Differential Geometry*, edited by H. Rund and W.F. Forbes (Academic, New York, 1976), p. 87.

[2] For notation see S.J. Aldersley, Phys. Rev. D **15**, 370 (1977). The definition of a tensorial concomitant used here is that on p. 35 of Ref. 9 in G.W. Horndeski, Utilitas Math. **9**, 3 (1976). For example, a scalar density concomitant of a pseudo-Riemannian metric g_{ab} is defined by a real-valued function ϕ of the form $\phi = \phi(g_{ab}; g_{ab,c}; \dots; g_{ab,c_1 \cdots c_p})$. The concomitant ϕ is said to be of class C^κ provided the function ϕ is of class C^κ at every metric in its domain. (For example, $g^{1/2}$ and $g^{1/2}R_{ij}$ are of class C^∞ but $g^{1/2}R^{1/2}$ is only continuous since it is not differentiable in any neighborhood of a metric for which $R = 0$ at some point.)

[3] É. Cotton, "Sur les varietés à trois dimensions," Ann. Fac. d. Sc. Toulouse (II) **1**, 385 (1899).

[4] J.W. York has used this tensor density with reference to the initial-value problem in general relativity. See Phys. Rev. Lett. **26**, 1656 (1971), and C.W. Misner, K.S. Thorne, and J.A. Wheeler, *Gravitation* (Freeman, San Francisco, 1973), pp. 541 and 550.

[5] If $L = L(g_{ab}; g_{ab,c}; \dots; g_{ab,c_1 \cdots c_\alpha})$, then

$$E^{ij}(L) := \sum_{\beta = 0}^{\alpha} (-1)^\beta \frac{d^\beta}{dx^{c_1} \cdots dx^{c_\beta}} \frac{\partial L}{\partial g_{ij,c_1 \cdots c_\beta}}.$$

[6] G.W. Horndeski, Tensor **28**, 303 (1974).

[7] I.M. Anderson Aeq. Math. (to be published). Also see Ref. 11.

[8] K. Kuchař [J. Math. Phys. **15**, 708 (1974)] has considered the tensor density $\beta^{ij} := \epsilon^{iab} R^j{}_{a|b}$ and shown that it is not an Euler–Lagrange expression. [Note that β^{ij} is *not* symmetric unless one is constrained to metrics for which $R_{|i} = 0$. Hence the fact that $\beta^{ij} \neq E^{ij}(L)$ for any L is obvious.] His method of proof does not apply to the Cotton tensor density in view of Eqs. (4) and the existence of a Lagrangian of the form (5). (That is, the functional curl of the Cotton tensor density vanishes identically.)

[9] T.Y. Thomas, *The Differential Invariants of Generalized Spaces* (Cambridge U.P., Cambridge, 1934), p. 109.

[10] H. Weyl, *The Classical Groups* (Princeton U.P., Princeton, 1939), p. 52 noting p. 65.

[11] These points are discussed further in S.J. Aldersley, "High Euler Operators and some of their Applications," preprint available from the author.

[12] J.W. York, J. Math. Phys. **14**, 456 (1973).

[13] Generalizations of the Cotton tensor density for higher dimensional spaces will be discussed in S.J. Aldersley and G.W. Horndeski (in preparation).

Plane symmetric static fields in Brans–Dicke theory of gravitation

A. Banerjee and D. Bhattacharya

Department of Physics, Jadavpur University, Calcutta 700032, India

(Received 13 October 1978)

An explicit form of the relationship between g_{00} and the Brans–Dicke scalar potential ϕ in the interior of a perfect fluid with an equation of state $p = \epsilon\rho$ as well as in the matter free space, is obtained assuming functional dependence of g_{00} on ϕ. Plane symmetric static perfect fluids in Brans–Dicke theory of gravitation are discussed. Explicit solutions are also obtained for a fluid with $\epsilon = 1/3$, that is, for a disordered radiation and some of its properties studied.

INTRODUCTION

In a recent communication Bruckman and Kazes (1977)[1] have studied perfect fluid with the equation of state $p = \epsilon\rho$ for the matter content in Brans-Dicke[2] theory of gravitation. They have arrived at a simple relation between g_{00} and the scalar potential ϕ for the spherically symmetric case in the interior as well as in the matter free exterior space, which they used to obtain some explicit solutions interpreted to be those for static spheres made of cold ultrahigh density perfect fluid.

In the first part of this note the relations between g_{00} and the scalar potential ϕ are generalized to both in the interior and in the exterior, where we have assumed a functional relationship between them instead of their being restricted by any special kind of symmetry. In the second part it is shown that field equations may be reduced to a nonlinear differential equation involving only one metric coefficient and the solution of this equation, once obtained will determine the others in view of the other existing relations. Explicit solutions for a very special case of disordered radiation ($\epsilon = \frac{1}{3}$) in Brans–Dicke[2] theory are obtained and some of their properties studied. These solutions, however, reduce to those of vacuum in Brans–Dicke theory when any of the arbitrary constants appearing there is set equal to zero, while on the other hand one recovers the solutions of Teixeira *et al.* (1977)[3] for disordered radiation in Einstein's theory in a natural way for the absence of the scalar field.

g_{00}–ϕ RELATION IN THE INTERIOR AND EXTERIOR REGIONS

For static space–time in Brans–Dicke theory one can write from the appropriate field equation (Brans–Dicke, 1961)[2]

$$R^0_0 + \frac{8\pi}{\phi}(T^0_0 - \tfrac{1}{2}T) = -\frac{1}{\phi}\phi^{\,0}_{\;;0} - \frac{1}{2\phi}\Box\phi \qquad (1)$$

with

$$\Box\phi = \frac{8\pi}{(3+2\omega)}T, \qquad (2)$$

where T^μ_ν stands for the energy–momentum tensor of the perfect fluid with the equation of state $p = \epsilon\rho$, being represented in comoving coordinates by

$$T^\mu_\nu = \rho\,\mathrm{diag}(1, -\epsilon, -\epsilon, -\epsilon), \quad \epsilon = \text{const}.$$

Now from (1) and (2) one can obtain the relation (see Bruckman and Kazes)

$$[(-g)^{1/2}(\ln\phi)^{,k}\phi\,]_{,k} = \tfrac{1}{2}c[(-g)^{1/2}\phi\,(\ln g_{00})^{,k}\,]_{,k} \qquad (3)$$

with

$$c = \frac{3\epsilon - 1}{(2\omega+3) + (\omega+1)(3\epsilon-1)}. \qquad (4)$$

Now for a functional relationship $g_{00} = g_{00}(\phi)$ the relation (3) reduces to

$$\left\{(-g)^{1/2}\phi^{,k}\left[1 - \frac{c}{2}\phi\frac{\dot g_{00}}{g_{00}}\right]\right\}_{,k} = 0 \qquad (5)$$

where

$$\dot g_{00} = dg_{00}/d\phi \text{ and } \phi^{,k} = g^{ik}\phi_{,i}.$$

Now defining a function of ϕ as

$$\xi(\phi) = \int\left[1 - \frac{c}{2}\phi\frac{\dot g_{00}}{g_{00}}\right]d\phi \qquad (6)$$

Eq. (5) may be written as

$$[\xi^{,k}(-g)^{1/2}]_{,k} = 0. \qquad (7)$$

Again, using (7), one may write

$$[\xi\xi^{,k}(-g)^{1/2}]_{,k} = g^{ik}\xi_{,i}\xi_{,k}(-g)^{1/2}. \qquad (8)$$

Integrating both sides of (8) throughout the interior of the fluid distribution and transforming the left-hand side into a surface integral, one obtains

$$\oint\xi\xi^{,k}(-g)^{1/2}ds_k = \int g^{ik}\xi_{,i}\xi_{,k}dv. \qquad (9)$$

The left-hand side of (9) vanishes in view of (7) when we consider the distribution bounded by an equipotential surface where $\phi = \text{const}$ which is again equivalent to $\xi = \text{const}$ from (6). In consequence $\xi_{,i} = 0$, since g_{ik} is negative definite and finally in view of (6), we arrive at a relation between g_{00} and ϕ in an explicit form such as

$$\phi = \text{const}\times(g_{00})^{c/2}. \qquad (10)$$

 0022-2488/79/091908-03$01.00 © 1979 American Institute of Physics

In the matter free exterior region, where $\rho = p = 0$, Eq. (1) takes up a very simple form

$$R_0^0 = -\frac{1}{\phi}\phi^0{}_{;0}, \tag{11}$$

which in turn gives

$$[g^{ik}\phi g_{00,k} g^{00}(-g)^{1/2}]_{,i} = 0. \tag{12}$$

Assuming the same functional relationship $g_{00} = g_{00}(\phi)$ to hold and utilising the relation $\Box\phi = 0$ for matter free space one can immediately obtain from (12)

$$\left(\frac{g'_{00}\phi}{g_{00}}\right)' = 0. \tag{13}$$

Integration of (13) yields explicitly the functional relationship

$$\phi = \text{const} \times (g_{00})^{c/2}, \tag{14}$$

c being an arbitrary constant here. The relations (12) and (14) are identical with those obtained by Bruckman and Kazes for spherical symmetry.

PLANE SYMMETRIC PERFECT FLUID–FIELD EQUATIONS AND THEIR INTEGRATION

We consider the plane symmetric static metric in the form

$$ds^2 = e^{2\alpha}dt^2 - e^{2\beta}dx^2 - e^{(\beta-\alpha)}(dy^2 + dz^2), \tag{15}$$

where α and β are functions of x alone. The field equations in Brans–Dicke theory are now

$$\alpha_{11} = \frac{8\pi}{\phi}\left(\rho - \frac{(\omega+1)}{(2\omega+3)}T\right)e^{2\beta} - \alpha_1\frac{\phi_1}{\phi}, \tag{16}$$

$$\beta_{11} + \tfrac{1}{2}(3\alpha_1 + \beta_1)(\alpha_1 - \beta_1)$$

$$= -\frac{8\pi}{\phi}\left(\epsilon\rho + \frac{(\omega+1)}{(2\omega+3)}T\right)e^{2\beta}$$

$$-\omega\phi_1^2/\phi^2 + \beta_1\phi_1/\phi - \phi_{11}/\phi, \tag{17}$$

$$\tfrac{1}{2}(\alpha_{11} - \beta_{11}) = \frac{8\pi}{\phi}\left[\epsilon\rho + \frac{(\omega+1)}{(2\omega+3)}T\right]e^{2\beta} + \tfrac{1}{2}(\beta_1 - \alpha_1)\frac{\phi_1}{\phi},$$

and the wave equation for the scalar potential ϕ is

$$\phi_{11} = -\frac{8\pi}{(2\omega+3)}Te^{2\beta} \tag{19}$$

One can now utilize (16) and (18) to obtain

$$\frac{\alpha_{11} + \alpha_1\phi_1/\phi}{\tfrac{1}{2}(\alpha_{11} - \beta_{11}) - \tfrac{1}{2}(\beta_1 - \alpha_1)\phi_1/\phi}$$

$$= \frac{1 - [(\omega+1)/(2\omega+3)](1-3\epsilon)}{\epsilon + [(\omega+1)/(2\omega+3)](1-3\epsilon)}, \tag{20}$$

the integration of which yields a relation like

$$\beta_1 = -\frac{c_0}{d_0}\alpha_1 + \frac{A}{\phi}, \tag{21}$$

where A is the integration constant and c_0, d_0 are constants being given by

$$c_0 = \left[\left(\epsilon - \frac{1}{2}\right) + \frac{3}{2}\frac{(\omega+1)}{(2\omega+3)}(1-3\epsilon)\right]$$

and

$$d_0 = \frac{1}{2}\left[1 - \frac{(\omega+1)}{(2\omega+3)}(1-3\epsilon)\right].$$

In view of (10) and (21) Eq. (17) can now be expressed as

$$A_1\alpha_{11} + A_2\alpha_1^2 + A_3e^{-c\alpha}\alpha_1 + A_4e^{-2c\alpha} = 0 \tag{22}$$

where A_1, A_2, A_3, A_4 are constants depending in general on the magnitude of ω, ϵ, and A. The Bianchi identity i.e., $T^{\mu\nu}{}_{;\nu} = 0$ give the relation

$$p = p_0e^{-k\alpha} \tag{23}$$

with $k = (1+\epsilon)/\epsilon$ and p_0 a constant.

Once we get a solution for α from (22), β can immediately be obtained from (21) in view of (10). It is not difficult to see that the solutions thus obtained along with ϕ being computed from (10) will satisfy all the field equations (16)–(18) and the wave equation (19).

We do not attempt here to obtain the general solution of (22), which is apparently not quite easy, except for a very simple solution in the special case for $A = 0$ in (21), so that $A_3 = A_4 = 0$ in (22). The explicit forms of such solutions are obtained after integration of (22) in the form

$$e^{\alpha} = (a_1x + a_2)^{1/h}, \quad e^{\beta} = (a_1x + a_2)^{m/h} \tag{24}$$

so that the matter density ρ is given by

$$\rho \propto (a_1x + a_2)^{c(1/h - 2/c) - 2m/h}, \tag{25}$$

where a_1, a_2 are integration constants; $m = -c_0/d_0$, $h = A_2/A_1$.

SPECIAL CASE OF $\varepsilon = 1/3$

It may be noted from (1) and (2) that for $\epsilon = 1/3$ the relation (10) is not valid and so the case of disordered radiation is to be considered separately. The Bianchi identity reduces to the relation

$$p = p_0e^{-4\alpha}. \tag{26}$$

One can integrate in this case Eq. (19) in view of $T = 0$, to obtain

$$\phi = (ax + b), \tag{27}$$

a and b being integration constants. Exact solutions are obtained on integrating the field equations and are given in the form

$$3\beta - \alpha = \frac{A}{a}\ln\left(1 + \frac{a}{b}x\right) + B. \tag{28}$$

and

$$\beta = \tfrac{1}{5}\ln\left(1 + \frac{a}{b}x\right)^{5\omega/2}\left[l\left(1 + \frac{a}{b}x\right)^{(1-\omega)} + n\right]. \tag{29}$$

The explicit form (29) is obtained, however, by making a choice $\omega = A/a$, where A, B, l, n are arbitrary constants.

It is now not difficult to see that one may get a class of empty space ($p = 0$) plane symmetric solutions in Brans–Dicke theory by setting any of l, n, and $(\omega - 1)$ equal to zero and they are found to satisfy the general relation (14) already derived for empty space in Brans–Dicke theory.

We now suitably adjust the constants in (28) and (29) so as to satisfy the regularity conditions on the plane of symmetry. Thus $\alpha = \beta = 0$ at $x = 0$ and further since the pressure gradient should vanish exactly at the central plane of symmetry the relation (26) shows that $\alpha_{,1} = 0$ at $x = 0$ (see Teixeira *et al.*[3]). In view of the above regularity conditions one has to put $n = (1 - l) = 5\omega/6(\omega - 1)$ and $B = 0$ and thus finally obtain the solutions of the plane symmetric static distributions of disordered radiation in Brans–Dicke theory in the form

$$\beta = \tfrac{1}{5}\ln\left[\frac{5\omega}{6(\omega-1)}\left(1+\frac{a}{b}x\right)^{(1+3\omega/2)} + \frac{(\omega-6)}{6(\omega-1)} \times\left(1+\frac{a}{b}x\right)^{5\omega/2}\right] \tag{30}$$

and

$$\alpha = \tfrac{3}{5}\ln\left[\frac{5\omega}{6(\omega-1)}\left(1+\frac{a}{b}x\right)^{(1-\omega/6)} + \frac{(\omega-6)}{6(\omega-1)} \times\left(1+\frac{a}{b}x\right)^{5\omega/6}\right]. \tag{31}$$

(26) now gives using the field equations the value of p_0 so that

$$24\pi p_0 = \frac{a^2\omega(\omega-6)}{12\cdot c}. \tag{32}$$

For a particular case of $\omega = 6$, the matter density vanishes everywhere. (26) and (31) indicate that for $\omega > 6$ the density and pressure vanish as $x \to \infty$ whereas for $\omega < 6$ the pressure and density become infinitely large as $x \to \infty$.

Finally it is not difficult to show that when in the limit the scalar field is absent, that is, when $a \simeq 0$, which is equivalent to $\omega \simeq \infty$ in view of the relation $\omega = A/a$, (30) and (31) reduce to those of Teixeira *et al.*[3] mentioned earlier by utilizing in (31) the well-known elementary result $\lim_{n\to\infty}(1 + y/n)^n = e^y$.

[1] W.E. Bruckman and E. Kazes, Phys. Rev. D **16**, 261 (1977).
[2] C. Brans and R.H. Dicke, Phys. Rev. **124**, 925 (1961).
[3] A.F. da. F. Teixeira, I. Wolk, and M.M. Som, J. Phys. A **10**, 1699 (1977).

Neutrino, Maxwell, and scalar fields in the cylindrical magnetic or plasm universe

David F. Cattell and M. A. Melvin
Department of physics, Temple University, Philadelphia, Pennsylvania 19122

(Received 9 October 1978; revised manuscript received 29 December 1978)

In this paper we discuss the solutions of the c-number quantum-mechanical equations of neutrino, Maxwell, and scalar fields in the background metric of a static, cylindrically symmetric magnetic or plasm universe. The magnetic universe consists of a parallel magnetic field distribution held together by the gravitational field of its own energy and stress. It is an exact solution of the Einstein–Maxwell equations for the gravitational field associated with a stationary source-free cylindrically symmetric magnetic field. The solution does not depend specifically upon the assumption that the seat of the energy-stress distribution is an electromagnetic field. A "plasm of index 2" would have the same metric. The main instrument of the present perturbation theory, valid in all algebraically special spacetimes, is the Cohen–Kegeles (CK) scalar wave equation for test fields of helicity h (or $-s$). This equation, written in the updated Geroch–Held–Penrose (GHP) formalism, is made explicit for the chosen null tetrad by determining the spin coefficients and the operators of that formalism. Exact solutions of the CK wave equation are obtained for the case of zero orbital angular momentum. For the case of nonzero orbital angular momentum the JWKB approximation is used to obtain the quadratures for the orbits in the magnetic universe. It is shown that the neutrino orbits precess about the magnetic (or plasm) universe axis in a manner different from orbits of particles with helicity zero and helicity one. The expression for the neutrino precessional frequency (due to nonvanishing helicity) is given explicitly. The photon precessional frequency is twice that for neutrinos. Finally, the CK procedure is compared with that of Teukolsky for finding the neutrino and Maxwell fields in type D spacetimes.

I. INTRODUCTION

A. The CK procedure

Cohen and Kegeles [1] have considered the problem of zero rest mass fields in a certain broad class of curved spacetimes (defined in the following). They showed that the components of certain zero rest mass fields in a fixed background geometry can be obtained by straightforward differentiation of a single complex function which obeys a linear scalar wave equation determined by the helicity of the zero mass field and by the geometry. The Cohen–Kegeles (CK) procedure is analogous to the Hertz–Debye potential scheme used in the solution of certain problems in classical electrodynamics where, of course, the spacetime is flat. For the CK procedure to be applicable the geometry must satisfy two conditions. First is the *algebraic special condition*: The Weyl conformal tensor which is associated with the geometry must have a repeated principal null vector ("propagation vector"); such a geometry is said to be algebraically special. [2] Second is the *congruence condition*: There must exist a congruence of shear-free null geodesics along this null vector. This second condition is automatically equivalent to the first in the vacuum geometries, but is an independent requirement in the case where there are stress–energy sources ("matter") present. Whenever both conditions are satisfied we shall call the geometry "principally anastigmatic." Thus, principally anastigmatic geometries are just members of the generalized Goldberg–Sachs class of spaces. [3]

To envisage a principally anastigmatic geometry we consider the cylindrical band formed by parallel transport of a repeated principal null vector of the Weyl tensor an equal small proper distance in all possible directions transverse to itself. [4] We do this at all points along the geodesic to which the principal null vector is tangent and thus form a world tube of "light rays." If the spacetime is principally anastigmatic, the shape of the cross section of the world tube does not change as we move along the world tube. For example, if the cross section is a circle at one place along the world tube, it will be a circle at all other places and so on. The shape of the cross section will change in spacetimes which are not principally anastigmatic. In the following we shall discuss and apply the treatment of the neutrino spinor field and other zero rest-mass fields by a single scalar wave equation in a principally anastigmatic spacetime.

B. The magnetic universe as background geometry

In this paper we apply the CK procedure to a background geometry discussed by one of the authors; [5,6] this geometry is an exact solution of the Einstein-Maxwell equations for the gravitational field assoicated with a stationary source-free cylindrically symmetric magnetic field. Interpreted in Newton-Maxwell language the solution represents a "universe" consisting of a parallel magnetic field distribution held together by the gravitational field of its own energy and stress. The magnetic universe solution has a number of interesting features [5,7] which make it especially suitable to serve as a test metric for new techniques in Relativity. In the present paper we assue that there is also a neutrino field present (or a photon field [8] or a zero-mass scalar field), but not in such strength as to affect the geometry appreciably. Since the magnetic universe geometry is of Petrov type D, and has shear-free congruences of null geodesics along each

 0022-2488/79/091911-14$01.00 © 1979 American Institute of Physics

of the two repeated principal directions of the Weyl tensor (see Appendix A), it is "principally anastigmatic" in our terminology, and we can apply the CK procedure.

C. The two-component neutrino and other zero-mass fields

Two-component neutrinos are described by a simple spinor satisfying the Weyl equation. This equation is an eigenvalue equation which determines allowed energy and momentum states associated with each of the two helicity values which are possible for neutrinos, $\frac{1}{2}$ and $-\frac{1}{2}$. The equation actually amounts to two coupled differential equations, each involving both neutrino spinor components. To solve the Weyl equations and interpret the solutions one must decouple the pair of two-component equations. Now the equations themselves, originally given in Minkowski spacetime, can easily be generalized to apply in any curved spacetime, which, of course, is associated with an Einstein gravitational field. The generalization consists of replacing ordinary derivatives with covariant derivatives in the Weyl equation ("minimal gravitational coupling"). This involves putting in extra terms involving the "spinor affine connection." But, in curved spacetimes, the decoupling of the two component equations is not an easy task because of the additional connection terms. The CK procedure gives a general prescription for the decoupling in principally anastigmatic spacetimes by deriving a single wave equation to replace the curved spacetime Weyl equations. [1] This technique makes use of the Newman–Penrose (NP) formalism. [9] In Sec. II we give a brief review of the CK procedure for obtaining the neutrino spinor field; in this review, however, we rewrite the CK procedure in the more compact formalism developed by Geroch, Held, and Penrose (GHP). [10] With the help of the spin coefficients (NP or GHP coefficients), and operators calculated in that formalism, one constructs a wave equation for a single complex function ψ. Once found the function ψ, differentiated, gives the spin dyad components of the neutrino spinor field. It is of course the case that a field of dyad components, or a "dyad frame," defines a unique null tetrad frame. [10] These provide two alternative reference frames for the components of a field.

In Sec. III we discuss the CK equations for the Maxwell field and for the scalar zero mass field. We write these equations in the GHP formalism and show that these wave equations, together with the CK wave equation for neutrinos, can be summarized by a single wave equation where the helicity enters as a parameter. This generalized CK equation is written in a simplified form involving the scalar field operator plus an addend involving the helicity.

Section IV lists NP spin coefficients for the magnetic universe. In Sec. V we give the wave equation for the neutrino "potential" ψ and discuss the probability density. Section VI contains solutions for the case of zero orbital angular momentum: The wave equation is solved, and the dyad components of the spinor field are found. In Sec. VII the solutions of the wave equation, the neutrino spinor field components, and the asymptotic probability density of the neutrino spinor field are discussed for the case of nonzero orbital angular momentum. We use the JWKB approximation to obtain an appropriate expression for the energy eigenvalues as well as the neutrino field in the classical limit. We show that the classical orbits of the neutrinos (i.e., orbits obtained in the short wavelength limit) are essentially the same as those obtained assuming the neutrinos to be zero rest mass point particles, without helicity, moving along geodesics (discussed in Ref. 7). From the JWKB approximation to the wave equation we show the first order correction to these orbits to be of a precessional nature. They can be taken to define the corrections to the motion of a classical particle possessing helicity. Section VIII contains a brief discussion of the scalar field in the magnetic universe and the absence of extraneous orbital precession is shown. Conclusions are given in Sec. IX along with a comparison of the results with those obtained by following a procedure for finding neutrino spinor fields in background metrics given by Teukolsky. [11]

II. THE NEUTRINO SPINOR FIELD IN A FIXED BACKGROUND SPACETIME

The key problem is the decoupling of the two spinor components appearing in the Weyl pair of equations in the case of the presence of a gravitational field. [1] A necessary condition for the applicability of the CK procedure for solving the generalized Weyl equation is that the spacetime be principally anastigmatic (as defined in the Introduction).

We have redeveloped the CK procedure using the improved formalism of Geroch, Held, and Penrose (GHP). [10] We use the GHP formalism in the procedure because the notation is more compact and allows one to shorten the derivation of the wave equation for ψ. We devote the remainder of this section to the derivation of the wave equation for ψ. The spinor notation follows that of Pirani, [12] Kegeles, [13] GHP, [10] and Plebański [4]

The generalized Weyl equation reads

$$\nabla^{AX'}\Phi_A = 0 \quad (A = 1,2;\ X' = 1',2'). \tag{2.1}$$

Let ϕ_a $(a = 1,2)$ be the scalar projections of Φ_A along the dyad spinors $o^A = \binom{1}{0}$, $i^A = \binom{0}{1}$. Equation (2.1), when written in GHP notation, [10] becomes

$$(\eth' - \tau')\phi_1 - (\text{Þ} - \rho)\phi_2 = 0, \tag{2.2a}$$

$$(\text{Þ}' - \rho')\phi_1 - (\eth - \tau)\phi_2 = 0. \tag{2.2b}$$

If the spacetime is principally anastigmatic and the tetrad vector l^μ is a repeated principal null vector of the Weyl tensor, we have the GHP commutator (the complex conjugate of Eq. (2.31) of Ref. 10) as applied to a scalar function ψ of type $\{0, -1\}$:

$$(\eth'\text{Þ} - \text{Þ}\eth')\psi = (\tau'\text{Þ} - \rho\eth')\psi, \tag{2.3}$$

which may be rewritten

$$(\eth' - \tau')\text{Þ}\psi - (\text{Þ} - \rho)\eth'\psi = 0. \tag{2.4}$$

Equation (2.2a) will then be satisfied automatically if we define a single complex scalar field ψ such that

$$\phi_1 = -\text{Þ}\psi, \tag{2.5}$$

$$\phi_2 = -\eth'\psi, \tag{2.6}$$

and (2.2b) becomes

$$[(\text{Þ}' - \rho')\text{Þ} - (\eth - \tau)\eth']\psi = 0. \tag{2.7}$$

The definition of ψ by Eq. (2.5) and (2.6) is consistent with the spin- and boost-weights of ϕ_1 and ϕ_2 given in Ref. 10. Once (2.7) is solved for ψ, the dyad components of the neutrino spinor field can be found by using (2.5) and (2.6).

At this point we should note that Teukolsky [11] has given a procedure for finding the solutions of Eq. (2.2a) and (2.2b) in type D spacetimes (the magnetic universe is a type D space–time as explained in Appendix A). This procedure involves solving two decoupled second order partial differential equations for the spin dyad components ϕ_1 and ϕ_2 of the neutrino spinor field. In the procedure discussed above it is necessary to solve only one second order partial differential equation (2.7). Furthermore, (2.7) is valid for a wider class of spacetimes than are Teukolsky's equations, in that the Weyl tensor associated with the spacetime only need be algebraically special, whereas the requirement for the validity of Teukolsky's procedure is that the spacetime be type D. [11]

Further advantages of the CK procedure over Teukolsky's are discussed in Appendix D, where a simple example is given.

We again note that it is assumed in the CK procedure for solving the Weyl equation that the neutrinos themselves do not alter the background geometry. This should be a good approximation, except under conditions such as in the early stages of the universe in a "big bang" theory.

III. THE MAXWELL AND SCALAR FIELDS IN A FIXED BACKGROUND SPACETIME

In this section we write the source-free Maxwell field equations in GHP notation. We then give the CK equation for the Debye function ψ of the Maxwell field and the prescription for finding the Maxwell field spinor by differentiating ψ. Next we give the scalar field equation in GHP notation and show that the neutrino, Maxwell, and scalar field equations for the Debye function ψ can be generalized to a single wave equation where the helicity of each specific field enters as a parameter.

Maxwell's source-free equations, when written in GHP notation, are

$$Þ\phi_1 - ð'\phi_0 = -\tau'\phi_0 + 2\rho\phi_1 - \kappa\phi_2, \tag{3.1a}$$

$$Þ\phi_2 - ð'\phi_1 = \sigma'\phi_0 - 2\tau'\phi_1 + \rho\phi_2, \tag{3.1b}$$

together with their primed versions (the prime operation is defined in Ref. 10).

Equations (3.1) are the analog of equations (2.2) for the neutrino field. Here the scalars ϕ_0, ϕ_1, and ϕ_2 are projections of the symmetric spinor Φ_{AB} which is related to the Maxwell field tensor or spinor. The projections are onto the dyad i^A, o^A defined in Sec. II:

$$\phi_0 \equiv o^A o^B \Phi_{AB}, \quad \{2,0\}, \tag{3.2a}$$

$$\phi_1 \equiv o^A i^B \Phi_{AB}, \quad \{0,0\}, \tag{3.2b}$$

$$\phi_2 \equiv i^A i^B \Phi_{AB}, \quad \{-2,0\}, \tag{3.2c}$$

where the numbers in brackets are the $\{p,q\}$ types of the ϕ's (Ref. 10).

The symmetric spinor Φ_{AB} is related to the Maxwell spinor $F_{ABA'B'} \leftrightarrow F_{\mu\nu}$ $(AA' \leftrightarrow \mu, BB' \leftrightarrow \nu)$ by [10]

$$F_{ABA'B'} = \Phi_{AB}\epsilon_{A'B'} + \epsilon_{AB}\Phi_{A'B'}, \tag{3.3}$$

where ϵ_{AB} is the Levi–Civita alternating symbol. [14]

Cohen and Kegeles have shown [13] that (3.1) and their primed versions can be decoupled in principally anastigmatic spacetimes when one takes the tetrad vector l^μ as a repeated principal null vector of the Weyl tensor. We give the CK results rewritten in GHP notation.

The dyad components (3.2) of Φ_{AB} are given by

$$\phi_0 = -(Þ - \bar\rho)(Þ + \bar\rho)\psi, \tag{3.4a}$$

$$\phi_1 = -[Þ(ð' + \bar\tau) - (\bar\tau - \tau')(Þ + \bar\rho)]\psi, \tag{3.4b}$$

$$\phi_2 = -[(ð' - \bar\tau)(ð' + \bar\tau) + \sigma'(Þ + \bar\rho)]\psi. \tag{3.4c}$$

The scalar ψ (or type $\{0, -2\}$) satisfies the equation

$$[(Þ' - \rho')(Þ + \bar\rho) - (ð - \tau)(ð' + \bar\tau)]\psi = 0. \tag{3.5}$$

Once (3.5) is solved for ψ, Eq. (3.4) can be employed to find the ϕ's. Equations (3.2) and (3.3) may then be utilized to find the Maxwell field tensor from which one can read off the electric and magnetic fields.

The original derivation of (3.5) by Cohen and Kegeles may be found in Ref. 13.

We conclude this section with a discussion of the field equation for scalar quanta of zero rest mass. This equation will be used in Sec. VIII to show that the additional orbital precession of neutrino orbits in the magnetic universe is due entirely to the neutrino's nonvanishing helicity.

Using the telescoped GHP notation of Ref. 15, the wave equation for scalar quanta may be written as

$$(A_0 + \mathring{A}_0)\bar\psi = 0, \tag{3.6}$$

where

$$A_0 \equiv Þ'Þ - \bar\rho'Þ - \rho Þ' \tag{3.7a}$$

and

$$\mathring{A}_0 \equiv -ð'ð + \bar\tau ð + \tau ð'. \tag{3.7b}$$

Substituting Eq. (3.7) into (3.6) and taking the complex conjugate of the result gives

$$[(Þ' - \rho')Þ - (ð - \tau)ð' - (\bar\rho Þ' - \bar\tau ð)]\psi = 0, \tag{3.8}$$

where here ψ:$\{0,0\}$.

We note the similarity of Eqs. (2.7), (3.5), and (3.8). Cohen and Kegeles [1] have pointed out that (2.7) and (3.5) together with the equation for gravitational perturbations (not discussed here) can be summarized by an equation which we rewrite here in GHP notation:

$$\{(Þ' - \rho')[Þ - (2s+1)\bar\rho] - (ð - \tau)[ð' - (2s+1)\bar\tau] - (s+1)(2s+1)\bar\psi_2\}\psi^{(s)} = 0,^{16} \quad \psi^{(s)}:\{0,2s\}, \tag{3.9}$$

where $s = -\frac{1}{2}$, -1, or -2 for neutrino, Maxwell and gravitational perturbation fields, respectively. ψ_2 is an NP component of the Weyl tensor and is defined in (A1). It is the only nonvanishing component in type D spacetimes.

If we substitute $s = 0$ in (3.9), we get

$$[(Þ' - \rho')(Þ - \bar\rho) - (ð - \tau)(ð' - \bar\tau) - \bar\psi_2]\psi^{(0)} = 0. \tag{3.10}$$

Using Eqs. (2.16) and (2.26) of Ref. 10 (complex-conjugated) this equation can be written in principally anastigmatic spacetimes as

$$\{(Þ' - \rho')Þ - (ð - \tau)ð' - (\bar\rho Þ' - \bar\tau ð) + \tfrac{1}{12}R\}\psi^{(0)}$$

$$= 0, \tag{3.11}$$

where R is the Riemann curvature scalar.

(3.11) agrees with (3.8) except for the extra term $\frac{1}{12}R$. For spacetimes where $R = 0$ (i.e., for spacetimes with a trace-free stress–energy tensor, like the magnetic universe and vacuum spacetimes), no ambiguity can arise in using (3.8) to treat scalar particles. For spacetimes where R does not vanish Graham [17] gives an argument for retaining this extra term. We may call this term an effective mass for the test particle in a gravitational field.

IV. THE CYLINDRICAL MAGNETIC UNIVERSE AND THE SPIN COEFFICIENTS

The cylindrical magnetic universe [5,7] is an exact solution of the Einstein–Maxwell equations for the gravitational field associated with a stationary source-free magnetic field. This hypothetical universe consists of a bundle of parallel magnetic field lines held together by their own gravitational field. Further discussion can be found in Refs. 5 and 7. The reason for working with this solution here is threefold. First, it is mathematically simple and serves as a good illustrative example of a spacetime for which the Cohen–Kegeles procedure for finding the neutrino spinor field is applicable. Second, there exist magnetic fields which are very strong (neutron stars) or very extensive (interstellar and perhaps intergalactic magnetic fields) and it is of interest to know how they affect the motions of elementary particles; some inferences may be drawn from our results. The third, and perhaps the most important reason, is that the solution illustrates effectively how the helicity of the neutrino interacts with the background geometry to modify the orbits from those obtained assuming neutrinos to be zero rest mass point particles following well-defined trajectories. This effect appears to be quite general and is discussed in Sec. VII below.

In this section we give the line element for the magnetic universe and the null tetrad used. The spin coefficients and the NP and GHP operators are given.

The line element of the cylindrical magnetic universe is [5]

$$ds^2 = W^2dt^2 - W^2dr^2 - (r/W)^2d\phi^2 - W^2dz^2 \tag{4.1}$$

$(\hbar = c = 1,\ t = x^0,\ r = x^1,\ \phi = x^2,\ z = x^3;\ W \equiv 1 + r^2)$,

corresponding to a sourceless electromagnetic field with only

$$F_{12} = -F_{21} = B_0 r/W^2 \neq 0. \tag{4.2}$$

Implicit in this expression is $\bar{a}$, a real-valued natural scale length (the "range radius") of the magnetic universe. [5] The coordinates t, r, and z, and the time–space interval ds are all expressed as multiples of the fundamental length. [18]

Following Newman and Penrose [9] we introduce four linearly independent 4-vectors $l^\mu, n^\mu, m^\mu, \bar{m}^\mu$ satisfying

$$l_\mu n^\mu = 1, \quad m_\mu \bar{m}^\mu = -1. \tag{4.3}$$

All other inner products between these vectors vanish. l^μ and n^μ are real and m^μ is complex. A bar above a symbol indicates complex conjugation, unless noted to the contrary.

We take the tetrad to be

$$l^\mu = \frac{1}{\bar{a}\sqrt{2W}}\{1,0,0,-1\}, \quad l_\mu = \frac{\bar{a}W}{\sqrt{2}}\{1,0,0,1\}, \tag{4.4a}$$

$$n^\mu = \frac{1}{\bar{a}\sqrt{2W}}\{1,0,01\}, \quad n_\mu = \frac{\bar{a}W}{\sqrt{2}}\{1,0,0,-1\}, \tag{4.4b}$$

$$m^\mu = \frac{1}{\bar{a}\sqrt{2W}}\left\{0,-1,\frac{iW^2}{r},0\right\}, \quad m_\mu = \frac{\bar{a}W}{\sqrt{2}}\left\{0,1,\frac{-ir}{W^2},0\right\}. \tag{4.4c}$$

In Appendix A it is shown that l^μ and n^μ are repeated principal null vectors of the Weyl tensor classifying the magnetic universe as Petrov type D. [2]

The spin coefficients, which can be calculated by a procedure given in Appendix B of Ref. 13 are found to be

$$\kappa = 0, \quad \sigma = 0, \quad \rho = 0, \quad \tau = \frac{\sqrt{2}\,r/\bar{a}}{W^2},$$
$$\beta = \frac{1}{\bar{a}r2\sqrt{2}}\frac{(r^2-1)}{W^2}, \quad \epsilon = 0 \tag{4.5a}$$

and

$$\kappa' = 0, \quad \sigma' = 0, \quad \rho' = 0, \quad \tau' = \frac{\sqrt{2}\,r/\bar{a}}{W^2},$$
$$\beta' = \frac{1}{\bar{a}r2\sqrt{2}}\frac{(r^2-1)}{W^2}, \quad \epsilon' = 0. \tag{4.5b}$$

The NP directional derivatives along the tetrad vectors are

$$D = \frac{1}{\bar{a}W\sqrt{2}}(\partial_t - \partial_z), \quad D' = \frac{1}{\bar{a}W\sqrt{2}}(\partial_t + \partial_z), \tag{4.6a}$$

$$\delta = \frac{1}{\bar{a}W\sqrt{2}}\left(-\partial_r + \frac{iW^2}{r}\partial_\phi\right),$$

$$\delta' = \bar{\delta} = \frac{1}{\bar{a}W\sqrt{2}}\left(-\partial_r - i\frac{W^2}{r}\partial_\phi\right),$$

and the GHP operators are[10]

$$þ\eta = (D - p\epsilon - q\bar{\epsilon})\eta, \quad þ'\eta = (D' + p\epsilon' + q\bar{\epsilon}')\eta,$$
$$ð\eta = (\delta - p\beta + q\beta')\eta, \quad ð'\eta = (\delta' + p\beta' - q\bar{\beta})\eta, \tag{4.6b}$$

where η is any scalar of type $\{p,q\}$.

These and the only nonvanishing component of the Weyl tensor (Appendix A),

$$\psi_2 = \frac{2}{\bar{a}^2}\frac{r^2-1}{W^4} \tag{4.7}$$

are needed to make Eqs. (2.7) and (3.9) explicit.

V. THE WAVE EQUATION FOR ψ, THE NEUTRINO SPIN DYAD COMPONENTS AND THE NEUTRINO PROBABILITY DENSITY

The wave equation for ψ (2.7) can be written as

$$[(D' + \epsilon' - \rho')(D + \bar{\epsilon}) - (\delta + \beta - \tau)(\delta' + \bar{\beta})]\psi = 0, \tag{5.1}$$

where Eqs. (4.6b) and the fact that ψ is of type $\{0, -1\}$ have been used.

Making use of (4.5) and (4.6a), (5.1) becomes

$$\frac{\partial^2\psi}{\partial r^2} - \frac{(r^2-1)}{rW}\frac{\partial\psi}{\partial r} - \left(\frac{\partial^2\psi}{\partial t^2} - \frac{\partial^2\psi}{\partial z^2}\right) + \frac{W^4}{r^2}\frac{\partial^2\psi}{\partial\phi^2} + \frac{i(3r^2-1)W}{r^2}\frac{\partial\psi}{\partial\phi} + \frac{\mathscr{P}_{1/2}(r)}{r^2W^2}\psi = 0, \quad (5.2)$$

where $W = 1 + r^2$ and $\mathscr{P}_{1/2}(r) \equiv \frac{1}{4}(3r^4 - 10r^2 - 1)$.

We assume ψ to be of the form

$$\psi(t,r,\phi,z) = R_{1/2}(r)e^{-i\omega t}e^{im\phi}e^{ikz} \quad (\omega > 0,\ m = 0, \pm 1, \pm 2, \cdots). \quad (5.3)$$

This separation of variables is quite natural and is discussed in Sec. VIII below.

The equation for $R(r)$ is

$$\frac{d^2R_{1/2}}{dr^2} - \frac{(r^2-1)}{rW}\frac{dR_{1/2}}{dr} + \left((\omega^2 - k^2) - \frac{W^4m^2}{r^2} - \frac{(3r^2-1)}{r^2}Wm + \frac{\mathscr{P}_{1/2}(r)}{r^2W^2}\right)R_{1/2}(r) = 0. \quad (5.4)$$

Using Eqs. (4.6a) and (4.6b) together with (5.3), (2.5) and (2.6) become

$$\phi_1 = \frac{i}{\bar{a}\sqrt{2}}\frac{(\omega + k)}{W}\psi, \quad (5.5a)$$

$$\phi_2 = \frac{1}{\bar{a}\sqrt{2}W}\left(\frac{R'_{1/2}}{R_{1/2}} - \frac{mW^2}{r} - \frac{1}{2r}\frac{(r^2-1)}{W}\right)\psi, \quad (5.5b)$$

where a prime indicates differentiation with respect to r.

Making the substitution

$$R_{1/2}(r) = (W/r)^{1/2}Q(r) \quad (5.6)$$

in (5.4) we obtain an equation where no first derivative term appears:

$$Q''(r) + \left(\lambda^2 - \frac{W^4m^2}{r^2} - \frac{(3r^2-1)}{r^2}Wm\right)Q(r) = 0, \quad (5.7)$$

where $\lambda^2 \equiv \omega^2 - k^2$.

The spin dyad components (5.5a) and (5.5b) become

$$\phi_1 = \frac{i(\omega + k)}{\bar{a}\sqrt{2}\sqrt{rW}}Qe^{-i\omega t}e^{im\phi}e^{ikz}, \quad (5.8a)$$

$$\phi_2 = \frac{1}{\bar{a}\sqrt{2}\sqrt{rW}}\left(\frac{Q'}{Q} - \frac{mW^2}{r}\right)Qe^{-i\omega t}e^{im\phi}e^{ikz}. \quad (5.8b)$$

In accord with (9.2) in the following the formal curved space probability density for neutrinos is

$$p(x') = \sqrt{|\det g_{ab}|}\,(|\phi_1|^2 + |\phi_2|^2)\sqrt{2}\bar{a}W, \quad (5.9)$$

where g_{ab} is the metric and $i, = 1, 2, 3$. For the magnetic universe with metric given by (4.1) we have

$$\sqrt{|\det g_{ab}|} = \bar{a}^4 rW^2 \quad (5.10)$$

and

$$p(r) = \frac{\bar{a}^3W^2}{\sqrt{2}}\left((\omega + k)^2 + \left|\frac{Q'}{Q} - \frac{mW^2}{r}\right|^2\right)|Q|^2. \quad (5.11)$$

We shall require that $p(r)$ given by (5.11) not diverge in the domain $0 \leqslant r < \infty$. We will seek solutions of (5.7) that satisfy this condition. For bound states $[p(r) \to 0, r \to \infty]$ we make the stronger requirement

$$\int_0^\infty p(r)dr < \infty \quad (5.12)$$

so that the neutrino spinor field can be normalized.

VI. THE CASE OF ZERO ORBITAL ANGULAR MOMENTUM

This section deals with the "meridian plane" orbits of Ref. 7, treated quantum mechanically for neutrinos. In what follows the wave equation for this case is solved, the neutrino spin dyad components are found, and the neutrino states are analyzed.

In this case $m = 0$ and (5.7) becomes

$$Q''(r) + \lambda^2 Q(r) = 0 \quad (m = 0), \quad (6.1)$$

where $\lambda^2 = \omega^2 - k^2$

For the case $\lambda = 0$ ($\omega = \pm k$ corresponding to neutrino motion parallel to the axis of the magnetic universe [7]). (6.1) further simplifies to

$$Q''(r) = 0 \quad (6.2)$$

with the general solution

$$Q(r) = ar + b \quad (m = 0,\ \lambda = 0,\ a,b \text{ constants}). \quad (6.3)$$

For the case $\lambda \neq 0$ the general solution of (6.1) is

$$Q(r) = Ae^{i\lambda r} + Be^{-i\lambda r} \quad (m = 0, \lambda \neq 0, A, B \text{ constants}). \quad (6.4)$$

In either case Eqs. (5.8) become

$$\phi_1 = \frac{i(\omega + k)}{\bar{a}\sqrt{2}\sqrt{rW}}Q(r)e^{-i\omega t}e^{ikz} \quad (m = 0), \quad (6.5a)$$

$$\phi_2 = \frac{1}{\bar{a}\sqrt{2}\sqrt{rW}}Q'(r)e^{-i\omega t}e^{ikz}, \quad (m = 0). \quad (6.5b)$$

Now the 2×2 spin matrices $\sigma_\mu^{bc'}$ connecting the spin dyad frame with the orthonormal frame can be determined from [9]

$$l_\mu = \sigma_\mu^{11'}, \quad m_\mu = \sigma_\mu^{12'}, \quad \bar{m}_\mu = \sigma_\mu^{21'}, \quad n_\mu = \sigma_\mu^{22'}, \quad (6.6)$$

where the tetrad vectors are given by (4.4) and the upper indices indicate dyad components.

Use of (4.4) and translating from dyad to spinor components in (6.6) gives

$$\sigma_0^{AB'} = \frac{\bar{a}W}{\sqrt{2}}\begin{pmatrix}1 & 0\\ 0 & 1\end{pmatrix}, \quad \sigma_1^{AB'} = \frac{\bar{a}W}{\sqrt{2}}\begin{pmatrix}0 & 1\\ 1 & 0\end{pmatrix},$$

$$\sigma_2^{AB'} = \frac{\bar{a}r}{W\sqrt{2}}\begin{pmatrix}0 & -i\\ i & 0\end{pmatrix}, \quad \sigma_3^{AB'} = \frac{\bar{a}W}{\sqrt{2}}\begin{pmatrix}1 & 0\\ 0 & -1\end{pmatrix}. \quad (6.7)$$

Consider the case $\lambda = 0, \omega = \pm k$ when $Q(r)$ is given by

(6.3). In order that the neutrino probability density (5.11) not diverge as $r \to \infty$ we must set $a = 0$ in (6.3). Then if $\omega = k$, Eqs. (6.5) become

$$\phi_1 = \frac{i\sqrt{2\omega}}{\bar{a}\sqrt{rW}} b e^{i\omega(z-t)} \quad (m = 0, \omega = k), \tag{6.8a}$$

$$\phi_2 = 0 \quad (m = 0, \omega = k), \tag{6.8b}$$

or in standard notation

$$\Phi_A = \frac{i\sqrt{2\omega}}{\bar{a}\sqrt{rW}} b \begin{pmatrix} 1 \\ 0 \end{pmatrix} e^{i\omega(z-t)} \quad (m = 0,\ \omega = k). \tag{6.9}$$

This solution corresponds to neutrino motion parallel to the z axis of the magnetic universe in the $+z$ direction. The solution is an eigenstate of $\sigma_3^{AB'}$ given in (6.7) with eigenvalue $+\bar{a}W/(2)^{1/2}$.

The nontrivial solution for neutrino motion in the $-z$ direction may be obtained, reworking the problem, by making the interchanges

$$l_\mu \leftrightarrow n_\mu, \quad m_\mu \leftrightarrow \bar{m}_\mu \tag{6.10}$$

in the NP formalism. (This corresponds to interchanging unprimed with primed quantities in GHP.) In this case $-z$ replaces z in (6.9) and the spin state now has an eigenvalue $-\bar{a}W/(2)^{1/2}$ of $\sigma_3^{AB'}$. In each case one obtains null results upon looking for neutrino motion in the opposite z direction.

For the case $\lambda \neq 0$ $Q(r)$ is given by (6.4). Equations (6.5) then become

$$\phi_1 = \frac{i(\omega + k)}{\bar{a}\sqrt{2}\sqrt{rW}} (A e^{i\lambda r} e^{-i\omega t} e^{ikz} + B e^{-i\lambda r} e^{-i\omega t} e^{ikz}) \tag{6.11a}$$

$$\phi_2 = \frac{i\lambda}{\bar{a}\sqrt{2}\sqrt{rW}} (A e^{i\lambda r} e^{-i\omega t} e^{ikz} - B e^{-i\lambda r} e^{-i\omega t} e^{ikz}). \tag{6.11b}$$

In the case $k = 0$ (upper sign $B = 0$, $A \equiv C$; lower sign $A = 0$, $B \equiv C$), Eqs. (6.11) simplify to

$$\phi_1 = \frac{i\omega}{\bar{a}\sqrt{2}\sqrt{rW}} C e^{i\omega(\pm r - t)} \quad (m = 0, k = 0), \tag{6.12a}$$

$$\phi_2 = \frac{\pm i\omega}{\bar{a}\sqrt{2}\sqrt{rW}} C e^{i\omega(\pm r - t)} \quad (m = 0, k = 0), \tag{6.12b}$$

or

$$\Phi_A = \frac{i\omega}{\bar{a}\sqrt{2}\sqrt{rW}} C \begin{pmatrix} 1 \\ \pm 1 \end{pmatrix} e^{i\omega(\pm r - t)}, \tag{6.13}$$

where the coefficients A, B, and C are constants.

The solution corresponds to neutrino motion directed radially (plus: outward; minus: inward) with respect to the axis of the magnetic universe. It is an eigenstate of $\sigma_1^{AB'}$ given in (6.7) with eigenvalue $+\bar{a}W/(2)^{1/2}$ $(-\bar{a}W/(2)^{1/2})$.

The most general meridian plane state given by (6.11) with A, B, k, and λ nonzero corresponds to a superposition of neutrino motions of the individual states in the subcases discussed above.

Since the meridian plane states ($m = 0$ states) are unbound, the constant prefactors appearing in the separate solutions (6.9) and (6.13) may be fixed by any sort of standard normalization, such as either box or delta function normalization.

The individual solutions (6.9) and (6.13) indicate a typical feature of the Cohen–Kegeles procedure, namely, that it leads directly to the neutrino spinor states. This is further discussed with a simple example in Appendix D.

Finally, we note that we are dealing with neutrinos of positive helicity, since in each of the subcases discussed above the neutrino's motion is in the same direction as its intrinsic angular momentum.

VII. THE CASE OF NONZERO ORBITAL ANGULAR MOMENTUM

In this section we discuss the nature of the solutions of (5.7) for the case $m \neq 0$. In subsection A behavior of these solutions near $r = 0$ and as $r \to \infty$ is given along with the asymptotic neutrino spinor field components. The asymptotic probability density of the neutrino spinor field is then given. It is deduced that bound neutrino states exist, and the reason for their existence is discussed. In subsection B use is made of the JWKB approximation to obtain an approximate expression for the energy eigenvalues and a comparison with the classical orbits of Ref. 7.

A. Behavior of *Q(r)* and the asymptotic neutrino spinor field

$Q(r)$ satisfies the differential equation

$$Q''(r) + \mathscr{R}^2(r) Q(r) = 0, \tag{7.1}$$

where

$$\mathscr{R}^2(r) \equiv \lambda^2 - \frac{W^4 m^2}{r^2} - \frac{(3r^2 - 1)}{r^2} Wm \tag{7.2}$$

and $\lambda^2 = \omega^2 - k^2$, $W = 1 + r^2$.

Near the origin

$$Q(r) = a_0' r^{-m+1} \left(1 + \frac{[\lambda^2 - 4m(m + \frac{1}{2})]}{4(m - \frac{3}{2})} r^2 + \mathscr{O}(r^4) \right) + a_0 r^m \left(1 - \frac{[\lambda^2 - 4m(m + \frac{1}{2})]}{4(m + \frac{1}{2})} r^2 + \mathscr{O}(r^4) \right) \tag{7.3}$$

($r \to 0$, a_0', a_0 constants).

Asymptotic solutions are of the form

$$Q(r) \sim c_0 r^{-(m+3)} \exp\left[-\left(\frac{mr^4}{4} + mr^2 \right) \right] \times \left[1 - \frac{(\lambda^2 + 8m)}{4mr^2} + \mathscr{O}\left(\frac{1}{r^4} \right) \right] + d_0 r^m \exp\left(\frac{mr^4}{4} + mr^2 \right) \left[1 + \frac{\lambda^2}{4mr^2} + \mathscr{O}\left(\frac{1}{r^4} \right) \right] \tag{7.4}$$

($r \to \infty$, c_0, d_0 constants, $m \neq 0$).

Physically acceptable solutions must have $d_0 = 0$ in (7.4) and $a_0' = 0$ in (7.3) so that the probability density (5.12) remains finite in the domain $0 \leqslant r < \infty$ when $m > 0$. For the case $m<0$, we must have $c_0=0$ in (7.4) and $a_0'=0$ in (7.3).

The asymptotic neutrino spinor field (5.8) for the case $m>0$ is

$$\phi_1 \sim \frac{i(\omega+k)}{\bar{a}\sqrt{2}\sqrt{rW}} r^{-(m+3)} \exp\left[-\left(\frac{mr^4}{4}+mr^2\right)\right] \times e^{-i\omega t} e^{im\phi} e^{ikz} \quad (m>0, r\to\infty), \tag{7.5a}$$

$$\phi_2 \sim \frac{-\sqrt{2}}{\bar{a}\sqrt{rW}} \frac{(mW^2+\frac{3}{2})}{r} r^{-(m+3)} \times \exp\left(-\left(\frac{mr^4}{4}+mr^2\right)\right) e^{-i\omega t} e^{im\phi} e^{ikz} \quad (m>0, r\to\infty). \tag{7.5b}$$

Equations (7.5) clearly indicate the existence of bound states for the case $m>0$.

The asymptotic neutrino probability density is, from (5.11) and (7.4) with $d_0=0$,

$$p(r) \sim \frac{\bar{a}^3|C_0|^2 W^2}{\sqrt{2}} \left[(\omega+k)^2 + 4\frac{(mW^2+\frac{3}{2})^2}{r^2}\right] r^{-2(m+3)} \times \exp\left[-\left(\frac{mr^4}{2}+2mr^2\right)\right] \quad (m>0, r\to\infty). \tag{7.6}$$

Results analogous to (7.5) and (7.6) are obtained for the case $m<0$.

The strong exponential damping factor in (7.6) and the condition $a_0'=0$ in (7.3) guarantees that the condition (5.12) is satisfied.

The existence of bound states in a consequence of the fact that the cylindrical magnetic universe is not asymptotically flat.[5] In fact, the rapid fall off of ϕ_1 and ϕ_2 can be attributed to the "tight belt" of the geometry of the cylindrical magnetic universe, since the $-m^2r^6$ behavior of $\mathscr{R}^2(r)$ in Eq. (7.1) as $r\to\infty$ can be traced back to the coefficient of $\partial^2\psi/\partial\phi^2$ in the unseparated wave equation (5.2) for the complex scalar function ψ.

B. Approximate eigenvalues and the classical orbits

Physically acceptable states exist if and only if

$$\lambda^2 > 4m(m+\tfrac{1}{2}), \quad m=\pm1,\pm2,\pm3, \tag{7.7}$$

For the case $m=1$ $\mathscr{R}^2(r)$ defined in (7.2) has one real positive zero which we denote by a. In this case the eigenvalues are approximately determined by the condition

$$\int_0^a [\lambda^2-(r^6+4r^4+9r^2+6)]^{1/2}dr = (n+\tfrac{3}{4})\pi \quad (m=1, n=0,1,2,\cdots). \tag{7.8}$$

For the cases $m>1, m<0$, $\mathscr{R}^2(r)$ has two real positive zeros which we denote by a and b where $a<b$. In this case the eigenvalues are approximately determined by the formula

$$\int_a^b \left(\lambda^2 - \frac{W^4m^2}{r^2} - \frac{(3r^2-1)}{r^2}Wm\right)^{1/2} dr = (n+\tfrac{1}{2})\pi \tag{7.9}$$

$$(m>1 \text{ or } m<0,\ n=0,1,2,\cdots).$$

Appendix B contains a summary of the JWKB functions for (7.1), their regions of validity, and the eigenvalue spectra of λ for the cases $m=1$ and $m\neq0$ or 1.

If we define

$$E\equiv\hbar\omega, \quad P\equiv\hbar k, \quad L\equiv\hbar m, \quad U\equiv\hbar\lambda, \tag{7.10}$$

where $\hbar$ is Planck's constant divided by 2π, then it is shown in Appendix C that the surfaces of constant phase of the neutrino spinor field move along trajectories determined from

$$t=\frac{E}{2} \times \int_{r_0}^{r} \frac{dx}{[U^2x-(1+x)^4L^2-(3x-1)(x+1)\hbar L]^{1/2}}, \tag{7.11a}$$

$$Z=\frac{P}{2} \times \int_{r_0}^{r} \frac{dx}{[U^2x-(1+x)^4L^2-(3x-1)(x+1)\hbar L]^{1/2}}, \tag{7.11b}$$

$$\phi=\frac{1}{2}\int_{r_0}^{r}\frac{1}{x} \times \frac{[(1+x)^4L+\frac{1}{2}(3x-1)(x+1)\hbar]dx}{[U^2x-(1+x)^4L^2-(3x-1)(x+1)\hbar L]^{1/2}} \quad (L\neq0), \tag{7.11c}$$

where r_0 is a suitable point in the r plane.

The trajectories of a classical point particle (without helicity) in the magnetic universe are given by[7]

$$t=\frac{E}{2}\int_{r_0}^{r}\frac{dx}{[U^2x-(1+x)^4L^2]^{1/2}}, \tag{7.12a}$$

$$Z=\frac{P}{2}\int_{r_0}^{r}\frac{dx}{[U^2x-(1+x)^4L^2]^{1/2}}, \tag{7.12b}$$

$$\phi=\frac{L}{2}\int_{r_0}^{r}\frac{1}{x}\frac{(1+x)^4dx}{[U^2x-(1+x)^4L^2]^{1/2}}. \tag{7.12c}$$

As explained in Appendix C, the principal difference between Eqs. (7.11) and (7.12) is the presence of a precession not present classically. To first order in $\hbar$, the neutrino orbits precess (as compared with the classical orbits) with frequency

$$\Omega_\nu = \frac{\hbar}{2Er^2}(3r^2-1)(r^2+1) \tag{7.13}$$

provided $L\neq0$. (7.13) is derived in Appendix C.

VIII. THE SCALAR FIELD IN A CYLINDRICAL MAGNETIC UNIVERSE

In this section we determine the wave equation for a zero rest mass particle without helicity in a cylindrical magnetic universe using the GHP formalism as outlined in Ref. 10. We find that the precessional corrections discussed in the preceding section are absent. This shows that these are due entirely to the helicity of the neutrino.

The wave equation for the scalar field $\bar{\psi}$ has the form (3.6),

$$(A_0 + \mathring{A}_0)\bar{\psi}=0, \tag{8.1}$$

where the operators A_0 and A_0^* are defined in (3.7).

Using (4.6b) together with the spin coefficients (4.5a) and (4.5b) we get

$$A_0 = D'D, \tag{8.2}$$

$$A_0^* = -\delta'\delta + \frac{1}{\bar{a}r\sqrt{2}}\frac{(1-r^2)}{W^2}\delta + \frac{\sqrt{2}\,r/\bar{a}}{W^2}(\delta+\delta'), \tag{8.3}$$

$W = 1 + r^2$, and $\bar{a}$ is a scale length of the magnetic universe defined in Ref. 5.

Use of (4.6a) then gives

$$\left(\frac{\partial^2}{\partial r^2} + \frac{1}{r}\frac{\partial}{\partial r} - \frac{\partial^2}{\partial t^2} + \frac{\partial^2}{\partial z^2} + \frac{W^4}{r^2}\frac{\partial^2}{\partial \phi^2}\right)\bar{\psi} = 0. \tag{8.4}$$

Setting

$$\bar{\psi}(t,r,\phi,z) = R_0(r)e^{-i\omega t}e^{im\phi}e^{ikz} \tag{8.5}$$

we get

$$\frac{d^2R_0}{dr^2} + \frac{1}{r}\frac{dR_0}{dr} + \left[(\omega^2 - k^2) - \frac{W^4m^2}{r^2}\right]R_0(r) = 0 \tag{8.6}$$

for the radial equation.

For the meridian plane ($m = 0$) orbits the solutions of (8.6) are of the form

$$R_0(r) = AJ_0(\lambda r) + BY_0(\lambda r) \quad (\lambda \neq 0), \tag{8.7a}$$

$$R_0(r) = C_1 \ln r + C_2 \quad (\lambda = 0), \tag{8.7b}$$

where J_0 and Y_0 denote Bessel functions of order zero, $\lambda^2 \equiv \omega^2 - k^2$, and A, B, C_1 and C_2 are constants.

To employ the JWKB approximation we make the substitution

$$R_0(r) = Q_0(r)/\sqrt{r} \tag{8.8}$$

to obtain a differential equation with no first derivative term:

$$Q_0''(r) + \left(\lambda^2 - \frac{W^4m^2}{r^2} + \frac{1}{4r^2}\right)Q_0(r) = 0. \tag{8.9}$$

At this point we note that the precessional correction to the classical neutrino orbits as derived in Appendix C is due to the appearance of the term linear in m in (5.7). No such term appears in (8.9) and it should therefore be concluded that zero rest mass, zero helicity particles do not have a precessional correction in the classical limit. This establishes that the differences between Eqs. (7.11) and (7.12) are due entirely to the helicity of the neutrino.

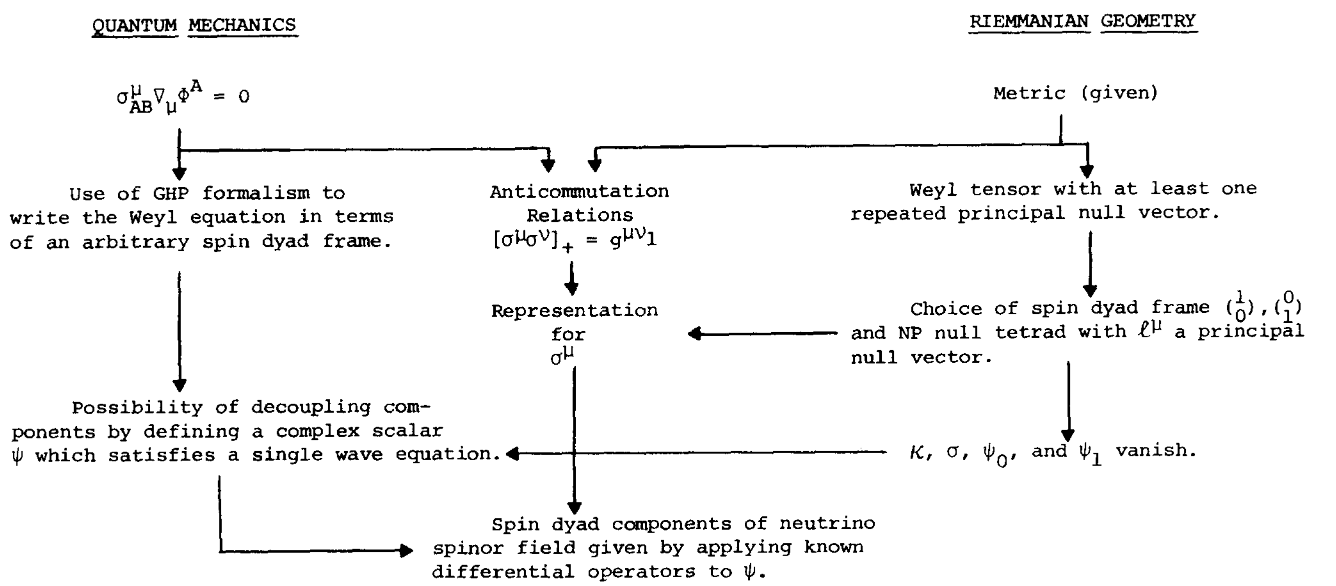
IX. CONCLUSIONS

In this section we summarize the contents of this paper and discuss the major results. We first give a condensed review of the formal steps which comprise the Cohen–Kegeles (CK) procedure. We then discuss briefly the momentum operator in curved space. This leads to a motivation and validation of the wave function separation (5.3). Three zero rest-mass fields are then compared: the scalar, neutrino, and Maxwell fields. We compare the radial equations for each of these cases and deduce the presence or absence of orbital precession. A detailed summary of the results for neutrinos is subsequently given along with an enumeration of the differences between the CK and Teukolsky [11] procedures.

The CK procedure may be summarized as in Fig. 1 where arrows indicate the logical order of the formal steps which lead to the solution of the Weyl equation in the generalized Goldberg–Sachs class of spacetimes. We explicitly note that we need not have a vacuum spacetime for the CK procedure to be applicable (as indeed the cylindrical magnetic universe is not a vacuum spacetime), but merely that the spacetime admit a congruence of shear-free null geodesics along the repeated principal null vector of the Weyl tensor (a "principally anastigmatic" spacetime). Two other conditions must also be met, namely, that the neutrinos are not present in such strength as to affect the background geometry appreciably and that the only influence on the neutrinos is gravity.

The CK procedure is discussed in detail in Sec. II above and a simple example of the procedure outlined in Fig. 1 is given in Appendix D below.

A discussion of elementary particles moving in an arena



of a curved spacetime background does not seem complete without reference being made as to what one uses for momentum operators under such conditions. We make use of a prescription discussed by Pauli [19] for nonrelativistic scalar particles which may readily be extended to relativistic particles with spin or nonzero helicity.

The following assumptions are made:

(1) $p_\mu x^\nu - x^\nu p_\mu = -i\hbar\delta^\nu_\mu$,

(2) $\langle\Psi, p_\mu\Psi\rangle = \overline{\langle\Psi, p_\mu\Psi\rangle}, \quad \mu,\nu = 1,2,3,$

where x^μ and p_μ are conjugate coordinates and momenta, respectively.

δ^ν_μ is the Kronecker delta and the brackets in (2) refer to the expectation value of the operators p_μ (for any given state Ψ).

$\Psi\overline{\Psi}$ is interpreted as the probability density for locating a particle with wave function Ψ. However, the probability density p is not invariant under coordinate changes since the volume element dV is not. It can be shown [20] that DdV, where $D = (|g|)^{1/2}$ and $|g|$ is the determinant of the metric, is an invariant. Thus if one requires that $\Psi\overline{\Psi}$ be an invariant we should have in the general case

$$p = D\Psi\overline{\Psi}. \tag{9.1}$$

[In the case of a two-component neutrino field

$$p = 2D\overline{\Phi}_{B'}\sigma_0^{AB'}\Phi_A, \tag{9.2}$$

where $\sigma_0^{AB'}$ for the tetrad used in this paper is given in (6.7).]

The expectation of p_μ in the state ψ is then

$$\langle\Psi, p_\mu\Psi\rangle = \int D\overline{\Psi}(p_\mu\Psi)dV, \tag{9.3}$$

where the integration is to be taken over all space.

Requiring that (2) hold and assuming the absence of a vector potential (which would be present with an electromagnetic interaction) one finds that for suitable boundary conditions [19] imposed on the state Ψ,

$$p_\mu = -i\hbar\frac{1}{\sqrt{D}}\frac{\partial}{\partial x^\mu}\sqrt{D}\,\Psi \quad (\mu = 1,2,3). \tag{9.4}$$

The literature on momentum operators (and Hermitian operators in general) in quantum mechanics in curvilinear coordinates and curved space is very great. Podolsky [21] arrives at (9.4) in a manner different than Pauli,

Assuming the prescription (9.4) for p_μ holds in the magnetic universe with metric given by (4.1), one finds that

$$p_\phi = -i\hbar\frac{\partial}{\partial\phi} \tag{9.5a}$$

$$p_z = -i\hbar\frac{\partial}{\partial z}. \tag{9.5b}$$

If we further assume that

$$p_t \equiv -i\hbar\frac{\partial}{\partial t} \tag{9.5c}$$

is equivalent to the Hamiltonian, particle states in the magnetic universe must be of the form

$$\psi(t,r,\phi,z) = R_h(r)e^{-i\omega t}e^{im\phi}e^{ikz} \tag{9.6}$$

in order to be simultaneous eigenstates of the mementa p_t, p_ϕ, and p_z. This justifies (5.3).

The correctness of (9.6) or (5.3) is borne out further when on examines the classical limit of scalar particles, neutrinos, and photons. We have primarily dealt with neutrinos in the paper but we have also briefly considered scalar particles in Sec. VIII above. This was done to verify that the precessional corrections in the neutrino orbits (for nonzero m) is due to their nonvanishing helicity. Indeed, scalar particles show no such corrections.

Below we give the equation for the radial functions $R_h(r)$ ($h = 0, \frac{1}{2}, 1$, for scalar particles, neutrinos, and Maxwell fields, respectiely.) If $h \neq 0$, a term linear in m appears in the radial equation in the coefficient of the last term, and as explained in Sec. VIII, this leads to precessional corrections in the classical limit.

Using Eqs. (4.5), (4.6), and (4.7) in the general CK equation (3.9) and applying separation of variables we find that the radial function satisfies

$$\frac{d^2R_h}{dr^2} - \left[\frac{(4h-1)r^2-1}{rW}\right]\frac{dR_h}{dr} + \left[\lambda^2 - \frac{W^4m^2}{r^2} - \frac{2hm(3r^2-1)}{r^2} + \frac{\mathscr{P}_h(r)}{r^2W^2}\right]R_h = 0, \tag{9.7}$$

where $\lambda^2 \equiv \omega^2 - k^2$, $W = 1 + r^2$, and $\mathscr{P}_h(r) = 3h^2r^4 - 2h(h+2)r^2 - h^2$. (In this notation $-h$ replaces s).

For neutrinos, $h = \frac{1}{2}$: $\mathscr{P}_{1/2}(r) = \frac{1}{4}(3r^4 - 10r^2 - 1)$.

For Maxwell fields, $h = 1$, in a (nonelectromagnetic) "plasm of index 2" universe: [8] $\mathscr{P}_1(r) = 3r^4 - 6r^2 - 1$.

As is evident from Eq. (9.7) and the discussion in Appendix C, the precessional frequency for photons is twice that for neutrinos as given in (7.13).

For meridian plane orbits ($m = 0$) (9.7) becomes

$$\frac{d^2R_h}{dr^2} - \left(\frac{(4h-1)r^2-1}{rW}\right)\frac{dR_h}{dr} + \left[\lambda^2 + \frac{\mathscr{P}_h(r)}{r^2W^2}R_h\right] = 0. \tag{9.8}$$

If in (9.8) we make the substitution $R_h = W^hG_h$ (no sum implied) we find that G_h satisfies

$$\frac{d^2G_h}{dr^2} + \frac{1}{r}\frac{dG_h}{dr} + \left(\lambda^2 - \frac{h^2}{r^2}\right)G_h = 0. \tag{9.9}$$

This is Bessel's equation of order h. If $\lambda \neq 0$, the general solution of (9.8) is therefore

$$R_h = W^hZ_h(\lambda r) \quad \text{(no sum implied)}, \tag{9.10}$$

where Z_h denotes a general cylinder function of order h.

For the case $m = 0, \lambda = 0$ the general solution of (9.8) is (8.7b),

$$R_0(r) = C_1\ln(r) + C_2 \tag{9.11a}$$

or

$$R_h(r) = C_3(Wr)^h + C_4(W/r)^h \quad (h \neq 0), \tag{9.11b}$$

where C_1, C_2, C_3, and C_4 are constants.

Equations (9.11) represent particle motion parallel to the magnetic or plasm universe z axis.

The results of the paper for neutrinos specifically may be summarized as follows. The energy eigenvalue spectrum is continuous for the meridian plane case ($m = 0$) and discrete otherwise ($m \neq 0$). The existence of bound states for the case $m \neq 0$ is due to the fact that the magnetic universe is not asymptotically flat. The fact that a spacetime which *is* asymptotically flat cannot in general give rise to bound neutrino states was emphasized by Brill and Wheeler.[22] In contrast the fact that a spacetime is not asymptotically flat does not, of course, *guarantee* that all neutrino states will be bound, as is evidenced by the meridian plane orbits.

In comparing the neutrino orbits in the classical limit with the orbits calculated from a completely classical standpoint one discovers a precession for $m \neq 0$ that is not present classically. As explained above, this precession is due to the nonvanishing helicity of the neutrino.

Since the magnetic universe is Petrov type D (cf. Appendix A) one may also solve the Weyl equation in the spacetime using a procedure given by Teukolsky.[11] The CK procedure turns out to be somewhat more straightforward as is generally the case (cf. Appendix D). By way of comparison, we list below the four major differences between the Cohen–Kegles and Teukolsky procedure.

(1) The CK procedure is valid in all principally anastigmatic spacetimes whereas the Teukolsky procedure is valid only in type D spacetimes.[11]

(2) The CK procedure solves explicitly for the spinor components as well as their spacetime dependence whereas Teukolsky's procedure yields only the spacetime dependence explicitly and requires another step to determine the ratio of the spinor components. This is illustrated in Appendix D below.

(3) The CK procedure involves the solution of one differential equation whereas the Teukolsky procedure involves the solution of two. We note that the two Teukolsky equations may be identical, or the second may be obtainable from the first by simply changing the sign of a parameter. In general, however, this may not be the case.

(4) Because of the nature of their derivation, the solution space of the Teukolsky equations is generally larger than that of the original Weyl equations. For this reason solutions of the Teukolsky equations must be substituted back into the Weyl equations to check on their validity. This is not necessary with the CK procedure.

APPENDIX A

In this appendix the NP Weyl tensor components are defined and calculated using the tetrad (4.4). l_μ and n_μ are shown to be repeated principal null vectors of the Weyl tensor classifying the cylindrical magnetic universe as Petrov type D.[9]

The NP dyad components of the Weyl tensor $C_{\alpha\beta\gamma\delta}$ are given individual symbols as follows:[9]

$$\begin{aligned}
\Psi_0 &= -C_{\alpha\beta\gamma\delta} l^\alpha m^\beta l^\gamma m^\delta,\\
\Psi_1 &= -C_{\alpha\beta\gamma\delta} l^\alpha n^\beta l^\gamma m^\delta,\\
\Psi_2 &= -\tfrac{1}{2} C_{\alpha\beta\gamma\delta}(l^\alpha n^\beta l^\gamma n^\delta - l^\alpha n^\beta m^\gamma \bar{m}^\delta),\\
\Psi_3 &= -C_{\alpha\beta\gamma\delta} \bar{m}^\alpha n^\beta l^\gamma n^\delta,\\
\Psi_4 &= -C_{\alpha\beta\gamma\delta} \bar{m}^\alpha n^\beta \bar{m}^\gamma n^\delta.
\end{aligned} \tag{A1}$$

If the only nonvanishing NP dyad component of the Weyl tensor is Ψ_2, then l^μ and n^μ are repeated principal null vectors of the Weyl tensor.[2,9] This is now shown to be the case.

The coordinate frame components of the Weyl tensor for the cylindrical magnetic universe are

$$\begin{aligned}
C_{0101} &= 2\bar{a}^2(r^2-1), \quad C_{0202} = 2\bar{a}^2\frac{r^4(r^2-1)}{(1+r^2)^4},\\
C_{0303} &= -4\bar{a}^2(r^2-1), \quad C_{1212} = 4\bar{a}^2\frac{r^4(r^2-1)}{(1+r^2)^4},\\
C_{1313} &= -2\bar{a}^2(r^2-1), \quad C_{3232} = -2\bar{a}^2\frac{r^4(r^2-1)}{(1+r^2)^4}.
\end{aligned} \tag{A2}$$

All others not obtainable from the symmetries of the Weyl tensor vanish. Use of (4.4) and (A2) in (A1) gives by straightforward calculation:

$$\begin{aligned}
\Psi_0 &= 0,\\
\Psi_1 &= 0,\\
\Psi_2 &= \frac{2}{\bar{a}^2}\frac{(r^2-1)}{(1+r^2)^4},\\
\Psi_3 &= 0,\\
\Psi_4 &= 0.
\end{aligned} \tag{A3}$$

This establishes the desired results, namely that l^μ and

TABLE I.

	$m = 1$	$m \neq 0,1$
Restriction on λ^2	> 6	$> 4m(m+\frac{1}{2})$
Real Positive Zeros of $\mathscr{R}(r)$	One called "a"	Two called "a" and "b" with $a < b$
JWKB functions	$\frac{1}{\lvert\mathscr{R}(r)\rvert^{1/2}}\exp\left(-\left\lvert\int_a^r \mathscr{R}(\xi)d\xi\right\rvert\right), (r \gg a)$ $\frac{2}{\lvert\mathscr{R}(r)\rvert^{1/2}}\cos\left[\left\lvert\int_a^r \mathscr{R}(\xi)d\xi\right\rvert - \frac{\pi}{4}\right], 0 \leqslant r < a$	$\frac{1}{\lvert\mathscr{R}(r)\rvert^{1/2}}\exp\left(-\left\lvert\int_b^r \mathscr{R}(\xi)d\xi\right\rvert\right), \quad (r \gg b)$ $\frac{2}{\lvert\mathscr{R}(r)\rvert^{1/2}}\cos\left[\left\lvert\int_a^r \mathscr{R}(\xi)d\xi\right\rvert - \frac{\pi}{4}\right], a < r < b$
Eigenvalue spectrum of λ	Discrete (7.8) $\int_0^a \mathscr{R}(r)dr = (n+\frac{3}{4})\pi$ $n = 0,1,2,\cdots$	Discrete (7.9) $\int_a^b \mathscr{R}(r)dr = (n+\frac{1}{2})\pi,$ $n = 0,1,2,\cdots$
Region of validity of the JWKB functions	Everywhere	Everywhere except $r = 0$

n^{μ} are repeated principal null directions of the Weyl tensor and that the magnetic universe is Petrov type D.

Further, since κ, σ and κ', σ' are all zero we also have that l^{μ} and n^{μ} are the tangent vectors of shear-free null geodesics. Since both the algebraic special condition and the congruence condition are satisfied, it follows that the magnetic universe is principally anastigmatic.

APPENDIX B

Table I contains a summary of the JWKB functions for (7.1), their regions of validity, and the eigenvalue spectra of λ. The table lists this information, together with the number of real positive zeros of $\mathscr{R}(r)$ [defined in (7.2)], for the cases $m = 1$ and $m \neq 0$ or 1. (For the case $m = 0$, the exact solution is given in Sec. VI).

APPENDIX C

In this appendix we determine the orbits of the surfaces of constant phase of the neutrino spinor field in the short wavelength limit. These are the orbits the neutrino would have if it were a well-localized point particle. We show that the orbits are similar to those of a classical point particle (without helicity) but possess a precession not found classically when $m \neq 0$.

We define the following:

$$E \equiv \hbar\omega, \quad P \equiv \hbar k, \quad L \equiv \hbar m, \; U \equiv \hbar\lambda. \tag{C1}$$

A suitable approximate solution to (7.1) in the short wavelength limit, is the standard JWKB function

$$Q(r) = \frac{1}{|\mathscr{R}(r)|^{1/2}} \exp\left(i \int^{r} \mathscr{R}(\xi) d\xi\right) \tag{C2}$$

provided we are not too close to a zero of $\mathscr{R}(r)$ (a classical turning point).

We can also write

$$\int^{r} \mathscr{R}(\xi) d\xi = \frac{1}{2} \int^{r^2} \frac{\mathscr{R}(\sqrt{\xi}\,)}{\sqrt{\xi}} d\xi \tag{C3}$$

as can be verified by differentiating both sides of this expression with respect to r.

Using (7.2) and (C1) in (C3) we get, changing the dummy variable of integration from ξ to x,

$$\int^{r} \mathscr{R}(\xi) d\xi = \frac{1}{2\hbar} \int^{r^2} \frac{1}{x} [U^2 x - (1+x)^4 L^2 - (3x-1)(1+x)\hbar L\,]^{1/2} dx. \tag{C4}$$

The dyad components of the neutrino spinor field are given by (5.8a)

$$\text{(5.8a)} \quad \phi_1 = \frac{i}{\bar{a}\sqrt{2}} \frac{(\omega + k)}{\sqrt{rW}} Q(r) e^{-i\omega t} e^{im\phi} e^{ikz}, \tag{C5a}$$

$$\text{(5.8b)} \quad \phi_2 = \frac{i}{\bar{a}\sqrt{2}} \frac{1}{\sqrt{rW}} \left(\frac{Q'(r)}{Q(r)} - \frac{mW^2}{r} \right) \times Q(r) e^{-i\omega t} e^{im\phi} e^{ikz}. \tag{C5b}$$

Surfaces of constant phase of the spinor field move along trajectories determined from

$$i \arg Q(r) - i\omega t + im\phi + ikz = S, \tag{C6}$$

where S is a constant.

From the approximate solution to (7.1) (C2), we obtain, using (C4),

$$\arg Q(r) = \frac{1}{2\hbar} \int_{r_0^2}^{r^2} \frac{1}{x} [U^2 x - (1+x)^4 L^2 - (3x-1)(1+x)\hbar L\,]^{1/2} dx, \tag{C7}$$

where r_0 is a suitable point in the r plane.

The time integral can be obtained from (C6) by considering the "constructive interference of ideal wave trains" and setting $\partial S / \partial E = 0$:[7]

$$\frac{\partial S}{\partial E} = 0 = -\frac{t}{\hbar} + \frac{1}{2\hbar} \int_{r_0^2}^{r^2} \frac{E dx}{[U^2 x - (1+x)^4 L^2 - (3x-1)(x+1)\hbar L\,]^{1/2}} \tag{C8}$$

or

(7.11a)

$$t = \frac{E}{2} \int_{r_0^2}^{r^2} \frac{dx}{[U^2 x - (1+x)^4 L^2 - (3x-1)(x+1)\hbar L\,]^{1/2}}. \tag{C9}$$

Similarly, the z and ϕ integrals are obtained by differentiating (C6) with respect to P and L, respectively:

$$\frac{\partial S}{\partial P} = 0 = \frac{z}{\hbar} - \frac{1}{2\hbar} \times \int_{r_0^2}^{r^2} \frac{P dx}{[U^2 x - (1+x)^4 L^2 - (3x-1)(x+1)\hbar L\,]^{1/2}} \tag{C10}$$

or

(7.11b)

$$z = \frac{P}{2} \int_{r_0^2}^{r^2} \frac{dx}{[U^2 x - (1+x)^4 L^2 - (3x-1)(x+1)\hbar L\,]^{1/2}}. \tag{C11}$$

$$\frac{\partial S}{\partial L} = 0 = \frac{\phi}{\hbar} - \frac{1}{2\hbar} \times \int_{r_0^2}^{r^2} \frac{1}{x} \frac{[(1+x)^4 L + \frac{1}{2}(3x-1)(x+1)\hbar] dx}{[U^2 x - (1+x)^4 L^2 - (3x-1)(x+1)\hbar L\,]^{1/2}} \tag{C12}$$

or

$$\text{(7.11c)} \; \phi = \frac{1}{2} \times \int_{r_0^2}^{r^2} \frac{1}{x} \frac{[(1+x)^4 L + \frac{1}{2}(3x-1)(x+1)\hbar] dx}{[U^2 x - (1+x)^4 L^2 - (3x-1)(x+1)\hbar L\,]^{1/2}} \quad (L \neq 0). \tag{C13}$$

Equations (C9), (C11), and (C13) are the equations determining the neutrino motion in the short wavelength limit. If in (C9), (C11), and (C13) we have, between the limits of integration,

$$U^2 x > (1+x)^4 L^2 \gg (3x-1)(x+1)\hbar L, \tag{C14}$$

then we can write

$$(7.12a) \quad t = \frac{E}{2}\int_{r_0^2}^{r^2} \frac{dx}{[U^2x - (1+x)^4L^2]^{1/2}}, \qquad \text{(C15a)}$$

$$(7.12b) \quad z = \frac{P}{2}\int_{r_0^2}^{r^2} \frac{dx}{[U^2x - (1+x)^4L^2]^{1/2}}, \qquad \text{(C15b)}$$

and

$$(7.12c) \quad \phi = \frac{L}{2}\int_{r_0^2}^{r^2} \frac{(1+x)^4dx}{x[U^2x - (1+x)^4L]^{1/2}}. \qquad \text{(C15c)}$$

Equations (C15) correspond to the motion of a classical point particle in the magnetic universe.

The major difference between Eqs. (C9), (C11), and (C13) and (C15) is the presence of a precession not present classically. To demonstrate this, we calculate the coordinate-time rate of change of the angle ϕ given by (C13).

From (C13) we get

$$\frac{d\phi}{dt} = \frac{d(r^2)}{dt}\frac{d}{d(r^2)}\frac{1}{2} \times \int_{r_0^2}^{r^2} \frac{1}{x}\frac{[(1+x)^4L + \frac{1}{2}(3x-1)(x+1)\hbar]dx}{[U^2x - (1+x)^4L^2 - (3x-1)(x+1)\hbar L]^{1/2}} \qquad \text{C16)}$$

or

$$\frac{d\phi}{dt} = \frac{1}{2}\frac{d(r^2)}{dt}\frac{1}{r^2} \times \frac{[(1+r^2)^4L^2 + \frac{1}{2}(3r^2-1)(r^2+1)\hbar]}{[U^2r^2 - (1+r^2)L^2 - (3r^2-1)(r^2+1)\hbar L]^{1/2}}. \qquad \text{(C17)}$$

From (C9) we obtain

$$\frac{dt}{d(r^2)} = \frac{E}{2}[U^2r^2 - (1+r^2)^4L^2 - (3r^2-1)(r^2+1)\hbar L]^{-1/2}. \qquad \text{(C18)}$$

Using the reciprocal of (C18) in (C17) we get

$$\frac{d\phi}{dt} = \frac{L}{E}\frac{(1+r^2)^4}{r^2} + \frac{\hbar}{2E}\frac{(3r^2-1)(r^2+1)}{r^2} \quad (L\neq 0). \qquad \text{(C19)}$$

The first term of (C19) is the result one obtains classically; the additional term is a precessional correction to the classical orbit. To first order in $\hbar$, the neutrino orbits precess (as compared with the classical orbits) with frequency

$$(7.13) \quad \Omega_\nu = \frac{\hbar}{2Er^2}(3r^2-1)(r^2+1) \qquad \text{(C20)}$$

provided $L\neq 0$.

The precession is due to the helicity of the neutrino, since a zero rest mass, zero helicity particle has no such corrections to the orbit in the classical limit as explained in Sec. VIII above.

APPENDIX D

In this appendix we solve the Weyl equation in flat space with Cartesian coordinates using the Cohen–Kegeles procedure. The purpose for doing this, besides for purely pedagogical reasons is twofold: first, it demonstrates the straightforward simplicity and inherent versatility of the Cohen–Kegeles procedure, and second, it demonstrates a further difference between this procedure and Teukolsky's procedure. This was mentioned in Sec. II above. These differences will be summarized below.

Here we simply have the line element

$$ds^2 = dt^2 - dx^2 - dy^2 - dz^2. \qquad \text{(D1)}$$

We wish to solve the Weyl equation

$$\sigma^\mu_{AB'}\nabla_\mu\Phi^A = 0. \qquad \text{(D2)}$$

Here, as throughout this paper, Greek indices (referring to coordinate space) run from 0 to 3 and capital Latin indices (referring to spin space) take on the values 1 and 2 (or 1′ and 2′.) Summation is understood for all repeated indices.

In order to solve (D2) we must first choose a representation for the matrices $\sigma^\mu_{AB'}$. In the Cohen–Kegeles procedure this is achieved through the choice of the "spin dyad frame" $\binom{1}{0}$, $\binom{0}{1}$ along with an NP null tetrad l_μ, n_μ, m_μ, $\bar{m}_\mu$. For curved space, which is restricted to be algebraically special, l_μ must be a repeated principal null vector of a Weyl tensor. For the case under consideration here this condition is trivially satisfied for any reasonable l_μ.

The standard representation for the $\sigma^{AB'}_\mu$ are the Pauli matrices ($\mu = 1, 2, 3$) and the identity matrix ($\mu = 0$.) To achieve this representation we take the tetrad to be

$$l_\mu = \frac{1}{\sqrt{2}}\{1,0,0,1\}, \quad n_\mu = \frac{1}{\sqrt{2}}\{1,0,0,-1\},$$
$$m_\mu = \frac{1}{\sqrt{2}}\{0,1,-i,0\}, \quad \bar{m}_\mu = \frac{1}{\sqrt{2}}\{0,1,i,0\}. \qquad \text{(D3)}$$

Then, since we have *in the spin dyad* frame,[9]

$$l_\mu = \sigma^{11'}_\mu, \quad n_\mu = \sigma^{22'}_\mu, \quad m_\mu = \sigma^{12'}_\mu, \quad \bar{m}_\mu = \sigma^{21'}_\mu, \qquad \text{(D4)}$$

$$\sigma^{ab'}_0 = \frac{1}{\sqrt{2}}\begin{pmatrix}1 & 0\\ 0 & 1\end{pmatrix}, \quad \sigma^{ab'}_1 = \frac{1}{\sqrt{2}}\begin{pmatrix}0 & 1\\ 1 & 0\end{pmatrix},$$
$$\sigma^{ab'}_2 = \frac{1}{\sqrt{2}}\begin{pmatrix}0 & -i\\ i & 0\end{pmatrix} \quad \sigma^{ab'}_3 = \frac{1}{\sqrt{2}}\begin{pmatrix}1 & 0\\ 0 & -1\end{pmatrix}, \qquad \text{(D5)}$$

which are essentially the unit matrix and the three standard Pauli matrices. (Lower case latin indices refer to the spin dyad frame).

All the NP spin coefficients vanish. The Cohen–Kegeles equation (5.1) becomes

$$(D'D - \delta\delta')\psi(t,x,y,z) = 0, \qquad \text{(D6)}$$

where

$$D' = n^\mu\frac{\partial}{\partial x^\mu}, \quad D = l^\mu\frac{\partial}{\partial x^\mu}, \quad \delta = m^\mu\frac{\partial}{\partial x^\mu}, \quad \delta' = \bar{m}^\mu\frac{\partial}{\partial x^\mu}, \qquad \text{(D7)}$$

so that (D6) becomes, using (D3),

$$\nabla^2\psi = \frac{\partial^2\psi}{\partial t^2}. \qquad \text{(D8)}$$

According to (2.5), (2.6), and (4.6b) the spin dyad components of the neutrino spinor field are

$$\phi_1 = -D\psi, \quad \phi_2 = -\delta'\psi, \tag{D9}$$

or, using (D3) and (D7)

$$\phi_1 = \frac{1}{\sqrt{2}}\left(-\frac{\partial}{\partial t} + \frac{\partial}{\partial z}\right)\psi, \quad \phi_2 = \frac{1}{\sqrt{2}}\left(\frac{\partial}{\partial x} + i\frac{\partial}{\partial y}\right)\psi. \tag{D10}$$

We seek plane wave solutions of (D8):

$$\psi(t,x,y,z) = Ae^{i(\mathbf{p}\cdot\mathbf{x} - \omega t)}, \tag{D11}$$

where A is a constant.

Substitution of (D11) into (D8) gives the relativistic dispersion relation

$$\omega^2 = \mathbf{p}^2 \tag{D12}$$

for zero rest mass particles.

Writing $\mathbf{p} = \{p_x, p_y, p_z\}$, (D10) becomes

$$\phi_1 = \frac{i}{\sqrt{2}}(\omega + p_z)\psi(t,x,y,z),$$
$$\phi_2 = \frac{1}{\sqrt{2}}(ip_x - p_y)\psi(t,x,y,z). \tag{D13}$$

If we have neutrino motion in the $+z$ direction, $p_x = p_y = 0$ and $p_z = \omega$ from (D12). Thus in this case (D13) becomes

$$\Phi_A = i\sqrt{2}\omega\psi\begin{pmatrix}1\\0\end{pmatrix}, \tag{D14}$$

which is seen to be an eigenstate of $\sigma_3^{AB'}$ of (D5) with eigenvalue $+1/(2)^{1/2}$.

For neutrino motion in the $-z$ direction, $p_x = p_y = 0$ and $p_z = -\omega$ $(\omega > 0)$ so that (D13) reduces to the null spinor field. Although this is a legitimate solution of the Weyl equation (D2) it is physically uninteresting.

A null result of this sort is encountered in the analysis of the meridian plane orbits of the neutrino in the cylindrical magnetic universe as discussed in Sec. V. As explained there, the nontrivial solution can be found by making the interchanges $l^\mu \leftrightarrow n^\mu$, $m^\mu \leftrightarrow \bar{m}^\mu$ in the NP null tetrad and reworking the probelm.

For the cases of neutrino motion parallel to and in the direction of the x axis ($p_y = p_z = 0, p_x = \omega$) and parallel to and in the direction of the y axis ($p_x = p_z = 0, p_y = \omega$) the neutrino spinor fields from (D13) are, respectively,

$$\Phi_A = \frac{i\omega}{\sqrt{2}}\psi\begin{pmatrix}1\\1\end{pmatrix} \tag{D15}$$

and

$$\Phi_A = \frac{\omega}{\sqrt{2}}\psi\begin{pmatrix}i\\-1\end{pmatrix}, \tag{D16}$$

which are, respectively, eigenstates of $\sigma_1^{AB'}$ and $\sigma_2^{AB'}$ of (D5) each with eigenvalue $+1/(2)^{1/2}$.

For this simple example Teukolsky's two equations [11] become for the same choice of tetrad,

$$\nabla^2\phi_a = \frac{\partial^2\phi_a}{\partial t^2} \quad (a = 1,2), \tag{D17}$$

which are equations directly involving the spinor field without the use of the "potential" ψ.

(D17) has the plane wave solution

$$\Phi_A = \begin{pmatrix}C_1\\C_2\end{pmatrix}e^{i(\mathbf{p}\cdot\mathbf{x} - \omega t)} \quad (C_1, C_2 \text{ constants}), \tag{D18}$$

which is essentially the Cohen–Kegeles result (D13). The Cohen–Kegeles result, however, automatically gives the solution in spin space in terms of the particle momentum $\mathbf{p}$. Teukolsky's procedure gives (D18), which gives only the spacetime dependence of the spinor field. The constants C_1 and C_2 remain to be determined under the restriction that Φ_A be an eigenstate of the helicity $\sigma^k_{AB'}p_k$ where here $k = 1,2,3$.

In conclusion we list the four major differences between the Cohen–Kegeles and Teukolsky procedures.

First, the Cohen–Kegeles procedure is valid in all algebrically special spacetimes whereas the Teukolsky procedure is valid only in type D spacetimes. [11]

Second, the Cohen–Kegeles procedure solves the problem directly in spin space as well as in coordinate space whereas the Teukolsky procedure does not.

Third, the Cohen–Kegeles procedure involves the solution of one differential equation whereas the Teukolsky procedure involves the solution of two. We note that the two Teukolsky equations may be identical (as in the above example) or the second may be obtainable from the first by simply changing the sign of a parameter. In general, however, this may not be the case.

Fourth, the Teukolsky equations are derived by applying specified NP operators to (2.2a) and (2.2b) and then adding these equations. [11] For this reason the solution space of the Teukolsky equations is generally larger than that of the original Weyl equations [Eqs. (2.2)]. It then becomes necessary to substitute solutions of the Teukolsky equations back into the Weyl equation to check on their validity. With the Cohen–Kegeles procedure this is not necessary.

Note added in proof: The completeness of the set of radial eignfunctions $R_{1/2}(r)$ for a given value of m may be demonstrated by substituting $R_{1/2}(r) \equiv WS(r)$ in Eq. (5.4). The result is

$$\frac{d}{dr}\left(rW\frac{dS}{dr}\right) - v(r)S(r) + \lambda^2 rWS(r) = 0 \quad (m \neq 0)$$

where

$$v(r) \equiv \frac{W^2 m}{r}(W^3 m + 3r^2 - 1) - \frac{1}{4rW}(3r^4 + 6r^2 - 1).$$

This differential equation, together with the B.C. $S(0) = 0$, $\lim_{r\to\infty} S(r) = 0$, constitutes a Sturm–Liouville problem on the interval $[0,\infty)$. The weight function is rW. Denoting the normalized eigenfunctions by $S^\lambda(r)$ we have that

$$\int_0^\infty rW\bar{S}^\lambda(r)S^{\lambda'}(r)\,dr = \delta^{\lambda\lambda'}.$$

ACKNOWLEDGEMENT

We wish to thank Professor Jürgen Ehlers for reading and commenting on the manuscript of this paper.

[1]J.M. Cohen and L.S. Kegeles, "Space–Time Perturbations," Phys. Lett. A **54**, 5 (1975).

[2]A.I. Janis and E.T. Newman, "Structure of Gravitational Sources," J. Math. Phys. **6**, 902 (1965).

[3]J. Wainright, "A class of Algebraically Special Perfect Fluid Space–Times," Commun. Math. Phys. **17**, 42 (1970).

[4]J. Plebański, *Spinors, Tetrads, and Forms* (Centro de Investigación y Estudios Avanzados del I.P.N., Mexico City, 1975). A detailed treatment of the optics of congruences of null geodesics is given in Sec. VI. 4 of this book.

[5]M.A. Melvin, "Pure Magnetic and Electric Geons," Phys. Lett. **8**, 65 (1964).

[6]M.A. Melvin, "Dynamics of Cylindrical Electromagnetic Universes," Phys. Rev. **139**, B225 (1965).

[7]M.A. Melvin and J.S. Wallingford, "Orbits in a Magnetic Universe," J. Math. Phys. **7**, 333 (1956).

[8]The simple perturbative theory may not be applied to photon fields (helicity 1) in a literal magnetic universe. In this case the background metric of the same form is ascribed instead to a nonelectromagnetic plasm of index 2 [5] so that the simple perturbative procedure applied for helicites $h = 0$ and $h = \frac{1}{2}$ also applies to $h = 1$. Whereas the CK method applies also when a test field is of the same nature as the background field, provided the latter is corrected for the perturbation cross term, the equations in this case are considerably more complicated. The case of radial electromagnetic perturbations in the literal magnetic universe was worked out in detail by one of the authors in an early paper in connection with a stability analysis of the magnetic universe. [6] This was done directly using the vector potential, which had only one component in that case, rather than by the use of the Debye potential as in the CK procedure.

[9]E. Newman and R. Penrose, "An Approach to Gravitational Radiation by a Method of Spin Coefficients," J. Math. Phys. **3**, 566 (1962).

[10]R. Geroch, A. Held, and R. Penrose, "A Spacetime Calculus Based on Pairs of Null Directions," J. Math. Phys. **14**, 874 (1973).

[11]S.A. Teukolsky, "Perturbations on a Rotating Black Hole. I. Fundamental Equations for Gravitational, Electromagnetic, and Neutrino Perturbations," Astrophys. J. **185**, 635 (1973).

[12]F.A.E. Pirani, *Brandeis Summer Institute*, 1964 (Prentice-Hall, Englewood Cliffs, N.J., 1965), p. 305.

[13]J.M. Cohen and L.S. Kegeles, "Electromagnetic Fields in Curved Spaces: A Constructive Procedure," Phys. Rev. **10**, 1070 (1974).

[14]Alternatively one may define the ϕ 's directly in terms of the Maxwell field tensor projected onto the null tetrad as follows: [2]

$$\phi_0 \equiv F_{\mu\nu} l^\mu n^\nu,$$

$$\phi_1 \equiv \tfrac{1}{2} F_{\mu\nu} (l^\mu n^\nu + \bar{m}^\mu n^\nu),$$

$$\phi_2 \equiv F_{\mu\nu} \bar{m}^\mu n^\nu.$$

[15]J. Stewart and M. Walker, Springer Tracts (astrophysics) **69**, 69 (1973).

[16]We note that (3.9) may be written using the following generalizations of the operators A_0 and A_0^* defined in Ref. 15:

$$A_s \equiv \text{Þ}'\text{Þ} - \bar{\rho}'\text{Þ} - (2s+1)\rho\text{Þ}' - s(s+\tfrac{1}{2})\psi_2 + \tfrac{1}{12}(s+\tfrac{1}{2})R,$$

$$A_s^* \equiv -\eth'\eth + \bar{\tau}\eth + (2s+1)\tau\eth' - s(s+\tfrac{1}{2})\psi_2 + \tfrac{1}{12}(s+\tfrac{1}{2})R.$$

Using these definitions the complex conjugate of (3.9) becomes

$$(A_s + A_s^*)\bar{\psi}^{(s)} = 0.$$

[17]R. Graham, "Lagrangian for Diffusion in Curved Phase Space,"Phys. Rev. Lett. **38**, 51 (1977).

[18]In Ref. 5, τ, ρ, and ζ are used in place of t, r, and z. We use the latter here to avoid confusion with the spin coefficients τ and ρ.

[19]W. Pauli, *Die Allgemeinen Prinzipien der Wellenmechanik* Handbuch d. Physik, Springer, 1933), 2nd ed., p. 120.

[20]Albert Einstein, *The Meaning of Relativity* (Princeton U.P., Princeton, N.J., 1956), 5th ed., p. 68.

[21]B. Podolsky, "Quantum-Mechanically Correct Form of Hamiltonian Function for Conservative Systems," Phys. Rev. **32**, 812 (1928).

[22]D.R. Brill and J.A. Wheeler, "Interaction of Neutrinos and Gravitational Fields," Rev. Mod. Phys. **29**, 465 (1957).

Expansions of the affinity, metric and geodesic equations in Fermi normal coordinates about a geodesic[a)]

Wann-Quan Li and Wei-Tou Ni
Department of Physics, National Tsing Hua University Hsinchu, Taiwan, Republic of China

(Received 14 February 1979)

Fermi normal coordinates about a geodesic form a natural coordinate system for the nonrotating geodesic (freely falling) observer. Expansions of the affinity, metric, and geodesic equations in these coordinates in powers of proper distance normal to the geodesic are calculated here to third order, fourth order, and third order, respectively. An iteration scheme for calculation to higher orders is also given. For generality, we compute the affinity and the geodesic equations in an arbitrary affine manifold, and compute the metric in a Riemannian manifold with arbitrary signature.

I. INTRODUCTION

Fermi[1] showed that, given any curve in a Riemannian manifold, it is possible to introduce coordinates near this curve in such a way that the Christoffel symbols vanish along the curve (Fermi condition), leaving the metric there rectangular. If the curve is a geodesic, one way to construct such a coordinate system is to set the acceleration to zero in Synge's[2] construction of a natural nonrotating coordinate system for an accelerated observer. Manasse and Misner[3] called these coordinates Fermi normal coordinates. They form a natural coordinate system for a freely falling observer; we use them throughout this paper. Fermi normal coordinates satisfy the Fermi condition along the geodesic. The constant "time" hypersurfaces are normal to the geodesic. The "space" coordinates on these hypersurfaces form normal coordinate systems.

Fermi normal coordinates determine expansions of the affinity, metric and geodesic equations in powers of proper distance normal to the geodesic. Manasse and Misner[3] derived the first-order expansion of the affinity and the second-order expansion of the metric. Using a result of Hodgkinson,[4] Mashhoon[5] obtained the first-order expansion of the geodesic equations. In a more recent paper, Mashhoon[6] also obtained the second-order expansion of the geodesic equations. By setting acceleration and rotation to zero in Ref. 7, we derived the second-order expansion of the affinity and the third-order expansion of the metric.

In this paper, we calculate the third-order expansion of the affinity, the fourth-order expansion of the metric and the third-order expansion of the geodesic equations. For the sake of generality and in view of the recent interests in higher dimensional superspace, we calculate these formulas in an arbitrary dimensional manifold with arbitrary signature. In fact, all formulas in Sec. II are derived for affine manifolds (without torsion). We also indicate an iteration scheme for higher-order calculations.

The present results would be useful in the path-integral formulation and in the calculation of effective action and energy–momentum tensor in quantum gravity. They would also be useful in obtaining physical effects in binary pulsars.

II. EXPANSIONS OF THE AFFINITY AND GEODESIC EQUATIONS

Let V_N be an N-dimensional affine manifold. Consider a geodesic $P_0(\tau)$ in V_N. τ is an affine parameter for $P_0(\tau)$. At a fixed point τ_0 on the geodesic $P_0(\tau)$, pick a basis N-ad $\{\mathbf{e}_\alpha(\tau_0)$: $\alpha = 0,1,2,\ldots,N-1\}$ with $\mathbf{e}_0(\tau_0) = (\partial/\partial\tau)_{\tau=\tau_0}$. Parallel transport $\mathbf{e}_\alpha(\tau_0)$ along $P_0(\tau)$ to obtain a basis N-ad $\{\mathbf{e}_\alpha(\tau)\}$ all along $P_0(\tau)$. Since $P_0(\tau)$ is a geodesic and τ an affine parameter, $\mathbf{e}_0(\tau) = \partial/\partial\tau$.

Throughout this paper we shall use Greek indices to vary from 0 to $N-1$ and Latin indices to vary from 1 to $N-1$.

At any point $P_0(\tau)$ we send out geodesic $P(\tau;\mathbf{n};s)$ with $\mathbf{n} = n^i\mathbf{e}_i(\tau)$, i.e., $n^0 = 0$, where $\mathbf{n} = (\partial/\partial s)_{P_0(\tau)}$ and s is an affine parameter for $P(\tau;\mathbf{n};s)$. At the point $P(\tau;\mathbf{n};s)$, assign the coordinates $x^0 \equiv \tau$, $x^i \equiv sn^i$. It is easy to show that these coordinates are well-defined in the neighborhood of $P_0(\tau)$. We shall call them Fermi normal coordinates in an affine manifold. This coordinate system is good for

$$s \ll \min\left\{\frac{1}{|R^\mu{}_{\nu\alpha\beta}|^{1/2}}, \frac{|R^\mu{}_{\nu\alpha\beta}|}{|R^\mu{}_{\nu\alpha\beta,\gamma}|}\right\},$$

since within this distance curvature has not yet caused geodesics to cross ($s \ll 1/|R^\mu{}_{\nu\alpha\beta}|^{1/2}$), and the Riemann tensor has not yet changed much from its value on $P_0(\tau)$ ($s \ll |R^\mu{}_{\nu\alpha\beta}|/|R^\mu{}_{\nu\alpha\beta,\gamma}|$).

Since the curve $x^0 = \tau = \text{const}$, $x^i = \alpha^i s$, satisfies the geodesic equation

$$\frac{d^2x^\mu}{ds^2} + \Gamma^\mu{}_{\alpha\beta}\frac{dx^\alpha}{ds}\frac{dx^\beta}{ds} = 0, \quad (1)$$

we have

$$\Gamma^\mu{}_{ij}|_{P_0(\tau)}\alpha^i\alpha^j = 0. \quad (2)$$

Because α^i can be arbitrary

$$\Gamma^\mu{}_{ij}|_{P_0(\tau)} = 0. \quad (3)$$

From the fact that each of the vectors $\mathbf{e}_\mu(\tau)$ satisfies the equation of parallel displacement along the geodesic $x^0 = \tau$,

[a)]Supported in part by the National Science Council of the Republic of China.

 0022-2488/79/091925-05$01.00 © 1979 American Institute of Physics

$x^i = 0$, we obtain

$$\Gamma^{\mu}{}_{\sigma 0}|_{P_0(\tau)} = 0. \quad (4)$$

(3) and (4) show that the Fermi condition

$$\Gamma^{\mu}{}_{\alpha\beta}|_{P_0(\tau)} = 0, \quad (5)$$

is satisfied.

Differentiating (5) along $P_0(\tau)$ with respect to τ, we get

$$\Gamma^{\sigma}{}_{\mu\nu,0}|_{P_0(\tau)} = 0. \quad (6)$$

From the definition of the Riemann tensor,

$$\Gamma^{\alpha}{}_{\mu 0,\nu} = R^{\alpha}{}_{\mu\nu 0} + \Gamma^{\alpha}{}_{\mu\nu,0} + (\Gamma^{\sigma}{}_{\mu\nu}\Gamma^{\alpha}{}_{\sigma 0} - \Gamma^{\sigma}{}_{\mu 0}\Gamma^{\alpha}{}_{\sigma\nu}). \quad (7)$$

Hence

$$\Gamma^{\alpha}{}_{\mu 0,\nu} = R^{\alpha}{}_{\mu\nu 0} \quad \text{along } P_0(\tau). \quad (8)$$

To calculate the Latin-indexed derivatives of the Γ's, we use the geodesic deviation equation

$$\frac{d^2K^{\mu}}{ds^2} + 2\frac{dK^{\sigma}}{ds}\Gamma^{\mu}{}_{\sigma\alpha}U^{\alpha} + K^{\sigma}U^{\alpha}U^{\beta}R^{\mu}{}_{\alpha\sigma\beta}$$
$$+ K^{\sigma}U^{\alpha}U^{\beta}(\Gamma^{\mu}{}_{\sigma\alpha,\beta} + \Gamma^{\tau}{}_{\sigma\alpha}\Gamma^{\mu}{}_{\tau\beta} - \Gamma^{\mu}{}_{\sigma\tau}\Gamma^{\tau}{}_{\alpha\beta}) = 0, \quad (9)$$

where $\mathbf{K} = \partial/\partial K$ and $\mathbf{U} = \partial/\partial s$ of a one-parameter family of geodesics $R\,(K,s)$, and where s is an affine parameter along the geodesic $R\,(K,s)$ for K fixed. The family of geodesics we want to consider is $P\,(\tau;\alpha^i;s) \equiv P\,(\tau;\mathbf{n};s)$ where $\mathbf{n} = \alpha^i\mathbf{e}_i$. The case $\mathbf{K} = \partial/\partial\alpha^i$ leads to the desired results. In this case $\mathbf{K} = \partial/\partial\alpha^i = s(\partial/\partial x^i)$, hence $K^{\mu} = s\delta_i^{\mu}$. Expanding terms in the geodesic deviation equation in powers of s, we have

$$2\delta_i^{\sigma}\Gamma^{\mu}{}_{\sigma j}\alpha^j = 2s\Gamma^{\mu}{}_{ij,k}|_{P_0(\tau)}\alpha^j\alpha^k + s^2\Gamma^{\mu}{}_{ij,kl}|_{P_0(\tau)}\alpha^j\alpha^k\alpha^l$$
$$+ \frac{s^3}{3}\Gamma^{\mu}{}_{ij,klm}|_{P_0(\tau)}\alpha^j\alpha^k\alpha^l\alpha^m + O(s^4), \quad (10)$$

$$K^{\sigma}U^{\alpha}U^{\beta}R^{\mu}{}_{\alpha\sigma\beta}$$
$$= s\alpha^j\alpha^kR^{\mu}{}_{jik}|_{P_0(\tau)} + s^2\alpha^j\alpha^k\alpha^lR^{\mu}{}_{jik;l}|_{P_0(\tau)}$$
$$+ \tfrac{1}{2}s^3\alpha^j\alpha^k\alpha^l\alpha^m(R^{\mu}{}_{jik,l})_{,m}|_{P_0(\tau)} + O(s^4), \quad (11)$$

$$K^{\sigma}U^{\alpha}U^{\beta}\Gamma^{\mu}{}_{\sigma\alpha,\beta}$$
$$= s\Gamma^{\mu}{}_{ij,k}|_{P_0(\tau)}\alpha^j\alpha^k + s^2\Gamma^{\mu}{}_{ij,kl}|_{P_0(\tau)}\alpha^j\alpha^k\alpha^l$$
$$+ \frac{s^3}{2}\Gamma^{\mu}{}_{ij,klm}|_{P_0(\tau)}\alpha^j\alpha^k\alpha^l\alpha^m + O(s^4), \quad (12)$$

$$K^{\sigma}U^{\alpha}U^{\beta}(\Gamma^{\tau}{}_{\sigma\alpha}\Gamma^{\mu}{}_{\tau\beta} - \Gamma^{\mu}{}_{\sigma\tau}\Gamma^{\tau}{}_{\alpha\beta})$$
$$= s^3(\Gamma^{\tau}{}_{ij,l}\Gamma^{\mu}{}_{\tau k,m} - \Gamma^{\mu}{}_{i\tau,l}\Gamma^{\tau}{}_{jk,m})|_{P_0(\tau)}$$
$$\times\alpha^j\alpha^k\alpha^l\alpha^m + O(s^4). \quad (13)$$

Substituting all these equations into Eq. (9), every order in s must vanish separately. Therefore,

$$(3\Gamma^{\mu}{}_{ij,k} + R^{\mu}{}_{jik})|_{P_0(\tau)}\alpha^j\alpha^k = 0, \quad (14)$$

$$(2\Gamma^{\mu}{}_{ij,kl} + R^{\mu}{}_{jik;l})|_{P_0(\tau)}\alpha^j\alpha^k\alpha^l = 0, \quad (15)$$

$$[\tfrac{5}{6}\Gamma^{\mu}{}_{ij,klm} + \tfrac{1}{2}(R^{\mu}{}_{jik,l})_{,m} + \Gamma^{\tau}{}_{ij,l}\Gamma^{\mu}{}_{\tau k,m}$$
$$- \Gamma^{\mu}{}_{i\tau,l}\Gamma^{\tau}{}_{jk,m}]|_{P_0(\tau)}\alpha^j\alpha^k\alpha^l\alpha^m = 0. \quad (16)$$

(14) leads to the results of Manasse and Misner[3]:

$$\Gamma^{\mu}{}_{ij,k} = -\tfrac{1}{3}(R^{\mu}{}_{ijk} + R^{\mu}{}_{jik}) \quad \text{along } P_0(\tau). \quad (17)$$

(15) leads to the results of Ref. 7 with $\mathbf{a} = \boldsymbol{\omega} = 0$ there, i.e.,

$$\Gamma^{\mu}{}_{ij,kl} = -\tfrac{1}{3}(R^{\mu}{}_{ijk;l} + R^{\mu}{}_{ijl;k}) - \tfrac{1}{12}\underset{jkl}{P}R^{\mu}{}_{jik;l}, \quad (18)$$

where the symbol $\underset{i_1\cdots i_r}{P}$ indicates that the expression following it is to be summed over all $r!$ permutations of $i_1,\ldots,i_r$.

Differentiating (6), (8), and (17) along $P_0(\tau)$ with respect to τ, we get

$$\Gamma^{\sigma}{}_{\mu\nu,00} = 0,$$
$$\Gamma^{\alpha}{}_{\mu 0,\nu 0} = R^{\alpha}{}_{\mu\nu 0;0}, \quad (19)$$
$$\Gamma^{\mu}{}_{ij,k0} = -\tfrac{1}{3}(R^{\mu}{}_{ijk;0} + R^{\mu}{}_{jik;0}), \quad \text{all along } P_0(\tau).$$

Differentiating (7) with respect to x^k and choosing appropriate indices, we have

$$\Gamma^{\mu}{}_{0i,jk} = \Gamma^{\mu}{}_{ij,k0} + R^{\mu}{}_{ij0;k}$$
$$= -\tfrac{1}{3}(R^{\mu}{}_{ijk;0} + R^{\mu}{}_{jik;0}) + R^{\mu}{}_{ij0;k}$$
$$\text{along } P_0(\tau). \quad (20)$$

Summarizing the first-order and second-order results, we obtain

$$\Gamma^{\sigma}{}_{\mu\nu,0} = 0,$$
$$\Gamma^{\alpha}{}_{\mu 0,\nu} = R^{\alpha}{}_{\mu\nu 0}, \quad (21)$$
$$\Gamma^{\mu}{}_{ij,k} = -\tfrac{1}{3}(R^{\mu}{}_{ijk} + R^{\mu}{}_{jik}), \quad \text{all along } P_0(\tau),$$

and

$$\Gamma^{\sigma}{}_{\mu\nu,00} = 0,$$
$$\Gamma^{\alpha}{}_{\mu 0,\nu 0} = R^{\alpha}{}_{\mu\nu 0;0},$$
$$\Gamma^{\mu}{}_{ij,k0} = -\tfrac{1}{3}(R^{\mu}{}_{ijk;0} + R^{\mu}{}_{jik;0}),$$
$$\Gamma^{\mu}{}_{00,jk} = R^{\alpha}{}_{0j0;k} + R^{\alpha}{}_{jk0;0}, \quad (22)$$
$$\Gamma^{\mu}{}_{0i,jk} = -\tfrac{1}{3}(R^{\mu}{}_{ijk;0} + R^{\mu}{}_{jik;0}) + R^{\mu}{}_{ij0;k},$$
$$\Gamma^{\mu}{}_{ij,kl} = -\tfrac{1}{3}(R^{\mu}{}_{ijk;l} + R^{\mu}{}_{ijl;k}) - \tfrac{1}{12}\underset{jkl}{P}R^{\mu}{}_{jik;l},$$
$$\text{all along } P_0(\tau).$$

To derive an expression for the third-order derivatives $\Gamma^{\mu}{}_{ij,klm}$, we use Eq. (16). Since α^i can be arbitrary, (16) is equivalent to

$$5\Gamma^{\mu}{}_{ij,klm} + 5\Gamma^{\mu}{}_{ik;jlm} + 5\Gamma^{\mu}{}_{il,jkm} + 5\Gamma^{\mu}{}_{im,jkl}$$
$$+ \tfrac{1}{2}\underset{jklm}{P}R^{\mu}{}_{jik,lm} + \underset{jklm}{P}\Gamma^{\tau}{}_{ij,l}\Gamma^{\mu}{}_{\tau k,m} - \underset{jklm}{P}\Gamma^{\mu}{}_{i\tau,l}\Gamma^{\tau}{}_{jk,m}$$
$$= 0 \quad \text{along } P_0(\tau). \quad (23)$$

From the definition of Riemannian tensor,

$$\Gamma^{\mu}{}_{\alpha\gamma,\beta} = R^{\mu}{}_{\alpha\beta\gamma} + \Gamma^{\mu}{}_{\alpha\beta,\gamma} + (\Gamma^{\sigma}{}_{\alpha\beta}\Gamma^{\mu}{}_{\sigma\gamma} - \Gamma^{\sigma}{}_{\alpha\gamma}\Gamma^{\mu}{}_{\sigma\beta}). \quad (24)$$

Differentiating (24) twice and using (5), we get

$$\Gamma^{\mu}{}_{\alpha\gamma,\beta\delta\epsilon}$$
$$= R^{\mu}{}_{\alpha\beta\gamma,\delta\epsilon} + \Gamma^{\mu}{}_{\alpha\beta,\gamma\beta\epsilon} + \Gamma^{\sigma}{}_{\alpha\beta,\delta}\Gamma^{\mu}{}_{\sigma\gamma,\epsilon} + \Gamma^{\sigma}{}_{\alpha\beta,\epsilon}$$
$$\times\Gamma^{\mu}{}_{\sigma\gamma,\delta} - \Gamma^{\sigma}{}_{\alpha\gamma,\delta}\Gamma^{\mu}{}_{\sigma\beta,\epsilon} - \Gamma^{\sigma}{}_{\alpha\gamma,\epsilon}\Gamma^{\mu}{}_{\sigma\beta,\delta}$$
$$\text{along } P_0(\tau). \quad (25)$$

Choosing appropriate indices and using (21), we derive

$$\Gamma^{\mu}{}_{ik,jlm} = \Gamma^{\mu}{}_{ij,klm} + R^{\mu}{}_{ijk,lm} - \tfrac{1}{3}(R^{0}{}_{ijl} + R^{0}{}_{jil})R^{\mu}{}_{km0}$$
$$- \tfrac{1}{3}(R^{0}{}_{ijm} + R^{0}{}_{jim})R^{\mu}{}_{kl0} + \tfrac{1}{3}(R^{0}{}_{ikl} + R^{0}{}_{kil})$$
$$\times R^{\mu}{}_{jm0} + \tfrac{1}{3}(R^{0}{}_{ikm} + R^{0}{}_{kim})R^{\mu}{}_{jl0} + \tfrac{1}{9}(R^{p}{}_{ijl}$$

$$
\begin{aligned}
&+R^{p}{}_{jil})(R^{\mu}{}_{pkm}+R^{\mu}{}_{kpm})\\
&+\tfrac{1}{9}(R^{p}{}_{ijm}+R^{p}{}_{jim})(R^{\mu}{}_{pkl}\\
&+R^{\mu}{}_{kpl})-\tfrac{1}{9}(R^{p}{}_{ikl}+R^{p}{}_{kil})\\
&\times(R^{\mu}{}_{pjm}+R^{\mu}{}_{jpm})-\tfrac{1}{9}(R^{p}{}_{ikm}+R^{p}{}_{kim})\\
&\times(R^{\mu}{}_{pjl}+R^{\mu}{}_{jpl})\quad\text{along } P_0(\tau),
\end{aligned}\tag{26}
$$

$$
\begin{aligned}
\Gamma^{\mu}{}_{0j,klm}={}&\Gamma^{\mu}{}_{jk,lm0}+R^{\mu}{}_{jk0,lm}-\tfrac{1}{3}(R^{\sigma}{}_{jkl}+R^{\sigma}{}_{kjl})R^{\mu}{}_{\sigma m0}\\
&-\tfrac{1}{3}(R^{\sigma}{}_{jkm}+R^{\sigma}{}_{kjm})R^{\mu}{}_{\sigma l0}-R^{0}{}_{jl0}R^{\mu}{}_{km0}\\
&+\tfrac{1}{3}(R^{\mu}{}_{pkm}+R^{\mu}{}_{kpm})R^{p}{}_{jl0}-R^{0}{}_{jm0}R^{\mu}{}_{kl0}\\
&+\tfrac{1}{3}(R^{\mu}{}_{pkl}+R^{\mu}{}_{kpl})R^{p}{}_{jm0}\quad\text{along } P_0(\tau),
\end{aligned}\tag{27}
$$

$$
\begin{aligned}
\Gamma^{\mu}{}_{00,klm}={}&\Gamma^{\mu}{}_{0k,0lm}+R^{\mu}{}_{0k0,lm}+R^{\sigma}{}_{kl0}R^{\mu}{}_{\sigma m0}\\
&+R^{\sigma}{}_{km0}R^{\mu}{}_{\sigma l0}+\tfrac{1}{3}R^{p}{}_{0l0}(R^{\mu}{}_{pkm}+R^{\mu}{}_{kpm})\\
&+\tfrac{1}{3}R^{p}{}_{0m0}(R^{\mu}{}_{pkl}+R^{\mu}{}_{kpl})\quad\text{along } P_0(\tau).
\end{aligned}\tag{28}
$$

From the definition of covariant derivatives and Eq. (5),

$$
\begin{aligned}
R^{\alpha}{}_{\beta\gamma\delta;\lambda\epsilon}={}&R^{\alpha}{}_{\beta\gamma\delta,\lambda\epsilon}+R^{\mu}{}_{\beta\gamma\delta}\Gamma^{\alpha}{}_{\mu\lambda,\epsilon}-R^{\alpha}{}_{\mu\gamma\delta}\Gamma^{\mu}{}_{\beta\lambda,\epsilon}\\
&-R^{\alpha}{}_{\beta\mu\delta}\Gamma^{\mu}{}_{\gamma\lambda,\epsilon}-R^{\alpha}{}_{\beta\gamma\mu}\Gamma^{\mu}{}_{\delta\lambda,\epsilon}\\
&\text{along } P_0(\tau).
\end{aligned}\tag{29}
$$

Using (26), (21), and (29) to solve Eq. (23) for $\Gamma^{\mu}{}_{ij,klm}$, we obtain, after some straightforward computations,

$$
\begin{aligned}
\Gamma^{\mu}{}_{ij,klm}={}&-\frac{1}{40}\underset{ij}{P}\,\underset{klm}{P}\,(3R^{\mu}{}_{ijk;lm}+R^{\mu}{}_{mik;lj})\\
&-\frac{1}{360}\underset{ij}{P}\,\underset{klm}{P}\,(23R^{\sigma}{}_{ijk}R^{\mu}{}_{l\sigma m}\\
&-9R^{\sigma}{}_{lim}R^{\mu}{}_{\sigma jk}-15R^{\sigma}{}_{lim}R^{\mu}{}_{j\sigma k})\\
&-\tfrac{1}{9}\underset{ij}{P}\,\underset{klm}{P}\,R^{0}{}_{ijk}R^{\mu}{}_{l0m}.
\end{aligned}\tag{30}
$$

Differentiating (22) with respect to τ and using (5), we derive

$$
\begin{aligned}
&\Gamma^{\sigma}{}_{\mu\nu,000}=0,\\
&\Gamma^{\sigma}{}_{\mu0,\nu00}=R^{\alpha}{}_{\mu\nu0;00},\\
&\Gamma^{\mu}{}_{ij,k00}=-\tfrac{1}{3}(R^{\mu}{}_{ijk;00}+R^{\mu}{}_{jik;00}),\\
&\Gamma^{\mu}{}_{00,jk0}=R^{\mu}{}_{0j0;k0}+R^{\alpha}{}_{jk0;00},\\
&\Gamma^{\mu}{}_{0i;jk0}=-\tfrac{1}{3}(R^{\mu}{}_{ijk;00}+R^{\mu}{}_{jik;00})+R^{\mu}{}_{ij0;k0},\\
&\Gamma^{\mu}{}_{ij,kl0}=-\tfrac{1}{3}(R^{\mu}{}_{ijk;l0}+R^{\mu}{}_{ijl;k0})\\
&\qquad-\frac{1}{12}\underset{jkl}{P}\,R^{\mu}{}_{jik;l},\quad\text{all along } P_0(\tau).
\end{aligned}\tag{31}
$$

Using (29), (31), and (21) to compute $\Gamma^{\mu}{}_{0j,klm}$, $\Gamma^{\mu}{}_{00,klm}$ in (27) and (28), we have

$$
\begin{aligned}
\Gamma^{\mu}{}_{0j,klm}={}&-\frac{1}{12}\underset{klm}{P}\,(R^{\mu}{}_{kjl;m0}+2R^{\mu}{}_{j0k;lm})+\frac{1}{6}\underset{klm}{P}\\
&(R^{\mu}{}_{\alpha0k}R^{\alpha}{}_{ljm}-R^{\mu}{}_{j\alpha k}R^{\alpha}{}_{lm0}-R^{\alpha}{}_{j0k}R^{\mu}{}_{l\alpha m})\\
&-\frac{1}{3}\underset{klm}{P}\,R^{0}{}_{j0k}R^{\mu}{}_{l0m}\quad\text{along } P_0(\tau),
\end{aligned}\tag{32}
$$

$$
\begin{aligned}
\Gamma^{\mu}{}_{00,klm}={}&\tfrac{1}{6}\underset{klm}{P}\,(R^{\mu}{}_{0k0;lm}+R^{\mu}{}_{kl0;m0})\\
&-\tfrac{1}{6}\underset{klm}{P}\,(R^{\mu}{}_{0k\sigma}R^{\sigma}{}_{l0m}+3R^{\sigma}{}_{l0m}R^{\mu}{}_{\sigma k0})\\
&+\tfrac{1}{6}\underset{klm}{P}\,R^{p}{}_{0k0}R^{\mu}{}_{lpm}\quad\text{along } P_0(\tau).
\end{aligned}\tag{33}
$$

(30), (31), (32), and (33) include all third order derivatives of Γ's.

Substituting (5), (21), (22), (30), (31), (32), and (33), into the Taylor expansions of $\Gamma^{\mu}{}_{\alpha\beta}(x^0;x^i)$ with $x^0=\tau$

$$
\begin{aligned}
\Gamma^{\mu}{}_{\alpha\beta}(x^0;x^i)={}&\Gamma^{\mu}{}_{\alpha\beta}(P_0(\tau))+\Gamma^{\mu}{}_{\alpha\beta,i}(P_0(\tau))x^i+\frac{1}{2!}\\
&\times\Gamma^{\mu}{}_{\alpha\beta,ij}(P_0(\tau))x^ix^j+\frac{1}{3!}\Gamma^{\mu}{}_{\alpha\beta,ijk}(P_0(\tau))\\
&\times x^ix^jx^k+O((x^i)^4),
\end{aligned}\tag{34}
$$

we obtain the third-order expansion of the affine connections.

To derive expansions of the geodesic equations

$$
\frac{d^2x^\mu}{ds^2}+\Gamma^{\mu}{}_{\alpha\beta}\frac{dx^\mu}{ds}\frac{dx^\nu}{ds}=0,\tag{35}
$$

in Fermi normal coordinates, we first put (35) into the following form:

$$
\frac{d^2x^i}{(dx^0)^2}+\left(\Gamma^{i}{}_{\mu\nu}-\Gamma^{0}{}_{\mu\nu}\frac{dx^i}{dx^0}\right)\frac{dx^\mu}{dx^0}\frac{dx^\nu}{dx^0}=0.\tag{36}
$$

Denoting dx^i/dx^0 by v^i, and substituting the expansion of the affinities into (36), we obtain the third-order expansion of the geodesic equations:

$$
\begin{aligned}
\frac{d^2x^i}{(dx^0)^2}={}&-x^lR_{0i0l}+2x^lR_{0j0l}v^iv^j+2x^{\hat{l}}R_{ij0l}v^j+\tfrac{2}{3}x^lR_{0jkl}v^iv^kv^j+\tfrac{2}{3}x^lR_{ijkl}v^jv^k\\
&-\tfrac{1}{2}(R_{i0l0;m}+R_{ilm0;0})x^lx^m+\tfrac{1}{3}R_{iljm;0}x^lx^mv^j-R_{ijl0;m}x^lx^mv^j\\
&+\tfrac{1}{12}(5R_{ikjl;m}+R_{iljm;k})x^lx^mv^jv^k+\tfrac{1}{2}R_{0l0m;0}x^lx^mv^i+R_{0j0l;m}v^iv^jx^lx^m+\tfrac{1}{3}R_{0ljm;0}x^lx^mv^iv^j\\
&+\tfrac{1}{12}(5R_{0kjl;m}+R_{0ljm;k})x^lx^mv^iv^jv^k+\tfrac{1}{6}(R^{i}{}_{0k\sigma}R^{\sigma}{}_{l0m}+3R^{\sigma}{}_{l0m}R^{i}{}_{\sigma k0}-R^{p}{}_{0k0}R^{i}{}_{lpm})x^lx^mx^k\\
&-\tfrac{1}{3}(R^{i}{}_{\alpha0k}R^{\alpha}{}_{ljm}-R^{i}{}_{j\alpha k}R^{\alpha}{}_{lm0}-R^{\alpha}{}_{j0k}R^{i}{}_{l\alpha m})x^lx^mx^kv^j+\tfrac{2}{3}R^{0}{}_{j0k}R^{i}{}_{l0m}x^lx^mx^kv^j\\
&+\tfrac{1}{180}(23R^{\lambda}{}_{jkl}R^{i}{}_{m\lambda n}-9R^{\lambda}{}_{mjn}R^{i}{}_{\lambda kl}-15R^{\lambda}{}_{ljm}R^{i}{}_{k\lambda n})x^lx^mx^nv^jv^k\\
&+\tfrac{2}{9}R^{0}{}_{jkl}R^{i}{}_{m0n}x^lx^mx^nv^jv^k+\tfrac{2}{3}R^{\sigma}{}_{l0m}R^{0}{}_{\sigma0k}x^lx^mx^kv^i+\tfrac{1}{3}(R^{0}{}_{\alpha0k}R^{\alpha}{}_{ljm}-R^{0}{}_{j\alpha k}R^{\alpha}{}_{lm0}\\
&-R^{\alpha}{}_{j0k}R^{0}{}_{l\alpha m})x^lx^mx^kv^iv^j-\tfrac{2}{3}R^{0}{}_{j0k}R^{0}{}_{l0m}x^lx^mx^kv^iv^j-\tfrac{1}{180}(37R^{\lambda}{}_{jkl}R^{0}{}_{m\lambda n}\\
&-9R^{\lambda}{}_{mjn}R^{0}{}_{\lambda kl}-15R^{\lambda}{}_{ljm}R^{0}{}_{k\lambda n})x^lx^mx^nv^iv^jv^k-\tfrac{2}{9}R^{0}{}_{jkl}R^{0}{}_{m0n}x^lx^mx^nv^iv^jv^k
\end{aligned}
$$

$$-\tfrac{1}{6}(R^{i}{}_{0k0;lm}+R^{i}{}_{kl0;m0})x^{l}x^{m}x^{k}+\tfrac{1}{6}(R^{i}{}_{kjl;m0}+2R^{i}{}_{j0k;lm})x^{l}x^{m}x^{k}v^{j}$$
$$+\tfrac{1}{20}(6R^{i}{}_{jkl;mn}+R^{i}{}_{njl;mk})x^{l}x^{m}x^{n}v^{j}v^{k}+\tfrac{1}{6}R^{0}{}_{kl0;m0}x^{l}x^{m}x^{k}v^{i}$$
$$-\tfrac{1}{6}(R^{0}{}_{kjl;m0}+2R^{0}{}_{j0k;lm})x^{l}x^{m}x^{k}v^{i}v^{j}-\tfrac{1}{20}(6R^{0}{}_{jkl;mn}+R^{0}{}_{njl;mk})x^{l}x^{m}x^{n}v^{i}v^{j}v^{k}. \quad (37)$$

III. THE METRIC TENSORS

In this section, we suppose V_N to be a (pseudo) Riemannian manifold of signature $r-s$ with $r+s=N$. In addition to $\Gamma^{\alpha}{}_{\beta\gamma}$, there is a metric $g_{\alpha\beta}$. In the construction of the Fermi normal coordinates in Sec. II, we now require the basis N-ad$\{\mathbf{e}_{\alpha}(\tau_0)\}$ we pick at $P_0(\tau_0)$ to be orthonormal, i.e.,

$$\mathbf{e}_{\alpha}(\tau_0)\cdot\mathbf{e}_{\beta}(\tau_0)=\eta_{\alpha\beta}, \quad (38)$$

where $\eta_{\alpha\beta}=\text{diagonal}\{\pm 1,\pm 1,\ldots,\pm 1\}$ with r plus signs and s minus signs. Parallel transport along a geodesic perserves the relations, i.e.,

$$\mathbf{e}_{\alpha}(\tau)\cdot\mathbf{e}_{\beta}(\tau)=\eta_{\alpha\beta}. \quad (39)$$

Substituting (5), (21), (22), (30), (31), (32) into the relation

$$g_{\mu\nu,\alpha}=g_{\mu\sigma}\Gamma^{\sigma}{}_{\nu\alpha}+g_{\sigma\nu}\Gamma^{\sigma}{}_{\mu\alpha} \quad (40)$$

and its successive differentiations, we obtain successively

$$g_{\alpha\beta,\gamma}=0 \quad \text{along } P_0(\tau), \quad (41)$$

$$g_{\alpha\beta,\gamma0}=0,\quad g_{00,jk}=-2R_{0j0k},$$
$$g_{0i,jk}=-\tfrac{2}{3}(R_{0jik}+R_{0kij}),$$
$$g_{lm,ij}=-\tfrac{1}{3}(R_{iljm}+R_{imjl}),$$
$$\text{all along } P_0(\tau), \quad (42)$$

$$g_{\alpha\beta,\gamma00}=0,\quad g_{00,jk0}=-2R_{0j0k;0},$$
$$g_{0i,jk0}=-\tfrac{2}{3}(R_{0kij;0}+R_{0jik;0}),$$
$$g_{lm,ij0}=-\tfrac{1}{3}(R_{iljm;0}+R_{imjl;0}),$$
$$g_{00,jkl}=-\tfrac{1}{3}\underset{jkl}{P}R_{0j0k;l},\quad g_{0i,klm}=-\tfrac{1}{4}\underset{klm}{P}R_{0kil;m},$$
$$g_{ij,lmn}=-\tfrac{1}{6}\underset{lmn}{P}R_{iljm;n},\quad \text{all along } P_0(\tau), \quad (43)$$

$$g_{\alpha\beta,\gamma000}=0,\quad g_{00,jk00}=-2R_{0j0k;00},$$
$$g_{0i,jk00}=-\tfrac{2}{3}(R_{0kij;00}+R_{0jik;00}),$$
$$g_{lm,ij00}=-\tfrac{1}{3}(R_{iljm;00}+R_{imjl;00}),$$
$$g_{00,lmn0}=-\tfrac{1}{3}\underset{lmn}{P}R_{0l0n;m0},$$
$$g_{0j,lmn0}=-\tfrac{1}{4}\underset{lmn}{P}R_{0ljm;n0},$$
$$g_{ij,lmn0}=-\tfrac{1}{6}\underset{lmn}{P}R_{iljm;n0},$$
$$g_{00,jklm}=-\tfrac{1}{12}\underset{jklm}{P}R_{0j0k;lm}+\tfrac{1}{3}\underset{jklm}{P}R_{0k\sigma m}R^{\sigma}{}_{j0l},$$
$$g_{0j,klmn}=-\tfrac{1}{15}\underset{klmn}{P}R_{0kjl;mn}+\tfrac{2}{15}\underset{klmn}{P}R^{\sigma}{}_{kjm}R_{0l\sigma n},$$
$$g_{ij,klmn}=-\tfrac{1}{20}\underset{klmn}{P}R_{ikjl;mn}+\tfrac{2}{45}\underset{klmn}{P}R^{\mu}{}_{kjl}R_{\mu min},$$
$$\text{all along } P_0(\tau). \quad (44)$$

From Eqs. (41)–(44), the line element at a point $P(x^0;x^i)$ near the geodesic $P_0(\tau)$ with $x^0=\tau$, is

$$\begin{aligned}ds^2=&(dx^0)^2(\eta_{00}-R_{0l0m}x^lx^m\\&-\tfrac{1}{3}R_{0l0m;n}x^lx^mx^n-\tfrac{1}{12}R_{0n0k;lm}x^lx^mx^nx^k\\&+\tfrac{1}{3}R_{0n\sigma k}R^{\sigma}{}_{l0m}x^lx^mx^nx^k)\\&+dx^0dx^i(-\tfrac{2}{3}R_{0lim}x^lx^m-\tfrac{1}{4}R_{0lim;n}x^lx^mx^n\\&-\tfrac{1}{15}R_{0kim;ln}x^kx^mx^lx^n+\tfrac{2}{15}R^{\sigma}{}_{kim}R_{0l\sigma n}x^kx^mx^lx^n)\\&+dx^idx^j(\eta_{ij}-\tfrac{1}{3}R_{iljm}x^lx^m\\&-\tfrac{1}{6}R_{iljm;n}x^lx^mx^n-\tfrac{1}{20}R_{ikjl;mn}x^kx^lx^mx^n\\&+\tfrac{2}{45}R^{\sigma}{}_{kjl}R_{\sigma min}x^kx^lx^mx^n)\\&+O[dx^{\alpha}dx^{\beta}x^kx^lx^mx^nx^p].\end{aligned} \quad (45)$$

Using (41)–(44), the formula

$$g^{\mu\nu}{}_{,\alpha}=-g^{\beta\mu}g_{\beta\gamma,\alpha}g^{\gamma\nu}, \quad (46)$$

and its successive differentiations, we calculate the derivatives of the contravariant fundamental tensor to be

$$g^{\mu\nu}{}_{,\alpha}=0,\quad \text{all along } P_0(\tau), \quad (47)$$

$$g^{\alpha\beta}{}_{,\gamma0}=0,\quad g^{00}{}_{,jk}=2R^{0}{}_{j}{}^{0}{}_{k},$$
$$g^{0i}{}_{,jk}=\tfrac{2}{3}(R^{0}{}_{k}{}^{i}{}_{j}+R^{0}{}_{j}{}^{i}{}_{k}),\quad g^{ij}{}_{,lm}=\tfrac{1}{3}(R^{i}{}_{l}{}^{j}{}_{m}+R^{i}{}_{m}{}^{j}{}_{l}),$$
$$\text{all along } P_0(\tau), \quad (48)$$

$$g^{\alpha\beta}{}_{,\gamma00}=0,\quad g^{00}{}_{,jk0}=2R^{0}{}_{j}{}^{0}{}_{k;0},$$
$$g^{0i}{}_{,jk0}=\tfrac{2}{3}(R^{0}{}_{k}{}^{i}{}_{j;0}+R^{0}{}_{j}{}^{i}{}_{k;0}),$$
$$g^{ij}{}_{,lm0}=\tfrac{1}{3}(R^{i}{}_{l}{}^{j}{}_{m;0}+R^{i}{}_{m}{}^{j}{}_{l;0}),$$
$$g^{00}{}_{,lmn}=\tfrac{1}{3}\underset{lmn}{P}R^{0}{}_{l}{}^{0}{}_{n;m},$$
$$g^{0j}{}_{,lmn}=\tfrac{1}{4}\underset{lmn}{P}R^{0}{}_{l}{}^{j}{}_{m;n},\quad g^{ij}{}_{,lmn}=\tfrac{1}{6}\underset{lmn}{P}R^{i}{}_{l}{}^{j}{}_{m;n},$$
$$\text{all along } P_0(\tau), \quad (49)$$

$$g^{\alpha\beta}{}_{,\gamma000}=0,\quad g^{00}{}_{,jk00}=2R^{0}{}_{j}{}^{0}{}_{k;00},$$
$$g^{0i}{}_{,jk00}=\tfrac{2}{3}(R^{0}{}_{k}{}^{i}{}_{j;00}+R^{0}{}_{j}{}^{i}{}_{k;00}),$$
$$g^{ij}{}_{,lm00}=\tfrac{1}{3}(R^{i}{}_{l}{}^{j}{}_{m;00}+R^{i}{}_{m}{}^{j}{}_{l;00}),\quad g^{00}{}_{,lmn0}=+\tfrac{1}{3}\underset{lmn}{P}R^{0}{}_{l}{}^{0}{}_{n;m0},$$
$$g^{0j}{}_{,lmn0}=+\tfrac{1}{4}\underset{lmn}{P}R^{0}{}_{l}{}^{j}{}_{m;n0},\quad g^{ij}{}_{,lmn0}=+\tfrac{1}{6}\underset{lmn}{P}R^{i}{}_{l}{}^{j}{}_{m;n0},$$

$$g^{00}{}_{,jklm}=\tfrac{1}{12}\underset{jklm}{P}R^{0}{}_{j}{}^{0}{}_{k;lm}+\tfrac{1}{9}\underset{jklm}{P}R^{0}{}_{k}{}^{\lambda}{}_{l}R^{0}{}_{j\lambda m}$$
$$+\tfrac{5}{9}\underset{jklm}{P}R^{0}{}_{k}{}^{0}{}_{l}R^{0}{}_{j0m},$$

$$g^{0i}{}_{,jklm}=\tfrac{1}{15}\underset{jklm}{P}R^{0}{}_{j}{}^{i}{}_{k;lm}+\tfrac{4}{45}\underset{jklm}{P}R^{0}{}_{j\lambda k}R^{\lambda}{}_{l}{}^{i}{}_{m}$$
$$+\tfrac{2}{9}\underset{jklm}{P}R^{0}{}_{k0l}R^{0}{}_{j}{}^{i}{}_{m},$$

$$g^{ij}{}_{,klmn} = \frac{1}{20} \underset{klmn}{P} R^{i}{}_{k}{}^{j}{}_{l;mn} + \frac{1}{15} \underset{klmn}{P} R^{\lambda}{}_{l}{}^{i}{}_{m} R_{\lambda k}{}^{j}{}_{n} + \frac{1}{3} \underset{klmn}{P} R^{0}{}_{l}{}^{i}{}_{m} R_{0k}{}^{j}{}_{n}, \quad \text{all along } P_0(\tau). \tag{50}$$

Using Eqs. (41)–(44), (5), (21), (22), (30)–(33), we can calculate $\Gamma_{\mu\alpha\beta}$ and their derivatives through the formula

$$\Gamma_{\mu\alpha\beta} = g_{\mu\nu} \Gamma^{\nu}{}_{\alpha\beta} \tag{51}$$

and its differentiations as follows:

$$\Gamma_{\mu\alpha\beta} = 0, \quad \text{along } P_0(\tau), \tag{52}$$

$$\Gamma_{\mu\alpha\beta,0} = 0, \quad \Gamma_{\alpha\mu 0,\nu} = R_{\alpha\mu\nu 0},$$
$$\Gamma_{\mu ij,k} = -\frac{1}{3}(R_{\mu ijk} + R_{\mu jik}), \quad \text{all along } P_0(\tau), \tag{53}$$

$$\Gamma_{\mu\alpha\beta,00} = 0, \quad \Gamma_{\alpha\mu 0,\nu 0} = R_{\alpha\mu\nu 0;0},$$

$$\Gamma_{\mu ij,k0} = -\frac{1}{3}(R_{\mu ijk;0} + R_{\mu jik;0}),$$

$$\Gamma_{\mu 00,lm} = \frac{1}{2} \underset{lm}{P} (R_{\mu 0l0;m} + R_{\mu lm0;0}),$$

$$\Gamma_{\mu 0j,lm} = \frac{1}{2} \underset{lm}{P} (R_{\mu jl0;m} - \frac{1}{3} R_{\mu ijm;0}),$$

$$\Gamma_{\mu ij,kl} = \frac{1}{24} \underset{ij}{P} \underset{kl}{P} (5R_{\mu ikj;l} - R_{\mu kil;j}), \quad \text{all along } P_0(\tau), \tag{54}$$

$$\Gamma_{\mu\alpha\beta,000} = 0, \quad \Gamma_{\alpha\mu 0,\nu 00} = R_{\alpha\mu\nu 0;00},$$

$$\Gamma_{\mu ij,k00} = -\frac{1}{3}(R_{\mu ijk;00} + R_{\mu jik;00}),$$

$$\Gamma_{\mu 00,lm0} = \frac{1}{2} \underset{lm}{P} (R_{\mu 0l0;m0} + R_{\mu lm0;00}),$$

$$\Gamma_{\mu 0j,lm0} = \frac{1}{2} \underset{lm}{P} (R_{\mu jl0;m0} - \frac{1}{3} R_{\mu ljm;00}),$$

$$\Gamma_{\mu ij,kl0} = \frac{1}{24} \underset{ij}{P} \underset{kl}{P} (5R_{\mu ikj;l0} - R_{\mu kil;j0}),$$

$$\Gamma_{000,klm} = -\frac{1}{6} \underset{klm}{P} R_{0k0l;m0},$$

$$\Gamma_{j00,klm} = \frac{1}{6} \underset{klm}{P} (R_{j0k0;lm} + R_{jkl0;m0}) - \frac{1}{6} \underset{klm}{P} (R_{j0k\sigma} R^{\sigma}{}_{l0m} + 3R^{\sigma}{}_{l0m} R_{j\sigma k0} + R^{\sigma}{}_{0k0} R_{jl\sigma m}),$$

$$\Gamma_{00j,klm} = -\frac{1}{12} \underset{klm}{P} (R_{0kjl;m0} + 2R_{0j0k;lm}) + \frac{1}{6} \underset{klm}{P} \times (R_{0\alpha 0k} R^{\alpha}{}_{ljm} - R_{0j\alpha k} \times R^{\alpha}{}_{lm0} + 3R^{\alpha}{}_{j0k} R_{0l\alpha m}),$$

$$\Gamma_{i0j,klm} = -\frac{1}{12} \underset{klm}{P} (R_{ikjl;m0} + 2R_{ij0k;lm}) + \frac{1}{6} \underset{klm}{P} (R_{i\alpha 0k} R^{\alpha}{}_{ljm} - R_{ij\alpha k} R^{\alpha}{}_{lm0} + R^{\alpha}{}_{j0k} R_{\alpha mil}),$$

$$\Gamma_{0nj,klm} = -\frac{1}{40} \underset{nj}{P} \underset{klm}{P} (3R_{0njk;lm} + R_{0mnk;lj}) + \frac{1}{360} \underset{nj}{P} \underset{klm}{P} (57R^{\lambda}{}_{njk} R_{0l\lambda m} + 9R^{\lambda}{}_{lnm} R_{0\lambda jk} + 15R^{\lambda}{}_{lnm} R_{0j\lambda k}),$$

$$\Gamma_{inj,klm} = -\frac{1}{40} \underset{nj}{P} \underset{klm}{P} (3R_{injk;lm} + R_{imnk;lj}) + \frac{1}{360} \underset{nj}{P} \underset{klm}{P} (17R^{\lambda}{}_{njk} R_{il\lambda m} + 9R^{\lambda}{}_{lnm} R_{i\lambda jk} + 15R^{\lambda}{}_{lnm} R_{ij\lambda k}), \quad \text{all along } P_0(\tau). \tag{55}$$

IV. ITERATION SCHEME AND DISCUSSIONS

(i) Suppose we know the derivatives of Γ's along $P_0(\tau)$ up to nth order. Differentiation with respect to τ gives expressions for $\Gamma^{\sigma}{}_{\mu\nu,\alpha_1\alpha_2\cdots\alpha_n 0}$. Using nth order differentiation of (24), $\Gamma^{\sigma}{}_{\mu 0,\alpha_1\alpha_2\cdots\alpha_n\nu}$ can be expressed in terms of $\Gamma^{\sigma}{}_{\mu\nu,\alpha_1\alpha_2\cdots\alpha_n 0}$ and lower derivatives of Γ's. To solve for $\Gamma^{\sigma}{}_{kl,i_1i_2\cdots i_{n+1}}$, we use the geodesic deviation equation as in Sec. II. From the vanishing of $(n+1)$th order in the expansion of the geodesic deviation equation in s, and from nth order differentiation of (24), we can obtain an expression for $\Gamma^{\sigma}{}_{kl,i_1i_2\cdots i_{n+1}}$. Thus, we obtain all $(n+1)$th derivatives of Γ's. Substituting these into (36), we obtain $(n+1)$th order expansion of the geodesic equations. Following the method used in Sec. III, we can derive expressions for the $(n+2)$th derivatives of the metric tensors. This forms an iteration scheme for computing higher-order expansions.

(ii) The above iteration scheme applies equally well to Riemann normal coordinates just by dropping the 0 index and going to $(N-1)$-dimensional manifolds. If we do these, we obtain the results in Ref. 8. Note that the differences in the expressions for the Latin-indexed quantities such as $\Gamma^{\mu}{}_{ij,klm}$, etc., in this paper and those in Ref. 8 come from the difference in the definitions of Riemannian tensors, etc., in different dimensional manifolds.

(iii) Specializing the results in this paper to $N = 4$ with signature 2 and letting $P_0(\tau)$ be a timelike geodesic, we obtain the third-order expansion of the coordinate acceleration and the fourth-order expansion of the metric in the proper frame of an observer moving along $P_0(\tau)$. For an accelerated rotating observer, the third-order effects in coordinate acceleration and the fourth-order effects in the metric can be classified into purely gravitational and coupled inertial–gravitational terms. The present paper gives purely gravitational terms to this order.

[1] E. Fermi, Atti Acad. Naz. Lincei Rend. Cl. Sci. Fiz. Mat. Nat. **31**, 21, 51 (1922).
[2] J.L. Synge, *Relativity: The General Theory* (North-Holland, Amsterdam, 1960).
[3] F.K. Manasse and C.W. Misner, J. Math. Phys. **4**, 735 (1963).
[4] D.E. Hodgkinson, Gen. Rel. Grav. **3**, 351 (1972).
[5] B. Mashhoon, Astrophys. J. **197**, 705 (1975).
[6] B. Mashhoon, Astrophys. J. **216**, 591 (1977).
[7] W.-Q. Li and W.-T. Ni, "Coupled inertial and gravitational effects in the proper reference frame of an accelerated, rotating observer," J. Math. Phys. **20**, xxx (1979).
[8] W.-T. Ni, "Geodesic triangles and expansion of metrics in normal coordinates," Chinese J. Phys. **16**, No. 4, 206 (1978).

Euclidean and Minkowski space formulations of linearized gravitational potential in various gauges

S. C. Lim[a)]

International Centre for Theoretical Physics, Trieste, Italy
(Received 18 July 1978)

We show that there exists a unitary map connecting linearized theories of gravitational potential in vacuum, formulated in various covariant gauges and noncovariant radiation gauge. The free Euclidean gravitational potentials in covariant gauges satisfy the Markov property of Nelson, but are nonreflexive. For the noncovariant radiation gauge, the corresponding Euclidean field is reflexive but it only satisfies the Markov property with respect to special half spaces. The Feynman–Kac–Nelson formula is established for the Euclidean gravitational potential in radiation gauge.

I. INTRODUCTION

The analysis carried out by Bracci and Strocchi,[1-3] using Wightman's axiomatic framework[4] on the quantized theory of linearized gravitational field, showed that the gauge problems that exist in massless spin-2 particles also give rise to difficulties similar to those in quantum electrodynamics (QED). A local and covariant quantization of the linearized Einstein equations is possible only in a Hilbert space with indefinite metric, i.e., in the Gupta–Bleuler formalism.[5,6] The basic mathematical framework in this formalism is a set of three Hilbert spaces $\mathscr{H}'' \subset \mathscr{H}' \subset \mathscr{H}$ and a nondegenerate, Hermitian sesquilinear form $\langle\cdot,\cdot\rangle$ on $\mathscr{H}$, which is semidefinite on $\mathscr{H}'$ and induces a definite inner product on the physical space $\mathscr{H}'/\mathscr{H}''$. Then linearized Einstein equations do not satisfy as operator equations but only as mean values in $\mathscr{H}'$. Any quantized theory of gravitational potential that does not involve unphysical states requires a nonlocal and noncovariant formalism analogous to the radiation (or Coulomb) gauge formalism in QED. It is only in this noncovariant theory that the linearized Einstein equations can be satisfied as operator equations.

The main purpose of this paper is to study the relationship between the linearized theories of free gravitational potential in covariant and noncovariant gauges in both Minkowski and Euclidean space–time, in particular the properties of the corresponding Euclidean fields. It is possible to show that there exists a unitary equivalence between the covariant and noncovariant gauge formalisms of gravitational potentials in vacuum just like the case in QED.[7-9] In the Euclidean region the gravitational potential is found to satisfy Nelson's Markov property[10,11] for a wide class of covariant gauges, though it does not satisfy the reflection property. For noncovariant gravitational radiation gauge, the corresponding Euclidean field is reflexive, but it only satisfies the Markov property with respect to special half-spaces (i.e., Hegerfeldt's Markov property of second kind[12]). It is possible to establish the Feynman–Kac–Nelson formula[13] for the Euclidean gravitational potential in radiation gauge.

Finally we note that a relationship similar to that in the relativistic case also holds for Euclidean gravitational potentials in covariant and noncovariant gauges.

II. LINEARIZED THEORY OF GRAVITATIONAL POTENTIALS IN VARIOUS GAUGES

Throughout this paper we shall confine our discussion only to the linearized theory of gravitational field in vacuum. First we shall recall some basic facts about the nonquantized theory of linearized gravitational field. In the weak field approximation, the Einstein equations in vacuum are[14]

$$R_{\mu\nu}(x) = 0, \quad R(x) = 0, \tag{1}$$

where

$$R_{\mu\nu} = g^{\lambda\rho} R_{\lambda\mu\nu\rho}, \quad R = g^{\mu\nu} R_{\mu\nu}, \tag{2}$$

$g^{\mu\nu}$ is the Minkowski metric tensor with $g^{0\mu} = \delta^{0\mu}$, $g^{ij} = -\delta^{ij}$, and $R_{\mu\nu\rho\sigma}(x)$ is the Riemann tensor or gravitational field. In addition to the above equations $R_{\mu\nu\rho\sigma}$ also satisfies the following identities:

$$R_{\lambda\mu\nu\rho} = -R_{\mu\lambda\nu\rho} = R_{\mu\lambda\rho\nu} = R_{\nu\rho\lambda\mu}, \tag{3}$$

$$R_{\lambda\nu\rho\sigma} + R_{\lambda\sigma\nu\rho} + R_{\lambda\rho\sigma\nu} = 0, \tag{4}$$

$$\epsilon^{\mu\nu\rho\sigma}\partial_\nu R_{\alpha\beta\rho\sigma} = 0 \quad \text{(Bianchi's identities).} \tag{5}$$

Equations (3), (4), and (5) imply that $R_{\mu\nu\rho\sigma}(x)$ can be expressed in terms of a symmetric rank-two tensor, or the gravitational potential $G_{\mu\nu}(x)$

$$\begin{aligned} R_{\mu\nu\rho\sigma}(x) &= \tfrac{1}{2}[\partial_\mu \partial_\sigma \delta^\alpha_\nu \delta^\beta_\rho - \partial_\mu \partial_\rho \delta^\alpha_\nu \delta^\beta_\sigma \\ &\quad + \partial_\nu \partial_\rho \delta^\alpha_\mu \delta^\beta_\sigma - \partial_\nu \partial_\sigma \delta^\alpha_\mu \delta^\beta_\rho] G_{\alpha\beta}(x) \\ &= \tfrac{1}{2}[\partial_\mu \partial_\sigma G_{\nu\rho}(x) - \partial_\mu \partial_\rho G_{\nu\sigma}(x) \\ &\quad + \partial_\nu \partial_\rho G_{\mu\sigma}(x) - \partial_\nu \partial_\sigma G_{\mu\rho}(x)] \\ &\equiv D^{\alpha\beta}_{\mu\nu\rho\sigma} G_{\alpha\beta}(x). \end{aligned} \tag{6}$$

This definition of $R_{\mu\nu\rho\sigma}(x)$ in terms of $G_{\mu\nu}(x)$ contains an arbitrariness of freedom corresponding to gauge transformation

[a)]On leave of absence from Department of Physics, National University of Malaysia, P.O. Box 1124, Kuala Lumpur, Malaysia.

 0022-2488/79/091930-08$01.00 © 1979 American Institute of Physics

$$G_{\mu\nu}(x)\rightarrow G_{\mu\nu}(x) + \partial_\mu B_\nu(x) + \partial_\nu B_\mu(x). \tag{7}$$

The four arbitrary vector gauge functions $B_\mu(x)$ can be fixed by four gauge conditions, which can be given in many ways just as in the case of electrodynamics. For example, the analog of the Lorentz gauge conditions for electromagnetic potential is

$$\partial^\mu G_{\mu\nu}(x) - \tfrac{1}{2}\partial_\nu G^\mu_\mu(x) = 0. \tag{8}$$

These four conditions are called gravitational "Lorentz" gauge conditions or simply Gupta gauge conditions. Together with Eq. (6) they give

$$R_{\mu\nu}(x) = -\Box G_{\mu\nu}(x) = 0, \tag{9}$$

implying that linearized Einstein's equations in vacuum describe essentially a free field theory. Similarly, one can also employ the four gravitational radiation gauge conditions

$$\partial_i G_{ij}(x) = 0, \quad G^\mu_\mu(x) = 0, \tag{10}$$

which are analogous to the electrodynamic radiation (or Coulomb) gauge condition.

In the quantized theory, even though not all the Wightman axioms are satisfied by $G_{\mu\nu}$, we can still follow a Wightman approach to the theory with some modifications.[8,3] We shall follow the general scheme of Wightman and Strocchi[8] in the following definition:

Definition 1 (see also Ref. 3):

A local and covariant gauge for $G_{\mu\nu}(x)$ is specified by $\{G_{\mu\nu}(x), \mathscr{H}, \langle\cdot,\cdot\rangle, \mathscr{H}'\}$, where:

(a) $G_{\mu\nu}(x)$ is an operator-valued tempered distribution in a Hilbert space $\mathscr{H}$;

(b) There exists a distinguished subspace $\mathscr{H}'\subset\mathscr{H}$ such that

(i) a nondegenerate, Hermitian sesquilinear form $\langle\cdot,\cdot\rangle$ exists in $\mathscr{H}$, with respect to which the representation U of the Poincaré group SL(2,C) is unitary;

(ii) $\langle\cdot,\cdot\rangle$, when restricted to $\mathscr{H}'$, is bounded and nonnegative;

(iii) the Riemann tensor $R_{\mu\nu\rho\sigma}\equiv D^{\alpha\beta}_{\mu\nu\rho\sigma} G_{\alpha\beta}(x)$ leaves $\mathscr{H}'$ invariant and the linearized Einstein equations hold as an expectation value in $\mathscr{H}'$,

$$\langle\varphi, g^{\mu\sigma} R_{\mu\nu\rho\sigma}(f)\psi\rangle = 0, \quad \forall\varphi,\psi\in\mathscr{H}',$$

with $f\in\mathscr{S}$, the Scwartz space of test functions, and ψ lies in the domain of $R_{\mu\nu\rho\sigma}$ in $\mathscr{H}'$;

(iv) the representation U leaves $\mathscr{H}'$ invariant, and there exists in $\mathscr{H}'$, a unique vacuum ψ_0, which is also invariant under U and is a cyclic vector for the operator $G_{\mu\nu}(f)$;

(c) $G_{\mu\nu}$ obeys local commutativity

$$[G_{\mu\nu}(x), G_{\rho\sigma}(y)] = 0 \quad \text{for}(x-y)^2 < 0$$

and transforms covariantly under U, i.e., $(a,\Lambda)\in$SL(2,C),

$$U(a,\Lambda)\, G_{\mu\nu}(x) U(a,\Lambda)^{-1} = \Lambda_\mu{}^{-1\rho} \Lambda_\nu{}^{-1\sigma} G_{\rho\sigma}(\Lambda x + a);$$

(d) The Fourier transform of the two-point vacuum expectation function $\langle\psi_0, G_{\mu\nu}(x)G_{\rho\sigma}(y)\psi_0\rangle$ has support in the forward light cone $\bar{V}_+$.

(e) The physical states are elements of the quotient space $\mathscr{H}'/\mathscr{H}''\equiv\mathscr{H}_{\rm ph}$, where $\mathscr{H}''$ is the subspace of vanishing norm; $U(a,\Lambda)$ is unitary on $\mathscr{H}_{\rm ph}$.

If, in addition to the above conditions, $G_{\mu\nu}(x)$ is required to satisfy the cluster property (see Refs. 3 and 4) and $R_{\mu\nu\rho\sigma}(x)$ is assumed to have the conventional normalization, then the most general two-point function for $G_{\mu\nu}$ is found to be (see Ref. 3)

$$\begin{aligned} W_{\mu\nu\rho\sigma}(x-y) &= \langle\psi_0, G_{\mu\nu}(x)G_{\rho\sigma}(y)\psi_0\rangle \\ &= -i(g_{\mu\rho}g_{\nu\sigma} + g_{\mu\sigma}g_{\nu\rho} - g_{\mu\nu}g_{\rho\sigma})D_+(x-y) \\ &\quad + (g_{\mu\nu}\partial_\rho\partial_\sigma + g_{\rho\sigma}\partial_\mu\partial_\nu)F_1(x-y) \\ &\quad + (g_{\mu\rho}\partial_\nu\partial_\sigma + g_{\mu\sigma}\partial_\nu\partial_\rho + g_{\nu\rho}\partial_\mu\partial_\sigma \\ &\quad + g_{\nu\sigma}\partial_\mu\partial_\rho)F_2(x-y) \\ &\quad + \partial_\mu\partial_\nu\partial_\rho\partial_\sigma F_3(x-y), \end{aligned} \tag{11}$$

where

$$D_+(x-y) = \frac{i}{2(2\pi)^3}\int e^{-ip(x-y)}\frac{d^3p}{|\mathbf{p}|}$$

and F_i, $i = 1,2,3$, are Lorentz invariant distributions. Here it is interesting to note that a subclass of the two-point functions given by Eq. (11) can be derived from a Lagrangian formulation. Consider the Lagrangian density

$$\mathscr{L} = \mathscr{L}_E + \mathscr{L}_G = G_{\mu\nu}(x)Q^{\mu\nu\rho\sigma}(\partial)G_{\rho\sigma}(x), \tag{12a}$$

where $Q^{\mu\nu\rho\sigma}(\partial)\equiv Q^{\mu\nu\rho\sigma}_E(\partial) + Q^{\mu\nu\rho\sigma}_G(\partial)$ is given by

$$\begin{aligned} Q^{\mu\nu\rho\sigma}_E(\partial) &= (g^{\mu\rho}g^{\nu\sigma} + g^{\mu\sigma}g^{\nu\rho} - 2g^{\mu\nu}g^{\rho\sigma})\Box \\ &\quad + (g^{\mu\nu}\partial^\rho\partial^\sigma + g^{\rho\sigma}\partial^\mu\partial^\nu) + (g^{\mu\rho}\partial^\nu\partial^\sigma \\ &\quad + g^{\mu\sigma}\partial^\nu\partial^\rho + g^{\nu\rho}\partial^\mu\partial^\sigma + g^{\nu\sigma}\partial^\mu\partial^\rho), \end{aligned} \tag{12b}$$

$$\begin{aligned} Q^{\mu\nu\rho\sigma}_G(\partial) &= -\tfrac{1}{2}a(g^{\mu\rho}\partial^\nu\partial^\sigma \\ &\quad + g^{\nu\sigma}\partial^\nu\partial^\rho + g^{\mu\rho}\partial^\mu\partial^\sigma + g^{\nu\sigma}\partial^\mu\partial^\rho) \\ &\quad + 4b\,(g^{\mu\nu}\partial^\rho\partial^\sigma + g^{\rho\sigma}\partial^\mu\partial^\nu) \\ &\quad + 4b^2 g^{\mu\nu}g^{\rho\sigma}\Box, \end{aligned} \tag{12c}$$

with a, b real constants and $\Box = \partial_\mu\partial^\mu$. $\mathscr{L}_E$ is just the usual linearized Einstein Lagrangian and $\mathscr{L}_G$ is the gauge-fixing term which can be expressed more compactly in the form $\mathscr{L}_G = \frac{1}{4}bC^\mu C_\mu$, where

$$C_\mu = C^{\rho\sigma}_\mu(a)G_{\rho\sigma} = \partial^\nu G_{\mu\nu} + a\partial_\mu G^\nu_\nu. \tag{13}$$

$Q^{\mu\nu\rho\sigma}(\partial)$ is a nonsingular matrix differential operator and so can be inverted to obtain the graviton propagator, which is equal to minus the Fourier transform of its inverse,

$$P_{\mu\nu\rho\sigma}(p) = \frac{1}{2p^2}\Big((g_{\mu\rho}\, g_{\nu\sigma} + g_{\mu\sigma}\, g_{\nu\rho} - g_{\mu\nu}\, g_{\rho\sigma}) + \frac{1+2a}{1+a}$$

$$\times (g_{\mu\nu}\, p_\rho\, p_\sigma + g_{\rho\sigma}\, p_\mu\, p_\nu)\frac{1}{p^2}\Big)$$

$$- \frac{1+b}{b}(g_{\mu\rho}\, p_\nu\, p_\sigma + g_{\mu\sigma}\, p_\nu p_\rho$$

$$+ g_{\nu\rho}\, p_\mu\, p_\sigma + g_{\nu\sigma}\, p_\mu\, p_\rho)\frac{1}{p^2} + \frac{1+2a}{b(1+a)^2}$$

$$\times (3 + a + b - 2ab)\frac{p_\mu\, p_\nu\, p_\rho\, p_\sigma}{p^4}. \qquad (14)$$

The corresponding two-point function of $G_{\mu\nu}$ is

$$W_{\mu\nu\rho\sigma}(x-y)$$

$$= \frac{1}{2(2\pi)^3}\int P_{\mu\nu\rho\sigma}(p) e^{-i\,|\mathbf{p}|(x_0 - y_0)} e^{i\mathbf{p}\cdot(\mathbf{x}-\mathbf{y})}\frac{d^3p}{|\mathbf{p}|}. \qquad (15)$$

This is the same as the two-point function given by Eq. (11) if we make the following identifications:

$$\widetilde{F}_1(p) = \frac{1+2a}{1+a}\frac{1}{p^4}, \quad \widetilde{F}_2(p) = -\left(\frac{1+b}{b}\right)\frac{1}{p^4},$$

$$\widetilde{F}_3(p) = \frac{1+2a}{b(1+a)^2}(3 + a + b - 2ab)\frac{1}{p^6},$$

where $\widetilde{F}_i(p)$, $i = 1,2,3$ are the Fourier transforms of $F_i(x)$. The propagator given in (14) reduces to the propagator with linearized harmonic condition as given by Fradkin and Tyutin[15] if we let $a = -\frac{1}{2}$. If we further require $b = -1$ then $P_{\mu\nu\rho\sigma}$ reduces to the familiar one with Gupta gauge conditions

$$P_{\mu\nu\rho\sigma}(p) = (2p^2)^{-1}(g_{\mu\rho}\, g_{\nu\sigma} + g_{\mu\sigma}\, g_{\nu\rho} - g_{\mu\nu}\, g_{\rho\sigma}). \qquad (16)$$

Two other important cases worth noting are the Landau-like gauges which arise when (i) $b \to \infty$ and $a = -\frac{1}{2}$, then

$$P_{\mu\nu\rho\sigma}(p)$$

$$= \frac{1}{2p^2}\left(d_{\mu\rho}\, d_{\nu\sigma} + d_{\mu\sigma}\, d_{\nu\rho} - g_{\mu\nu}\, g_{\rho\sigma} - \frac{2p_\mu\, p_\nu\, p_\rho\, p_\sigma}{p^4}\right), \qquad (17)$$

where $d_{\mu\nu} = -(g_{\mu\nu} - p_\mu\, p_\nu\, p^{-2})$. This gauge corresponds to the Landau gauge in QED in the sense that

$$\widetilde{C}^{\mu\nu}_{\lambda}(a = -\tfrac{1}{2})P_{\mu\nu\rho\sigma}(p) = 0 \qquad (18)$$

and $\partial^\mu G_{\mu\nu}(x) - \frac{1}{2}\partial_\nu G^\mu_\mu(x) = 0$ holds as an operator equation. Case (ii) occurs when $b \to \infty$ and $a = -\frac{1}{4}$, which gives

$$P_{\mu\nu\rho\sigma}(p) = \frac{1}{2p^2}\left(d_{\mu\rho}\, d_{\nu\sigma} + d_{\mu\sigma}\, d_{\nu\rho} - \tfrac{2}{3}d_{\mu\nu}\, d_{\rho\sigma}\right), \qquad (19)$$

satisfying $p^\mu P_{\mu\nu\rho\sigma}(p) = 0$ and $\partial^\mu G_{\mu\nu}(x) = 0$ holds as an operator equation. We shall exclude the case $a = -1$ since for such a value of a, $Q^{\mu\nu\rho\sigma}$ is singular and cannot be inverted.

If one chooses the gauge fixing term to be of the form

$$C_\mu = b\sum_{i=1}^{3}\partial_i G_{i\mu}, \quad \mu = 0,1,2,3 \quad \text{and } b \to \infty, \qquad (20)$$

then it specifies a noncovariant Prentki gauge[16] with the corresponding propagator in the form

$$P_{\mu\nu\rho\sigma}(p) = (2p^2)^{-1}[(\bar{d}^{0}_{\mu\rho}\, \bar{d}^{0}_{\nu\sigma} + \bar{d}^{0}_{\mu\sigma}\, \bar{d}^{0}_{\nu\rho} - \bar{d}^{0}_{\mu\nu}\, \bar{d}^{0}_{\rho\sigma})$$

$$+ \bar{d}^{0}_{\mu\rho}\, g^{0}_{\nu\sigma} + g^{0}_{\mu\sigma}\, \bar{d}^{0}_{\nu\rho} + \bar{d}^{0}_{\nu\rho}\, g^{0}_{\mu\sigma} + \bar{d}^{0}_{\nu\sigma}\, g^{0}_{\mu\rho})\frac{1}{\mathbf{p}^2}$$

$$- (\bar{d}^{0}_{\mu\nu}\, g^{0}_{\rho\sigma} + \bar{d}^{0}_{\rho\sigma}\, g^{0}_{\mu\nu})\frac{1}{\mathbf{p}^2} + \frac{p^2}{\mathbf{p}^4}\, g^{0}_{\mu\nu\rho\sigma}\Big], \qquad (21)$$

where

$$\bar{d}^{0}_{\mu\nu} = (1 - g_{\mu 0})(1 - g_{\nu 0})\left[-g_{\mu\nu} + \frac{p_\mu\, p_\nu}{-\mathbf{p}^2}\right],$$

$$g^{0}_{\mu\nu} = g_{\mu 0}\, g_{\nu 0},$$

and $g^{0}_{\mu\nu\rho\sigma} = g_{\mu 0}\, g_{\nu 0}\, g_{\rho 0}\, g_{\sigma 0}$. This reduces to the more familiar gravitational "Coulomb" gauge if $G_{\mu 0}(x) = 0$, $\mu = 0,1,2,3$. Then the propagator becomes

$$P_{ijmn}(p) = (2p^2)^{-1}(\bar{d}_{im}\, \bar{d}_{jn} + \bar{d}_{in}\, \bar{d}_{jm} - \bar{d}_{ij}\, \bar{d}_{mn}),$$

with $\bar{d}_{ij} = \delta_{ij} - p_i\, p_j \mathbf{p}^{-2}$. In this gauge the following operator equations hold:

$$\partial_i G_{ij}(x) = 0, \quad G^\mu_\mu(x) = 0. \qquad (22)$$

III. EQUIVALENCE OF FORMALISM IN VARIOUS GAUGES

Define a Hilbert space $\mathscr{H}$ with indefinite metric as the completion of Schwartz space of symmetry tensor valued test functions $\mathscr{S}(\mathbb{R}^4)\times\mathbb{C}^{10}$ with respect to the sesquilinear form

$$\langle f,g\rangle = \sum_{\substack{\mu,\nu \\ \rho,\sigma}}\int\int \overline{f_{\mu\nu}(x)}\, W_{\mu\nu\rho\sigma}(x-y) g_{\rho\sigma}(y) d^4x\, d^4y. \qquad (23)$$

This form becomes positive semidefinite on the closed subspace

$$\mathscr{H}' = \{f = (f_{\mu\nu}) \in \mathscr{H} \,|\, p^\mu \tilde{f}_{\mu\nu}(p) = 0 \text{ a.e on } C_+\},$$

where C_+ is the mantle of the forward light cone, and $\tilde{f}$ is the Fourier transform of f. Let $\mathscr{H}''$ be the kernel of the restricted form[17]

$$\mathscr{H}'' = \{f \in \mathscr{H}' \,|\, \tilde{f}_{\mu\nu}(p) = p_\mu\, \tilde{h}_\nu(p) + p_\nu\, \tilde{h}_\mu(p),$$

$$p^\mu \tilde{h}_\mu(p) = 0, \text{ and } \tilde{h}_\mu(p) \in \mathscr{S}(\mathbb{R}^4)\}.$$

Then we can define the physical one-particle space for the free graviton in covariant gauge as $\mathscr{H}_{\text{ph}} = \mathscr{H}'/\mathscr{H}''$.

It is interesting to note that $\mathscr{H}_{\text{ph}}$ defined above is independent of gauge parameters a and b (see also Ref. 3). In other words, the one-particle physical Hilbert spaces for free gravitons in various covariant gauges considered in Sec. II are the same and coincide with that of the Gupta gauge (hence we shall denote it by $\mathscr{H}_G$). This implies that the physical contents of the theory formulated in different covariant gauges are the same.

The corresponding one-particle space for the graviton in radiation or "Coulomb" gauge is given by

$$\mathcal{H}_C \equiv \mathcal{H}' \cap \{f \in \mathcal{H}' | f_{0\mu} = 0, \quad \mu = 0,1,2,3\}.$$

Note that the elements of $\mathcal{H}_C$ do not transform covariantly as a tensor under the Poincaré group. However, we can still show that this noncovariant formalism is physically equivalent (in the sense expressed below) to the covariant formalism.

Proposition 1: There exists a unitary equivalence $\mathcal{H}_C \cong \mathcal{H}_G = \mathcal{H}'/\mathcal{H}''$ given by the unitary map

$$\gamma : \tilde{f}_{\mu\nu}(p) \to \tilde{f}_{\mu\nu}(p) - \frac{p_\mu \tilde{f}_{0\nu}(p)}{p_0} - \frac{p_\nu \tilde{f}_{\mu 0}(p)}{p_0} - \frac{p_\mu p_\nu \tilde{f}_{00}(p)}{p_0^2}. \tag{24}$$

Proof: First we note that $(\gamma \tilde{f})_{0\mu} = 0, \mu = 0,1,2,3$, for all $f \in \mathcal{H}'$. Therefore $\mathcal{H}_C = \gamma \mathcal{H}'$. But $(I-\gamma)$ maps $\mathcal{H}'$ into $\mathcal{H}''$. γ vanishes on $\mathcal{H}''$ and $\gamma \tilde{h}(p) = 0$ implies $\tilde{h}(p) \in \mathcal{H}''$. Hence $\mathcal{H}''$ is the kernel of γ. Furthermore, γ is well defined, as can be seen by restricting to $\bar{V}^+$ the Taylor expansion about $p = 0$. Thus

$$\tilde{f}(p) \to \gamma \tilde{f}(p) + (1-\gamma)\tilde{f}(p)$$

defines a unique decomposition $\mathcal{H}' = \mathcal{H}_C \oplus \mathcal{H}''$ or $\mathcal{H}_C \cong \mathcal{H}_G = \mathcal{H}'/\mathcal{H}''$. Q.E.D.

We can generalize the above result to Borchers' field algebra[18] for the free graviton. Denote by $\mathcal{S}_n$ the space of the n-fold tensor product of $\mathcal{S}(\mathbb{R}^4) \times \mathbb{C}^{10}$, with $\mathcal{S}_0 = \mathbb{C}$ corresponding to the subspace of the vacuum state and for $n \geqslant 1$ the elements of $\mathcal{S}_n$ are symmetric tensor-valued test functions $f^{(n)}(x) = (f^{(n)}_{(\mu_1\nu_1),\dots,(\mu_n\nu_n)}(x_1,\dots,x_n))$ symmetric in x and satisfying for $1 \leqslant j \leqslant n$,

$$f^{(n)}_{(\mu_1\nu_1),\dots,(\mu_j\nu_j),\dots,(\mu_n\nu_n)}(x_1,\dots,x_n) = f^{(n)}_{(\mu_1\nu_1),\dots,(\nu_j\mu_j),\dots,(\mu_n\nu_n)}(x_1,\dots,x_n).$$

Denote by $\mathfrak{a}$ the locally convex direct sum of these spaces, $\overset{n}{\oplus} \mathcal{S}_n$. If $\mathfrak{a}$ is equipped with the product defined by

$$(\mathbf{f} \times \mathbf{g})^{(n)}(x_1,\dots,x_n) = \sum_{j=0}^{n} f^{(j)}(x_1,\dots,x_j) g^{(n-j)}(x_{j+1},\dots,x_n)$$

for all $\mathbf{f},\mathbf{g} \in \mathfrak{a}$, and the involution * defined by

$$(\mathbf{f}^*)^{(n)}(x_1,\dots,x_n) = \overline{f^{(n)}(x_n,\dots,x_1)},$$

where the overbar denotes complex conjugate, then $\mathfrak{a}$ is a * test function (or Borcher's) algebra, carrying a natural topology induced by the Schwartz topology of the spaces $\mathcal{S}_n$

The two-point function $W_{\mu\nu\rho\sigma}$ defines a positive (semidefinite) linear functional on $\mathfrak{a}$ if one imposes the following transversality condition on $\mathfrak{a}$:

$$(p \cdot \tilde{\mathbf{f}})^{(n)}(p_1,\dots,p_n) = (p_j)^{\mu_j} \tilde{f}_{\mu_1,\dots,\mu_j,\dots,\mu_n}(p_1,\dots,p_n) = 0,$$
$$\forall p_j \in \bar{V}_+, j = 1,\dots,n; \text{ and } \mathbf{f} \in \mathfrak{a}. \tag{25}$$

We shall denote by $\mathfrak{a}_1$ the Borcher's algebra satisfying Eq. (25). The the two-sided ideal $\mathcal{I}$ of $\mathfrak{a}_1$ is contained in the kernel of W_n, the n-point functions of $G_{\mu\nu}$, and is given by

$$\mathcal{I} = \mathfrak{a}_1 \cap \{(\tilde{\mathbf{f}})^{(0)} = 0 \text{ and} (\tilde{\mathbf{f}})^{(n)} = \tilde{f}^{(n)}_{(\mu_1\nu_1),\dots,(\mu_j\nu_j),\dots,(\mu_n\nu_n)}(p_1,\dots,p_j,\dots,p_n)$$
$$= (p_{\mu_j} \tilde{h}_{\mu_j}(p) + p_{\nu_j} \tilde{h}_{\nu_j}(p)) \tilde{f}_{(\mu_1\nu_1),\dots,\widehat{(\mu_j\nu_j)},\dots,(\mu_n\nu_n)}(p_1,\dots,\hat{p}_j,\dots,p_n),$$
$$\text{for at least one } j, \text{ and } p^{\mu_j} \tilde{h}_{\mu_j}(p) = 0 \forall h_{\mu_j} \in \mathcal{S}\}. \tag{26}$$

The physical test function algebra for the free graviton in covariant gauge is then given by the quotient algebra $\mathfrak{a}_G = \mathfrak{a}_1/\mathcal{I}$. Then through the Gel'fand–Naimark–Segal construction[19] the positive linear functionals on $\mathfrak{a}_G$ determine a unique theory for the gravitational potential in covariant gauge.

The corresponding test function algebra for the graviton in the "Coulomb" gauge is then given by

$$\mathfrak{a}_C = \mathfrak{a}_1 \cap \{\mathbf{f} \in \mathfrak{a} | (\mathbf{f})^{(n)} = f^{(n)}_{\mu_1,\dots,\mu_n} = 0 \text{ if any } \mu_j = 0 \forall j\}.$$

Now let Γ denote the natural algebraic generalization of the map γ defined in Proposition 1, we then have

$$\mathfrak{a}_C = \text{Range}[\Gamma(\mathfrak{a}_1)],$$

which leads to

Proposition 2: Γ defines a *-algebraic isomorphism

$$\mathfrak{a}_C \cong \mathfrak{a}_G = \mathfrak{a}_1/\mathcal{I}.$$

IV. EUCLIDEAN GRAVITATIONAL POTENTIALS IN COVARIANT GAUGES

Euclidean formulation of the linearized gravitational potential in vacuum as a generalized Gaussian Markov field was first considered by Lim[20] and Guerra.[21] In this section we shall give a more detailed discussion of Euclidean gravitational potentials in various covariant gauges, and we shall see how this can lead to the noncovariant case, which will be given in the next section.

The Euclidean (or Schwinger) two-point function for the covariant gravitational potential can be obtained by applying a matrix transformation, in addition to the usual analytic continuation to pure imaginary time, to the relativistic two-point function

$$S_{ijmn}(x_E - y_E) = (A_{i\mu} A_{j\nu} - \tfrac{1}{4}\delta_{ij} g_{\mu\nu})(A_{m\rho} A_{n\sigma} - \tfrac{1}{4}\delta_{mn} g_{\rho\sigma}) \times W^{\mu\nu\rho\sigma}(\mathbf{x} - \mathbf{y}, (i(x_0 - y_0)), \tag{27}$$

where $A_{i\mu} = 1$ for $i = \mu = 1,2,3$; $A_{40} = i$ and $A_{i\mu} = 0$ otherwise, and x_E, y_E are the Euclidean 4-vectors. The matrix transformation is necessary not only to change all $g_{\mu\nu}$ into δ_{ij}, it also preserves the tracelessness of the factor $(g_{\mu\rho} g_{\nu\sigma} + g_{\mu\sigma} g_{\nu\rho} - g_{\mu\nu} g_{\rho\sigma})$ by changing it to $(\delta_{im}\delta_{jn} + \delta_{in}\delta_{jm} - \tfrac{1}{2}\delta_{ij}\delta_{mn})$. The Fourier transform, of S_{ijmn} then has the following general form:

$$\tilde{S}_{ijmn}(p_E) = \frac{1}{2p_E^2}\Big((\delta_{im}\delta_{jn} + \delta_{in}\delta_{jm} - \tfrac{1}{2}\delta_{ij}\delta_{mn}) + \frac{\mathcal{F}_1(p_E^2)}{p_E^2}(\delta_{ij} p_m p_n + \delta_{mn} p_i p_j)$$

$$+\frac{\mathcal{F}_2(p_E^2)}{p_E^2}(\delta_{im}\,p_j\,p_n+\delta_{in}\,p_j\,p_m+\delta_{jm}\,p_i\,p_n$$

$$+\delta_{jn}\,p_i\,p_m)+\frac{\mathcal{F}_3(p_E^2)}{p_E^4}p_i\,p_j\,p_m\,p_n\Big). \qquad (28)$$

In order for S_{ijmn} to be the Schwinger function for some Euclidean field, we require $\mathcal{F}_i(p_E^2)$, $i=1,2,3$, to be nonnegative measurable functions.

Define a Hilbert space $\mathcal{H}$ as the completion of the real symmetric tensor-valued test function space $\mathcal{S}(\mathbb{R}^4)\times\mathbb{R}^{10}$ with respect to the topology given by the inner product

$$\langle f,g\rangle_{\mathcal{H}}=\sum_{i,j,m,n}\iint f_{ij}(x_E)S_{ijmn}(x_E-y_E)g(y_E)d^4x_E\,d^4y_E.$$

We can then define the Euclidean gravitational potential in the usual manner.

Definition 2: The Euclidean gravitational potential $\mathcal{G}$ with gauge functions $\mathcal{F}_i(p_E^2)$, $i=1,2,3$ is the real Gaussian random field over $\mathcal{H}$ with mean zero and covariance given by

$$E[\mathcal{G}(f)\mathcal{G}(g)]=\langle f,g\rangle_{\mathcal{H}}.$$

The Euclidean one-particle space is then given by the quotient space $\mathcal{H}/\text{kernel}\|\cdot\|_{\mathcal{H}}$. It is clear from our definition that $\mathcal{G}$ transforms covariantly under the Euclidean group. However, it does not satisfy the reflection property.

Proposition 3: $\mathcal{G}$ is nonreflexive.

Proof: Note that the $4j-4n$ component (i.e., S_{4j4n}) of the two-point Schwinger function contains terms with factors $p_4^2\,p_E^{-4}$ and $p_4^4\,p_E^{-6}$, which allow test functions of finite $\mathcal{H}$ norm localized at the hyperplane $x_4=0$ of the form $f_{ij}(x^E)=\mathbf{f}_{ij}(\mathbf{x})\otimes\delta(x_4)$, with $f_{4j}\neq 0$ and $\mathbf{f}_{ij}(\mathbf{x})\in\mathcal{S}(\mathbb{R}^3)$. Clearly, for such a test function $\theta f_{4j}(x_E)=-f_{4j}(x_E)$, where θ is the unitary time-reflection operator. Therefore, we conclude that $\theta\mathcal{G}(f)\theta^{-1}\neq\mathcal{G}(f)$. Q.E.D.

This result is as expected because in Nelson's theory the reflection property (together with the Markov property and other regularity conditions) is essential for a Euclidean field to have a Wightman theory in the Minkowski region. For the covariant gravitational potential the relativistic Hilbert space has an indefinite metric, hence it cannot be a Wightman theory (see additional remarks on this point at the end of this section).

Definition 3: A covariant gauge for the Euclidean gravitational potential is called a Markov gauge if the matrix inverse of the Fourier transform of its two-point Schwinger function is a polynomial in p_E^2 and the components p_j.

This definition includes many interesting gauges such as Gupta and harmonic gauges as the Markov gauge. The reason for such a definition is clear from the following result.

Proposition 4: The Euclidean gravitational potential $\mathcal{G}$ in the Markov gauge satisfies Nelson's Markov property

Proof: If we write Eq. (27) as a matrix equation $\widetilde{S}(p_E)=A\widetilde{W}(\mathbf{p},ip_0)$, then $\widetilde{S}^{-1}(p_E)=\widetilde{W}^{-1}(\mathbf{p},ip_0)A^{-1}$. Since A is just a constant nonsingular matrix, its inverse is again another constant matrix. From Eqs. (12), (14) and (15) we get $\widetilde{W}^{-1}(\mathbf{p},ip_0)=\widetilde{Q}(p,ip_0)$ as a matrix polynomial in p_E and p_j. Now Q is just a matrix local differential operator hence the argument of Nelson for the scalar case applies (see Refs. 10 and 11).

We note that Definition 3 does not include Landau-like gauges as Markov gauges. However we shall show that for such gauges the Euclidean gravitational potential is also Markovian in Nelson's sense.

Proposition 5: The Euclidean gravitational potential $\mathcal{G}$ in the Landau-like gauge satisfies Nelson's Markov property.

Proof: The correct two-point Schwinger functions in Landau-like gauges are obtained in a slightly different way

$$S_{ijmn}(x_E-y_E)=A_{i\mu}\,A_{j\nu}\,A_{m\rho}\,A_{n\sigma}\,W^{\mu\nu\rho\sigma}(\mathbf{x}-\mathbf{y},i(x_0-y_0)),$$

where $A_{i\mu}$ is defined in the same way as before. Then we obtain for the two cases of Landau-like gauges the following:

$$\text{(i) }\widetilde{S}_{ijmn}(p)=\frac{1}{2p_E^2}\Big(d_{im}\,d_{jn}+d_{in}\,d_{jm}-\delta_{ij}\,\delta_{mn}-\frac{2p_i\,p_j\,p_m\,p_n}{p_E^4}\Big),$$

where $d_{ij}=\delta_{ij}-p_i\,p_j\,p_E^{-2}$. Like in the Minkowski region, we have

$$\sum_i(p_i-p_j\,\delta_{ij})\widetilde{S}_{ijmn}(p_E)=0.$$

$$\text{(ii) }\widetilde{S}_{ijmn}(p_E)=\frac{1}{2p_E^2}\Big[d_{im}\,d_{jn}+d_{in}\,d_{jm}-\tfrac{2}{3}\,d_{ij}\,d_{mn}\Big],$$

which satisfies

$$\sum_i p_i\,\widetilde{S}_{ijmn}(p_E)=0.$$

For Case (i) we shall consider one-particle space with positive metric as $\mathcal{H}_1\subset\mathcal{H}$ with elements satisfying $\Sigma_i[\,p_i\,\tilde f_{ij}(p_E)-\frac12 p_j\,\tilde f_{ii}(p_E)]=0$. Then for any element $h\in C^\infty(\mathcal{O})\times\mathbb{R}^{10}$, where $\mathcal{O}\subset\mathbb{R}^4$ is an open set,

$$\sum_i p_i\,\widetilde{S}_{ijmn}(p_E)\tilde h_{mn}(p_E)-\tfrac12\sum_i p_j\,\widetilde{S}_{iimn}(p_E)\tilde h_{mn}(p_E)=0,$$

i.e., S_{ijmn} maps every element of $C^\infty(\mathcal{O})\times\mathbb{R}^{10}$ into an element of $\mathcal{H}_1(\mathcal{O})$. Furthermore, for $f,g\in\mathcal{H}$,

$$\langle\tilde f,\tilde g\rangle_{\mathcal{H}}=\sum_{i,j,m,n}\int\tilde f_{ij}(p_E)\frac{\delta_{im}\,\delta_{jn}}{p_E^2}\tilde g_{mn}(p_E)\,d^4p_E$$

$$=\sum_{i,j}\int|\tilde f_{ij}(p_E)|^2\frac{d^4p_E}{p_E^2}.$$

Clearly, we can again apply Nelson's argument to complete the proof.

The proof is similar for case (ii). We consider the one-particle space with positive metric as $\mathcal{H}_2\subset\mathcal{H}$ with elements satisfying $\Sigma_i\,p_i\,\tilde f_{ij}(p_E)=0$ and $\Sigma_i\,\tilde f_{ii}(p_E)=0$. Again for any $h\in C^\infty(\mathcal{O})\times\mathbb{R}^{10}$,

$$\sum_i p_i \tilde{S}_{ijmn}(p_E)\tilde{h}_{mn}(p_E) = 0 \text{ and}$$

$$\sum_i S_{iimn}(p_E)\tilde{h}_{mn}(p_E) = 0.$$

Thus S_{ijmn} maps every $h \in C^\infty(\mathcal{O}) \times \mathbb{R}^{10}$ into an element in $\mathcal{K}_2(\mathcal{O})$. Finally we note that for any $f,g, \in \mathcal{K}_2$,

$$\langle f,g\rangle_{\mathcal{K}} = \sum_{i,j} \int |\tilde{f}_{ij}(p_E)|^2 \frac{d^4 p_E}{p_E^2},$$

so that Markovicity of the field follows by Nelson's argument. Q.E.D.

Before we discuss the Euclidean gravitational potential in noncovariant radiation gauge, some remarks on the covariant case will be given. First we note that the failure of the reflection property is an essential feature of the Euclidean gravitational potential in a covariant gauge. The effect of the reflection property may be considered so as to prevent the theory from being too regular in its ultraviolet behavior. It can be seen in Nelson's theory that the reflection property excludes scalar boson fields with covariance functions of the form $(-\Delta + m^2)^n$, $n > 1$, which are regularized propagators without ultraviolet divergences.[22] However, such fields either give rise to Hilbert spaces with indefinite metric or nonlocal theories without unphysical states, thus they cannot be Wightman theories. Furthermore, the reflection property is necessary for the proof of the self-adjointness of the free Hamiltonian; however, this can still be achieved if one restricts to physical space with positive metric.

Since both free Euclidean photon and graviton potentials in covariant gauges are Markovian and nonreflexive, we want to find a more general property to describe these Euclidean massless fields in covariant gauges, yet still exclude nonlocal theories such as those with propagator of the form $(-\Delta + m^2)^n$, $n > 1$. This can be done as follows. Let (Ω,Σ,μ) be the underlying probability space for the Euclidean random field Φ with one-particle space $\mathcal{K}$. For any open set $\mathcal{O} \subset \mathbb{R}^4$, let $\mathcal{K}(\mathcal{O})$ be the closed subspace generated by $\{f \in \mathcal{K} | \operatorname{supp} f \subset \mathcal{O}\}$, and let $\Sigma_{\mathcal{O}}$ be the σ algebra generated by $\{\Phi(f) | f \in \mathcal{K}(\mathcal{O})\}$. Denote by $\mathcal{K}^0(\mathcal{O})$ the subspace of $\mathcal{K}(\mathcal{O})$, consisting of measures, and let $\Sigma^0_{\mathcal{O}}$ be the Borel σ ring generated by $\{\Phi(f) | f \in \mathcal{K}^0(\mathcal{O})\}$. For any subset $\mathcal{M} \subset \mathbb{R}^4$, let $\Sigma^0_{\mathcal{M}}$ be the intersection $\cap\{\Sigma^0_{\mathcal{O}} | \mathcal{O} \supset \mathcal{M}, \mathcal{M} \text{ open}\}$. Then we have

Definition 4: The Euclidean field Φ is said to satisfy the classical Markov property if, for every function $\mathcal{E}: \Omega \rightarrow \mathbb{R}$ which is $\Sigma_{\mathcal{O}}$ measurable and for every open set $\mathcal{O} \subset \mathbb{R}^4$

$$E[\mathcal{E}|\Sigma_{\mathcal{O}'}] = E[\mathcal{E}|\Sigma^0_{\partial\mathcal{O}}]$$

is valid, where $\mathcal{O}'$ is the complement of $\mathcal{O}$ in $\mathbb{R}^4$, $\partial\mathcal{O}$ is the boundary of $\mathcal{O}$ and $E[\cdot|\cdot]$ is the conditional expectation.

The Euclidean electromagnetic and gravitational potentials in various covariant Markov gauges satisfy the classical Markov property. In fact, for these massless fields, the classical and Nelson's definition of the Markov property coincide since $\Sigma^0_{\partial\mathcal{O}}$ coincides with

$$\Sigma_{\partial\mathcal{O}} = \cap\{\Sigma_{\mathcal{O}_i} | \mathcal{O}_i \subset \partial\mathcal{O}\}.$$

Actually one can also work with the Euclidean gravitational field tensor $\mathcal{R}_{ijmn}$ (for the linearized theory in vacuum), which is related to $\mathcal{G}_{ij}$ by

$$\mathcal{R}_{ijmn} = \tfrac{1}{2}[\partial_j\partial_m \mathcal{G}_{in} - \partial_j\partial_n \mathcal{G}_{im} + \partial_i\partial_n \mathcal{G}_{jm} - \partial_i\partial_m \mathcal{G}_{jn}].$$

One can show that $\mathcal{R}_{ijmn}$ satisfies Nelson's Markov property by using an argument similar to that given in Refs. 23 and 24 (see also Ref. 25 for a simpler proof). Now $\mathcal{R}_{ijmn}$ is reflexive and it leads to a Wightman theory in the Minkowski region. However, the underlying probability space for $\mathcal{R}_{ijmn}$ is in general smaller than that for $\mathcal{G}_{ij}$ because all the elements of the σ algebra generated by $\mathcal{G}_{ijmn}$ correspond to physical states in relativistic theory, whereas the elements of the σ algebra generated by $\mathcal{R}_{ij}$ also correspond to unphysical states in addition to the physical ones.

V. EUCLIDEAN GRAVITATIONAL POTENTIAL IN RADIATION GAUGE

A natural question which arises from the above discussion is whether there exists a suitable subspace of $\mathcal{K}$ for which both the reflection and Markov property are fulfilled. The answer to this question leads to the consideration of the Euclidean gravitational potential in noncovariant "Coulomb" or radiation gauge.

Consider the subspace $\mathcal{K}_2$ of $\mathcal{K}$ considered in the previous section (see Proposition 5) with elements satisfying the tracelessness and transversality conditions $\Sigma_i f_{ii} = 0$ and $\Sigma_i \partial_i f_{ij} = 0$. We can define the one-particle Hilbert space for Euclidean gravitational potential in "Coulomb" gauge as the closed subspace

$$\mathcal{K}_C = \mathcal{K}_2 \cap \{f \in \mathcal{K} | f_{4j} = 0, \quad j = 1,2,3,4\}.$$

Note that we can decompose $\mathcal{K}_2$ as follows:

$$\mathcal{K}_2 = \mathcal{K}_C \oplus \mathcal{K}_L,$$

where $\mathcal{K}_L$ is the subspace of "longitudinal" elements of the form $(p_i \tilde{f}_{4j} + p_j \tilde{f}_{i4})p_4^{\ -1} - p_i p_j \tilde{f}_{44} p_4^{\ -2}$. Clearly such a decomposition is not Euclidean invariant.

Proposition 6: The Euclidean gravitational potential in radiation gauge $\mathcal{G}_C$ satisfies the reflection property.

Proof: The proof is straighforward, which follows from the fact that $\mathcal{G}_{C4j} = 0$ for $j = 1,2,3,4$, and the definition of the reflection property.

Q.E.D.

This result agrees with our earlier remark that the reflection property is closely related to the positivity of the metric of the relativistic one-particle space, which for the case of $\mathcal{G}_C$ is positive.

Proposition 7: The Euclidean gravitational potential in radiation gauge $\mathcal{G}_C$ satisfies the Markov property with respect to the half-spaces bounded by $x_4 =$ constant.

Proof: Let $E_\pm$ and E_0 be the projections onto the subspaces of $\mathcal{K}_2$ with supports in the hyperplanes $\mathbb{R}^4_\pm = \{\mathbb{R}^4 | x_4 \lessgtr 0\}$ and $R^4_0 = \{\mathbb{R}^4 | x_4 = 0\}$, respectively, and $E^C_\pm$ and E^C_0 the corresponding projections in $\mathcal{K}_C$. Let θ be the unitary time-reflection operator

$$\theta: \tilde{f}_{ij}(p) \rightarrow (-1)^{\delta_{i4}+\delta_{j4}} \tilde{f}_{ij}(\mathbf{p}, -p_4).$$

Then to show that $\mathcal{G}_C$ satisfies the Markov property with respect to the half-spaces $x_4 = 0$, one needs to show that $E_+^C \theta E_+^C$ is a projection.[12,26]

Consider the decomposition

$$E_+ \theta E_+ = E_+^C \theta E_+^C \oplus E_+^L \theta E_+^L.$$

Then for $f \in \mathscr{K}_2$

$$\langle f, E_+ \theta E_+ f \rangle_{\mathscr{K}} = \langle f_C^+, \theta f_C^+ \rangle_{\mathscr{K}} + \langle f_L^+, \theta f_L^+ \rangle_{\mathscr{K}}.$$

First we want to show that $\langle f_L^+, \theta f_L^+ \rangle_{\mathscr{K}} = 0$. Since each $f_L \in \mathscr{K}_L$ is of the form $(p_i \tilde{f}_{4j} + p_j \tilde{f}_{i4}) p_4^{-1} - p_i p_j \tilde{f}_{44} p_4^{-2}$, and $\mathscr{K}_L \subset \mathscr{K}_2$, so using the tracelessness condition we get

$$\sum_i \frac{p_i \tilde{f}_{44}}{p_4^2} = 2 \sum_i \frac{\tilde{f}_{i4}}{p_4}. \quad (29)$$

Therefore,

$$\langle \tilde{f}_L^+, \theta \tilde{f}_L^+ \rangle_{\mathscr{K}}$$

$$= \sum_{i,j} \int_{\mathbf{R}^4} \frac{d^4 p_E}{p_E^2} \left[\left(\frac{p_i \tilde{f}_{4ij}(p_E) + p_j \tilde{f}_{i4}^+(p_E)}{p_4} - \frac{p_i p_j \tilde{f}_{44}^+(p_E)}{p_4^2} \right) \theta \left(\frac{p_i \tilde{f}_{4j}^+(p_E) + p_j \tilde{f}_4^+(p_E)}{p_4} - \frac{p_i p_j \tilde{f}_{44}^+(p_E)}{p_4^2} \right) \right]$$

$$= \sum_{i,j} \int_{\mathbf{R}^4} \frac{d^4 p_E}{p_E^2} \left(\frac{p_i \tilde{f}_{4j}^+(p_E) + p_j \tilde{f}_{i4}^+(p_E)}{p_4} - \frac{p_i p_j \tilde{f}_{44}^+(p_E)}{p_4^2} \right) \left(\frac{-p_i \tilde{f}_{4j}^+(\mathbf{p}, -p_4) - p_j \tilde{f}_{i4}^+(\mathbf{p}, -p_4)}{p_4} - \frac{p_i p_j \tilde{f}^+(\mathbf{p}, -p_4)}{p_4^2} \right)$$

$$= -\sum_{i,j} \int_{\mathbf{R}_0^4} \frac{d^4 p_E}{p_E^2} \left[\left(\frac{p_i \tilde{f}_{4j}^+(p_E) + p_j \tilde{f}_{i4}^+(p_E)}{p_4} \right)^2 - \left(\frac{p_i p_j \tilde{f}_{44}^+(p_E)}{p_E^2} \right)^2 \right]$$

$$= -\sum_{i,j} \int_{\mathbf{R}_0^4} \frac{d^4 p_E}{p_E^2} \left[\left(\frac{2 p_j \tilde{f}_{i4}^+(p_E)}{p_4} \right)^2 - \left(\frac{2 p_j \tilde{f}_{i4}^+(p_E)}{p_4} \right)^2 \right] = 0,$$

where we have used Eq. (29) for the last step. Thus we have obtained $E_+ \theta E_+ = E_+^C \theta E_+^C$. Using the fact that $\mathscr{K}_C$ is invariant under θ, it is not difficult to show that $E_+^C \theta E_+^C \geqslant 0$. Finally we want to show that $(E_+^C \theta E_+^C)^2 = E_+^C \theta E_+^C$. Again, making use of the time-reflection invariance of $\mathscr{K}_C$,

$$\langle f_C^+, (E_+^C \theta E_+^C)^2 f_C^+ \rangle_{\mathscr{K}} = \langle f_C^+, \theta E_+^C \theta f_C^+ \rangle_{\mathscr{K}}$$

$$= \langle f_C^+, \theta E_+^C f_C^+ \rangle_{\mathscr{K}}$$

$$= \langle f_C^+, E_+^C \theta E_+^C f_C^+ \rangle_{\mathscr{K}}.$$

Then, following the same argument as given in Ref. 12, one obtains the result $E_+^C \; E_-^C = E_0^C$.

Q.E.D.

Note that for the construction of relativistic fields the Markov property is needed for open half-spaces $\{x_E | x_4 \lessgtr s\}$ only. Actually Nelson's Markov property is a rather strong condition and up to now there still does not exist any nontrivial model which satisfies this property. Thus the correct condition for Euclidean fields seems to be the Osterwalder–Schrader positivity,[27,29] which is weaker than Nelson's Markov property (for further details on this point see Refs. 26 and 29).

One can recover the relativistic one-particle space $\mathscr{H}_C$ from $\mathscr{K}_C$ in a similar way as given by Osterwalder and Schrader (see Refs. 26 and 28). We note that $\langle \theta f_C^+, g_C^+ \rangle_{\mathscr{K}}$ defines a positive semidefinite form on $\mathscr{K}_C^+ \times \mathscr{K}_C^+$, This is none other than the positivity condition of Osterwalder and Schrader. If we denote by $\mathscr{H}_C^r$ the real $\mathscr{H}_C$, then we have

Proposition 8: $\mathscr{H}_C^r$ is isomorphic to the closure of $\mathscr{K}_+^C$ / kernel$\|\cdot\|_{\mathscr{K}_+^C}$, with the topology given by $\langle f_+, g_+ \rangle_{\mathscr{K}_+^C} = \langle \theta f_+, g_+ \rangle_{\mathscr{K}}$.

Now we can establish the Feynman–Kac–Nelson formula for the free gravitational potential in radiation gauge. Let $\mathscr{L}^2(M)$ be the space of square-integrable functions on the sample space M of $\mathscr{G}_C$, which are measurable with respect to the σ algebra generated by $\{\mathscr{G}_C(f) | f \in M\}$. Denote by J the projection of $\mathscr{L}^2(\mathscr{K}_2)$ onto $\mathscr{L}^2(\mathscr{H}_C^r)$, and T_s the induced unitary action of time translation in $\mathscr{K}_2$ with $T_s f(\bar{x}, x_4) = f(\mathbf{x}, x_4 - s)$. Then we have

Proposition 9:

$$e^{-sH_0} = J T_s u, \quad u \in \mathscr{L}^2(\mathscr{H}_C^r),$$

and H_0 is the free Hamiltonian in $\mathscr{L}^2(\mathscr{H}_C^r)$.

Proof: This can be considered as a result for $\mathscr{K}_C$ rather than $\mathscr{K}_2$ because $\mathscr{H}_C^r \subset \mathscr{K}_C$ and $\mathscr{K}_C$ is closed in $\mathscr{K}_2$ and also closed under time translation as well as under complex conjugation. Consider those $f \in \mathscr{K}_C$ which are C^∞ functions with compact support such that their Fourier transforms and their derivatives vanish if $|p_E| < \epsilon$ for some $\epsilon > 0$. Such f are dense in $\mathscr{K}_C$. If f and g are two such functions, then for each real s, the functions $f(\mathbf{x}) \to f(\mathbf{x}, s)$, $g(\mathbf{x}) \to g(\mathbf{x}, t)$ are in $\mathscr{H}_C^r$. Then by a direct computation

$$\langle f, g \rangle_{\mathscr{K}}$$

$$= \sum_{i,j} \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} \tfrac{1}{2} [e^{-|t-s| \, |\mathbf{p}|} f_{ij}(\mathbf{x}, s) g_{ij}(\mathbf{x}, t)]_{\mathscr{H}_C^r} \, dt \, ds.$$

Since $\mathscr{K}_C$ is the closure of the space of such functions, it follows that the time-translation group in $\mathscr{K}_C$ is the minimal unitary extension of the semigroup $e^{-t|\mathbf{p}|}$ in $\mathscr{H}^r_C$. The rest of the proof follows just as in the scalar case (see Ref. 13).

Q.E.D.

Finally we should like to consider whether or not a result similar to Proposition 1 holds in the Euclidean region. This is in fact the case if we define the Euclidean one-particle "physical" space for the gravitational potential in covariant gauge as $\mathscr{H}_G = \mathscr{H}'/\mathscr{H}''$, where $\mathscr{H}'$ is the closed subspace of $\mathscr{H}$ with vanishing four-divergence $\Sigma_i \partial_i f_{ij}(x_E) = 0$, and $\mathscr{H}''$ is the subspace of vanishing norm. Then we have

Proposition 10: There exists a unitary map given by

$$\gamma_E: \tilde{f}_{ij} \to \tilde{f}_{ij} - \frac{p_i \tilde{f}_{4j}}{p_4} - \frac{p_j \tilde{f}_{i4}}{p_4} - \frac{p_i p_j \tilde{f}_{44}}{p_4^2},$$

which defines a unitary equivalence $\mathscr{H}_C \cong \mathscr{H}_G = \mathscr{H}'/\mathscr{H}''$. The proof is similar to Proposition 1, therefore we shall omit it. This result can be generalized to Euclidean field algebra (or Schwinger algebra) in the same manner as in the relativistic case.

VI. CONCLUSION

The Euclidean formulation of the linearized gravitational potential in covariant gauges seems to have some nice features. The difficulties due to gauge problems do not arise in the Euclidean field, here the covariance and locality (in the sense of the Markov structure) properties are compatible with the positive metric of Euclidean one-particle space. Even in noncovariant radiation gauge, the Euclidean gravitational potential does have some kind of local structure in the form of the Markov property with respect to special half-spaces. It would be interesting to see whether the Euclidean method can be of any use in other gauge theories such as the Yang–Mills field.

ACKNOWLEDGMENTS

The author would like to thank Professor Abdus Salam, the International Atomic Energy Agency, and UNESCO for hospitality at the International Centre for Theoretical Physics, Trieste. Thanks are also due to Professor R.F. Streater for stimulating discusison, and to Professor F. Guerra for correspondence.

[1]L. Bracci and F. Strocchi, J. Math. Phys. **13**, 1151 (1972).
[2]L. Bracci and F. Strocchi, Commun. Math. Phys. **24**, 289 (1972).
[3]L. Bracci and F. Strocchi, J. Math. Phys. **16**, 2522 (1975).
[4]R. Streater and A. Wightman, *PCT, Spin and Statistics and All That* (Benjamin, New York, 1964).
[5]S. Gupta, Proc. Phys. Soc. (London) A **63**, 681 (1950).
[6]K. Blueler, Helv. Phys. Acta. **23**, 567 (1950).
[7]R. Ferrari, L. Picasso, and F. Strocchi, Nuovo Cimento A **39**, 1 (1974).
[8]F. Strocchi and A. Wightman, J. Math. Phys. **15**, 2198 (1974).
[9]S.C. Lim, "Free Electromagnetic Potentials in Minkowski and Euclidean Regions," to appear in J. Phys. A (1979).
[10]E. Nelson, J. Func. Anal. **12** 97 (1973).
[11]E. Nelson, J. Func. Anal. **12** 211 (1973).
[12]G.C. Hegerfeldt, Commun. Math. Phys **35**, 155 (1974).
[13]B. Simon, *The $P(\phi)_2$ Euclidean (Quantum) Field Theory* (Princeton U.P., Princeton, N.J., 1974).
[14]L.D. Landau and E.M. Lifshitz, *Classical Theory of Fields* (Pergamon, New York, 1975), 4th ed.
[15]E.S. Fradkin and I.V. Tyutin, Phys. Rev. D **2**, 2841 (1970).
[16]G. 't Hooft and M. Veltman, Ann. Inst. Inst. Henri Poincaré A **20**, 69 (1974).
[17]J. Bertrand, Nuovo Cimento A **23** 703 (1974).
[18]H.J. Borchers, Nuovo Cimento **24**, 214 (1962).
[19]G.G. Emch, *Algebraic Methods in Statistics Mechanics and Quantum Field Theory* (Wiley-Interscience, New York, 1972).
[20]S.C. Lim, J. Phys. A **9**, 1751 (1976).
[21]F. Guerra (private communication).
[22]A. Pais and G.E. Unlenbeck, Phys. Rev. **79**, 145 (1950).
[23]S.C. Lim, thesis, London University (1975).
[24]T.H. Yao, J. Math. Phys. **17**, 241 (1976).
[25]S.C. Lim (paper in preparation).
[26]J.L. Challifour, J. Math. Phys. **10**, 1889 (1976).
[27]K. Osterwalder and R. Schrader, Commun. Math. Phys. **31**, 83 (1973).
[28]K. Osterwalder and R. Schrader, Commun. Math. Phys. **42**, 281 (1975).
[29]A. Klein, J. Func. Anal. **27**, 277 (1978).

The classification of all $\mathscr{H}$ spaces admitting a Killing vector

J. D. Finley, III[a)]
Department of Physics and Astronomy, The University of New Mexico, Albuquerque, New Mexico, U.S.A.
J. F. Plebański[b)]
Departamento de Física, Centro de Investigación y de Estudios Avanzados, Instituto Politécnico Nacional, México 14, D.F., Mexico

(Received 2 March 1979; accepted for publication 27 April 1979)

We show that all $\mathscr{H}$ spaces (self-dual solutions of the complex Einstein vacuum equations) that admit (at least) one Killing vector may be gauged in such a way as to be divided into only five types, characterized by the type of equation which determines their potential function. In four of these types we show that this knowledge is sufficient to reduce the requirement of being an $\mathscr{H}$ space to a linear equation whose solutions are well known. The fifth case is reduced considerably and a large class of special solutions is given.

I. INTRODUCTION

This paper is a continuation of our studies on the structure of heavens ($\mathscr{H}$ spaces). We recall that these $\mathscr{H}$ spaces are solutions of the *complex* vacuum Einstein equations with a Riemannian curvature whose anti-self-dual part vanishes. These spaces have been studied extensively (by different approaches) by groups associated with Newman,[1] Penrose,[2] and Plebański.[3] We follow here the notation and approach of Boyer and Plebański.[4]

In coordinates $\{x,y,u,v\}$, all $\mathscr{H}$ spaces may be thought of as determined by solutions to the single equation[3] for the potential function Θ:

$$\Theta_{xx}\Theta_{yy} - (\Theta_{xy})^2 + \Theta_{xu} + \Theta_{yv} = 0. \tag{1.1}$$

This is clearly a difficult equation, whose general solution is not known, although many interesting families of solutions are in fact known.[5-7] It has also been shown[4] that the general solution is determined by *two* arbitrary functions of *three* complex variables.

Quite often in the past, solutions of Einstein's equations have been generated by the desire to have particular symmetries. Therefore, it is very interesting to better understand the relation between the solutions and the allowed Killing vectors. In Ref. 8 a single master equation has been given which gives the required correlation between the potential Θ and any allowed Killing vector. However, as is usual in problems in general relativity, the quantities in it may be subjected to various gauge conditions. This fact causes unknown functions of two variables to appear in the master equation which, in any given case, could be gauged away. In particular, we know that there are certainly not more than ten Killing vectors in our (four-dimensional) space even though there appear to be arbitrary functions in the master equation. These functions merely indicate the gauge freedom available in determining a space by giving a potential Θ in a specific set of coordinates.

Therefore, in this article we first look in detail at the group of gauge transformations[9] which leave invariant the form of the tetrad in an $\mathscr{H}$ space and, thereby, the form of all results obtained from it. Then we utilize these gauge transformations to separate out the distinct kinds of single Killing vectors allowed. This is basically a quotient of the infinite group of symmetries of the manifold possessing a single Killing vector by the infinite group of gauge transformations. This quotient results in a *finite* number of distinct types. We then consider, in turn, each of these types and find the constraint which the existence of a single Killing vector of the type specified puts on the potential function. Incorporation of this information into Eq. (1.1) allows an explicit determination of the class of allowed solutions in all cases but one. We show how these solutions may be determined and discuss in some detail the one irresolvable case. In each of the cases we point out the (complex) Petrov types which are allowed. We also show that this is very closely related to current questions of interest involving Yang–Mills and gravitational instantons.

II. TETRADIAL GAUGE TRANSFORMATIONS IN $\mathscr{H}$ SPACES

One of the most useful facts about an $\mathscr{H}$ space is that it admits two congruences of totally null (two-dimensional) surfaces. These congruences determine complementary foliations of the total $\mathscr{H}$ space, providing us with three natural coordinatizations, each of which is most conveniently expressed in a (two-component) spinor formalism. Since each of these sets of coordinates will be useful, we will describe them here, their relation to each other, and convenient tetrads formed from them.

We denote the 2-form describing one of these congruences of totally null surfaces by Σ, which must be closed and simple. Therefore we may pick a pair of coordinates q_A which label the leaves of this congruence

$$\Sigma = dq^A \wedge dq_A = 2dq_1 \wedge dq_2. \tag{2.1}$$

[a)]Work supported in part by the Fomento Educacional, A.C., México 5, D.F., México, and by the C.I.E.A. del I.P.N., México 14, D.F.
[b)]On leave of absence from the University of Warsaw, Warsaw, Poland.

 0022-2488/79/091938-08$01.00 © 1979 American Institute of Physics

Similarly, the other congruence, $\tilde{\Sigma}$, allows us to have another pair, $\tilde{q}_B$, such that

$$\tilde{\Sigma} = d\tilde{q}^B \wedge d\tilde{q}_B. \tag{2.2}$$

(The spinor indices are raised and lowered via ϵ^{AB} and ϵ_{AB} as usual.[3,9]) These two congruences completely determine the space, so that

$$\Sigma \wedge \tilde{\Sigma} \neq 0, \tag{2.3}$$

which says that the two pairs q_A, $\tilde{q}_B$ form a coordinate system for the space.[3,4,10] We may express the metric g in terms of the symmetric tensor product of coordinate 1-forms, a null tetrad basis, or a spinor basis of 1-forms:

$$g = g_{\mu\nu}\, dx^\mu \otimes dx^\nu = 2e^1 \underset{s}{\otimes} e^2 + 2e^3 \underset{s}{\otimes} e^4$$
$$= -\tfrac{1}{2} g^{A\dot{B}} \underset{s}{\otimes} g_{A\dot{B}}, \tag{2.4a}$$

where

$$g^{A\dot{B}} = \sqrt{2}\begin{pmatrix} e^4 & e^2 \\ e^1 & -e^3 \end{pmatrix}. \tag{2.4b}$$

It is shown in Ref. 3 that the space is an $\mathscr{H}$ space when

$$g = 2P^{AB}\, dq_A \underset{s}{\otimes} d\tilde{q}_B, \tag{2.5}$$

$$P_{AB} = \Omega_{q^A \tilde{q}^B}, \quad \text{and} \quad \tfrac{1}{2}\Omega_{q^A\tilde{q}^B}\Omega_{q_A\tilde{q}_B} = 1 = \det P_{AB},$$

where the notation X_{q^A} is consistently used for $\partial X/\partial q^A$.

It is convenient and natural to choose a tetrad which takes advantage of the fact that these congruences are integrable. We therefore find two different tetrads to be of use:

$$e^A = 2^{-1/2} g^{A\dot{1}} = dq^A, \quad E^A = 2^{-1/2} g^{A\dot{2}} = P^{AB}\, d\tilde{q}_B,$$

or

$$\tilde{e}^A = 2^{-1/2} g^{A\dot{2}} = d\tilde{q}^A, \quad \tilde{E}^A = 2^{-1/2} g^{A\dot{1}} = P^{BA}\, dq_B, \tag{2.6a}$$

with

$$g = 2E^A \underset{s}{\otimes} e_A = 2\tilde{E}_A \underset{s}{\otimes} \tilde{e}_A. \tag{2.6b}$$

Noting that P^{AB} itself may be considered as an element of SL(2,C), and looking at the complex Lorentz transformations as SL(2,C) ⊗ $\overline{SL(2,C)}$, we see that these two tetrads are related by the complex Lorentz transformation generated by $P^A{}_B$ in SL(2,C) and the identity in $\overline{\mathrm{SL}(2,C)}$. In the two cases, then, we may, as well, work out the usual (null) bases for anti-self-dual and self-dual 2-forms:

$$S^{\dot{1}\dot{1}} = e^A \wedge e_A = \tilde{E}^B \wedge \tilde{E}_B = \Sigma,$$
$$S^{\dot{1}\dot{2}} = e^A \wedge E_A - \tilde{E}^B \wedge \tilde{e}_B,$$
$$S^{\dot{2}\dot{2}} = E^A \wedge E_A = \tilde{e}^B \wedge \tilde{e}_B = \tilde{\Sigma}, \tag{2.7}$$
$$S^{AB} = 2e^{(A} \wedge E^{B)} = 2\tilde{E}^{(A} \wedge \tilde{e}^{B)}.$$

This choice of tetrad also has the very convenient feature that $\Gamma_{\dot{A}\dot{B}} = 0$, by virtue of the above anti-self-dual 2-forms all being closed. We see that these two tetrads are related by the complex Lorentz transformations generated by $P^A{}_B$ in SL(2,C) and the identity in $\overline{\mathrm{SL}(2,C)}$.

Two (parallel) alternate coordinate sets are suggested by the constraint equation satisfied by Ω, which may be interpreted as saying that

$$1 = \det P_{AB} = \frac{\partial(\Omega_{q^A})}{\partial(\tilde{q}^B)} = \frac{\partial(\Omega_{\tilde{q}^B})}{\partial(q^A)}. \tag{2.8}$$

This suggests that either of the sets

$$\{q_A, p_B \equiv \Omega_{q^B}\}, \quad \{\tilde{p}_A \equiv \Omega_{\tilde{q}^A}, \tilde{q}_B\} \tag{2.9}$$

may also be used as coordinates. It is easily shown that the p_A (alternatively the $\tilde{p}_A$) are affine parameters along any member of the congruence Σ (alternatively $\tilde{\Sigma}$). Therefore one set is associated only with properties of Σ, the other with $\tilde{\Sigma}$.

From Eq. (2.6) we find that

$$E^A = \Omega_{q_A\tilde{q}_B}\, d\tilde{q}_B = d\Omega_{q_A} - \Omega_{q_A q_B}\, dq_B.$$

Again, following Ref. 3 we define a new function $\Theta = \Theta(q_A, p_B)$ such that

$$\Theta_{p_A p_B} = -\Omega_{q_A q_B} = -Q^{AB}, \tag{2.10}$$

and have a tetrad based on Σ,

$$e^A = dq^A, \quad E^A = -dp^A - Q^{AB}\, dq_B. \tag{2.11}$$

Similarly we can have a function $\tilde{\Theta} = \tilde{\Theta}(\tilde{p}_A, \tilde{q}_B)$ such that Eqs. (2.10) are repeated, but with tildes everywhere. In Ref. 3 the constraint equation for Θ—analogous to Eq. (2.5) for Ω—is found to be

$$\tfrac{1}{2}\Theta_{p_A p_B}\Theta_{p^A p^B} + \Theta_{p_A q^A} = 0. \tag{2.12}$$

There is of course an identical equation for $\tilde{\Theta}$ in terms of $\tilde{p}_A$, $\tilde{q}_B$. In this form of the coordinates we record the simple form of the connections and the curvature.

$$\Gamma_{\dot{A}\dot{B}} = 0, \quad \Gamma_{AB} = -\Theta_{p^A p^B p^C} e^C,$$
$$C_{\dot{A}\dot{B}\dot{C}\dot{D}} = 0, \quad C_{AB\dot{C}\dot{D}} = 0, \quad R = 0, \quad C_{ABCD} = \Theta_{p^A p^B p^C p^D}. \tag{2.13}$$

We now wish to point out that any spinor coordinate q_A, maintaining $\Sigma \propto dq^A \wedge dq_A$ is as good as any other. Therefore we wish to consider a transformation[9] to new parameters $q'_R = q'_R(q_A)$,

$$dq'_R = D_R{}^A\, dq_A, \quad \Delta \equiv \det D_R{}^A \neq 0, \tag{2.14a}$$

which is only a relabeling of the leaves of the congruence. Since the particular form of the tetrad has been important in the derivation and final form of both the equation which the potential Θ satisfies and the master Killing equation, we endeavor to determine a new set of affine parameters for each leaf which will maintain this form. It is well known, of course, that affine parameters are only defined up to linear transformations. However, there is no reason why the linear transformation in question might not be different for each leaf. With this motivation it is easily seen that the desired transformation properties for the coordinates p^A are given by

$$p'^R = D^{-1}{}_A{}^R p^A + \sigma^R, \tag{2.14b}$$

where the σ^R are arbitrary functions of the q_A, only. The set of equations given by (2.14), when inserted into Eq. (2.11) for the tetrad show that this is the gauge group of transformations which are generated by the relabeling hypothesis in Eq. (2.14a) and which preserve the form of the tetrad. However, we had also arranged for the $\Gamma_{\dot{A}\dot{B}}$ to vanish. This is a very reasonable choice and should be maintained by this transformation. In order to determine this constraint we calculate the $SL(2,C)\otimes\overline{SL(2,C)}$ (complex Lorentz) transformation generated by the transformations (2.14). They are easily expressed by the transformation properties of the S^{AB} and $S^{\dot{A}\dot{B}}$

$$S'^{RS} = L^R{}_A L^S{}_B S^{AB}, \quad S'^{\dot{R}\dot{S}} = L^{\dot{R}}{}_{\dot{A}} L^{\dot{S}}{}_{\dot{B}} S^{\dot{A}\dot{B}},$$
$$L^{\dot{R}}{}_{\dot{A}} = M^{\dot{R}}{}_{\dot{A}} \equiv \begin{pmatrix} \Delta^{+1/2} & , 0 \\ h\Delta^{1/2} & , \Delta^{-1/2} \end{pmatrix}, \qquad (2.15)$$

where $h \equiv \frac{1}{2}\sigma^R{}_{q^R}$. Since

$$\Gamma'_{\dot{R}\dot{S}} = L_{\dot{R}}{}^{\dot{A}} L_{\dot{S}}{}^{\dot{B}} \Gamma_{\dot{A}\dot{B}} + L_{\dot{R}\dot{A}}\, dL_{\dot{S}}{}^{\dot{A}}$$
$$= L^{-1\dot{A}}{}_{\dot{R}} L^{-1\dot{B}}{}_{\dot{S}} \Gamma_{\dot{A}\dot{B}} + L^{-1}{}_{\dot{A}\dot{R}}\, dL^{-1\dot{A}}{}_{\dot{S}}, \qquad (2.16)$$

it is clear that the necessary and sufficient condition to maintain $\Gamma'_{\dot{R}\dot{S}} = 0$ is that Δ and h should be constants. We therefore re-collect the transformation equations here with that proviso, denoting Δ and h now by Δ_0 and h_0

$$dq'_R = D_R{}^A\, dq_A = \Delta_0^{1/2} L^{-1A}{}_R\, dq_A, \quad L^{-1A}{}_R \in SL(2,C), \qquad (2.14')$$
$$p'^R = D^{-1}{}_A{}^R p^A + \tfrac{1}{4} h_0 q'^R + \rho_{q'_R},$$

where the degrees of freedom are given by $L^{-1A}{}_R \in SL(2,C)$, ρ an arbitrary function of q'_R and two constants Δ_0 and h_0. The transformation laws can be completed by noting that the potential function Θ transforms in the following way:

$$\Delta_0^2 \Theta' = \Theta + \tfrac{1}{6} L_R{}^C (L^{RB})_{q_A} p_A p_B p_C$$
$$- \tfrac{1}{2}\rho_{q_A q_B} p_A p_B + \lambda^A p_A + \nu \qquad (2.17)$$

where λ^A and ν are new arbitrary functions of q_A only, which generate gauge transformations of Θ only, in its role as a potential function.

III. CANONICAL FORMS OF MASTER KILLING VECTOR EQUATION

We now recall the master Killing equation from Ref. 8, adapted to the special case of an $\mathscr{H}$ space. Any allowed Killing vector may be written in the form

$$K = (\alpha_0 p^A + \delta^A)\frac{\partial}{\partial q^A} + (-2\alpha_0 \Theta_{p_A} + \delta^B_{q_A} p_B + \epsilon^A)\frac{\partial}{\partial p^A}, \qquad (3.1a)$$

with

$$\delta^A = \tfrac{1}{2}\rho_0 q^A + \varphi_{q_A}, \quad \epsilon^A = -\tfrac{1}{2}\gamma_0 q^A + \sigma_{q_A}, \qquad (3.1b)$$

where $\alpha_0, \rho_0, \gamma_0$ are constants while φ and σ are arbitrary functions of q_A only. In order for a manifold to permit any particular Killing vector the potential function Θ must satisfy the master equation

$$K\Theta = 2\rho_0 \Theta + 2\alpha_0 \Lambda + \tfrac{1}{6}\varphi_{q^A q^B q^C} p^A p^B p^C$$
$$+ \tfrac{1}{2}\sigma_{q^A q^B} p^A p^B + \psi_{q^A} p^A + \eta, \qquad (3.2)$$

where ψ and η are some new arbitrary functions of q_A only, while Λ is a higher-order potential function, defined by

$$\tfrac{1}{2}\Lambda_{p^A} = \Theta_{q^A} + \tfrac{1}{2}\Theta_{p_B}\Theta_{p^B p^A}, \qquad (3.3)$$

whose existence is guaranteed by the heavenly equation (2.12).

These constraint equations guarantee[8] the satisfaction of the complete set of Killing equations, which can most easily be written for our purposes in spinor form[11] (where $K_B{}^{\dot{B}} = -\frac{1}{2} g_B{}^{\dot{B}\alpha} K_\alpha$ with K_α being the components of the Killing 1-form)

$$\nabla_A{}^{\dot{A}} K_B{}^{\dot{B}} = \epsilon_{AB}\, l^{\dot{A}\dot{B}} + \epsilon^{\dot{A}\dot{B}}\, l_{AB}, \qquad (3.4)$$

where $l^{\dot{A}\dot{B}}$ and l_{AB} are symmetric and satisfy the constraints

$$\nabla_A{}^{\dot{A}} l_{BC} = -2C_{ABC}{}^N K_N{}^{\dot{A}}, \qquad (3.5a)$$
$$\nabla_A{}^{\dot{A}} l^{\dot{B}\dot{C}} = -2C_{\dot{N}}{}^{\dot{A}\dot{B}\dot{C}} K_A{}^{\dot{N}} = 0. \qquad (3.5b)$$

Equations (3.5) tell us that l_{BC} and $l^{\dot{B}\dot{C}}$ are very important in the description of our Killing vector. In particular Eq. (3.5b) shows that $l^{\dot{B}\dot{C}}$ must be constant in an $\mathscr{H}$ space. Our solution, determined by Eqs. (3.1) and (3.2), of course already satisfies this constraint, but it will nonetheless be of considerable use to us,

$$l^{\dot{B}\dot{C}} = \begin{pmatrix} 2\alpha_0 & \rho_0 \\ \rho_0 & \gamma_0 \end{pmatrix}. \qquad (3.6)$$

There is of course a matrix $l_{\dot{B}\dot{C}}$ for every Killing vector allowed by a particular solution. Under an arbitrary $\overline{SL(2,C)}$ gauge transformation, $L^{\dot{R}}{}_{\dot{A}}$, $l_{\dot{B}\dot{C}}$ transforms as a pure spinor quality,

$$l'^{\dot{R}\dot{S}} = L^{\dot{R}}{}_{\dot{A}} L^{\dot{S}}{}_{\dot{B}}\, l^{\dot{A}\dot{B}}. \qquad (3.7)$$

The quantity $l^{\dot{B}\dot{C}} l_{\dot{B}\dot{C}} = 2\det(l_{\dot{B}\dot{C}})$ is clearly invariant under such a transformation. Therefore by using the transformations $L^{\dot{R}}{}_{\dot{A}}$ an $l_{\dot{B}\dot{C}}$ for any one given Killing vector can always be reduced to one of the following three canonical types:

$$\text{I: } l^{\dot{B}\dot{C}} = 0, \quad \text{II: } l^{\dot{B}\dot{C}} = \begin{pmatrix} 0 & 0 \\ 0 & \gamma_0 \end{pmatrix}, \quad \text{III: } l^{\dot{B}\dot{C}} = \begin{pmatrix} 0 & \rho_0 \\ \rho_0 & 0 \end{pmatrix}. \qquad (3.8)$$

Note that the fact that $l_{\dot{B}\dot{C}}$ is constant is quite important here because we must look only at constant $\overline{SL(2,C)}$ gauge transformations so as to preserve $\Gamma'_{\dot{R}\dot{S}} = 0$. [See Sec. 2.] We have also deliberately made choices of the canonical forms which do *not* involve α_0 since it multiplies Λ in Eq. (3.2) which is a more complicated quantity than Θ itself. Such a choice eliminates the need to work with Λ for the specific Killing vector under consideration.

Unfortunately, the set of tetradial gauge transformations considered in Sec. II do not include all possible elements of $\overline{SL(2,C)}$ since they only generate $L^{\dot{R}}{}_{\dot{A}}$ which are lower triagonal [Eq. (2.15)]. However, we also consider now the possibility of starting with the coordinate system $\{\tilde{p}^A, \tilde{q}_B\}$ based on $\tilde{\Sigma}$. A completely analogous set of tetradial gauge transformations based on $\tilde{\Sigma}$ is generated by

$$d\tilde{q}'_R = \tilde{D}_R{}^A\, d\tilde{q}_A, \quad \tilde{p}'^R = \tilde{D}^{-1}{}_A{}^R \tilde{p}^A + \tilde{\sigma}^R,$$
$$\det\tilde{D}_R{}^A = \tilde{\Delta}_0, \quad \tfrac{1}{2}\tilde{\sigma}^R{}_{\tilde{q}^R} = \tilde{h}_0, \text{ constants}, \qquad (3.9)$$
$$L^R{}_A = \Delta_0^{1/2} \tilde{D}_A{}^R,$$
$$L^{\dot{R}}{}_{\dot{A}} = \tilde{M}^{\dot{R}}{}_{\dot{A}} \equiv \begin{pmatrix} \tilde{\Delta}_0^{-1/2} & \tilde{h}_0 \tilde{\Delta}_0^{1/2} \\ 0 & , \tilde{\Delta}_0^{1/2} \end{pmatrix}.$$

Clearly an arbitrary product of $\widetilde{M}^{R}{}_{\dot{A}}$ and $M^{\dot{S}}{}_{B}$ would generate an arbitrary $L^{R}{}_{\dot{A}}\in\overline{\mathrm{SL}(2,C)}$, so that the canonical form of an $l_{\dot{B}\dot{C}}$ may always be effectuated by a combination of such transformations. By looking in more detail one sees that an $l_{\dot{B}\dot{C}}$ with $\alpha_0 = 0$ may always be put into the forms in Eq. (3.8) simply by a transformation of the type $M^{R}{}_{\dot{A}}$, i.e., by relabeling the leaves of only the one congruence, Σ. However, for a Killing vector with $\alpha_0 \neq 0$, one notes that

$$\begin{pmatrix} 2\alpha_0 & \rho_0 \\ \rho_0 & \gamma_0 \end{pmatrix} = \begin{pmatrix} \tilde{\gamma}_0 & \tilde{\rho}_0 \\ \tilde{\rho}_0 & 2\tilde{\alpha}_0 \end{pmatrix} \tag{3.10}$$

so that a transformation to $\gamma_0 = 0$ (if necessary) and then passing to the coordinate system based on $\widetilde{\Sigma}$ causes $\tilde{\alpha}_0$ to be zero and the passage to the appropriate canonical form may be completed in these coordinates. We therefore always assume that this has been done for the single Killing vector of interest here. However, we do not write any tildes since, of course, the two sets are isomorphic. In Appendix A we give an explicit procedure by which one may proceed from $\Theta = \Theta(p^A, q_B)$ to the corresponding $\widetilde{\Theta} = \widetilde{\Theta}(\tilde{p}^A, \tilde{q}_B)$. (Also, in Appendix B we determine the explicit forms for the ten Killing vectors in the flat space corresponding to $\Theta = 0$, showing that nonzero α_0 is essential when dealing with sufficiently many Killing vectors at once.)

We now show that the terms due to the functions ψ and η in Eq. (3.2) are completely gauge dependent, i.e., they can always be gauged away. To see how this occurs, as well as some other gauge transformations we will need, we include here the transformations of the quantities φ, σ, etc. which appear in the master equation. These transformations are of course defined in such a way that the equations (3.1) and (3.2) remain *form invariant* under the gauge transformations of Sec. II. Under the assumption that α_0 has already been transformed to 0, we obtain

$$\delta'_R = D_R{}^A \delta_A,$$
$$\epsilon'^S = D^{-1}{}_A{}^S \epsilon^A + (\rho_{q'^R} \delta'^R)_{q'_S} + 2h_0 \delta'^S, \tag{3.11}$$

while, of course, ρ_0 and γ_0, separately transform as indicated in Eq. (3.7).

We first consider a transformation

$$\Theta' = \Theta + \xi_{q^A} p^A + \nu \tag{3.12}$$

which causes a transformation of ψ and η:

$$\psi' = \psi + \rho_0 \xi - \delta^A \xi_{q^A}, \quad \eta' = \eta + \epsilon^A \xi_{q^A} - \delta^A \nu_{q^A}. \tag{3.13}$$

It is clear that these equations can be considered as first-order partial differential equations to determine ξ and ν in such a way that ψ' and η' vanish and that they always have (local) solutions so long as $\delta^A \neq 0$. Further, now consider the translation

$$p'^A = p^A + \rho_{q'_A}. \tag{3.14}$$

We find that σ transforms as

$$\sigma' = \sigma + \delta^A \rho_{q^A}. \tag{3.15}$$

Clearly when $\delta^A \neq 0$ we may always choose ρ in such a way as to guarantee $\sigma' = 0$.

On the other hand, if $\delta^A = 0$, $\epsilon^A \neq 0$, then Eq. (3.13) still assures us that we can arrange for $\eta' = 0$. Then the translation [Eq. (3.14)] now leaves ϵ^A invariant, but causes

$$\Theta' = \Theta - \tfrac{1}{2} \rho_{q^A q^B} p^A p^B,$$

which transforms ψ to

$$\psi' = \psi + \epsilon^A \rho_{q^A} - \gamma_0 \rho.$$

It is again clear that ρ may always be found so as to cause ψ' to vanish. Clearly if both δ^A and ϵ^A vanish (along with α_0) then there is no Killing vector. Therefore we have shown that we may always gauge away the terms in Eq. (3.2) generated by ψ and η. (We refer to this result as Lemma 1.) Moreover (when $\alpha_0 = 0$) σ must be retained only when $\delta^A = 0$ (Lemma 2). In the process we have used up the gauge freedom embodied in a λ^A of the form of ξ_{q^A}, and ν from Eq. (2.17) as well as the translations generated by ρ.

One more lemma is required before actually making the reduction to canonical form. Suppose that β^A is some function of q_B (for instance δ^A or ϵ^A, whichever may be nonzero). Then we can use a choice of $D_R{}^A$ to align the coordinates in the direction β^A in one of two ways, depending on whether $\beta^A{}_{q^A}$ vanishes or not. First, if $\beta^A \neq 0$, $\beta^A{}_{q^A} = 0$, then there exists b such that $\beta^A = b_{q_A}$ and we may surely always choose new coordinates q_R such that one of them is b, say q_Z, where Z is a fixed choice of index, which sets $\beta^A = \delta^A{}_Z$ the Kronecker delta (Lemma 3a). On the other hand, if $\beta^A{}_{q^A} = b_0$, a nonzero constant, then there always exist two functions S^A such that $\beta^A dq_A = b_0 S^1 dS^2$. Furthermore $\beta^A{}_{q^A} \neq 0$ implies that we can always find a transformation (just a two-dimensional canonical transformation) such that $q'^R = S^R$. This implies $\beta'^1 = b_0 q'^1$, $\beta'^2 = 0$ (Lemma 3b).

Assuming that all of the above transformations have been done we are ready to see that there are in fact only five independent situations for the existence of an $\mathscr{H}$ space with (at least) one Killing vector. First we consider case I, in which all of α_0, ρ_0, and γ_0 vanish for the Killing vector in question [see Eq. (3.8)]. If both δ^A and ϵ^A were to vanish as well, then there would be no Killing vector at all; therefore, one must be nonzero. If $\delta^A \neq 0$, we gauge away ϵ^A and note that $\delta^A{}_{q^A} = \rho_0 = 0$. Therefore Lemma 3a allows us to choose coordinates so that $\varphi = q_Z$ (here Z is a specific choice of index, either 1 or 2) and the master equation reduces to

$$K\Theta = \Theta_{q'} = 0 \quad \text{(Case Ia)}. \tag{3.16}$$

On the other hand, if $\delta^A = 0$, $\epsilon^A \neq 0$, $\epsilon^A{}_{q^A} = \gamma_0 = 0$, then again Lemma 3a may be invoked to pick coordinates so that $\sigma = q_Z$ (Z fixed to be either 1 or 2) and the master equation reduces to

$$K\Theta = \Theta_{p'} = 0 \quad \text{(Case Ib)}. \tag{3.17}$$

We also note that these Killing vectors of Case I are just exactly those Killing vectors which can be generated from a $D(1,0)$ [or $D(0,1)$] Killing spinor.[8] Therefore it is shown what at least some portion of the role of such Killing spinors is in $\mathscr{H}$ spaces.

The second case, where $l^{\dot{B}\dot{C}} \neq 0$ but $\det l^{\dot{B}\dot{C}} = 0$ is described canonically by $\alpha_0 = 0 = \rho_0$, $\gamma_0 \neq 0$, which implies

$\delta^A{}_{q'} = 0$, $\epsilon^A{}_{q'} \neq 0$. In the situation that $\delta^A \neq 0$, then Lemmas 2 and 3b may be invoked, and a choice of Δ_0 may be made to normalize γ_0, so that we obtain

$$K\Theta = \left(\frac{\partial}{\partial q^Z} + q^A \frac{\partial}{\partial p^A} \right)\Theta = 0 \quad [\text{Case IIa}]. \quad (3.18)$$

In case $\delta^A = 0$, then σ may nontheless be gauged away because $\gamma_0 \neq 0$. This follows from Eq. (3.11) which can be manipulated to read

$$(\sigma' - \sigma)_{q'_S} = \tfrac{1}{2}\gamma_0(\Delta_0{}^{-1} q'^S - D^{-1}{}_A{}^S q^A).$$

The integrability condition, then, to find $D^{-1}{}_A{}^S$ such that $\sigma'_{q'^K} = 0$ is simply that the right-hand-side of the equation should be a gradient, which is equivalent to the requirement that $(\Delta_0{}^{-1}{}_{q'}S - D^{-1}{}_A{}^S q^A)_{q'^S} = 0$, which is seen to be the case upon explicit calculation (Lemma 3c). Therefore we are left with

$$K\Theta = q^A \frac{\partial}{\partial p^A} \Theta = 0 \quad [\text{Case IIb}]. \quad (3.19)$$

Lastly in the case III, where $\det l^{BC} \neq 0$, the canonical form of Eq. (3.8) has $\alpha_0 = 0 = \gamma_0, \rho_0 \neq 0$, which tells us that $\delta^A \neq 0$. Therefore we may always gauge away σ and use Lemma 3b to acquire

$$K\Theta = \left(q^1 \frac{\partial}{\partial q^1} - p^2 \frac{\partial}{\partial p^2} \right)\Theta = -2\Theta \quad [\text{Case III}]. \quad (3.20)$$

We have thus shown that the arbitrary functions of two variables appearing in the master equation had two roles. One was to simply allow for all the possible gauge freedom in choice of coordinates and potential function Θ. The other much more important function was to distinguish these five types (six if you also include the possibility of the nonadmissibility of any Killing vector). That is, the "quotient" of the two (infinite) gauge groups is finite.

IV. APPLICATION TO THE HEAVENLY EQUATION

We now proceed to show that the knowledge that a manifold admits a Killing vector allows one of course to determine a form for Θ in terms of a function of only three variables and that this information inserted into the heavenly equation (2.12) is sufficient to allow a determination of all such functions Θ in cases I and II. In case III the equation simplifies but still does not yield completely, as will be indicated.

Case Ia evidently has a simple solution—that Θ is independent of one of the labeling variables, q^A. Arbitrarily we pick that one to be v. Since this breaks the spinor symmetry we write the corresponding version of Eq. (2.12) in the form given in Eq. (1.1),

$$\Theta_{xx}\Theta_{yy} - \Theta_{xy}{}^2 + \Theta_{xu} = 0. \quad (4.1)$$

This equation can be modified so as to become linear[4] by means of a Legendre transformation. Set

$$p = \Theta_x, \quad \psi \equiv px - \Theta = \psi(p,y,u), \quad (4.2a)$$

which implies

$$x = \psi_p, \quad \Theta_y = -\psi_y, \quad \Theta_u = -\psi_u. \quad (4.2b)$$

Equation (4.1) may be written in terms of the 3-form

$$d\Theta_x \wedge d\Theta_y \wedge du + d\Theta_x \wedge dx \wedge dy = 0.$$

Inserting Eq. (4.2) into this we find that

$$\psi_{yy} + \psi_{pu} = 0, \quad (4.3)$$

which is just the three-dimensional Laplace equation. Therefore every $\mathscr{H}$ space having a Killing vector of type Ia is characterized by a solution of the three-dimensional Laplace equation, whose solutions are all well known. For an approach particularly relevant to the philosophy used here see Ref. 12 where applicability of complex-valued integral representations is discussed. We point out that these solutions are rather general and include all possible Petrov types.

Case Ib also has a very simple solution; Θ is independent of one of the affine variables, p^A, say x. In that case, Eq. (2.12) becomes simply

$$\Theta_{yv} = 0,$$

the two-dimensional Laplace equation, the solution of which is of course just $\Theta = F(y,u) + G(v,u)$. However, in this case we may still use the gauge freedom of the arbitrary function $\nu(q_A)$ in Eq. (2.17) to eliminate G. Therefore

$$\Theta = F(y,u) \quad (4.4)$$

is the form of Θ for all Killing vectors of type Ib. These are just the $[N] \otimes [-]$ spaces already discussed in some detail in Ref. 3. (An affine parameter is a Killing variable only for an $\mathscr{H}$ space of Petrov type $[N] \otimes [-]$.)

For case IIa the form given in Eq. (3.18) for the Killing vector has considerable symmetry. However a form which is much more convenient for calculation is obtained by not eliminating σ and using Lemma 3b on ϵ^A. We demonstrate this explicitly by setting $p'^A = p^A + \rho_{q_A}$ and choose $\rho = \tfrac{1}{2}vu^2$ where we have chosen $Z = 1$ and have labeled $q_1 = v$, $q_2 = u, p^1 = y, p^2 = x$. Then Eq. (3.18) is converted to

$$K\Theta = \Theta_u + 2u\Theta_y = 0, \quad (4.5)$$

whose solution is $\Theta = K(y - u^2, x, v)$, where $K = K(z,x,v)$ is an arbitrary holomorphic function of three variables. When inserted into Eq. (1.1) the constraint becomes

$$K_{zz}K_{xx} - K_{xz}{}^2 - 2uK_{xz} + K_{zv} = 0, \quad (4.6)$$

which is gauge equivalent to Eq. (4.6). However K is not a function of u. Therefore, any valid solution (one which determines a Θ) must have $K_{xz} = 0$. With that additional constraint the general solution is easily found to determine

$$\Theta = H[y - u^2 - f(v)] + \tfrac{1}{2}f'(v)x^2, \quad \text{Case IIa}, \quad (4.7)$$

where f and H are arbitrary holomorphic functions of one variable. Such a space is of Petrov type $[N] \otimes [-]$. Case IIb [Eq. (3.19)] simply says that we can choose coordinates and Θ so that

$$\Theta = F(q_A p^A, q_B), \quad (4.8)$$

where $G = G(\omega, q_B)$ is an arbitrary holomorphic function of three variables. The heavenly equation (1.1) reduces to $(q_A F_\omega)_{q_1} = 0$, whose solution is immediate. All $\mathscr{H}$ spaces which admit a Killing vector of the form $K = q^A(\partial/\partial p^A)$ allow a Θ of the form

$$\Theta = \omega^{-1} H(q_A/\omega), \quad \omega \equiv q_B p^B, \quad (4.9)$$

where H is an arbitrary holomorphic function of two variables. All such $\mathscr{H}$ spaces are of type $[N] \otimes [-]$ and have been discussed in Ref. 5. They have null strings which are expanding.

Case III again generates a complicated equation. The general solution of Eq. (3.20) may be written in the form

$$\Theta = u^{-1} P(ux,y,v), \tag{4.10}$$

where $P = P(z,y,v)$ must satisfy the equation

$$P_{zz}P_{yy} - (P_{yz})^2 + P_{yv} + zP_{zz} - P_z = 0. \tag{4.11}$$

We have now presented the reductions obtained by the requirement that the $\mathscr{H}$ space admit (at least) one Killing vector. There are only five distinct cases. Four of those have already been reduced to linear equations whose solutions are well known. The other one still has irresolvable nonlinearities. We will soon demonstrate considerable simplification of the form of Eq. (4.11). However, it is worth pointing out first that it can be solved in the event that there is no dependence on v. In such a situation we would have two Killing vectors labeled by $u\partial/\partial u$ and $\partial/\partial v$ in the coordinates z, y, v, u. In this case Eq. (4.11) may be considered without the P_{yv} terms. Writing $p \equiv P_y$, $q \equiv P_z$ the equation is equivalent to the vanishing of the pair of 2-forms

$$\begin{aligned} &dp\wedge dq + z\,dy\wedge dq - q\,dy\wedge dz = 0,\\ &dp\wedge dy + dq\wedge dz = 0. \end{aligned} \tag{4.12}$$

Choosing p and z as functions of q and y (considered as independent variables) we acquire

$$\begin{aligned} &z = -\Lambda_q, \quad p = \Lambda_y,\\ &\Lambda_{yy} + q\Lambda_{qq} - \Lambda_q = 0. \end{aligned} \tag{4.13}$$

This equation is clearly linear and separable. It is in fact equivalent to the Euler–Poisson–Darboux equation, as can be seen by putting $\alpha = 2q^{1/2} - iy$, $\beta = 2q^{1/2} + iy$

$$(\alpha+\beta)\Lambda_{\alpha\beta} = \tfrac{3}{2}(\Lambda_\alpha + \Lambda_\beta). \tag{4.13'}$$

This equation is well known[13] and integral representations of the solutions are discussed in Ref. 14.

In the more general case where dependence on v is allowed, we first rewrite Eq. (4.11) in terms of 3-forms, when $r \equiv P_y$, $s \equiv P_z$:

$$\begin{aligned} &dr\wedge ds\wedge dv + dr\wedge dy\wedge dz\\ &\quad - z\,ds\wedge dy\wedge dv - s\,dy\wedge dz\wedge dv = 0,\\ &dr\wedge dy\wedge dv + ds\wedge dz\wedge dv = 0. \end{aligned} \tag{4.14}$$

Setting $\Upsilon = -dr + s\,dv - z\,dy$, we see that the first of this pair of 3-forms is simply $\Upsilon\wedge d\Upsilon$. Therefore, Frobenius' theorem guarantees us the existence of functions F, G of y, z, v such that $\Upsilon = F\,dG$. In terms of these new functions Eqs. (4.14) become

$$\begin{aligned} &dF\wedge dG\wedge dv + dz\wedge dy\wedge dv = 0,\\ &F\,dG\wedge dy\wedge dv + dF\wedge dG\wedge dz = 0. \end{aligned} \tag{4.15}$$

Rewriting the second of Eqs. (4.14), first in terms of Υ, and picking G, z, v as independent variables, this pair is equivalent to

$$y_G = -F_z, \quad F_{zz} - (\ln F)_{vG} = 0. \tag{4.16}$$

Writing $F = e^H$ our constraint equation is now simply the equation

$$(e^H)_{zz} - H_{vG} = 0. \tag{4.16'}$$

The general solution of this equation is not known. However a large class of solutions is determined by requiring the two parts to vanish separately. The general solution in that case is given by both of the two forms given below:

$$e^H = h(v)[m(G)z + l(G)],$$

or

$$e^H = k(G)[j(v)z + n(v)], \tag{4.17}$$

where h, m, l, k, j and n are arbitrary holomorphic functions of one variable. This set already includes solutions of all possible Petrov types.

Lastly, with respect to the general solution of Eq. (4.11), the general theory of Cartan[15] may be applied to construct the regular integral manifolds that are appropriate to it. In this way we determine, at each step, their arbitrariness; that is, we are able to determine on how many arbitrary functions of how many variables the general manifold depends. To do this we start with the pair of 1-forms generated by Υ and the definitions of r and s given just prior to Eq. (4.14), namely $r = P_y$, $s = P_z$, and, as well, $w = P_v$:

$$\begin{aligned} &\omega_1 \equiv -dr + s\,dv - z\,dy - F\,dG,\\ &\omega_2 \equiv dP - r\,dy - s\,dz - w\,dv. \end{aligned} \tag{4.18}$$

An integral manifold is a submanifold of Euclidean nine-dimensional space on which these two Pfaffian 1-forms vanish. The procedure[16] is given in detail in Ref. 15 and is straightforward. We simply note here that the result is that the general solution of Eq. (4.11) depends on two arbitrary functions of two complex variables. [This clearly includes, as a special case, the special solutions given in Eq. (4.17) which depend on three arbitrary functions of one complex variable.]

V. CONCLUSIONS

This problem was generated by an attempt to factor out the gauge dependence of the form of the Killing vectors in an arbitrary $\mathscr{H}$ space, in an attempt to uncover the underlying geometry. This effort was successful in that we were able to show that all $\mathscr{H}$ spaces which admit (at least) one Killing vector can be divided into only five distinct types in which the majority of the gauge freedom has been removed. In each of these types there is a simple first-order linear partial differential equation which is the mechanism whereby a symmetry is permitted into the functions which define the geometry. When this symmetry is entered explicitly into the heavenly equation (which determines that the manifold is an $\mathscr{H}$ space) a specific equation is obtained which determines all possible spaces with the specified symmetry. In four of these cases (those in which the Killing equation explicitly depended only on the derivatives of Θ, rather than Θ itself) we were able to reduce the problem to the solution of a known *linear* differential equation. This is therefore a complete exemplification of *all* such solutions since whichever particular solutions of these linear equations are desired may be used, or the known general integral representations may be used.

In the last case an irresolvable nonlinearity remains. We have given a number of different kinds of solutions for it even though the general solution has not been reduced to known quantities. In addition we have shown that its general solution depends on two arbitrary functions of two complex variables.

Finally we would like to point out that there is a very intimate relation between these $\mathscr{H}$ spaces and gravitational instantons. It has already been shown by Tod[17] that the general self-dual Yang–Mills equation can be put into a form which is basically the same as our heavenly equation [see the form in Eq. (2.12)]. In addition there has been much interest lately in the general problem of gravitational instantons.[18] Such objects are simply Euclidean slices of an $\mathscr{H}$ space with appropriate topological properties. Since our approach to $\mathscr{H}$ spaces is local a number of explicit calculations must still be done to determine these Euclidean slices and to confirm the desired topological properties. However, with the explicit *linear* equations given here this should be a relatively straightforward matter. Additionally we mention that the topological quantity of most interest is an integral over the entire manifold of the quantity $C_{ABCD}C^{ABCD}$, which should be of the form of a divergence for such calculations. It is trivially seen from Eq. (2.13) that, in terms of the potential function Θ,

$$C_{ABCD}C^{ABCD} = [\Theta\Theta_{p^A p^B p^C p^D}]_{p^A p^B p^C p^D}. \tag{5.1}$$

ACKNOWLEDGMENTS

The authors would like to express thanks for useful discussions with C.P. Boyer. Also one of us (JDF) appreciates discussions with Stanly Steinberg which led to the special class of solutions given in Eq. (4.17).

APPENDIX A: SWITCHING NULL CONGRUENCES

We show a method which allows for calculation of the coordinatization based on one congruence of null surfaces, given the other. In particular, suppose that we are given a complete description based on Σ: Coordinates q_A which label the leaves of Σ, affine parameters p^A along them, and $\Theta = \Theta(q_A, p^B)$ which determines the metric. Then we desire to obtain labels for the congruence $\widetilde{\Sigma}$, $\tilde{q}_B = \tilde{q}_B(q_C, p^D)$ and a function $\Omega = \Omega(q_A, \tilde{q}_B)$ satisfying the constraint equation (2.5). To do this we note that, in the Ω approach, the tetrad based on Σ has $\partial^{A2} = 2^{1/2}(\partial/\partial q_A)$ and therefore $\partial^{A2}\tilde{q}_B = 0$. But the tetrad is the same in the Θ-form. Therefore our requirements may be stated in the following way. We desire to find functions $\tilde{q}_B$ satisfying

$$2^{-1/2}\partial^{A2}\tilde{q}_B = \left(\frac{\partial}{\partial q_A} + Q^{AC}\frac{\partial}{\partial p^C}\tilde{q}_B\right) = 0. \tag{A1}$$

In general there is no reason to believe that there would exist two independent solutions to this pair of first-order equations. However, the integrability conditions are easily shown to be simply the Eq. (2.12) which Θ must satisfy in order to produce an $\mathscr{H}$ space. Therefore, there always exist two independent solutions of this pair of equations. In determining the $\tilde{q}_B$ it is to be noted that they are, of course, only well defined within a canonical transformation among themselves. This freedom will express itself in the arbitrary functions which come into play in the solution of these first-order partial differential equations. Since each of Eqs. (A1) consists of partial derivatives with respect to three different variables, it follows that there exist a pair $C_A(q_B, p^C)$ of characteristic variables such that $\tilde{q}_B = F_B(C_A)$, where the F_B are arbitrary holomorphic functions of two variables except that they must be chosen in such a way that $\det[(F_B)_{,p^C}] = 1$ [Eq. (2.5)]. For example, we may always pick $F_1(C_A) = C_1$ and then use the determinant condition to pick F_2. Having thus chosen a particular set of $\tilde{q}_B$, we may write Eq. (2.10) in the form

$$\Omega_{q_A \tilde{q}_B} = -\Theta_{p^A p^B}[q_C, p^D(q_E, \tilde{q}_F)]$$

and integrate this to determine an Ω, noting that two choices of Ω which differ by a term of the form $f(q_A) + g(\tilde{q}_A)$ generate the same metric and therefore are to be considered equivalent.

Once one has an $\Omega = \Omega(q_A, \tilde{q}_B)$ the construction of $\tilde{p}^A$ and $\widetilde{\Theta}(\tilde{p}^A, \tilde{q}_B)$ is completely straightforward via the tilded versions of Eqs. (2.9) and (2.10). See also Appendix B for a description of how the Killing vectors (in flat space as an example) are correlated in the three different descriptions.

APPENDIX B: FLAT SPACE HEAVENLY KILLING VECTORS

These Killing vectors are given here simply as a complete set which can generate some intuitive feeling about the physical meaning of the parameters in the Killing master equation. We pick $\Theta = 0$ which generates flat space in local Minkowski coordinates. We note that a reasonable choice of identification with the usual x, y, z, t coordinates is

$$p^A = \frac{1}{\sqrt{2}}\begin{pmatrix} -x+iy \\ t+z \end{pmatrix}, \quad q_A = \frac{1}{\sqrt{2}}\begin{pmatrix} x+iy \\ t-z \end{pmatrix}, \tag{B1}$$

which makes $\Sigma = (dz - dt)\wedge(dx + idy)$, $\widetilde{\Sigma} = (dz + dt)\wedge(dx - idy)$ the congruences of null surfaces on which the properties of $\mathscr{H}$ space are based. Then the master equation is easily solved and gives the following forms for the parameters of interest

$$\begin{aligned} \delta^A &= \tfrac{1}{2}\rho_0 q^A + A_0{}^{AB}q_B + B_0{}^A, \\ \epsilon^A &= -\tfrac{1}{2}\gamma_0 q^A + C_0{}^A, \quad \alpha_0 = \alpha_0, \end{aligned} \tag{B2}$$

where $A_0{}^{AB}$ is symmetric. Therefore there are exactly ten independent parameters, as desired. The subgroup generated by $\{B_0{}^A, C_0{}^B\}$ is Abelian and clearly corresponds to the translations, $\partial/\partial q_A$, and $\partial/\partial p^B$. [See the form of a Killing vector in terms of δ^A, ϵ^A and α_0 in Eq. (3.1a).] There are two more subalgebras, which are not Abelian but which do commute between themselves, generated by $\{\alpha_0, \rho_0, \gamma_0\}$ and

$A_0{}^{AB}$ respectively. We take the usual Lie algebra of the homogeneous Lorentz group in terms of the generators of rotations, $\mathscr{L} = \mathbf{x}\times\nabla$, and boosts, $\mathscr{K} = \mathbf{x}(\partial/\partial t) + t\nabla$, which form the components of a skew 4×4 matrix of 1-vectors, $L_{\mu\nu}$, where $L_{ij} = \epsilon_{ijk}\mathscr{L}^k$, $L_{i4} = \mathscr{K}_i$. We then expand this matrix in terms of the basis of (self-dual and anti-self-dual) 2-forms

$$L_{\mu\nu} = L_{AB}S^{AB}{}_{\mu\nu} + L_{\dot A\dot B}S^{\dot A\dot B}{}_{\mu\nu}. \tag{B3}$$

Then we may easily characterize these two (commuting) subgroups. They are simply those generated by $A_0{}^{AB}L_{AB}$—corresponding to the generator $\mathscr{K} - i\mathscr{L}$—and $l^{\dot A\dot B}L_{\dot A\dot B}$—corresponding to $\mathscr{K} + i\mathscr{L}$—where $l^{\dot A\dot B}$ is composed of α_0, ρ_0, γ_0 as given in Eq. (3.6).

Having these Killing vectors it is of some interest to understand their form in terms of the Ω and $\widetilde{\Theta}$ coordinatizations. In order to do that we include here a very brief sketch of the derivation of the Killing master equation in the Ω approach. The equation is of course formally identical in the $\widetilde{\Theta}$ approach, but with all variables tilded. Note also that the Ω-coordinatization is of considerable interest in its own right (for example, the explicit complex Kähler structure is seen here). Writing the Killing vector in the form

$$K = L^A\frac{\partial}{\partial q^A} + \widetilde{M}^A\frac{\partial}{\partial \tilde q^A}, \tag{B4}$$

and taking the (coordinate) metric in the form

$$g(\partial_{q_i}, \partial_{q_n}) = 0 = g(\partial_{\tilde q_i}, \partial_{\tilde q_n}), \quad g(\partial_{q_i}, \partial_{\tilde q_n}) = \Omega_{q_i\tilde q_n}, \tag{B5}$$

we find that Killing's equations split into three sets:

$$\begin{aligned} &\Omega_{\tilde q_n q_i}\widetilde{M}_{B q_c} = 0, \quad \Omega_{\tilde q_c q_i}L_{C\tilde q_n} = 0,\\ &L^D\Omega_{q^D q_i\tilde q_n} + \widetilde{M}^D\Omega_{q_i q_n D} + \Omega_{q_i\tilde q_D}\widetilde{M}_{D\tilde q_n} + \Omega_{q_D\tilde q_n}L_{D\tilde q_i}\\ &\quad = 2\chi_0\Omega_{q_i\tilde q_n}. \end{aligned} \tag{B6}$$

The two simpler equations are easily integrated to give

$$L^A = B\,\Omega_{q_i} + J^A, \quad \widetilde{M}^A = \widetilde{A}\,\Omega_{\tilde q_i} + \widetilde{N}^A, \tag{B7}$$

where B, J^A are functions of q_A only, while $\widetilde{A}$, $\widetilde{N}^A$ are functions of $\tilde q_A$ only. The rest of the equations and the integrability conditions for Killing structures—in this case, simply that $l^{\dot B\dot C}$ be constant—give us a final form:

$$K\Omega = 2\chi_0\Omega - a_0F - b_0G + \lambda + \tilde\mu, \tag{B8}$$

$$K = [(b_0\Omega + \zeta)_{q_i} + (\chi_0 - \tfrac12 c_0)q^A]\frac{\partial}{\partial q^A} + [(a_0\Omega + \tilde\eta)_{q_i} + (\chi_0 + \tfrac12 c_0)\tilde q^A]\frac{\partial}{\partial \tilde q^A},$$

where ζ, λ are arbitrary functions of q_A only, $\tilde\eta$, $\tilde\mu$ functions of $\tilde q_A$ only,

$$-l^{\dot B\dot C} = \begin{pmatrix} b_0 & c_0\\ c_0 & a_0\end{pmatrix}, \tag{B9}$$

and F and G are higher-order potentials obtained by noting that the constraint equation on Ω is completely equivalent to the assertion of the existence of two arbitrary functions F and G such that

$$\Omega_{\tilde q_n}\Omega_{q^i\tilde q^n} = q_A + F_{q_i} \quad\text{or}\quad \Omega_{q_i}\Omega_{q^i\tilde q^n} = \tilde q_B + G_{\tilde q^n}. \tag{B10}$$

The Ω that corresponds to $\Theta = 0$ (via the approach described in Appendix A) is $\Omega = q^A\tilde q_A$. One then easily finds that

$$\begin{aligned} L^A &= \alpha_0\tilde q^A + A_0{}^{AB}q_B + \tfrac12\rho_0 q^A + B_0{}^A,\\ \widetilde{M}^A &= -\tfrac12\gamma_0 q^A + A_0{}^{AB}\tilde q_B - \tfrac12\rho_0\tilde q^A + C_0{}^A, \end{aligned}$$

where the symbols denoting the ten parameters have the same meaning as before. One may then continue on to the situation $\widetilde{\Theta} = 0$ in coordinates $\tilde p^A$, $\tilde q_B$. Denoting the analogous quantities by tilded symbols we find that

$$\begin{aligned} &\widetilde{A}_0{}^{AB} = A_0{}^{AB}, \quad 2\tilde\alpha_0 = \gamma_0, \quad \tilde\rho_0 = -\rho_0,\\ &\tilde\gamma_0 = \tfrac12\alpha_0, \quad \widetilde{B}_0{}^A = C_0{}^A, \quad \widetilde{C}_0{}^A = B_0{}^A, \end{aligned}$$

as suggested in the main body of this article.

[1]E.T. Newman, Gen. Rel. Grav. 7, 107 (1976); E.T. Newman and K.P. Tod, Gen. Rel. Grav. 8, 679 (1977); M. Ko. M. Ludvigsen, E.T. Newman, and K.P. Tod, "The Theory of $\mathscr{H}$-space," in *Asymptotic Structure of Space–Time*, edited by F.P. Esposito and L. Witten (Plenum, N. Y., 1977).
[2]R. Penrose, Gen. Rel. Grav. 7, 31 (1976); R.S. Ward, Proc. Soc. (London) Ser. A **363**, 289 (1978).
[3]J.F. Plebański, J. Math. Phys. **16**, 2395 (1975).
[4]C.P. Boyer and J.F. Plebański, J. Math. Phys. **18**, 1022 (1977). See also Ref. 9 for a later discussion of the spinor notation in $\mathscr{H}$ spaces.
[5]C.P. Boyer, J.D. Finley, III and J.F. Plebański, "Complex General Relativity, $\mathscr{H}$ and $\mathscr{H}\mathscr{H}$-Spaces—A Survey of One Approach," to appear in the Einstein memorial volume published by the Society for General Relativity & Gravitation (Plenum, N.Y., 1979).
[6]J.D. Finley, III and J.F. Plebański, J. Math. Phys. **17**, 585 (1976).
[7]C.W. Fette, A.I. Janis, and E.T. Newman, Gen. Rel. Grv. **8**, 29 (1977).
[8]J.D. Finley, III and J.F. Plebański, J. Math. Phys. **19**, 760 (1978). Note that there are some printing errors there with respect to the sign of terms involving α_0.
[9]J.D. Finley, III and J.F. Plebański, J. Math. Phys. **17**, 2207 (1976).
[10]E.J. Flaherty, *Hermitian and Kählerian Geometry in Relativity* (Springer–Verlag, Berlin, 1976).
[11]E.J. Ernst and J.F. Plebański, Ann. Phys. (N.Y.) **107**, 266 (1977); S. Hacyan and J.F. Plebański, J. Math. Phys. **17**, 2203 (1976).
[12]R.P. Gilbert, *Function Theoretic Methods in Partial Differential Equations*, (Academic, N.Y., 1969).
[13]W. Miller, Jr., *Symmetry and Separation of Variables* (Addison–Wesley, Reading, Mass., 1977).
[14]N.S. Koshlyakov, M.M. Smirnov, and E.B. Gliner, *Differential Equations of Mathematical Physics* (North-Holland, Amsterdam, 1964).
[15]E. Cartan, *Les systèmes différentiels extérieurs et leurs applications géométriques* (Hermann, Paris, 1971), Chapter IV.
[16]See also Sec. 5 of Ref. 4 where a determination of the Cartan genus for our Eq. (1.1) is given in some detail.
[17]K.P. Tod, Twistor Newsletter, 1977 (unpublished).
[18]The literature on gravitational instantons is already quite large. We mention only a few selected papers. S.W. Hawking, Phys. Lett. A **60**, 81 (1977); G.W. Gibbons and S.W. Hawking, Phys. Lett. B **78**, 430 (1978); T. Eguchi, P.B. Gilkey, and A.J. Hanson, Phys. Rev. D **17**, 423 (1978); A. Back, P.G.O. Freund, and M. Forger, Phys. Lett. B **77**, 181 (1978).

The nondiverging and nontwisting type D electrovac solutions with λ

Jerzy F. Plebański

Centro de Investigación y Estudios Avanzados del I.P.N., México, México 14 — 740
(Received 15 May 1978; revised manuscript received 5 September 1978)

Assuming the type D of the metric and the alignment of the EM field along the double D–P directions all solutions of Einstein–Maxwell equations free of complex expansion in the presence of cosmological constant are studied. All solutions of this type are found equivalent to Carter's branch $[\tilde{B}(-)]$ in a previous paper and are derivable from the general 7-parametric D's by contractions. Various special cases are examined, and at least a 4-parameter group of symmetries of these solutions is exhibited and studied.

INTRODUCTION

The systematic progress in the study of the algebraically degenerate electrovac solutions is perhaps the most advanced for the D-type metrics. Carter[1] was the first to obtain an important subfamily of D's—with the cosmological constant λ—which contains the charged Kerr–NUT metrics.[2,3] His result, derived from the theory of separation of variables, consists of a basic metric and field—branch $[A]$—which through limiting processes can degenerate into his branches $[\tilde{B}(+)]$, $[\tilde{B}(-)]$, and $[D]$, all of them being described in remarkably simple and plausible coordinate charts. An essential step in the theory of D's was then accomplished by Kinnersley. In his thesis,[4] working with NP formalism, he presents all electrovac solutions—but without λ—under the assumption (A)≡that both real eigenvectors of the E.M. field are aligned along the double DP vectors. His vacuum results, with the explicit statement about completeness, have been published in Ref. 5. The most essential point in Kinnersley's work is that all of his metrics, including the charged ones, possess two commuting Killing vectors. (The deeper reasons for this were later explained by Hughston, *et al.*[6]) Most of the metrics listed in Ref. 4 were known—obtained by different methods before: Kinnersley states that "the only new one is (IV.51)" (of Ref. 4). I believe that also (V.13) of Ref. 4 was new at that time. However, of methodological importance was the identification of the previously known solutions on the list of Kinnersley subcases. In particular, Kinnersley identifies the Carter branch $[\tilde{B}(-)]$ (with $\lambda = 0$) with *all* charged divergenceless D's confined additionally to the assumption (A).[7] The form in which Refs. 4 and 5 state results was, however, not the optimal one: The unnecessary use of elliptic functions, the rather complicated coordinatization and incomplete interpretation of parameters of numerous subbranches asked for improvements. After the work of Debever in this direction,[8] we succeeded with Demiański to derive a seven-parameter family of charged D's with λ [where (A) applies], which—modulo limiting transitions—contains *all* branches of D's *with* λ previously integrated by other authors. Our results stated in Refs. 9–11 are obtained independently of Ref. 4, by techniques of Ref. 12 and describe the metric and field in a compact form (in coordinates generalizing those which appear in the Carter subbranch) permitting thus a natural interpretation for all parameters involved.

In this situation a conjecture was expressed in Ref. 10 that, provided (A) holds, all charged D's *with* λ are contained—modulo limiting transitions—in our seven-parametric family. (This opinion is apparently shared by Kinnersley, at least as far as his subcase of $\lambda = 0$ is considered; see Ref. 13, p. 129.) Indeed, in Refs. 9, 10, and 11 it was explicitly shown that the two complete branches of charged D's without λ, i.e., (1) the Carter branch (without the acceleration parameter), and (2) the charged C branch (without the rotation parameter), studied with more details in Ref. 14, are contractions of the seven-parametric solution. For the uncharged subcase of $\lambda = 0 = e_0 + ig_0$, the independent proof that Kinnersley's metrics are all contained in the seven-parametric solution is given by the direct integration of equations of Ref. 12 in the recent paper by Weir and Kerr.[15] Now, it is the intention of Kerr and the present author (working together with Alarcón Gutiérrez), to prove, following the technique of Ref. 15 the completeness of the seven-parametric solution in all generality, with both λ and $e_0 + ig_0$ being taken into account. Within this program, the present paper has only limited objectives. As the presence of λ forms evidently the main obstacle to the claims of completeness, it concentrates on the divergenceless charged D's (i.e., without the complex expansion) in *the presence of* λ and under the assumption (A) (this is a simplest subcase of what must be studied anyway within a more general program). Notice, that the solutions with $Z = 0$ were also studied long ago by Kundt[16] who established the basic results concerning the radiative type N and III solutions in the vacuum case. A formal proof is given—based on a direct integration process—that *all* charged D's with λ, submitted to assumption (A), which, moreover, have the geodesic and shearless double DP directions, overlap precisely with the $[\tilde{B}(-)]$ solutions of Carter[1] modulo the possibility of the contraction to the Bertotti–Robinson solutions[17,18] which correspond to

[a]On leave of absence from University of Warsaw, Warsaw, Poland.

 0022-2488/79/091946-17$01.00 © 1979 American Institute of Physics

Carter's $[D]$ case. When $\lambda < 0$ and the double DP vectors *are not* assumed geodesic and shearless (a subcase studied in a paper by Hacyan and the present author[19]), we have found an exceptional branch of charged D's which, in general, does not possess any Killing vectors at all, and have one of the DP vectors with nonvanishing geodesicity. (The existence of this exceptional branch illustrates the nontrivial nature of the inclusion of λ into the completeness program.) The basic form of the nonexceptional branch of D's with $Z = 0$ given in this paper exhibits a 2-space of constant curvature in the structure of the solution, does not contain spurious parameters which were present in Carter's $[\tilde{B}(-)]$ form, and moreover, permits one to read off the four Killing vectors with which these solutions are endowed.[20] A section of this paper analyzes also in detail how the nonexceptional branch of divergenceless D's with λ can be obtained by a series of contractions from the seven parameter metric. In doing so, one is able to establish how the parameters of the metric can be viewed as the "genetic descendents" of parameters with established interpretation. In particular, one finds that the studied metric arises as a result of two-step contraction: (1) after switching off the acceleration, and (2) after a contraction at the generalized *exceptional value* of the Kerr parameter (a generalization of $m^2 = a^2$).[21] With the interpretation of the parameters established, it was also possible to reexamine the Melvin "magnetic universe" as the special case of the discussed nonexceptional branch.[22,23] As to the formalism used in the present paper, it is the same as in Refs. 12, 9, 10, 11, and 15; for a more complete description see Ref. 24.

1. STATEMENT OF THE PROBLEM

We are interested in the search for solutions of the Einstein–Maxwell equations with λ which: (1) are of type D and (2) are characterized by such an alignment of the electromagnetic field (if it is nontrivial) with respect to the conformal curvature that the double DP vectors are its eigenvectors.

We shall work with the null tetrad formalism. Thus, with the metric given in the form

$$ds^2 = 2e^1 \underset{s}{\otimes} e^2 + 2e^3 \underset{s}{\otimes} e^4 \tag{1.1}$$

(with $e^2 = \bar{e}^1$ and e^3, e^4 real) we can choose the (cotangent) double DP vectors to coincide with e^3 and e^4. Then the spinorial components of the electromagnetic field, f_{AB}, can be assumed to have the values:

$$f_{11} = 0 = f_{22}, \quad f_{12} = \tfrac{1}{4}(\mathscr{E} + i\check{\mathscr{B}}) \tag{1.2}$$

so that

$$\begin{aligned}\omega: &= \tfrac{1}{2}(f_{\mu\nu} + \check{f}_{\mu\nu})\, dx^\mu \wedge dx^\nu \\ &= (\mathscr{E} + i\check{\mathscr{B}})(e^1 \wedge e^2 + e^3 \wedge e^4),\end{aligned} \tag{1.3}$$

where the scalar $\mathscr{E}$ and the pseudo-scalar $\check{\mathscr{B}}$ characterize the complex electromagnetic invariant

$$\mathscr{F} := \tfrac{1}{4} f_{\mu\nu} f^{\mu\nu} + \tfrac{1}{4}\check{f}_{\mu\nu} f^{\mu\nu} = -\tfrac{1}{2}(\mathscr{E} + i\check{\mathscr{B}})^2. \tag{1.4}$$

[we use the definition of the dual:

$$\check{f}^{\mu\nu} = (i/2\sqrt{-g})\epsilon^{\mu\nu\rho\sigma} f_{\rho\sigma}$$

assuming of course $\epsilon^{1234} = 1$.]

The Maxwell equations ($d\omega = 0$) can be easily seen to reduce to the statement that

$$\begin{aligned}&d \ln(\mathscr{E} + i\check{\mathscr{B}})^{1/2} + \Gamma_{314}e^1 + \Gamma_{423}e^2 - \Gamma_{312}e^3 - \Gamma_{421}e^4 \\ &\quad = 0,\end{aligned} \tag{1.5}$$

where $\Gamma_{abc} = \Gamma_{[ab]c}$ are the components of the connection forms $\Gamma_{ab} = \Gamma_{abc}e^c$ which are defined by the first structure equations,

$$de^a = e^b \wedge \Gamma^a{}_b. \tag{1.6}$$

Now, the Einstein equations with electromagnetic sources require [we work with the signature $(+\ +\ +\ -)$ and in units where $G = 1 = c$; with $R_{ab} = R^s{}_{abs}$ being the tetrad components of the Ricci tensor, $R = R^s{}_s$, we use as the symbol for the traceless Ricci tensor $C_{ab} := R_{ab} - \tfrac{1}{4}g_{ab}R$]:

$$\begin{aligned}&R = -4\lambda, \quad C_{ab} = 0 \text{ except for} \\ &C_{12} = -C_{34} = -\mathscr{E}^2 - \check{\mathscr{B}}^2.\end{aligned} \tag{1.7}$$

Moreover, with e^3 and e^4 being double DP vectors, for the quantities $C^{(a)}$ which characterize the conformal curvature, we have:

$$C^{(a)} = 0, \quad C^{(3)} \neq 0. \tag{1.8}$$

Therefore, the second Cartan formulas with the Einstein equations incorporated into them amount to the conditions

$$\begin{aligned}\mathscr{A} &:= d\Gamma_{42} + \Gamma_{42} \wedge (\Gamma_{12} + \Gamma_{34}) = (\tfrac{1}{2}C^{(3)} + \lambda_0)e^3 \wedge e^1, \\ \mathscr{B} &:= d\Gamma_{31} + (\Gamma_{12} + \Gamma_{34}) \wedge \Gamma_{31} = (\tfrac{1}{2}C^{(3)} + \lambda_0)e^4 \wedge e^2,\end{aligned} \tag{1.9}$$

$$\begin{aligned}\mathscr{C} &:= d(\Gamma_{12} + \Gamma_{34}) + 2\Gamma_{42} \wedge \Gamma_{31} \\ &= [C^{(3)} - \lambda_0 - \mathscr{E}^2 - \check{\mathscr{B}}^2]e^1 \wedge e^2 \\ &\quad + [C^{(3)} - \lambda_0 + \mathscr{E}^2 + \check{\mathscr{B}}^2]e^3 \wedge e^4,\end{aligned}$$

where $3\lambda_0 := \lambda$ (we use here the abbreviation $\lambda_0 = \lambda/3$ in order to avoid the bothersome factor "1/3" in the subsequent formulas).

The differential equation of the problem considered thus amounts to (1.5), (1.6), and (1.9). The integrability conditions of this system of equations consists of the Bianchi identities; the special Bianchi identities are automatically fulfilled when Maxwell equations are assumed. The general Bianchi identities amount in the case considered to

$$\begin{aligned}&[\mathscr{E}^2 + \check{\mathscr{B}}^2 - \tfrac{3}{2}C^{(3)}]\cdot\begin{Bmatrix}\Gamma_{424} \\ \Gamma_{313}\end{Bmatrix} = 0, \\ &[\mathscr{E}^2 + \check{\mathscr{B}}^2 + \tfrac{3}{2}C^{(3)}]\cdot\begin{Bmatrix}\Gamma_{422} \\ \Gamma_{311}\end{Bmatrix} = 0\end{aligned} \tag{1.10}$$

and

$$\begin{aligned}\tfrac{1}{2}dC^{(3)} &= [\mathscr{E}^2 + \check{\mathscr{B}}^2 - \tfrac{3}{2}C^{(3)}]\cdot(\Gamma_{314}e^1 + \Gamma_{423}e^2) \\ &\quad + [\mathscr{E}^2 + \check{\mathscr{B}}^2 + \tfrac{3}{2}C^{(3)}]\cdot(\Gamma_{312}e^3 + \Gamma_{421}e^4).\end{aligned} \tag{1.11}$$

It is obvious that in the description of the problem presented above, the tetrad remains free up to the σ gauge, i.e., the transformations:

$$\begin{aligned}&e^1 \to e^{i\,\mathrm{Im}\sigma}\cdot e^1, \qquad e^3 \to e^{\mathrm{Re}\sigma}\cdot e^3, \\ &e^2 \to e^{-i\,\mathrm{Im}\sigma}\cdot e^2, \qquad e^4 \to e^{-\mathrm{Re}\sigma}\cdot e^4,\end{aligned} \tag{1.12}$$

where complex σ is arbitrary.

In the present text we fulfill the integrability conditions (1.10) by assuming that both e^3 and e^4 are geodesic and shearfree:

$$\Gamma_{424} = \Gamma_{422} = 0, \quad \Gamma_{313} = \Gamma_{311} = 0. \tag{1.13}$$

If only

$$(\mathscr{E}^2 + \check{\mathscr{B}}^2)^2 - (\tfrac{3}{2}C^{(3)})^2 \neq 0, \tag{1.14}$$

then (1.13) is not an independent assumption, but the necessary consequence of the previous postulates. The case of (1.14) being not valid, i.e., $(\mathscr{E}^2 + \check{\mathscr{B}}^2)^2 - (\frac{3}{2}C^{(3)})^2 = 0$, is studied in detail in a separate publication by Hacyan and the present author.[16]

Moreover, in the present text, we confine ouselves to the study of the *divergenceless* subcase, where both distinguished congruences have vanishing complex expansions, i.e.,

$$\Gamma_{421} = 0, \quad \Gamma_{312} = 0. \tag{1.15}$$

Within these restrictive assumptions, we intend, however, to integrate the problem completely, i.e., to determine *all* divergenceless D's *with* λ where the real eigenvectors of the electromagnetic field coincide with the double DP directions.

2. THE DETERMINATION OF THE TETRAD

Under our assumptions we have:

$$\Gamma_{42} = \Gamma_{423}e^3, \quad \Gamma_{31} = \Gamma_{314}e^4. \tag{2.1}$$

Theorem: We claim that necessarily

$$\Delta := |\Gamma_{423}|^2 - |\Gamma_{314}|^2 = 0. \tag{2.2}$$

Proof; Indeed, assume the contrary, i.e., $\Delta \neq 0$. Then, if $\mathscr{E} + i\check{\mathscr{B}} \neq 0$, the Maxwell equations (1.5) written in the form:

$$\begin{aligned} &d\ln(\mathscr{E} + i\check{\mathscr{B}})^{1/2} + \Gamma_{314}e^1 + \Gamma_{423}e^2 = 0, \\ &d\ln(\mathscr{E} - i\check{\mathscr{B}})^{1/2} + \Gamma_{324}e^2 + \Gamma_{413}e^1 = 0, \end{aligned} \tag{2.3}$$

could be solved for e^1 and e^2, thus expressing these forms as linear combinations of two differentials. But if so, it follows that

$$e^1 \wedge e^2 \wedge de^1 = 0 = e^1 \wedge e^2 \wedge de^2. \tag{2.4}$$

If $\mathscr{E} + i\check{\mathscr{B}} = 0$, and we cannot infer this from Maxwell's equations, then from (1.11) in the form

$$\begin{aligned} &\tfrac{1}{3}d\ln C^{(3)} + \Gamma_{314}e^1 + \Gamma_{423}e^2 = 0, \\ &\tfrac{1}{3}d\ln \bar{C}^{(3)} + \Gamma_{324}e^2 + \Gamma_{413}e^1 = 0, \end{aligned} \tag{2.5}$$

by the same argument, we again with $\Delta \neq 0$ reach the conclusion that (2.4) applies. On the other hand, from the first structure equations employing (2.1) we have:

$$\begin{aligned} de^1 &= -e^1 \wedge \Gamma_{12} + (\Gamma_{423} - \Gamma_{324})e^3 \wedge e^4, \\ de^2 &= e^2 \wedge \Gamma_{12} + (\Gamma_{413} - \Gamma_{314})e^3 \wedge e^4, \end{aligned} \tag{2.6}$$

and using (2.4) we obtain

$$e^1 \wedge e^2 \wedge e^3 \wedge e^4 \cdot \begin{cases} \Gamma_{423} - \Gamma_{324} \\ \Gamma_{413} - \Gamma_{314} \end{cases} = 0. \tag{2.7}$$

Therefore, $\Gamma_{423} = \Gamma_{324} (= \overline{\Gamma_{314}})$ which leads to the contradictory $\Delta = 0$. □

Besides (2.6) the two remaining first structure equations are:

$$\begin{aligned} de^3 &= (\Gamma_{413}e^1 + \Gamma_{423}e^2 + \Gamma_{34}) \wedge e^3, \\ de^4 &= (\Gamma_{314}e^1 + \Gamma_{324}e^2 - \Gamma_{34}) \wedge e^4. \end{aligned} \tag{2.8}$$

Therefore, $e^3 \wedge de^3 = 0 = e^4 \wedge de^4$, and, as it is well known for geodesic, shearless, and (complex) expansionless congruences, both forms e^3 and e^4 are surface orthogonal. Employing this fact, and using the freedom of the boost transformations from (1.12), we can thus so fix the tetrad that:

$$e^3 = A du, \quad e^4 = A dv \tag{2.9}$$

with the same (real) coefficient A.

We must now distinguish two cases: If $\Gamma_{423} \neq 0$, then consistently with (2.2) by using the remaining freedom of the phase transformations in (1.12) we can entirely freeze the tetrad by requiring

$$\Gamma_{423} = \Gamma_{314} =: B \neq 0 \rightarrow \Gamma_{42} = Be^3, \ \Gamma_{31} = Be^4, \tag{2.10}$$

where B is in general complex. There is also the pathological possibility that $\Gamma_{423} = 0$ and by (2.2) $\Gamma_{314} = 0$. Here we are still left with the freedom of the phase transformations. This pathological case with $\Gamma_{42} = 0 = \Gamma_{31}$ which leads to the Bertotti–Robinson (BR) solution,[17,18] will be discussed in a separate section. Right now, we shall thus proceed with the explicit assumption $B \neq 0$.

Maxwell equations and the Bianchi identities now take the form of:

$$d\ln(\mathscr{E} + i\check{\mathscr{B}})^{1/2} + B(e^1 + e^2) = 0, \tag{2.11}$$

$$\tfrac{1}{2}dC^{(3)} = [\mathscr{E}^2 + \check{\mathscr{B}}^2 - \tfrac{3}{2}C^{(3)}] \cdot B \cdot (e^1 + e^2). \tag{2.12}$$

The first structure equations can now be stated in the form of:

$$d(e^1 + e^2) = \Gamma_{12} \wedge (e^1 - e^2), \tag{2.13a}$$

$$d(e^1 - e^2) = \Gamma_{12} \wedge (e^1 + e^2), + 2(B - \bar{B})e^3 \wedge e^4, \tag{2.13b}$$

$$(-d\ln A + \bar{B}e^1 + Be^2 + \Gamma_{34}) \wedge e^3 = 0, \tag{2.13c}$$

$$(-d\ln A + Be^1 + \bar{B}e^2 - \Gamma_{34}) \wedge e^4 = 0. \tag{2.13d}$$

Moreover, with $\Gamma_{42} = ABdu$, $\Gamma_{31} = ABdv$, we easily see that the first two of the conditions (1.9) involving the forms $\mathscr{A}$ and $\mathscr{B}$ can be stated in the form of:

$$\begin{aligned} &[d\ln AB - \Gamma_{12} - \Gamma_{34} + B^{-1}(\tfrac{1}{2}C^{(3)} + \lambda_0)e^1] \wedge e^3 = 0, \\ &[d\ln AB + \Gamma_{12} + \Gamma_{34} + B^{-1}(\tfrac{1}{2}C^{(3)} + \lambda_0)e^2] \wedge e^4 = 0. \end{aligned} \tag{2.14}$$

It is then obvious that the Eq. (2.13c), (2.13d), and (2.14) are equivalent to:

$$\begin{aligned} -d\ln A + \bar{B}e^1 + Be^2 + \Gamma_{34} &= -\rho e^3, \\ -d\ln A + Be^1 + \bar{B}e^2 - \Gamma_{34} &= -\sigma e^4, \end{aligned} \tag{2.15}$$

and

$$\begin{aligned} &d\ln AB - \Gamma_{12} - \Gamma_{34} + B^{-1}(\tfrac{1}{2}C^{(3)} + \lambda_0)e^1 = \zeta e^3, \\ &d\ln AB + \Gamma_{12} + \Gamma_{34} + B^{-1}(\tfrac{1}{2}C^{(3)} + \lambda_0)e^2 = \eta e^4, \end{aligned} \tag{2.16}$$

where ρ and σ are real and ζ and η are in general complex.

We can now organize better the information contained in these relations: By adding and subtracting (2.15) and (2.16) we obtain:

$$d\ln A^2 = (B+\bar{B})(e^1+e^2) + \rho e^3 + \sigma e^4, \tag{2.17a}$$

$$d\ln A^2B^2 = -B^{-1}(\tfrac{1}{2}C^{(3)}+\lambda_0)(e^1+e^2) + \zeta e^3 + \eta e^4, \tag{2.17b}$$

and

$$\Gamma_{34} = \tfrac{1}{2}(B-\bar{B})(e^1-e^2) - \tfrac{1}{2}\rho e^3 + \tfrac{1}{2}\sigma e^4, \tag{2.18a}$$

$$\Gamma_{12}+\Gamma_{34} = \tfrac{1}{2}B^{-1}(\tfrac{1}{2}C^{(3)}+\lambda_0)(e^1-e^2) - \tfrac{1}{2}\zeta e^3 + \tfrac{1}{2}\eta e^4. \tag{2.18b}$$

The complex conjugate of the last relation is of course

$$-\Gamma_{12}+\Gamma_{34} = -\tfrac{1}{2}\bar{B}^{-1}(\tfrac{1}{2}\bar{C}^{(3)}+\lambda_0)(e^1-e^2) - \tfrac{1}{2}\bar{\zeta}e^3 + \tfrac{1}{2}\bar{\eta}e^4. \tag{2.19}$$

Therefore, (2.18b) amounts to the information that:

$$\Gamma_{12} = \tfrac{1}{4}\{B^{-1}(\tfrac{1}{2}C^{(3)}+\lambda_0)+\bar{B}^{-1}(\tfrac{1}{2}\bar{C}^{(3)}+\lambda_0)\}(e^1-e^2) - \tfrac{1}{4}(\zeta-\bar{\zeta})e^3 + \tfrac{1}{4}(\eta-\bar{\eta})e^4, \tag{2.20a}$$

$$\Gamma_{34} = \tfrac{1}{4}\{B^{-1}(\tfrac{1}{2}C^{(3)}+\lambda_0)-\bar{B}^{-1}(\tfrac{1}{2}\bar{C}^{(3)}+\lambda_0)\}(e^1-e^2) - \tfrac{1}{4}(\zeta+\bar{\zeta})e^3 + \tfrac{1}{4}(\eta+\bar{\eta})e^4, \tag{2.20b}$$

this immediately leads to a theorem.

Theorem: The form e^1+e^2 is closed, i.e., there exists a real function X such that:

$$e^1+e^2 = dX. \tag{2.21}$$

Proof: Indeed, feeding (2.20a) into (2.13a) we have

$$d(e^1+e^2) = -\tfrac{1}{4}(\zeta-\bar{\zeta})e^3\wedge(e^1-e^2) + \tfrac{1}{4}(\eta-\bar{\eta})e^4\wedge(e^1-e^2). \tag{2.22}$$

On the other hand, if $\mathscr{C}+i\check{\mathscr{B}}\neq 0$, (2.11) implies

$$(e^1+e^2)\wedge d(e^1+e^2) = 0. \tag{2.23}$$

If $\mathscr{C}+i\check{\mathscr{B}}=0$, then (2.23) also follows from (2.12). Therefore, (2.22) necessitates for consistency:

$$2e^1\wedge e^2\wedge\{-\tfrac{1}{4}(\zeta-\bar{\zeta})e^3+\tfrac{1}{4}(\eta-\bar{\eta})e^4\} = 0, \tag{2.24}$$

so that both ζ and η must be real:

$$\zeta=\bar{\zeta},\quad \eta=\bar{\eta}. \tag{2.25}$$

But if so, (2.22) reduces to:

$$d(e^1+e^2)=0 \tag{2.26}$$

which proves the thesis (2.21). □

We should like to note that X, u, and v can be considered as independent functions; indeed

$$dX\wedge du\wedge dv = A^{-2}(e^1+e^2)\wedge e^3\wedge e^4\neq 0. \tag{2.27}$$

Knowing (2.18a) and (2.20b) we can now compare the two expressions for Γ_{34}; remembering (2.25) we infer easily that:

$$\rho=\zeta,\quad \sigma=\eta \tag{2.28}$$

and that

$$\tfrac{1}{2}B^{-1}(\tfrac{1}{2}C^{(3)}+\lambda_0) - B = \tfrac{1}{2}\bar{B}^{-1}(\tfrac{1}{2}\bar{C}^{(3)}+\lambda_0]) - \bar{B} =: G, \tag{2.29}$$

where G is some real function.

We can now return to the Eqs. (2.17). Using (2.17a) in (2.17b) and remembering (2.28) we infer easily that

$$d\ln B^2 = -\{B^{-1}(\tfrac{1}{2}C^{(3)}+\lambda_0)+B+\bar{B}\}(e^1+e^2) \tag{2.30}$$

or, if we shall use (2.29) and (2.21),

$$dB = -B\{G+B+\tfrac{1}{2}(B+\bar{B})\}dX. \tag{2.31}$$

This important condition implies that B is a function of the real X only. But if so, (2.17a) means that

$$d\left\{\ln A^2 - \int(B+\bar{B})dX\right\} = \rho A\,du + \sigma A\,dv. \tag{2.32}$$

With $du\wedge dv\neq 0$, this means, however, that there exists a real function $g=g(u,v)$ such that:

$$\ln A^2 - \int(B+\bar{B})dX = \ln g(u,v) \tag{2.33}$$

and

$$\rho A = (\ln g)_u,\quad A\sigma = (\ln g)_v. \tag{2.34}$$

Therefore, it follows that A^2 has the form of

$$A^2 = g(u,v)\exp\left(\int(B+\bar{B})dX\right), \tag{2.35}$$

and is a product of a function dependent on X and a function dependent on u and v only.

The other important consequence of (2.31) is that it implies, with G real, that

$$d\ln(B/\bar{B}) + (B-\bar{B})dX = 0, \tag{2.36}$$

which can be easily integrated. Indeed, denoting the X derivative by a dot, we can seek the solution of the equation

$$[\ln(B/\bar{B})]^{\cdot} + B - \bar{B} = 0, \tag{2.37}$$

in the form of

$$B = (\ln F)^{\cdot} = \dot{F}/F, \tag{2.38}$$

where F is some complex function of X.

Then (2.37) reduces to $[\ln(BF/\overline{BF})]^{\cdot}=0$, i.e., to the simple

$$[\ln(\dot{F}/\dot{\bar{F}})]^{\cdot} = 0 \rightarrow \ln(\dot{F}/\dot{\bar{F}}) = 2i\phi_0, \tag{2.39}$$

where ϕ_0 is a constant. Consequently, $\dot{F}\exp(-i\phi_0) - \dot{\bar{F}}\times\exp(i\phi_0) = 0$ and there exists a constant (real) l such that

$$Fe^{-i\phi_0} - \bar{F}e^{i\phi_0} = 2il. \tag{2.40}$$

Since nothing changes if we replace F by $F\exp(i\phi_0)$ in (2.38) we thus conclude that the most general solution of the Eq. (2.34) has the form of

$$B = \dot{H}/(H+il) = [\ln(H+il)]^{\cdot}, \tag{2.41}$$

where l is a real constant and the function $H=H(X)$ is real. With B represented in this manner, we have

$$\int(B+\bar{B})dX = \int[\ln(H^2+l^2)]^{\cdot}dX = \ln(H^2+l^2)+\ln C_0, \tag{2.42}$$

where C_0 is a real constant, which can be set—without losing generality–equal to one, just by incorporating it into the definition of $g(u,v)$ in (2.33)

Thus, we find it convenient to work with $H=H(X)$ as the structural function of the considered class of metrics, because in terms of this function we have simply:

$$A^2 = g(u,v)(H^2+l^2), \quad B = \dot{H}/(H+il). \tag{2.43}$$

Notice that the standing assumption of this section, $B\neq 0$, necessitates of course:

$$\dot{H}\neq 0. \tag{2.44}$$

We can now express the pertinent quantities of the studied problem in terms of H and its derivatives. From (2.30) we have

$$\begin{aligned} B^{-1}(\tfrac{1}{2}C^{(3)}+\lambda_0) &= -\{(\ln B^2)^{\cdot} + B + \bar{B}\} \\ &= -\{\ln B^2(H^2+l^2)\}^{\cdot} \\ &= -\left(\ln \dot{H}^2 \frac{H-il}{H+il}\right)^{\cdot}. \end{aligned} \tag{2.45}$$

Notice that with $C^{(3)}$ understood as defined by this, Eq. (2.29) is now automatically valid and that (2.18b) can be now rewritten in the form of

$$\begin{aligned} \Gamma_{12}+\Gamma_{34} = &-\tfrac{1}{2}\left(\ln\frac{H-il}{H+il}\dot{H}^2\right)^{\cdot}(e^1-e^2) \\ &-\tfrac{1}{2}(\ln g)_u\, du + \tfrac{1}{2}(\ln g)_v\, dv. \end{aligned} \tag{2.46}$$

In particular, subtracting from this its complex conjutate, we obtain

$$\Gamma_{12} = -(\ln\dot{H})^{\cdot}(e^1-e^2). \tag{2.47}$$

The information gathered up to now permits us to integrate the remaining equation for $d(e^1-e^2)$ which follows from the first structure equations, i.e., (2.13b). We can now write it in the form

$$d(e^1-e^2) = -(e^1-e^2)\wedge d\ln\dot{H} + 2(B-\bar{B})A^2\, du\wedge dv. \tag{2.48}$$

However, we notice that

$$\begin{aligned} 2(B-\bar{B})A^2 &= 2\dot{H}\left(\frac{1}{H+il}-\frac{1}{H-il}\right) \\ &\quad\times(H^2+l^2)\cdot g(u,v) \\ &= -4il\dot{H}\cdot g(u,v). \end{aligned} \tag{2.49}$$

Therefore, being sure of $\dot{H}\neq 0$ and dividing (2.48) by $\dot{H}$ we have

$$\begin{aligned} &(1/\dot{H})d(e^1-e^2) - (d\dot{H}/\dot{H}^2)\wedge(e^1-e^2) \\ &\quad= -4ilg(u,v)du\wedge dv \end{aligned} \tag{2.50}$$

so that

$$d\{\dot{H}^{-1}(e^1-e^2)\} = -4ilg(u,v)du\wedge dv. \tag{2.51}$$

It follows that if we represent g in the form

$$g(u,v) = f_u(u,v), \tag{2.52}$$

we have

$$d\{\dot{H}^{-1}(e^1-e^2) + 4ilf(u,v)dv\} = 0. \tag{2.53}$$

Therefore, there exists a real function Y such that

$$e^1-e^2 = i\dot{H}(dY - 4lf(u,v)dv). \tag{2.54}$$

This formula together with:

$$\begin{aligned} &e^1+e^2 = dX, \quad e^3 = \sqrt{f_u}\cdot\sqrt{H^2+l^2}du, \\ &e^4 = \sqrt{f_u}\cdot\sqrt{H^2+l^2}dv, \end{aligned} \tag{2.55}$$

thus establishes the canonical form of the tetrad for the studied problem; the tetrad therefore depends on the two unknown functions $H = H(X)$, $f = f(u,v)$ with $\dot{H}\neq 0\neq f_u$ and on one constant l. It is also obvious that $dX\wedge dY\wedge du\wedge dv \neq 0$ so that $\{X,Y,u,v\}$ can serve as independent coordinates.

The coordinate X is however rather inconvenient for our purposes. It is much more convenient to employ the function H itself as the coordinate: This is always possible because with $\dot{H}\neq 0$, $dX = \dot{H}^{-1}dH$; $\dot{H}$ itself can be thus considered as a function of H. We will thus represent this quantity in the form of:

$$\dot{H}^2 = \mathscr{Q}(H)/(H^2+l^2) \tag{2.56}$$

and we shall consider $\mathscr{Q} = \mathscr{Q}(H)$ as the sought structural function; the tetrad is therefore now:

$$\begin{aligned} &e^1+e^2 = \sqrt{(H^2+l^2)/\mathscr{Q}}\ dH, \\ &e^1-e^2 = i\sqrt{\mathscr{Q}/(H^2+l^2)}\,(dY - 4lf(u,v)dv), \\ &\left.\begin{matrix} e^3 \\ e^4 \end{matrix}\right\} = \sqrt{f_u}\cdot\sqrt{H^2+l^2}\cdot\left\{\begin{matrix} du \\ dv \end{matrix}\right. . \end{aligned} \tag{2.57}$$

Of course, in the present parametrization of our problem

$$\left.\begin{matrix}\Gamma_{42} \\ \Gamma_{31}\end{matrix}\right\} = B\cdot\left\{\begin{matrix} e^3, \\ e^4, \end{matrix}\right. \quad B = \frac{1}{H+il}\sqrt{\frac{\mathscr{Q}}{H^2+l^2}}, \tag{2.58}$$

so that

$$\left.\begin{matrix}\Gamma_{42} \\ \Gamma_{31}\end{matrix}\right\} = \frac{\sqrt{f_u\mathscr{Q}}}{H+il}\cdot\left\{\begin{matrix} du \\ dv \end{matrix}\right. . \tag{2.59}$$

For $\Gamma_{12}+\Gamma_{34}$ we can use the expression (2.4) which now amounts to

$$\begin{aligned} \Gamma_{12}+\Gamma_{34} = &-\frac{i}{2}\left(\ln\frac{\mathscr{Q}}{(H+il)^2}\right)'\cdot\frac{\mathscr{Q}}{H^2+l^2}\cdot(dY-4lfdv) \\ &-\tfrac{1}{2}(\ln g)_u du + \tfrac{1}{2}(\ln g)_v dv, \end{aligned} \tag{2.60}$$

where prime now denotes differentiation with respect to H and $g\equiv f_u$. One of the basic advantages of the present parametrization of our problem consists in the fact that the Eq. (2.11) takes the form

$$d\ln[(\mathscr{E}+i\check{\mathscr{B}})^{1/2}(H+il)] = 0 \rightarrow \mathscr{E}+i\check{\mathscr{B}} = -z/(H+il)^2, \tag{2.61}$$

where z is a complex constant.

3. THE FINAL CONSEQUENCES OF THE DYNAMICAL EQUATIONS AND THE CANONICAL FORM OF THE METRIC

Until now we did not explore the consequences of the condition (1.9) which involves the form $\mathscr{C}$. According to (2.60) we have

$$\begin{aligned} \mathscr{C} = &-\frac{i}{2}\left[\frac{\mathscr{Q}}{H^2+l^2}\left(\ln\frac{\mathscr{Q}}{(H+il)^2}\right)'\right]' dH\wedge(dY-4lfdv) \\ &+(\ln g)_{uv} du\wedge dv + 2ilg\cdot\frac{\mathscr{Q}}{H^2+l^2} \\ &\times\left(\ln\frac{\mathscr{Q}}{H^2+il}\right)' du\wedge dv + 2g\cdot\frac{\mathscr{Q}}{(H+il)^2}du\wedge dv. \end{aligned} \tag{3.1}$$

On the other hand, according to (1.9) this should be equal to

$$\mathscr{C} = [C^{(3)} - \lambda_0 - \mathscr{E}^2 - \check{\mathscr{B}}^2] e^1 \wedge e^2 + [C^{(3)} - \lambda_0 + \mathscr{E} + \check{\mathscr{B}}^2] e^3 \wedge e^4. \tag{3.2}$$

Evaluating the last object, we can use the fact that

$$e^1 \wedge e^2 = -(i/2) dH \wedge (dY - 4lfdv),$$
$$e^3 \wedge e^4 = g(H^2 + l^2) du \wedge dv. \tag{3.3}$$

Moreover, according to (2.45) we can substitute for $C^{(3)}$:

$$C^{(3)} = -2\lambda_0 - \frac{2}{H+il} \cdot \frac{\mathscr{Q}}{H^2+l^2} \left(\ln \frac{\mathscr{Q}}{(H+il)^2} \right)', \tag{3.4}$$

and for $\mathscr{E}^2 + \check{\mathscr{B}}^2$ according to (2.61):

$$\mathscr{E}^2 + \check{\mathscr{B}}^2 = z\bar{z}/(H^2 + l^2)^2. \tag{3.5}$$

Therefore, remembering that $3\lambda_0 = \lambda$, we have for C from (3.2) the expression:

$$\mathscr{C} = -\frac{i}{2} dH \wedge (dY - 4lfdv) \left\{ -\lambda - \frac{2}{H+il} \frac{\mathscr{Q}}{H^2+l^2} \times \left(\ln \frac{\mathscr{Q}}{(H+il)^2} \right)' - \frac{z\bar{z}}{(H^2+l^2)^2} \right\} + g(H^2 + l^2) \times du \wedge dv \left\{ -\lambda - \frac{2}{H+il} \frac{\mathscr{Q}}{H^2+l^2} \times \left(\ln \frac{\mathscr{Q}}{(H+il)^2} \right)' + \frac{z\bar{z}}{(H^2+l^2)^2} \right\}. \tag{3.6}$$

Thus, by comparing (3.1) and (3.6) we arrive at the two equations:

$$\frac{1}{g} (\ln g)_{uv} = -2il \frac{\mathscr{Q}}{H^2+l^2} \left(\ln \frac{\mathscr{Q}}{(H+il)^2} \right)' - 2 \frac{\mathscr{Q}}{(H+il)^2} - 2 \frac{\mathscr{Q}}{H+il} \left(\ln \frac{\mathscr{Q}}{(H+il)^2} \right) - \lambda (H^2 + l^2) + \frac{z\bar{z}}{(H^2+l^2)} \tag{3.7}$$

and

$$\left[\frac{\mathscr{Q}}{H^2+l^2} \left(\ln \frac{\mathscr{Q}}{(H+il)^2} \right)' \right]' = -\lambda - \frac{2}{H+il} \frac{\mathscr{Q}}{H^2+l^2} \left(\ln \frac{\mathscr{Q}}{(H+il)^2} \right)' - \frac{z\bar{z}}{(H^2+l^2)^2}. \tag{3.8}$$

We shall first study the equation (3.7); ordering its right-hand member, we easily reduce it to

$$\frac{1}{g} (\ln g)_{uv} = -\frac{2H\mathscr{Q}'}{H^2+l^2} + \frac{2\mathscr{Q}}{H^2+l^2} - \lambda (H^2 + l^2) + \frac{z\bar{z}}{H^2+l^2} \tag{3.9}$$

Thus, this equation is real, as it should be. It now has very important consequences for our purposes; because its left hand member depends on u and v only, its right-hand member depends on H only, therefore their common value must be a constant which we shall denote by 2ϵ (of course, this constant is real). Thus

$$(\ln g)_{uv} = 2\epsilon g \tag{3.10}$$

and

$$2\epsilon = -[2/(H^2 + l^2)](H\mathscr{Q}' - \mathscr{Q}) - \lambda (H^2 + l^2) + z\bar{z}/(H^2 + l^2). \tag{3.11}$$

Equation (3.10) is however, when $\epsilon \neq 0$, the well-known Liouville's equation which has the most general solution of the form

$$\epsilon \neq 0 \rightarrow g = \frac{1}{\epsilon} \frac{\alpha'(u)\beta'(v)}{(\alpha+\beta)^2} \neq 0, \tag{3.12}$$

where α and β are arbitrary functions of u and v, respectively. If $\epsilon = 0$, the general solution can be of course represented in the form

$$\epsilon = 0 \rightarrow g = p'(u) q'(v) \neq 0. \tag{3.13}$$

For our purposes, however, it is convenient to represent the solution of (3.10) in an universal form valid for $\epsilon \neq 0$ and $\epsilon = 0$. This can be easily done. Let in (3.12) $\alpha = \epsilon\gamma(u)$. Then $g = \gamma'\beta' \cdot (\epsilon\gamma + \beta)^{-2} = -\gamma'(\beta^{-1})' \cdot (1 + \epsilon\gamma\beta^{-1})^{-2}$. Thus, denoting $p(u) = -\gamma(u)$, $q(v) = \beta^{-1}(v)$ we have the general solution of (3.10) in the form

$$g = p'(u) q'(v)/(1 - \epsilon pq)^2, \tag{3.14}$$

valid for all values of ϵ. In order to know the tetrad, however, we still need the $f(u,v)$ defined by $f_u = g$. We can take as the integral of this equation

$$f = pq'/(1 - \epsilon pq) \tag{3.15}$$

which applies for all values of ϵ. Having established this, we notice that the freedom of the arbitrary functions $p = p(u)$ and $q = q(v)$ is apparent and that these functions can be easily absorbed by a coordinate transformation. Indeed, these functions participate in the metric and the forms $e^1 \wedge e^2 + e^3 \wedge e^4$ (needed to construct the electromagnetic field) only in the combinations:

$$fdv = \frac{pdq}{1 - \epsilon pq}, \quad gdu \otimes dv = \frac{dp \otimes dq}{(1 - \epsilon pq)^2},$$
$$gdu \wedge dv = \frac{dp \wedge dq}{(1 - \epsilon pq)^2}. \tag{3.16}$$

It follows that without losing generality, but only by properly adjusting the coordinates, we can set from the very beginning

$$p = u, \quad q = v, \tag{3.17}$$

thus having

$$g = 1/(1 - \epsilon uv)^2, \quad fdv = udv/(1 - \epsilon uv). \tag{3.18}$$

Consequently, the dependence of our tetrad (2.57) on u and v becomes entirely specific:

$$e^1 + e^2 = \sqrt{(H^2 + l^2)/\mathscr{Q}}\, dH,$$
$$e^1 - e^2 = i\sqrt{\mathscr{Q}/(H^2 + l^2)}\, [dY - 4ludv/(1 - \epsilon uv)], \tag{3.19}$$
$$\left.\begin{matrix} e^3 \\ e^4 \end{matrix}\right\} = \frac{\sqrt{H^2 + l^2}}{1 - \epsilon uv} \cdot \left\{\begin{matrix} du \\ dv \end{matrix}\right.$$

We now integrate Eq. (3.11). One easily sees that it can be written in the form

$$\left(\frac{\mathscr{Q}}{H} \right)' = -\epsilon \left(1 + \frac{l^2}{H^2} \right) + \frac{z\bar{z}}{2H^2}$$

$$-\frac{1}{2}\lambda\left(H^2+2l^2+\frac{l^4}{H^2}\right), \tag{3.20}$$

so that

$$\frac{\mathscr{Q}}{H} = -\epsilon\left(H-\frac{l^2}{H}\right)-\frac{1}{2}\frac{z\bar{z}}{H}-\frac{1}{2}\lambda\left(\frac{1}{3}H^3+2l^2H-\frac{l^4}{H}\right)+n_0, \tag{3.21}$$

when n_0 is a constant of integration (of course real). Therefore, with

$$\mathscr{Q} = -\tfrac{1}{2}z\bar{z}+n_0H-\epsilon(H^2-l^2)-\tfrac{1}{2}\lambda\left(\tfrac{1}{3}H^4+2l^2H^2-l^4\right) \tag{3.22}$$

Our tetrad—and hence the metric—is now uniquely fixed.

In order to be sure of the consistency of the result derived above, we must still investigate (3.8). After ordering, one easily finds that it is equivalent to the real condition

$$(H^2+l^2)\mathscr{Q}''-2H\mathscr{Q}'+2\mathscr{Q} = -\lambda(H^2+l^2)^2-z\bar{z}. \tag{3.23}$$

Substituting here from (3.22) one easily finds that this equation becomes an identity, as it should.

We should now like to compute the conformal curvature of the obtained solution. For this purpose, we rewrite (3.4) in the form

$$C^{(3)} = \frac{2}{(H^2+l^2)(H+il)^2}\{-\tfrac{1}{3}\lambda(H^2+l^2)(H+il)^2-(H+il)\mathscr{Q}'+2\mathscr{Q}\}. \tag{3.24}$$

This expression is linear in $\mathscr{Q}$, therefore one can easily compute the contributions from the independent parameters of $\mathscr{Q}$, i.e, λ, ϵ, n_0, and $z\bar{z}$. The result after ordering can be represented in the form

$$C^{(3)} = \frac{2}{(H+il)^3}\left\{n_0+2il\left(\epsilon+\frac{2\lambda}{3}l^2\right)-\frac{z\bar{z}}{(H-il)}\right\}. \tag{3.25}$$

In the next step, we still need some additional information concerning the electromagnetic field which accompanies our solution. Its complex invariant is of course

$$\mathscr{F} = -\tfrac{1}{2}z^2\cdot 1/(H+il)^4 \tag{3.26}$$

The form of the electromagnetic field itself, (1.3), is

$$\omega = -\frac{z}{(H+il)^2}\left\{-\frac{i}{2}dH\wedge\left(dY-\frac{4ludv}{1-\epsilon uv}\right)+\frac{H^2+l^2}{(1-\epsilon uv)^2}du\wedge dv\right\}. \tag{3.27}$$

Remembering that

$$d\frac{udv}{1-\epsilon uv} = \frac{du\wedge dv}{(1-\epsilon uv)^2},$$

one easily shows that this form is indeed closed and can be explicitly represented by the real electric potentials A_μ and the pure imaginary magnetic potentials $\check{A}_\mu$ according to:

$$\omega = -d\chi,\qquad \chi := (A_\mu+\check{A}_\mu)dx^\mu \tag{3.28}$$

$$= z\left\{\frac{udv}{1-\epsilon uv}+\frac{i}{2}\frac{1}{H+il}\left(dY-\frac{4ludv}{1-\epsilon uv}\right)\right\}.$$

The complex potential 1-form given above corresponds, of course, to a specific choice for the electric and magnetic gauges: In general we can replace $\chi\to\chi+dS$, with S being complex without changing the electromagnetic field ω. It can be noticed that χ from (3.28) can be written as spanned by our tetrad,

$$\chi = z\left\{\frac{u}{(H^2+l^2)^{1/2}}e^4+\frac{1}{2}\frac{1}{\sqrt{\mathscr{Q}}}\left(\frac{H-il}{H+il}\right)^{1/2}(e^1-e^2)\right\}. \tag{3.29}$$

We can observe, however, that the result obtained can be represented slightly more symmetrically with respect to the variables u and v if we add to the form (3.28), $dS = -\frac{1}{2}\times[d(uv)/(1-\epsilon uv)]$. We then have

$$\chi = \frac{1}{2}z\left\{\frac{udv-vdu}{1-\epsilon uv}+\frac{i}{H+il}\left(dY'-2l\frac{udv-vdu}{1-\epsilon uv}\right)\right\}$$
$$= \frac{1}{2}z\left\{\frac{1}{(H^2+l^2)^{1/2}}(ue^4-ve^3)+\frac{1}{\sqrt{\mathscr{Q}}}\left(\frac{H-il}{H+il}\right)^{1/2}(e^1-e^2)\right\}, \tag{3.30}$$

where we introduced

$$dY' := dY-2ld(uv)/(1-\epsilon uv). \tag{3.31}$$

The variable Y' is also more adequate on the level of the tetrad. Using it, we will now execute some final adjustments of cosmetical nature in the notation for the result obtained. With the tetrad:

$$e^1+e^2 = \sqrt{(H^2+l^2)/\mathscr{Q}}\,dH,$$
$$e^1-e^2 = i\sqrt{\mathscr{Q}/(H^2+l^2)}\left(dY'+2l\frac{vdu-udv}{1-\epsilon uv}\right), \tag{3.32}$$
$$\left.\begin{matrix}e^3\\e^4\end{matrix}\right\} = \frac{\sqrt{H^2+l^2}}{1-\epsilon uv}\cdot\left\{\begin{matrix}du\\dv\end{matrix}\right.,$$

the final form of the metric is

$$ds^2 = \frac{1}{2}\left\{\frac{H^2+l^2}{\mathscr{Q}}dH^2+\frac{\mathscr{Q}}{H^2+l^2}\left(dY'+2l\times\frac{vdu-udv}{1-\epsilon uv}\right)^2\right\}+2\frac{H^2+l^2}{(1-\epsilon uv)^2}du\,dv. \tag{3.33}$$

Now, the coefficient $\frac{1}{2}$ in the front of the first group of the terms is unesthetical. In order to remove it, it suffices to introduce in place of $\mathscr{Q}$ the structural function $\mathscr{P} := 2\mathscr{Q}$; at the same time, we shall rename the variable H by the latter p; we have thus according to (3.22),

$$\mathscr{P} = -z\bar{z}+2n_0p-2\epsilon(p^2-l^2)-\lambda\left(\tfrac{1}{3}p^4+2l^2p^2-l^4\right). \tag{3.34}$$

The variable $Y'/2$ We will now name σ.

Then the metric assumes the more aesthetic form

$$ds^2 = \frac{p^2+l^2}{\mathscr{P}}dp^2+\frac{\mathscr{P}}{p^2+l^2}\left(d\sigma+l\frac{vdu-udv}{1-\epsilon uv}\right)^2$$

$$+ 2\frac{p^2+l^2}{(1-\epsilon uv)^2}du\,dv. \tag{3.35}$$

In the present notation our null tetrad is:

$$\left.\begin{matrix}e^1\\e^2\end{matrix}\right\} = \frac{1}{\sqrt{2}}\left\{\left(\frac{p^2+l^2}{\mathscr{P}}\right)^{1/2}dp \pm i\left(\frac{\mathscr{P}}{p^2+l^2}\right)^{1/2} \times\left(d\sigma + l\frac{vdu-udv}{1-\epsilon uv}\right)\right\},$$

$$\left.\begin{matrix}e^3\\e^4\end{matrix}\right\} = \frac{\sqrt{p^2+l^2}}{1-\epsilon uv}\cdot\left\{\begin{matrix}du\\dv\end{matrix}\right. . \tag{3.36}$$

The electromagnetic field is now given by

$$\omega = (\mathscr{E}+i\mathscr{B})(e^1\wedge e^2 + e^3\wedge e^4),$$
$$\mathscr{E}+i\mathscr{B} = -z/(p+il)^2, \tag{3.37}$$

or in the form

$$\omega = -d\chi,$$
$$\chi = z\left\{\frac{1}{2}\frac{udv-vdu}{1-\epsilon uv} + \frac{i}{p+il}\left(d\sigma + l\frac{vdu-udv}{1-\epsilon uv}\right)\right\}, \tag{3.38}$$

with the complex invariant

$$\mathscr{F} = -\tfrac{1}{2}z^2/(p+il)^4. \tag{3.39}$$

The object $C^{(3)}$ can be now written in the form

$$C^{(3)} = \frac{2}{(p+il)^3}\left\{n_0 + 2il\left(\epsilon + \frac{2\lambda}{3}l^2\right) - \frac{z\bar{z}}{p-il}\right\}. \tag{3.40}$$

Our motivation of the so-adjusted notation is not only based on esthetical reasons: We also intend to compare the result obtained with the corresponding D's endowed with nonvanishing divergence; for this purpose we will find that the present notation is quite adequate, when we study this problem in one of the subsequent sections.

We will close this section by noticing that in the present notation the quantities Γ_{ab} which accompany our tetrad are

$$\left.\begin{matrix}\Gamma_{42}\\\Gamma_{31}\end{matrix}\right\} = \frac{1}{\sqrt{2}}\frac{\sqrt{\mathscr{P}}}{p+il}\cdot\frac{1}{1-\epsilon uv}\cdot\left\{\begin{matrix}du\\dv\end{matrix}\right. \tag{3.41}$$

and

$$\Gamma_{12}+\Gamma_{34} = -i\frac{\mathscr{P}}{p^2+l^2}\left(\ln\frac{\sqrt{\mathscr{P}}}{p+il}\right)_{,p}\left(d\sigma + l\frac{udv-vdu}{1-\epsilon uv}\right) + \epsilon\frac{udv-vdu}{1-\epsilon uv}. \tag{3.42}$$

The same in the terms of the tetrad amounts to

$$\left.\begin{matrix}\Gamma_{42}\\\Gamma_{31}\end{matrix}\right\} = B\left\{\begin{matrix}e^3\\e^4,\end{matrix}\right.$$
$$B = \frac{1}{\sqrt{2}}\frac{\sqrt{\mathscr{P}}}{(p+il)\sqrt{p^2+l^2}}, \tag{3.43}$$

and

$$\Gamma_{12}+\Gamma_{34} = -\frac{1}{\sqrt{2}}\frac{\sqrt{\mathscr{P}}}{\sqrt{p^2+l^2}}\left(\ln\frac{\sqrt{\mathscr{P}}}{p+il}\right)_{,p}(e^1-e^2) + \frac{\epsilon}{\sqrt{p^2+l^2}}(ue^4-ve^3). \tag{3.44}$$

4. THE PATHOLOGICAL SUBCASES

Because our aim is the completeness of the result, we should now like to examine the only subcase within our assumptions which we have left undiscussed, i.e., the subcase $\Gamma_{423}=0 \rightarrow \Gamma_{314}=0$.

We have here

$$\Gamma_{42}=0, \quad \Gamma_{31}=0. \tag{4.1}$$

Therefore, from (1.5) necessarily

$$\mathscr{E}+i\mathscr{B} = \text{const.} \tag{4.2}$$

The first two of the structure formulas (1.9) then imply

$$C^{(3)} = -2\lambda_0 = -\tfrac{2}{3}\lambda = \text{const.} \tag{4.3}$$

Thus, without λ the considered subcase as a D solution is empty, and all arguments hold only for $\lambda\neq 0$. The real and the imaginary parts of the equation for the form C in (1.9) then give

$$d\Gamma_{12} = -(\lambda + \mathscr{E}^2 + \mathscr{B}^2)e^1\wedge e^2, \tag{4.4}$$
$$d\Gamma_{34} = -(\lambda - \mathscr{E}^2 - \mathscr{B}^2)e^3\wedge e^4, \tag{4.5}$$

The integrability conditions (1.10) and (1.11) are of course automatically satisfied in the present case. Because of (4.4) and (4.5) all vectors e^a are surface orthogonal. Therefore, by using the freedom of the gauge (1.12), we can certainly select the tetrad so that:

$$e^1 = \Omega d\xi, \quad e^3 = A du, \quad e^2 = \Omega d\bar{\xi}, \quad e^4 = A dv, \tag{4.6}$$

with both functions Ω and A being real.

Then (4.4) and (4.5) become equivalent to:

$$(d\ln\Omega - \Gamma_{12})\wedge e^1 = 0, \quad (d\ln A - \Gamma_{34})\wedge e^3 = 0,$$
$$(d\ln\Omega + \Gamma_{12})\wedge e^2 = 0, \quad (d\ln A + \Gamma_{34})\wedge e^4 = 0, \tag{4.7}$$

and consequently

$$d\ln\Omega - \Gamma_{12} = \rho d\xi, \quad d\ln A - \Gamma_{34} = \zeta\, du,$$
$$d\ln\Omega + \Gamma_{12} = \bar{\rho} d\bar{\xi}, \quad d\ln A + \Gamma_{34} = \eta\, dv, \tag{4.8}$$

when ρ is complex and ζ and η are real. Adding and substracting these equations we have:

$$d\ln\Omega^2 = \rho d\xi + \bar{\rho}d\bar{\xi} \rightarrow \Omega = \Omega(\xi,\bar{\xi}), \quad \rho = (\ln\Omega^2)_{\xi},$$
$$\bar{\rho} = (\ln\Omega^2)_{\bar{\xi}},$$
$$d\ln A^2 = \zeta du + \eta dv \rightarrow A = A(u,v), \quad \zeta = (\ln A^2)_u,$$
$$\eta = (\ln A^2)_v, \tag{4.9}$$

and

$$\Gamma_{12} = -\tfrac{1}{2}\rho d\xi + \tfrac{1}{2}\bar{\rho}d\bar{\xi} = -\tfrac{1}{2}(\ln\Omega^2)_{\xi}d\xi + \tfrac{1}{2}(\ln\Omega^2)_{\bar{\xi}}d\bar{\xi}, \tag{4.10}$$
$$\Gamma_{34} = -\tfrac{1}{2}\zeta du + \tfrac{1}{2}\eta dv = -\tfrac{1}{2}(\ln A^2)_u du + \tfrac{1}{2}(\ln A^2)_v dv.$$

Now feeding this into (4.4) we obtain:

$$(\ln\Omega^2)_{\xi\bar{\xi}}d\xi\wedge d\bar{\xi} = -(\lambda + \mathscr{E}^2 + \mathscr{B}^2)\Omega^2 d\xi\wedge d\bar{\xi}, \tag{4.11}$$

$$(\ln A^2)_{uv} du \wedge dv = -(\lambda - \mathscr{E}^2 - \check{\mathscr{B}}^2) A^2 du \wedge dv.$$

Therefore, the functions Ω^2 and A^2 have to fulfill a pair of field equations:

$$(\ln \Omega^2)_{\xi\bar{\xi}} = -(\lambda + \mathscr{E}^2 + \check{\mathscr{B}}^2)\Omega^2,$$
$$(\ln A)_{uv} = -(\lambda - \mathscr{E}^2 - \check{\mathscr{B}}^2) A^2. \tag{4.12}$$

These are Liouville equations of the same type as (3.10) and guided by (3.14) we can take as the most general solution of these:

$$\Omega^2 = f'(\xi)\bar{f}' \cdot [1 + \tfrac{1}{2}(\lambda + \mathscr{E}^2 + \check{\mathscr{B}}^2) f\bar{f}]^{-2},$$
$$A^2 = \alpha'(u)\beta'(v)[1 + \tfrac{1}{2}(\lambda - \mathscr{E}^2 - \check{\mathscr{B}}^2)\alpha\beta]^{-2}, \tag{4.13}$$

where $f = f(\xi)$ is an analytic function which together with the real $\alpha = \alpha(u)$ and $\beta = \beta(v)$ is arbitrary. Because $f(\xi)$, $\alpha(u)$, and $\beta(v)$ enter in the metric and the electromagnetic field only through their differentials and the denominators of (4.13), it is thus clear that, without loosing generality, but only by properly selecting the coordinates, we can assume from the very beginning that $f = \xi$, $\bar{f} = \bar{\xi}$, $\alpha = u$, and $\beta = v$ [compare the argument which led to (3.16)]. Consequently, defining

$$\phi := 1 + \tfrac{1}{2}(\lambda + \mathscr{E}^2 + \check{\mathscr{B}}^2)\xi\bar{\xi},$$
$$\psi := 1 + \tfrac{1}{2}(\lambda - \mathscr{E}^2 - \check{\mathscr{B}}^2)uv, \tag{4.14}$$

we have the null tetrad in the form

$$\left.\begin{matrix} e^1 \\ e^2 \end{matrix}\right\} = \phi^{-1}\left\{\begin{matrix} d\xi \\ d\bar{\xi} \end{matrix}\right., \quad \left.\begin{matrix} e^3 \\ e^4 \end{matrix}\right\} = \psi^{-1}\left\{\begin{matrix} du \\ dv \end{matrix}\right., \tag{4.15}$$

and the metric in the null coordinates $\{x^\mu\} = \{\xi, \bar{\xi}, u, v\}$ is simply

$$ds^2 = 2\phi^{-2} d\xi\, d\bar{\xi} + 2\psi^{-2} du\, dv. \tag{4.16}$$

The electromagnetic field which accompanies this metric is

$$\omega = (\mathscr{E} + i\check{\mathscr{B}})(e^1 \wedge e^2 + e^3 \wedge e^4)$$
$$= (\mathscr{E} + i\check{\mathscr{B}})(\phi^{-2} d\xi \wedge d\bar{\xi} + \psi^{-2} du \wedge dv)$$
$$= \tfrac{1}{2}(\mathscr{E} + i\check{\mathscr{B}}) d\{\phi^{-1}(\xi d\bar{\xi} - \bar{\xi} d\xi) + \psi^{-1}(u dv - v du)\}. \tag{4.17}$$

With $\mathscr{E} + i\check{\mathscr{B}} = \text{const}$, the field invariant is a constant, $\mathscr{F} = -\tfrac{1}{2}(\mathscr{E} + i\check{\mathscr{B}})^2$, and the conformal curvature of the metric is characterized entirely by $C^{(3)} = -\tfrac{2}{3}\lambda$.

Of course, the solution described above is the Bertotti–Robinson solution represented in null coordinates; for its discussion and interpretation see the original papers.[17,18] Notice also that for $\mathscr{E} + i\check{\mathscr{B}} \to 0$, $\lambda \neq 0$, it reduces to the little known Nariari solution,[25] which possesses for $\lambda > 0$, $O(3,\mathbb{R})$ as the subgroup of symmetries.

From the point of view of the present paper, however, we are interested in this solution only for reasons of completeness, as a pathological specimen of the investigated class of solutions. It is clear now that the general solution described in the previous section by the formulas (3.34)–(3.44) and the BR solution described above by (4.14)–(4.17), exhaust *all D* solutions of Einstein–Maxwell equations with λ, where the geodesic, shearless, and divergenceless double DP vectors are the eigenvectors of the electromagnetic field.

The question arises whether the BR solution can be obtained from our more general solution in the previous section by some limiting procedure.

In order to answer this question, we propose to examine the general solution of the previous section with the parameter l switched of. This amounts to the collection of formulas:

$$l \to 0, \quad \mathscr{P} = -z\bar{z} + 2n_0 p - 2\epsilon p^2 - \lambda p^4/3, \tag{4.18a}$$

$$\left.\begin{matrix} e^1 \\ e^2 \end{matrix}\right\} = \frac{1}{\sqrt{2}}\left\{\frac{p\,dp}{\sqrt{\mathscr{P}}} \pm i\frac{\sqrt{\mathscr{P}}}{p} d\sigma\right\}, \tag{4.18b}$$

$$\left.\begin{matrix} e^3 \\ e^4 \end{matrix}\right\} = \frac{p}{1 - \epsilon uv}\left\{\begin{matrix} du \\ dv \end{matrix}\right., \tag{4.18c}$$

$$\omega = -d\chi, \quad \chi = \tfrac{1}{2}z\left\{\frac{u dv - v du}{1 - \epsilon uv} + 2i\frac{d\sigma}{p}\right\}, \tag{4.18d}$$

$$\mathscr{E} + i\check{\mathscr{B}} = -z/p^2, \tag{4.18e}$$

$$C^{(3)} = (2/p^3)(n_0 - z\bar{z}/p). \tag{4.18f}$$

Now let ε be a positive parameter. We then select the constants in the solution (4.18) in a specific manner, setting

$$n_0 = z\bar{z} - \tfrac{1}{3}\lambda + \varepsilon^2, \quad \epsilon = \tfrac{1}{2}(z\bar{z} - \lambda), \tag{4.19}$$

i.e., we now remain only with z, λ (assumed independent of ε), and ε as the disposable parameters of the considered solution. Then, maintaining u and v as coordinates, we introduce the new coordinates p' and σ' through

$$p =: 1 + \varepsilon p', \quad \sigma = \varepsilon^{-1}\sigma'. \tag{4.20}$$

We would now like to investigate the limit of the formulas (4.18) when

$$\varepsilon \to +0, \tag{4.21}$$

with $\{x^\mu\} = \{u, v, p', \sigma'\}$ considered as independent of ε.

We have selected the constants as in (4.19) because with this choice

$$\mathscr{P} = 2\varepsilon^2[1 - \tfrac{1}{2}(\lambda + z\bar{z})p'^2] + O(\varepsilon^3). \tag{4.22}$$

We can now easily execute the limit, obtaining a finite result. Indeed, (4.18c), (4.18e), and (4.18f) all have the trivial limits:

$$\left.\begin{matrix} e^3 \\ e^4 \end{matrix}\right\} = [1 + \tfrac{1}{2}(\lambda - z\bar{z})uv]^{-1}\left\{\begin{matrix} du \\ dv \end{matrix}\right., \tag{4.23a}$$

$$\mathscr{E} + i\check{\mathscr{B}} = -z, \tag{4.23b}$$

$$C^{(3)} = -\tfrac{2}{3}\lambda. \tag{4.23c}$$

Thus, combining (4.23a) and (4.23b) we have precisely

$$\left.\begin{matrix} e^3 \\ e^4 \end{matrix}\right\} = \psi^{-1}\left\{\begin{matrix} du \\ dv \end{matrix}\right., \quad \psi = 1 + \tfrac{1}{2}(\lambda - \mathscr{E}^2 - \check{\mathscr{B}}^2)uv, \tag{4.24}$$

which coincides with the expressions for e^3, e^4 of the BR solution in our coordinates. Observing that $p^{-1}d\sigma = -p'd\sigma' + d(\varepsilon^{-1}\sigma') + O(\varepsilon)$ we also have an easy limit for (4.18d):

$$\omega = -d\chi,$$
$$\chi = -\tfrac{1}{2}(\mathscr{E} + i\check{\mathscr{B}})\{\psi^{-1}(u dv - v du) - 2ip' d\sigma'\}. \tag{4.25}$$

[We substituted here for z from (IV.23b).]

Finally, we execute the limit in (4.18b) by using (4.22), obtaining easily

$$\left.\begin{matrix} e^1 \\ e^2 \end{matrix}\right\} = \frac{1}{2}\frac{dp'}{\sqrt{1-\frac{1}{2}(\lambda+z\bar{z})p'^2}} \pm i\sqrt{1-\frac{1}{2}(\lambda+z\bar{z})p'^2}\,d\sigma'. \tag{4.26}$$

Of course, the range of p' in this contracted solution should be confined to the region where

$$1-\tfrac{1}{2}(\lambda+z\bar{z})p'^2>0. \tag{4.27}$$

We claim that the contracted solution obtained above coincides precisely with the BR solution. To show this, we first rewrite (4.26) in the form:

$$\left.\begin{matrix} e^1 \\ e^2 \end{matrix}\right\} = \sqrt{1-\tfrac{1}{2}(\lambda+z\bar{z})p'^2}\begin{cases} d\eta \\ d\bar{\eta}, \end{cases}$$
$$\eta := \frac{1}{2}\int\frac{dp'}{1-\frac{1}{2}(\lambda+z\bar{z})p'^2} + i\sigma'. \tag{4.28}$$

If the contracted solution coincides with the BR solution, then there exists a real function θ and a coordinate transformation $\xi=\xi(\eta)$ such that

$$e^1 = \sqrt{1-\tfrac{1}{2}(\lambda+z\bar{z})p'^2}\,d\eta = \frac{e^{i\theta}}{1+\frac{1}{2}(\lambda+z\bar{z})\xi\bar{\xi}}d\xi. \tag{4.29}$$

[This formula follows by comparing (4.28) with (4.15), and the observation that the phase of e^1 in the BR solutions and the general solutions which we have had contracted, was fixed according to different criteria.] In order to show that indeed a real θ and $\xi=\xi(\eta)$ can be found so that (4.29) is valid, we first eliminate θ replacing (4.29) by the condition

$$[1-\tfrac{1}{2}(\lambda+z\bar{z})p'^2]d\eta\wedge d\bar{\eta} = \frac{d\xi\wedge d\bar{\xi}}{[1+\frac{1}{2}(\lambda+z\bar{z})\xi\bar{\xi}]^2}. \tag{4.30}$$

Integrating this differential equation for the coordinate transformation $\xi=\xi(\eta)$ we must now distinguish three cases:

$$C^+ : a^2 := \tfrac{1}{2}(\lambda+z\bar{z})>0,\quad C^0 : \lambda+z\bar{z}=0,$$
$$C^- : -a^2 := \tfrac{1}{2}(\lambda+z\bar{z})<0. \tag{4.31}$$

By executing the integral (4.28) in these three cases, we find:

$$\begin{matrix} C^+: \\ C^0: \\ C^-: \end{matrix}\quad \eta = \begin{cases} \dfrac{1}{4a}\ln\dfrac{1+ap'}{1-ap'}+i\sigma' \\ \frac{1}{2}p'+i\sigma' \\ \dfrac{1}{2a}\arctan(ap')+i\sigma' \end{cases}$$
$$\rightarrow \begin{cases} p' = \dfrac{\tanh a(\eta+\bar{\eta})}{a} \\ p' = \eta+\bar{\eta} \\ p' = \dfrac{\tan a(\eta+\bar{\eta})}{a} \end{cases} \tag{4.32}$$

and (4.30) thus reduces correspondingly to:

$$\begin{aligned} &C^+: \quad \frac{d\eta\wedge d\bar{\eta}}{[\cosh a(\eta+\bar{\eta})]^2} = \frac{d\xi\wedge d\bar{\xi}}{(1+a^2\xi\bar{\xi})^2}, \\ &C: \quad d\eta\wedge d\bar{\eta} = d\xi\wedge d\bar{\xi}, \\ &C^-: \quad \frac{d\eta\wedge d\bar{\eta}}{[\cos a(\eta+\bar{\eta})]^2} = \frac{d\xi\wedge d\bar{\xi}}{(1-a^2\xi\bar{\xi})^2}, \end{aligned} \tag{4.33}$$

which now has the obvious solution

$$C^+ : \xi = \frac{\tanh a\eta}{a},\quad C^0 : \xi=\eta,\quad C^- : \xi = \frac{\tan a\eta}{a}. \tag{4.34}$$

Knowing this, we now easily determine that (4.29) is indeed satisfied with the phase factor $e^{i\theta}$ given correspondingly by:

$$\begin{matrix} C^+: \\ C^0: \\ C^-: \end{matrix}\quad e^{i\theta} = \begin{cases} \cosh a\eta/\cosh a\bar{\eta} = [(1-a^2\bar{\xi}^2)/(1-a^2\xi^2)]^{1/2}, \\ 1, \\ \cos a\eta/\cos a\bar{\eta} = [(1+a^2\bar{\xi}^2)/(1+a^2\xi^2)]^{1/2}. \end{cases} \tag{4.35}$$

Moreover, knowing from (4.32) the shape of p' for the three cases as a function of $\eta+\bar{\eta}$ and $2i\sigma'=\eta-\bar{\eta}$, we can now evaluate $d(-2ip'\,d\sigma')$ needed as the contribution to the limiting electromagnetic field in terms of the variables ξ and $\bar{\xi}$:

$$\begin{matrix} C^+: \\ C^0: \\ C^-: \end{matrix}\quad d(-2ip'\,d\sigma') = \begin{cases} \dfrac{2\,d\eta\wedge d\bar{\eta}}{\cosh^2 a(\eta+\bar{\eta})}, \\ 2d\eta\wedge d\bar{\eta}, \\ \dfrac{2\,d\eta\wedge d\bar{\eta}}{\cos^2 a(\eta+\bar{\eta})}. \end{cases} \tag{4.36}$$

But knowing that in the result of (4.34), (4.33) is valid, this simply amounts to:

$$\begin{matrix} C^+: \\ C^0: \\ C^-: \end{matrix}\quad d(-2ip'd\sigma') = 2\begin{cases} \dfrac{d\xi\wedge d\bar{\xi}}{(1+a^2\xi\bar{\xi})^2} \\ d\xi\wedge d\bar{\xi} \\ \dfrac{d\xi\wedge d\bar{\xi}}{(1-a^2\xi\bar{\xi})^2} \end{cases} = d\begin{cases} \dfrac{\xi\,d\bar{\xi}-\bar{\xi}d\xi}{1+a^2\xi\bar{\xi}}, \\ \xi d\bar{\xi}-\bar{\xi}d\xi, \\ \dfrac{\xi\,d\bar{\xi}-\bar{\xi}\,d\xi}{1-a^2\xi\bar{\xi}}. \end{cases} \tag{4.37}$$

This however means that (4.25) can be written in all three cases in the form

$$\begin{aligned} &\omega = -d\chi, \\ &\chi = -\tfrac{1}{2}(\mathscr{E}+i\check{\mathscr{B}})\{\psi^{-1}(udv-vdu) \\ &\qquad + \phi^{-1}(\xi d\bar{\xi}-\bar{\xi}d\xi)\}, \end{aligned} \tag{4.38}$$

where $\phi = 1+\frac{1}{2}(\lambda+\mathscr{E}^2+\check{\mathscr{B}}^2)\xi\bar{\xi}$. Therefore, in coordinates $u, v, \xi, \bar{\xi}$ the electromagnetic field also coincides precisely with the electromagnetic field of the BR metric. It follows that our contracted solution is precisely identical with the BR solution, with λ and $\mathscr{E}+i\check{\mathscr{B}}$ being arbitrary parameters.

This conclusion is methodically very important: It means that permitting the limiting transitions, *all* type *D* solutions of Einstein–Maxwell equations with λ where the double DP vectors are geodesic, shearless, and without complex expansion being at the same time the real eigenvectors of the electromagnetic field, *are* all *contained* in our general solution described by formulas (3.34)–(3.44). Thus, studying this type of solution it is sufficient to consider only this class of metrics—modulo their possible contractions.

We will return now to our class of solutions with the parameter switched off, described by (4.18), and we shall specialize it further, by setting:

$$\epsilon = 0, \quad \lambda = 0. \tag{4.39}$$

We obtain in this way a solution of the simple type which is described by the formulas:

$$l \to 0, \quad \epsilon \to 0, \quad \lambda \to 0, \tag{4.40}$$

$$\left.\begin{matrix} e^1 \\ e^2 \end{matrix}\right\} = \frac{1}{\sqrt{2}} \left\{ \frac{p dp}{\sqrt{\mathscr{P}}} \pm i \frac{\sqrt{\mathscr{P}}}{p} d\sigma \right\}, \tag{4.40a}$$

$$\left.\begin{matrix} e^3 \\ e^4 \end{matrix}\right\} = p \left\{ \begin{matrix} du \\ dv' \end{matrix} \right. \tag{4.40b}$$

$$\mathscr{P} = -z\bar{z} + 2n_0 p, \tag{4.40c}$$

$$C^{(3)} = (2/p^3)(n_0 - z\bar{z}/p), \tag{4.40d}$$

$$\omega = -d\chi, \quad \chi = \tfrac{1}{2} z\{udv - vdu + 2id\sigma/p\}, \tag{4.40e}$$

$$\mathscr{E} + i\check{\mathscr{B}} = -z/p^2. \tag{4.40f}$$

When the electromagnetic field is present, i.e., with $z \neq 0$, this solution amounts to Melvin's interesting cosmological model of the magnetic universe.[22]

In order to exhibit this explicitly, we notice first that with $z \neq 0$, $\mathscr{P}$ can be positive only in the presence of $n_0 \neq 0$. If one uses in the place of p the coordinate $p' := p - z\bar{z}/2n_0$, then $\mathscr{P} = 2n_0 p'$ must be positive; thus, we can now introduce the final coordinate which shall replace p through $\mathscr{P} = 2n_0 p' = (2n_0 \rho)^2$. At the same time, we shall introduce in the place of the remaining coordinates:

$$\phi := \sigma/4n_0, \quad u/\sqrt{2} = \xi - \tau, \quad v/\sqrt{2} = \xi + \tau. \tag{4.41}$$

Then, one easily sees that

$$ds^2/(4n_0)^2 = (\rho^2 + \beta^2)(d\rho^2 + d\xi^2 - d\tau^2) + [\rho/(\rho^2 + \beta^2)]^2 d\phi^2, \tag{4.42}$$

where

$$\beta^2 = z\bar{z}/(2n_0)^2. \tag{4.43}$$

We work in units with $G = 1 = c$; if ds^2 has the dimension $(\text{length})^2$ and the tetrad correspondingly is of dimension $(\text{length})^1$, then one should ascribe to the objects in (4.40) dimensions according to the scheme

$$p, \sigma, n_0, z \sim (\text{length}), \quad u, v \sim (\text{length})^0. \tag{4.44}$$

Therefore, ρ, ξ, ϕ, and τ are all dimensionless, and β^2 is a dimensionless constant. In terms of these coordinates we have:

$$C^{(3)} = \frac{1}{(2n_0)^2} \frac{\rho^2 - \beta^2}{(\rho^2 + \beta^2)^4},$$

$$\mathscr{E} + i\check{\mathscr{B}} = -\frac{z}{(2n_0)^2} \frac{1}{(\rho^2 + \beta^2)^2}. \tag{4.45}$$

$$\omega = -d\{2z[\xi d\tau - \tau d\xi + i\, d\phi/(\rho^2 + \beta^2)]\}.$$

Already at this point, we can observe a remarkable fact: The sign of n_0 has been shown to be irrelevant, it disappeared from all pertinent quantities as a result of the sequence of the transformations executed above. Is the constant β relevant? In order to answer this question, we rescale the coordinates according to

$$(\rho, \xi, \tau) = \beta(\rho', \xi', \tau'), \quad \phi = \beta^4 \phi', \tag{4.46}$$

which leads to

$$\frac{ds^2}{R^2} = (1 + \rho'^2)^2 (d\rho'^2 + d\xi'^2 - d\tau'^2) + \left(\frac{\rho'}{1 + \rho'^2}\right)^2 d\phi'^2,$$

$$C^{(3)} = \frac{4}{R^2} \frac{\rho'^2 - 1}{(\rho'^2 + 1)^4}, \quad \mathscr{E} + i\check{\mathscr{B}} = \frac{e^{i\theta}}{R} \frac{1}{(1 + \rho'^2)^2}, \tag{4.47}$$

$$\omega = Re^{i\theta} d\{\xi' d\tau' - \tau' d\xi' + i\, d\phi'/(1 + \rho'^2)\},$$

where we denoted by R

$$R := 4|n_0|\beta^3, \tag{4.48}$$

and

$$e^{i\theta} = -z/2|n_0|\beta. \tag{4.49}$$

(According to the definition of β, $|z/2n_0\beta| = 1$.)

Therefore, the solution depends essentially only on *one* real parameter R of the dimension of length and the trivial phase $e^{i\theta}$ which is the subject of duality rotations. In particular, if we choose $e^{i\theta} = i$ so that the electromagnetic field reduces to a pure magnetic field,

$$\mathscr{E} = 0, \quad \check{\mathscr{B}} = \frac{1}{R} \frac{1}{(1 + \rho'^2)^2}, \tag{4.50}$$

we obviously have $R = \check{\mathscr{B}}_{\max}^{-1}$, where $\check{\mathscr{B}}_{\max}$ is the maximal value of the magnetic field along the manifold.

The formulas (4.47), dropping primes, amount to the essence of Melvin's result in its canonical form. The remarkable scaling properties of this solution indicate that it would be of interest to investigate its conformal Killing vectors and, more generally, Killing spinors. For a further interpretation of these solutions, see Refs. 22 and 23.

The solution (4.40) is also of some interest in the vacuum case, when:

$$l \to 0, \quad \epsilon \to 0, \quad \lambda \to 0 \quad z \to 0, \quad n_0 \neq 0, \tag{4.51}$$

$$\left.\begin{matrix} e^1 \\ e^2 \end{matrix}\right\} = \frac{1}{\sqrt{2}} \left\{ \frac{p}{\sqrt{\mathscr{P}}} dp \pm i \frac{\sqrt{\mathscr{P}}}{p} d\sigma \right\}, \tag{4.51a}$$

$$\left.\begin{matrix} e^3 \\ e^4 \end{matrix}\right\} = p \left\{ \begin{matrix} du \\ dv' \end{matrix} \right. \tag{4.51b}$$

$$\mathscr{P} = 2n_0 p, \tag{4.51c}$$

$$C^{(3)} = 2n_0/p^3. \tag{4.51d}$$

Introducing the coordinates;

$$p = 2n_0 \rho^2, \quad \phi = \sigma/4n_0, \quad u/\sqrt{2} = \xi - \tau,$$

$$v/\sqrt{2} = \xi + \tau, \tag{4.52}$$

we have

$$ds^2/(4n_0)^2 = \rho^4 (d\rho^2 + d\xi^2 - d\tau^2) + \rho^{-2} d\phi,$$

$$C^{(3)} = \frac{1}{(2n_0)^2} \frac{1}{\rho^3}. \tag{4.53}$$

This metric is also insensitive to the sign of n_0, similarly to the Melvin metric, with which it is closely related. Notice that the change of variables:

$$x := \tfrac{1}{3}\rho^3, \quad y := 3^{2/3}\xi, \quad t = 3^{2/3}\tau, \quad z := 3^{-1/3}\phi, \tag{4.54}$$

brings the metric to the form

$$\frac{ds^2}{(4n_0)^2} = dx^2 + (x)^{4/3}(dy^2 - dt^2) + (x)^{-2/3}dz^2, \tag{4.55}$$

$$C^{(3)} = \frac{1}{(6n_0)^2}\cdot\frac{1}{x^2}.$$

Thus, this is a metric of the Kasner type,[26] with the distinguished variable being spacelike. Notice that it can be obtained from the well-known limit of the Schwarzschild metric for $m \to \infty$, $ds^2 = t^{4/3}(dx^2 + dy^2) + t^{-2/3}dz^2 - dt^2$ by complexification and then taking another real slice.

5. ALL DIVERGENCELESS D's AS THE LIMIT OF D's WITH NONTRIVIAL COMPLEX EXPANSION

We have demonstrated in the previous section that modulo limiting transitions, *all D* solutions of Einstein–Maxwell equations with λ (where the DP directions, geodesic, shearless, and divergenceless, are the eigenvectors of the electromagnetic field) are described in canonical form by the formulas (3.34)–(3.44).

It is natural to conjecture that these general solutions, in their turn, could be considered as a contraction of a more general family of solutions of the D type, where the geodesic and shearless DP vectors are permitted to possess nontrivial complex expansion. This section is devoted to the proof of this conjecture, whose validity was already pointed out—without details— in Carter's work.[1]

In Ref. 27 we studied a class of solutions of Einstein–Maxwell equations with λ of type D described by the collection of formulas:

$$ds^2 = \frac{p^2+q^2}{\mathscr{P}}dp^2 + \frac{\mathscr{P}}{p^2+q^2}(d\tau + q^2 d\sigma)^2 + \frac{p^2+q^2}{\mathscr{Q}}dq^2 - \frac{\mathscr{Q}}{p^2+q^2}(d\tau - p^2 d\sigma)^2, \tag{5.1a}$$

$$\omega := \tfrac{1}{2}(f_{\mu\nu} + \check{f}_{\mu\nu})dx^\mu \wedge dx^\nu = -d\left\{\frac{e_0 + ig_0}{q+ip}(d\tau - ipq\,d\sigma)\right\}, \tag{5.1b}$$

$$\mathscr{P} := b_0 - g_0^2 + 2n_0 p - \epsilon_0 p^2 - \frac{\lambda}{3}p^4, \tag{5.1c}$$

$$\mathscr{Q} := b_0 + e_0^2 - 2m_0 q + \epsilon_0 q^2 - \frac{\lambda}{3}q^4,$$

$$\left.\begin{matrix}e^1\\e^2\end{matrix}\right\} = \frac{1}{\sqrt{2}}\left\{\left(\frac{p^2+q^2}{\mathscr{P}}\right)^{1/2}dp \pm i\left(\frac{\mathscr{P}}{p^2+q^2}\right)^{1/2}(d\tau + q^2 d\sigma)\right\}, \tag{5.1d}$$

$$\left.\begin{matrix}e^3\\e^4\end{matrix}\right\} = \frac{1}{\sqrt{2}}\left\{\left(\frac{p^2+q^2}{\mathscr{Q}}\right)^{1/2}dq \pm \left(\frac{\mathscr{Q}}{p^2+q^2}\right)^{1/2}(d\tau - p^2 d\sigma)\right\},$$

$$\mathscr{F} = -\frac{1}{2}\frac{(e_0+ig_0)^2}{(q+ip)^4}, \tag{5.1e}$$

$$C^{(3)} = \frac{-2}{(q+ip)^2}\left\{\frac{m_0 + in_0}{q+ip} - \frac{e_0^2+g_0^2}{q^2+p^2}\right\}. \tag{5.1f}$$

The $\{x^\mu\} = \{pq\tau\sigma\}$ are real coordinates here, b_0, e_0, g_0, ϵ_0, m_0, n_0, and λ are constants. The interpretation of these constants has been established as follows: e_0 and g_0 have the interpretation of the electric and magnetic monopole charges, respectively, m_0 and n_0 correspond to the mass and NUT parameter (i.e., gravitational "electric" and "magnetic" charges), λ is a cosmological constant; ϵ_0 by rescaling of coordinates can be restricted to the values $\epsilon_0 = 1, 0, -1$; for $\epsilon_0 = 1$, the field complex has a rest frame, for $\epsilon_0 = 0$ in a sense it moves with the speed of light, for $\epsilon_0 = -1$ the field system permits a "tachyonic" interpretation (compare Ref. 28). The constant b_0 is related to the Kerr rotation parameter. With respect to the null tetrad selected as in (5.1d), the conformal curvature is described entirely by $C^{(3)}$ from (5.1f) and e^3 and e^4 are proportional to the pair of the double DP vectors:

$$\pm K_\mu^{(\pm)}dx^\mu := d\left(\tau \pm \int\frac{q^2dq}{\mathscr{Q}}\right) - p^2 d\left(\sigma \mp \int\frac{dq}{\mathscr{Q}}\right), \tag{5.2}$$

which are endowed with the nontrivial complex expansion (common in value)

$$Z := 1/(q+ip). \tag{5.3}$$

The vectors e^3, e^4 are here moreover the real eigenvectors of the electromagnetic field, if it is nontrivial.

This family of solutions which was first encountered by Carter[1] from the point of view of the theory of the separation of variables, is by itself a contraction of the more general family of D solutions with similar structural properties (DP vectors aligned along the eigenvectors of the electromagnetic field), which has one parameter more, this parameter being the Levi–Civita acceleration parameter, which is present in C metrics; see Refs. 9 and 10, for the definite publication see Ref. 11; compare also Ref. 15.

We claim now that *all* the metrics without the complex expansion considered in this paper, can be obtained by the proper contraction of the formulas (5.1).

Indeed, already in Ref. 27, a contraction of the formulas (5.1) was considered which was leading to the solution which we called "anti-NUT" type. (See Ref. 27, Sec. 10 and 11).

This contraction consists of the following: We replace the coordinates in (5.1) by the new coordinates p', q', σ', τ' defined by:

$$p = p', \quad q = q_0 + \varepsilon q', \quad \sigma = \varepsilon^{-1}\sigma', \quad \tau = \tau' - q_0^2\varepsilon^{-1}\sigma', \tag{5.4}$$

where q_0 is a constant, and ε is the contraction parameter. Then, we arrange the disposable constants of the solution so that

$$\epsilon_0 = \xi_0 + 2\lambda q_0^2, \quad m_0 = -\eta_0\varepsilon + \xi_0 q_0 + \frac{4\lambda}{3}q_0^3, \tag{5.5}$$

$b_0 = -2\eta_0 q_0\varepsilon + \zeta_0\varepsilon^2 - e_0^2 + \xi_0 q_0^2 + \lambda q_0^4$ (e_0, g_0, n_0, λ are independent of $\mathscr{E}$), and the constants ζ_0, η_0, and ξ_0 are independent of the contraction parameter ε.

Executing now in (5.1) the limiting transition

$$\varepsilon \to +0, \tag{5.6}$$

there was obtained in Ref. 27, Sec. 10, the finite result which we repeat here as the collection of formulas:

$$\left.\begin{matrix} e^1 \\ e^2 \end{matrix}\right\} = \frac{1}{\sqrt{2}}\left(\frac{dp'}{\sqrt{\phi}} \pm i\sqrt{\phi}\,[d\tau' + 2q_0 q' d\sigma']\right), \tag{5.7a}$$

$$\left.\begin{matrix} e^3 \\ e^4 \end{matrix}\right\} = \frac{1}{\sqrt{2}}\sqrt{q_0^2 + q'^2}(dq'/\sqrt{\jmath} \mp \sqrt{\jmath}\, d\sigma')$$
(5.7b)

$$\jmath := \zeta_0 + 2\eta_0 q' + \xi_0 q'^2,$$

$$\phi := -\xi_0 - \frac{\lambda}{3}(p'^2 + 5q_0^2) + 2\,\mathrm{Re}\left(\frac{m_0' + in_0}{q_0 + ip'}\right) - \left|\frac{e_0 + ig_0}{q_0 + ip'}\right|^2,$$
(5.7c)

$$m_0' = \xi_0 q_0 + \frac{4\lambda}{3} q_0^3,$$

$$ds^2 = \phi^{-1} dp'^2 + \phi\,(d\tau' + 2q_0 q' d\sigma')^2 + (q_0^2 + p'^2) \times (dq'^2/\jmath - \jmath\, d\sigma'^2), \tag{5.7d}$$

$$\omega = -d\left\{(e_0 + ig_0)\left(\frac{d\tau'}{q_0 + ip'} + \frac{q_0 - ip'}{q_0 + ip'} q' d\sigma'\right)\right\}, \tag{5.7e}$$

$$C^{(3)} = \frac{-2}{(q_0 + ip')^2}\left\{\frac{m_0' + in_0}{q_0 + ip'} - \frac{e_0^2 + g_0^2}{q_0^2 + p'^2}\right\}, \tag{5.7f}$$

$$\mathscr{F} = -\frac{1}{2}\frac{(e_0 + ig_0)^2}{(q_0 + ip')^2}. \tag{5.7g}$$

The DP vectors of this solution are proportional to

$$\pm K_\mu'^{(\pm)} dx^\mu = d\left(\sigma' \mp \int dq'/\jmath\right). \tag{5.8}$$

In Ref. 27, Sec. 11, some "trigonometrical" and "hyperbolical" parametrizations of the variables of this solution were considered.

We claim now that the collection of formulas (5.7)—modulo a choice of coordinates and notation—is entirely equivalent to our general solution for the divergenceless D's described in Sec. 3, by (3.34)–(3.44).

In order to prove this, we first consider a two-dimensional space of signature (+ −) which is present in the structure of (5.7)

$$dl^2 := \frac{dq'^2}{\jmath} - \jmath\, d\sigma'^2, \quad \jmath := \zeta_0 + 2\eta_0 q' + \xi_0 q'^2. \tag{5.9}$$

This is clearly a space of contant curvature. Indeed, defining

$$E^1 = \frac{1}{\sqrt{2}}\left(\frac{dq'}{\sqrt{\jmath}} - \sqrt{\jmath}\, d\sigma'\right),$$

$$E^2 = \frac{1}{\sqrt{2}}\left(\frac{dq'}{\sqrt{\jmath}} + \sqrt{\jmath}\, d\sigma'\right),$$

so that $dl^2 = 2E^1 \otimes E^2$, from the first structure equations $dE^a = E^b \wedge \Gamma^a{}_b$ we easily find that the connection form Γ_{12} is $\Gamma_{12} = \frac{1}{2}\jmath_{q'}\, d\sigma'$ so that $d\Gamma_{12} = +\frac{1}{2}\jmath_{q'q'} E^1 \wedge E^2 = \xi_0 E^1 \wedge E^2$.

We can now describe the same space in the new coordinates u, v, according to

$$dl^2 = (2du \otimes dv)/(1 - \epsilon uv)^2, \quad \epsilon \text{ to be determined.} \tag{5.10}$$

We compute the curvature of this line element by defining $E^1 = du/(1 - \epsilon uv)$, $E^2 = dv/(1 - \epsilon uv)$; from the first structure equations we now find that

$$\Gamma_{12} = \epsilon\frac{udv - vdu}{1 - \epsilon uv} \quad \text{and consequently} \quad d\Gamma_{12} = 2\epsilon E^1 \wedge E^2.$$

Because the curvatures of the two spaces coincide, we now have

$$\xi_0 = 2\epsilon. \tag{5.11}$$

The whole point of the present argument is that although we want to pass from the coordinates q' and σ' to the coordinates u, v for the discussed 2-space, we would like to avoid the irrelevant explicit determination of the coordinate transformation which, with ζ_0, η_0, ξ_0 arbitrary, has plenty of bothersome subcases. For this purpose we can use the following argument: If the condition (5.11) applies, then we know that the coordinate transformation such that

$$dl^2 = \left(\frac{dq'}{\sqrt{\jmath}} - \sqrt{\jmath}\, d\sigma'\right) \otimes \left(\frac{dq'}{\sqrt{\jmath}} + \sqrt{\jmath}\, d\sigma'\right) = 2\frac{du \otimes dv}{(1 - \epsilon uv)^2} \tag{5.12}$$

certainly exists, and the existance of a real function μ follows, such that:

$$\frac{1}{\sqrt{2}}\left(\frac{dq'}{\sqrt{\jmath}} - \sqrt{\jmath}\, d\sigma'\right) = \frac{e^{\mu} du}{1 - \epsilon uv},$$

$$\frac{1}{\sqrt{2}}\left(\frac{dq'}{\sqrt{\jmath}} + \sqrt{\jmath}\, d'\sigma\right) = \frac{e^{-\mu} dv}{1 - \epsilon uv}. \tag{5.13}$$

But then by taking the wedge product of both sides:

$$dq' \wedge d\sigma' = \frac{du \wedge dv}{(1 - \epsilon uv)^2}. \tag{5.14}$$

This however means that

$$d\left\{q' d\sigma' - \frac{1}{2}\frac{udv - vdu}{1 - \epsilon uv}\right\} = 0 \tag{5.15}$$

and there exists a (real) χ such that

$$q' d\sigma' = \frac{1}{2}\frac{udv - vdu}{1 - \epsilon uv} + d\chi. \tag{5.16}$$

But this is exactly that which is needed in order to execute the coordinate transformation in the formulas (5.7) consisting in replacing q' and σ' by u and v. Indeed, in the metric and in e^1, e^2 these coordinates appear only in the form of dl^2 known in terms of u, v as (5.12) and in the combination $d\tau' + 2q_0 q' d\sigma'$, which, if we introduce the new coordinate which replaces τ':

$$\tau' =: \rho - 2q_0 \chi \tag{5.17}$$

amounts according to (5.16) to

$$d\tau' + 2q_0q'd\sigma' = d\rho + q_0\frac{udv - vdu}{1-\epsilon uv}. \tag{5.18}$$

Then, q' and σ' still enter into (5.7) in the potential form for ω. But using (5.16) and (5.17) we easily find that

$$\frac{d\tau'}{q_0+ip'} + \frac{q_0-ip'}{q_0+ip'}q'd\sigma' = -\frac{1}{2}\frac{udv-vdu}{1-\epsilon uv} + \frac{1}{q_0+ip'}\left(d\rho + q_0 \times \frac{udv-vdu}{1-\epsilon uv}\right) - d\chi. \tag{5.19}$$

Therefore, in the new coordinates, the form ω amounts to

$$\omega = (e_0+ig_0)d\left\{\frac{1}{2}\frac{udv-vdu}{1-\epsilon uv} - \frac{1}{q_0+ip'} \times\left(d\rho + g_0\frac{udv-vdu}{1-\epsilon uv}\right)\right\}. \tag{5.20}$$

In the next step, we will discuss the structural function ϕ from (5.7c). In doing so, we shall replace ξ_0 by the presently more convenient 2ϵ; then defining $\mathscr{P}:=(q_0^2+p'^2)\phi$, we have

$$\begin{aligned}\mathscr{P} &= -2\epsilon(p'^2+q_0^2) - \frac{\lambda}{3}(p'^2+q_0^2)(p'^2+5q_0^2) - (e_0^2+g_0^2) \\ &\quad + (m_0'+in_0)(q_0-ip') + (m_0'-in_0)(q_0+ip') \\ &= -(e_0^2+g_0^2) + 2m_0'q_0 + 2n_0p' - 2\epsilon(p'^2+q_0^2) \\ &\quad -\lambda\left(\tfrac{1}{3}p'^4 + 2q_0^2p'^2 + \tfrac{5}{3}q_0^4\right).\end{aligned} \tag{5.21}$$

By using here the value $m_0' = 2\epsilon q_0 + \frac{4}{3}\lambda q_0^3$ and reordering we reduce this to

$$\mathscr{P} = -(e_0^2+g_0^2) + 2n_0p' - 2\epsilon(p'^2-q_0^2) - \lambda\left(\tfrac{1}{3}p'^4 + 2q_0^2p'^2 - q_0^4\right). \tag{5.22}$$

With this specific $\mathscr{P}$ our contracted metric thus has the form

$$ds^2 = \frac{p'^2+q_0^2}{\mathscr{P}}dp'^2 + \frac{\mathscr{P}}{p'^2+q_0^2}\left(d\rho + q_0\frac{udv-vdu}{1-\epsilon uv}\right)^2 + \frac{2(q_0^2+p'^2)}{(1-\epsilon uv)^2}du\,dv. \tag{5.23}$$

Now compare this result with (3.34) and (3.35). We recognize then that these formulas are identically the same if we identify the present coordinate p' with p, present ρ with σ, and the constants according to the scheme:

$$|z| = \sqrt{e_0^2+g_0^2}, \quad l = -q_0, \tag{5.24}$$

the remaining constants and coordinates being already denoted by the same symbols. We can also compare the ω from (3.38) with the present ω from (5.20). They also precisely agree if we identify

$$z = e_0 + ig_0. \tag{5.25}$$

With the metric and the electromagnetic field being identical in the same coordinates, all derived secondary quantities for both solutions must agree. E.g., the expressions for $C^{(3)}$ and $\mathscr{F}$ from (5.7), remembering our identifications of coordinates and constants, amount to:

$$\mathscr{F} = -\frac{1}{2}\frac{(e_0+ig_0)^2}{(q_0+ip')^4} = -\frac{1}{2}\frac{z^2}{(p+il)^4},$$
$$\begin{aligned}C^{(3)} &= \frac{-2}{(q_0+ip')^2}\left\{\frac{m_0'+in_0}{(q_0+ip')} - \frac{e_0^2+g_0^2}{q_0^2+p'^2}\right\} \\ &= \frac{2}{(p+il)^3}\left\{n_0 - im_0' - \frac{z\bar{z}}{p-il}\right\},\end{aligned} \tag{5.26}$$

[With $m_0' = 2\epsilon q_0 + \frac{4}{3}\lambda q_0^3 = -(2\epsilon l + \frac{4}{3}\lambda l^3)$ this agrees as it should with (3.40).]

[A minor technical point remains: Does the tetrad of the limiting solution (5.7) automatically agree with the tetrad which we have choosen describing—as is now proven—the same solution by (3.34)–(3.44)? The tetrad members e^1, e^2 surely coincide in both descriptions of the solution; but in (5.13) we have left the boosting factor undertermined. In order to answer this question, we must recall the criteria according to which the tetrad gauge has been fixed for both cases. For the divergenceless metrics the boost gauge was fixed by demanding that the coefficient A in $e^3 = Adu$, $e^4 = Adv$ must be the same, while the phase gauge was fixed by the condition that $\Gamma_{423} = \Gamma_{314}$. For the metrics (5.1) we have fixed both gauges in Ref. 27 by the conditions that $\Gamma_{421} = \Gamma_{312}$ and $\Gamma_{423} = \Gamma_{314}$. Thus, for the limiting metric the same criteria for the phase gauge lead to the agreement of e^1, e^2 in the two cases, but in general the e^3, e^4 of both cases *can differ* by some boost gauge. It is clear, however, that the gauge for e^3, e^4 from (3.36) is the most advantageous for the discussed class of solutions.]

The result which we have established: that *all* D's with λ but without complex expansion (with geodesic and shearless DP vectors being eigenvectors of the electromagnetic field) are contained in the sense of the limiting transition from (5.1) by means of (5.4), (5.5), (5.6) in the more general class of solutions with complex expansion [i.e., (5.1), coinciding in fact precisely with the anti-NUT degeneration of these solutions studied in Ref. 27, equivalent to Carter's $[B(+)]$ solutions][1], is of serious methodological importance. The divergenceless solutions are thus just *a special case of* (5.1), which in its turn is just a special case (contraction) of the most general D solutions known, the seven-parametric family,[9,10,11] after switching off the acceleration parameter. The recent work of Weir and Kerr[15] has established that these solutions in the vacuum case ($e_0 + ig_0 = 0 = \lambda$) exhaust *all* diverging D's. At the present time, a work is in progress (by Kerr, Alarcón, and the present author) concerned with providing the extension of the proof in Ref. 15 to the more general case: The conjecture is that all D's, diverging, with and the double DP vectors being eigenvectors of the electromagnetic field, and in the presence of λ are exhausted by the seven-parametric family.

Not denying the physical interest which some of the divergenceless D's may possess, one can thus say that they lie low in the hierachy of degenerations of the known D type solutions.

It is of particular interest to notice that because the physical interpretation of the parameters of the solution (5.1) is fairly well established, we can easily trace their degeneration into the free parameters of their genetic descendants, divergenceless D's—establishing in this manner the interpretation of this class of solutions.

Indeed, simply executing the limit $\varepsilon \to 0$ in (5.5) we obtain the set of relationships:

$$\lambda = \lambda, \tag{5.27a}$$

$$e_0 + ig_0 = z, \tag{5.27b}$$

$$n_0 = n_0, \tag{5.27c}$$

$$\epsilon_0 = 2(\epsilon + l^2\lambda), \tag{5.27d}$$

$$m_0 = -2l(\epsilon + \tfrac{2}{3}\lambda l^2), \tag{5.27e}$$

$$b_0 = -e_0^2 + l^2(2\epsilon + \lambda l^2), \tag{5.27f}$$

where the limiting values on the right side are the parameters of the divergenceless solutions. The relations (b) and (c) identify the interpretation of z as the complex charge and n_0 as the NUT parameter. In order to better understand the remaining relations, we should consider first the simpler case of λ absent, $\lambda = 0$, where b_0 is interpreted by (compare Ref. 27, Sec. 12)

$$b_0 = g_0^2 - n_0^2 + a_0^2, \tag{5.28}$$

where a_0 is the Kerr constant if $\epsilon_0 = 1$. In this case, Eq. (5.27) thus amount to:

$$1 = 2\epsilon, \quad m_0 = -2\epsilon l, \quad g_0^2 - n_0^2 + a_0^2 = -e_0^2 + 2\epsilon l^2. \tag{5.29}$$

This means, of course, that we confine ourselves here to the case where

$$l = -m_0, \quad \epsilon = \tfrac{1}{2} \tag{5.30}$$

and that the basic constants of (5.1) are submitted to the constraint condition

$$a_0^2 + e_0^2 + g_0^2 = m_0^2 + n_0^2. \tag{5.31}$$

In terms of the parameters of the divergenceless D's this condition means that

$$a_0^2 := l^2 + n_0^2 - z\bar{z} \geqslant 0. \tag{5.32}$$

This condition is not so strange as it looks: Indeed, with the present values of the constants $\mathscr{P}$ in (3.34) has the form

$$\begin{aligned}\mathscr{P} &= l^2 + n_0^2 - z\bar{z} - (p - n_0)^2 \\ &= m_0^2 + n_0^2 - e_0^2 - g_0^2 - (p - n_0)^2 \\ &= a_0^2 - (p - n_0)^2\end{aligned} \tag{5.33}$$

and $a_0^2 > 0$ in order to guarantee a nontrivial range for p.

Therefore, when $\lambda = 0$ and $\epsilon_0 = 1$, the Kerr-NUT charged solutions which can contract to divergenceless D's are only those where $a_0^2 + e_0^2 + g_0^2 = m_0^2 + n_0^2$ and consequently in (5.1),

$$\mathscr{P} = a_0^2 - (p - n_0)^2, \quad \mathscr{Q} = (q - m_0)^2. \tag{5.34}$$

[It may be observed that with λ present and the basic constants of (5.1) given in the form of (5.27), the polynomial $\mathscr{Q}$ also possesses a double root.]

The moral of these considerations is that at least with λ switched off the interpretation of the parameters of the divergenceless D's is such that when $\epsilon > 0$, which can be put equal to $\frac{1}{2}$ by rescaling the coordinates, then $l = -m_0$ corresponds to the mass parameter and n_0 is just the NUT parameter, while $z = e_0 + ig_0$ is to be interpreted in terms of the monopole electric and magnetic charges. Some residual rotation and nontrivial Kerr parameter according to (5.32) seems to be characteristic for the divergenceless D's. According to (5.31), the Kerr parameter present in the solutions discussed corresponds to its generalized critical value: when $e_0 + ig_0 = 0 = n_0$, (5.31) amounts to just $a_0^2 = m_0^2$.

6. THE SYMMETRIES OF DIVERGENCELESS *D*'s

According to the results of the previous sections, the metric and the electromagnetic field of *all* D's with λ, characterized by the vanishing of the complex expansion and the alignment of the real eigenvectors of the electromagnetic field along the double DP vectors, can be represented—modulo possible contractions to BR solutions—in the form

$$ds^2 = \frac{p^2 + l^2}{\mathscr{P}} dp^2 + \frac{\mathscr{P}}{p^2 + l^2}\left(d\sigma + l\,\frac{v\,du - u\,dv}{1 - \epsilon uv}\right)^2 + 2\,\frac{p^2 + l^2}{(1 - \epsilon uv)^2}\,du\,dv \tag{6.1}$$

and

$$\omega = -d\chi,$$
$$\chi = z\left\{\frac{1}{2}\,\frac{u\,dv - v\,du}{1 - \epsilon uv} + \frac{i}{p + il}\left(d\sigma + l\,\frac{v\,du - u\,dv}{1 - \epsilon uv}\right)\right\} \tag{6.2}$$

where $\{p,\sigma,u,v\}$ are real coordinates and

$$\mathscr{P} := -z\bar{z} + 2n_0 p - 2\epsilon(p^2 - l^2) - \lambda(\tfrac{1}{3}p^4 + 2l^2p^2 - l^4) \tag{6.3}$$

is a polynomial which depends on the constants

$$\epsilon,\ l,\ n_0,\ z,\ \lambda \tag{6.4}$$

with the interpretation established to some extent by the discussion at the end of the last section.

Our present objective will consist in investigating the symmetries of these solutions.

First of all, we notice that if the constants of the solutions are restricted by the conditions:

$$n_0 = 0, \quad \epsilon + \tfrac{2}{3}\lambda l^2 = 0, \quad z = 0 \to \mathscr{P} = -(\lambda/3)(p^2 + l^2)^2, \tag{6.5}$$

then according to (3.40) the conformal curvature vanishes, while the Einstein tensor is given by $G_{\mu\nu} = \lambda g_{\mu\nu}$. Consequently, this solution valid for $\lambda < 0$ (because of $\mathscr{P} > 0$) reduces to the de Sitter space–time with its symmetry group O(4,1). Thus denoting $l_0^{-2} := -\lambda/3$, we can assert that the metric

$$ds^2 = l_0^2 \frac{dp^2}{p^2 + l^2} + l_0^{-2}(p^2 + l^2)\left(d\sigma + l\,\frac{v\,du - u\,dv}{1 - \epsilon uv}\right)^2 + 2\,\frac{p^2 + l^2}{[1 - 2(l/l_0)^2 uv]^2}\,du\,dv \tag{6.6}$$

is, for every l, the de Sitter metric in the disguise of unfamiliar coordinates. One could, of course, work out the 10 Killing vectors which accompany this metric and close in the Lie algebra of O(4,1). But it would be rather pointless to do so, because the simple geometry of the de Sitter spaces is perfectly well understood and its special coordinatization (6.6) seems to present little interest.

Consider now our solution for arbitrary values of constants. We can then observe that the coordinates u, v, and σ enter into the structure of the metric and the electromagnetic field only in the two essential combinations:

$$dl^2 := \frac{2du\,dv}{(1-\epsilon uv)^2}\epsilon A^1 \otimes A^1, \qquad \beta := d\sigma + l\frac{v\,du - u\,dv}{1-\epsilon uv}\epsilon A^1. \tag{6.7}$$

As far as the metric is concerned, this statement is self-evident. As far as the electromagnetic field is concerned, we notice that when $l \neq 0$, then by gauging χ properly we can represent it simply as

$$\omega = -d\chi, \quad \chi = \frac{-z}{2l}\frac{p-il}{p+il}\cdot\beta. \tag{6.8}$$

We will proceed now for a while with $l\neq 0$. The space dl^2 of (6.7) is a space of constant curvature with a three-parametric Lie group of symmetries. One easily finds that all these symmetries—plus the translational symmetry of the variable σ—are also the symmetries of both objects of (6.7), dl^2, and the 1-form β. Indeed, it presents no difficulty to prove that the expressions for dl^2 and β remain invariant under the transformation

$$u = \frac{Au'+B}{\epsilon Cu'+D}, \quad v = \frac{Dv'+C}{\epsilon Bv'+A}, \qquad \sigma = \sigma' - \frac{l}{\epsilon}\ln\left(\frac{AD+\epsilon ACu'}{AD+\epsilon BDv'}\right) + E, \tag{6.9}$$

where A, B, C, D, and E are arbitrary constants restricted by the single condition:

$$AD - \epsilon BC = 1. \tag{6.10}$$

Notice that these formulas remain valid also in the limiting case of $\epsilon \to 0$.

Being symmetries of dl^2 and β, the transformations (6.9) are also symmetries of ds^2 and ω. Knowing these finite transformations, by passing to the infinitesimal transformations in the neighborhood of unity, one easily finds the corresponding Killing vectors. Indeed, by setting:

$$A = 1+\delta A, \quad D = 1-\delta A, \qquad B = \delta B, \quad C = \delta C, \quad E = \delta E, \tag{6.11}$$

and neglecting the second order quantities in $\delta(\cdots)$, we fulfill (6.10) and we obtain an infinitesimal transformation:

$$u' = u - \delta B - 2\delta A u + \delta C\cdot\epsilon u^2, \qquad v' = v - \delta C + 2\delta A v + \delta B\cdot\epsilon v^2, \qquad \sigma' = \sigma - \delta B\cdot lv + \delta C\cdot lu - \delta E. \tag{6.12}$$

From these formulas we can now read off the Killing vectors which correspond to the independent $\delta A,\ldots,\delta E$. These vectors are:

$$\delta A: -2(u\partial_u - v\partial_v), \quad \delta B: -(\partial_u - \epsilon v^2\partial_v + lv\partial_\sigma), \qquad \delta C: -(\partial_v - \epsilon u^2\partial_u - lu\partial_\sigma), \quad \delta E: -\partial_\sigma. \tag{6.13}$$

When $l\neq 0$ and $\epsilon\neq 0$, the most advantageous linear combinations of these vectors from the point of view of their Lie algebra are:

$$\Gamma_0 := \partial_\sigma, \qquad 2\Gamma_1 := (\partial_u - \epsilon v^2\partial_v + lv\partial_\sigma) + (\partial_v - \epsilon u^2\partial_u - lu\partial_\sigma), \qquad 2\Gamma_2 := (\partial_u - \epsilon v^2\partial_v + lv\partial_\sigma) - (\partial_v - \epsilon u^2\partial_u - lu\partial_\sigma), \qquad \Gamma_3 := \epsilon(u\partial_u - v\partial_v) + l\partial_\sigma. \tag{6.14}$$

Indeed, one easily finds that the commutation rules among these four generators of symmetries are

$$[\Gamma_0,\Gamma_i] = 0, \quad i = 1,2,3 \tag{6.15}$$

and

$$[\Gamma_1,\Gamma_2] = \Gamma_3, \quad [\Gamma_2,\Gamma_3] = \epsilon\Gamma_1, \quad [\Gamma_3,\Gamma_1] = -\epsilon\Gamma_2. \tag{6.16}$$

But if $\epsilon\neq 0$, by rescaling of the constants and coordinates of the solution studied it can be set—without loss of generality—equal to ± 1. Therefore, for $\epsilon = 1$, we recognize in (6.15) and (6.16) the commutation rules of the group $\mathbb{R}\times O(2,1)$, and when $\epsilon = -1$, of the group $\mathbb{R}\times O(1,2)$, respectively. One should observe, however, that with $l\neq 0$, according to (6.14), the orbits of the corresponding O groups are three dimensional. Therefore, the situation is parallel here to the well-known case of the Taub-NUT solution, where in the presence of the Lie algebra $T\times O(3,\mathbb{R})$, the O group cannot be interpreted as spherical symmetry in the usual sense.[29]

In the case of $\epsilon = 0$ but $l\neq 0$, the generators

$$\gamma_0 := \partial_\sigma, \quad 2\gamma_1 := (\partial_u + \partial_v) - l(u-v)\partial_\sigma, \qquad 2\gamma_2 := (\partial_u - \partial_v) + l(u+v)\partial_\sigma, \quad \gamma_3 := u\partial_u - v\partial_v, \tag{6.17}$$

have the Lie algebra

$$[\gamma_0,\gamma_i] = 0, \quad [\gamma_1,\gamma_2] = l\gamma_0, \qquad [\gamma_2,\gamma_3] = \gamma_1, \quad [\gamma_3,\gamma_1] = -\gamma_2. \tag{6.18}$$

This is the Lie algebra of the repulsive oscillator, according to the terminology of nuclear physics.

Consider now the case of $l = 0$; the situation is here much simpler: According to (6.1) and (6.2) the metric

$$ds^2 = \frac{p^2}{\mathscr{P}}dp^2 + \frac{\mathscr{P}}{p^2}d\sigma^2 + 2\frac{p^2}{(1-\epsilon uv)^2}du\,dv \tag{6.19}$$

and the electromagnetic field

$$\omega = -d\chi, \quad \chi := z\left\{\frac{1}{2}\frac{u\,dv - v\,du}{1-\epsilon uv} + \frac{i}{p}d\sigma\right\}, \tag{6.20}$$

endowed with the polynomial $\mathscr{P} = -z\bar{z} + 2n_0 p - 2\epsilon p^2 - (\lambda/3)p^4$, have the obvious group of symmetries consisting of translations of σ independent of the symmetries of $(1-\epsilon uv)^{-2}du\,dv$. For the Killing vectors we can choose here the objects (6.14) with $l = 0$, which then generate two-dimensional orbits and still fulfill the algebra (6.15) and (6.16), when $\epsilon\neq 0$. When $l = 0 = \epsilon$ the situation is the simplest: The Killing vectors

$$\partial_\sigma, \ \partial_u, \ \partial_v, \ u\partial_u - v\partial_v \tag{6.21}$$

constitute the generators of the σ translations and the 3-symmetries of the pseudo-Euclidean plane $du\wedge dv$.

The important conclusion of these considerations consists in the fact that the divergenceless D's studied always possess at least a four-dimensional group of symmetries, being endowed with four Killing vectors.

ACKNOWLEDGMENTS

I would like to express my best thanks to Dr. R.P. Kerr, Dr. M.A. Melvin, Dr. A. Krasiński, Dr. J. Stachel, and Mr. S. Alarcón Guitiérrez for useful discussions.

[1]B. Carter, Commun. Math. Phys. **10**, 280 (1968).
[2]R.P. Kerr, Phys. Rev. Lett. **11**, 237 (1963).
[3]E.T. Newman and M. Demiański, Bull. Acad. Pol. Sci., Ser. Math. Phys. **14**, 653 (1966).
[4]W. Kinnersley, thesis, Caltec (1969), submitted 1968, unpublished.
[5]W. Kinnersley, J. Math. Phys. **10**, 1195 (1969).
[6]L.P. Hughston, R. Penrose, P. Sommers, and M. Walker, Commun. Math. Phys. **27**, 303 (1972).
[7]His form for that branch, unlike Carter's, does not contain spurious parameters. However, the result as stated in (V.17) of Ref. 4 contains two numerical errors in the coefficients of $g_{\mu\nu}$. The original Carter result is correct.
[8]R. Debever, Bull. Soc. Math. de Belg. **23**, 360 (1971).
[9]J.F. Plebański, In "Gravitational Radiation and Gravitational Collapse," C. DeWitt-Morette, Ed., 188 (1974) [Copernican Conference, Warsaw, Sept. 1973].
[10]J.F. Plebański, Ann. N.Y. Acad. Sci. **262**, 246 (1975).
[11]J.F. Plebański, and M. Demiański, Ann. Phys. **98**, 98–127 (1976) and Caltec preprint 0AD-401 (April 1975).
[12]G. Debney, R.P. Kerr, and A. Schild, J. Math. Phys. **10**, 1195 (1969).
[13]W. Kinnersley, in *General Relativity and Gravitation*, edited by G. Shaviv and J. Rosen (Wiley, New York, 1975), p. 129.
[14]W. Kinnersley and M. Walker, Phys. Rev. D **2**, 1359 (1970).
[15]G. Wier and R.P. Kerr, Proc. R. Soc. London, Ser. A **355**, 31–52 (1977) (presented by R. Penrose, F.R.S.).
[16]W. Kundt, A. Phys. **163**, 77 (1961).
[17]B. Bertotti, Phys. Rev. **116**, 1331 (1959).
[18]I. Robinson, Bull. Acad. Pol. Sci., Ser. Math. Phys. **7**, 352 (1959).
[19]J. Plebański and S. Hacyan, to be published in J. Math. Phys. (1979).
[20]For $\lambda = 0$ this is equivalent, of course, to the Kinnersley's case (V.17) (see Ref. 4), modulo numerical errors in his metric coefficients.
[21]It is of interest to notice that the solutions which correspond to the contraction at this exceptional value have a four-dimensional Lie group, similarly to the NUT branch without the acceleration and with the Kerr rotation parameter switched off, $a = 0$.
[22]M.A. Melvin, Phys. Lett. **8**, 65 (1964).
[23]M.A. Melvin, Phys. Rev. **139**, B225 (1965).
[24]J.F. Plebański, "Spinors, Tetrads and Forms," a monograph of Centro de Investigación y de Estudios Avanzados del IPN, 297 pages (1974), unpublished.
[25]H. Hariai, Sci. Rep. Tohoku Univ. **34**, 160 (1950); **35**, 62 (1951).
[26]E. Kasner, Am. J. Math. **43**, 217 (1921).
[27]J.F. Plebański, Ann. Phys. **10**, 280 (1975).
[28]A. Pérez, Phys. Lett. A **31**, 361 (1970).
[29]C.W. Misner and A.H. Taub, Sov. Phys. JETP **28**, 122 (1969).

Bures distance and relative entropy

Stanley Gudder
Department of Mathematics, University of Denver
Jean-Paul Marchand
Departments of Mathematics and Physics, University of Denver
Walter Wyss
Department of Physics, University of Colorado, Boulder

(Received 2 October 1978)

We have previously constructed an entropy functional which characterizes statistical inference from partial measurement by maximum relative entropy. Here we discuss the mathematical properties of this concept in greater detail and establish its relation to the Bures distance and the Uhlmann transition probability.

1. INTRODUCTION

We have previously defined statistical inference relative to an *a priori* state v and a measured subalgebra $\mathscr{B}$ of observables as a state w_0 whose generalized Radon–Nikodym derivative with respect to v is in $\mathscr{B}$.[1] Subsequently we have investigated the relative entropy $H_v(w)$ which characterizes w_0 as a maximum entropy state compatible with the measurement of $\mathscr{B}$.[2] In this note we explore the mathematical properties of H_v in greater detail.

Section 2 contains the definitions of the statistical inference w_0 and the corresponding entropy function H_v. In Sec. 3 we establish the connection between H_v and Bures' ρ function and identify w_0 as the state which minimizes the Bures distance and maximizes the Uhlmann transition probability relative to v. In Sec. 4 it is shown that H_v is a concave function of the states (Theorem 4.1) that does not decrease under restriction (Theorem 4.2) or decoupling (Theorem 4.3) of the states. The concept of conditioning relative to $\mathscr{B}$ provides an alternative characerization of inference.[3] In Sec. 5 we show that H_v does not decrease under conditioning. Section 6 contains some remarks.

2. RELATIVE ENTROPY AND STATISTICAL INFERENCE

Let $\mathfrak{A}$ be a von Neumann algebra of operators on a separable Hilbert space H, $\mathfrak{A}_+$ its positive cone, and $N(\mathfrak{A})$ the set of normal states on $\mathfrak{A}$. Let $v,w \in N(\mathfrak{A})$. If w is majorized by λv, $\lambda \geqslant 1$, i.e., if

$$w(A^\dagger A) \leqslant \lambda v(A^\dagger A), \quad \forall A \in \mathfrak{A},$$

a theorem by Sakai[4] implies that there exists a unique $T \in \mathfrak{A}_+$, $0 < T \leqslant \lambda^{1/2} I$, such that

$$w(\cdot) = v(T \cdot T). \tag{2.1}$$

In Ref. 2 the concept of relative entropy has been defined as follows:

Definition 2.1: Let $v,w \in N(\mathfrak{A})$, $w \leqslant \lambda v$, $T \in \mathfrak{A}_+$ such that (2.1) holds, then the v entropy of w is

$$H_v(w) = v(T). \tag{2.2}$$

Next we define the concept of statistical inference relative to a partial measurement.[1] Let $\mathscr{B} \subset \mathfrak{A}$ be a von Neumann subalgebra, $\mathscr{B}_+$ its positive cone, $v \in N(\mathfrak{A})$ an arbitrary *a priori* state with restriction $v_{\mathscr{B}}$, and $w_{\mathscr{B}} \in N(\mathscr{B})$ the normal state obtained from a partial measurement of the subalgebra $\mathscr{B}$.

Definition 2.2: Let $w_{\mathscr{B}} \leqslant \lambda v_{\mathscr{B}}$ and $T_0 \in \mathscr{B}_+$ be the corresponding Sakai operator satisfying $w_{\mathscr{B}}(\cdot) = v_{\mathscr{B}}(T_0 \cdot T_0)$. Then the $(\mathscr{B},v)$ inference w_0 from $w_{\mathscr{B}}$ is defined by

$$w_0(A) = v(T_0 A T_0), \quad \forall A \in \mathfrak{A} \tag{2.3}$$

We note that $w_0 \in N(\mathfrak{A})$ is an extension of $w_{\mathscr{B}}$, i.e., that the inferred state is compatible with the partial measurement.

We have shown in Ref. 2 that, among all the extensions w of $w_{\mathscr{B}}$ admitting a Sakai operator $T \in \mathfrak{A}_+$, the $(\mathscr{B},v)$ inference w_0 is uniquely charaterized by maximum v entropy $H_v(w)$, and that up to normalization, $H_v(w)$ is the only functional of the form $v[F(T)]$, $F(T)$ analytic, which, for arbitrary subalgebras $\mathscr{B} \subset \mathfrak{A}$, characterizes w_0 as a maximum entropy state.

3. THE BURES DISTANCE FUNCTION

Let $\mathfrak{A}$ and $N(\mathfrak{A})$ be as above and π a $*$-representation of $\mathfrak{A}$ on some Hilbertspace H_π. For $z \in H_\pi$, $\|z\| = 1$, we denote by w_z the vector state defined on $\mathfrak{A}$ by $w_z(\cdot) = \langle z, \pi(\cdot) z \rangle$. Let, for any $v,w \in N(\mathfrak{A})$,

$$\rho_\pi(v,w) = \sup_{x,y} \{ |\langle x,y \rangle| : \quad w_x = v; \quad w_y = w \}$$

$$d_\pi(v,w) = \inf_{x,y} \{ \|x-y\| : \quad w_x = v; \quad w_y = w \}$$

and

$$\rho(v,w) = \sup_\pi \rho_\pi(v,w)$$

$$d(v,w) = \inf_\pi d_\pi(v,w). \tag{3.1}$$

According to Bures,[5] $d(v,w)$ is a distance function on $N(\mathfrak{A})$ satisfying

$$d^2(v,w) = 2[1 - \rho(v,w)]. \tag{3.2}$$

Furthermore, we see that, for $v = w_x$ and $w = w_y$ vector states on $\mathfrak{A} = B(H)$, the quantity $\rho^2(v,w) = |\langle x,y \rangle|^2$ can be interpreted as the quantum-mechanical transition probability between the pure states $x,y \in H$, as pointed out by Uhl-

 0022-2488/79/091963-04$01.00 © 1979 American Institute of Physics

mann.[6] This suggests the following:

Definition 3.1: Let $v,w \in N(\mathfrak{A})$. Then the quantities $d(v,w)$ and $\rho^2(v,w)$ defined by (3.1) are called the Bures distance and the Uhlmann transition probability between the states v and w.

We now establish the relation between the Bures distance and the relative entropy. Let $v,w \in N(\mathfrak{A})$, $w \leqslant \lambda v$, and $T \in \mathfrak{A}_+$ the Sakai operator satisfying $w(\cdot) = v(T \cdot T)$. It then follows from Lemma 2 of Araki[7] that

$$d^2(v,w) = 2[1 - v(T)]. \tag{3.3}$$

Combining (3.3) with (2.2) and (3.2) we obtain

$$H_v(w) = \rho(v,w), \tag{3.4}$$

which shows that the v entropy of w is the square root of Uhlmann's transition probability. Hence. we see that, among all extensions w of $w_{\mathscr{B}}$, *the $(\mathscr{B},v)$ inference w_0 from $w_{\mathscr{B}}$ is characterized by maximum transition probability or, eqivalently, minimum Bures distance relative to the a priori state v.*

We conclude this section by computing (3.4) in the case where $v,w \in N(\mathfrak{A})$ admit density operators V,W such that $\mathrm{Tr}(V\cdot) = v(\cdot)$; $\mathrm{Tr}(W\cdot) = w(\cdot)$. Using the relations[8] $W = TVT$ and $V^{1/2}TV^{1/2} = (V^{1/2}WV^{1/2})^{1/2}$ we obtain in this special case

$$H_v(w) = v(T) = \mathrm{Tr}(V^{1/2}TV^{1/2}) = \mathrm{Tr}[(V^{1/2}WV^{1/2})^{1/2}].$$

4. PROPERTIES OF $H_v(w)$

According to (3.4) the discussion of the v entropy $H_v(w)$ amounts to a discussion of the Bures function $\rho(v,w)$ in the case where w is majorized to λv. Most of our proofs will be based on the following result by Araki[7] which shows that the extrema in (3.1) are actually reached:

Araki's Lemma: For any $v \in N(\mathfrak{A})$ there exist a representation π on a Hilbert space H_π and a vector $x \in H_\pi$ such that $\omega_x = v$ and such that, for any $w \in N(\mathfrak{A})$, there is a $y \in H_\pi$ satisfying $\omega_y = w$ and $|\langle x,y \rangle| = \rho(v,w)$.

Our first theorem shows that $H_v(w)$ is a concave function of w.

Theorem 4.1: Let $w_\lambda \in N(\mathfrak{A})$ be the convex linear combination $w_\lambda = \lambda w_1 + (1-\lambda)w_2$, $0 \leqslant \lambda \leqslant 1$, of $w_1,w_2 \in N(\mathfrak{A})$. Then, for any $v \in N(\mathfrak{A})$ which majorizes w_1 and w_2, we have

$$H_v(w_\lambda) \geqslant \lambda H_v(w_1) + (1-\lambda)H_v(w_2). \tag{4.1}$$

Proof: It is clear that v majorizes w_λ. By Araki's Lemma there exist a representation π on H_π and unit vectors $x,y_1,y_2 \in H_\pi$ such that $v = \omega_x$, $w_i = \omega_{y_i}$, $\rho(v,w_i) = |\langle x,y_i \rangle|$, $i = 1,2$. Multiplying y_i by proper phase factors, we can assume that $\rho(v,w_i) = \langle x,y_i \rangle \geqslant 0$. Consider now in the direct sum $\hat{H} = H_\pi \oplus H_\pi$ the representation $\hat{\pi} = \pi \oplus \pi$ and the vectors

$$\hat{x} = (\lambda^{1/2}x, \quad (1-\lambda)^{1/2}x),$$

$$\hat{y}_1 = (y_1,0), \quad \hat{y}_2 = (0,y_2).$$

Denoting the scalar product in $\hat{H}$ by $\langle\langle \cdot\,,\cdot \rangle\rangle$ we have

$$\begin{aligned}
&\langle\langle \hat{x},\hat{\pi}(\cdot)\hat{x} \rangle\rangle\\
&= \langle\langle (\lambda^{1/2}x,(1-\lambda)^{1/2}x),\hat{\pi}(\cdot)(\lambda^{1/2}x,(1-\lambda)^{1/2}x) \rangle\rangle\\
&= \langle\langle (\lambda^{1/2}x,(1-\lambda)^{1/2}x),(\lambda^{1/2}\pi(\cdot)x,(1-\lambda)^{1/2}\pi(\cdot)x) \rangle\rangle\\
&= \lambda\langle x,\pi(\cdot)x \rangle + (1-\lambda)\langle x,\pi(\cdot)x \rangle\\
&= \omega_x(\cdot) = v(\cdot)
\end{aligned}$$

and

$$\begin{aligned}
&\langle\langle \lambda^{1/2}\hat{y}_1 + (1-\lambda)^{1/2}\hat{y}_2,\hat{\pi}(\cdot)[\lambda^{1/2}\hat{y}_1 + (1-\lambda)^{1/2}\hat{y}_2] \rangle\rangle\\
&= \langle\langle (\lambda^{1/2}y_1,(1-\lambda)^{1/2}y_2),(\lambda^{1/2}\pi(\cdot)y_1,(1-\lambda)^{1/2}\pi(\cdot)y_2) \rangle\rangle\\
&= \lambda\omega_{y_1}(\cdot) + (1-\lambda)\omega_{y_2}(\cdot)\\
&= \lambda w_1(\cdot) + (1-\lambda)w_2(\cdot) = w_\lambda(\cdot).
\end{aligned}$$

It then follows that

$$\begin{aligned}
\rho(v,w_\lambda) &\geqslant |\langle\langle \hat{x},\lambda^{1/2}\hat{y}_1 + (1-\lambda)^{1/2}\hat{y}_2 \rangle\rangle|\\
&= |\langle\langle (\lambda^{1/2}x,(1-\lambda)^{1/2}x),(\lambda^{1/2}y_1,(1-\lambda)^{1/2}y_2) \rangle\rangle|\\
&= |\lambda\langle x,y_1 \rangle + (1-\lambda)\langle x,y_2 \rangle| = \lambda\rho(v,w_1)\\
&\quad + (1-\lambda)\rho(v,w_2). \blacksquare
\end{aligned}$$

Next we show that the v entropy $H_v(w)$ does not decrease when the states v,w are restricted to a subalgebra $\mathscr{B} \subset \mathfrak{A}$

Theorem 4.2: Let $\mathscr{B} \subset \mathfrak{A}$ be a von Neumann subalgebra and $v_{\mathscr{B}},w_{\mathscr{B}} \in N(\mathscr{B})$ the restrictions of $v,w \in N(\mathfrak{A})$. Then

$$H_{v_{\mathscr{B}}}(w_{\mathscr{B}}) \geqslant H_v(w). \tag{4.2}$$

Proof: By Araki's Lemma there exists a representation π on H_π and unit vectors $x,y \in H_\pi$ such that $v = \omega_x$, $w = \omega_y$, $\rho(v,w) = |\langle x,y \rangle|$. Let $\pi_{\mathscr{B}}$, $v_{\mathscr{B}}$, $w_{\mathscr{B}}$ be the restrictions of π, v, w to $\mathscr{B}$. Then $\pi_{\mathscr{B}}$ is a *-representation of $\mathscr{B}$ on H_π and $v_{\mathscr{B}} = \omega_x^{\mathscr{B}}, w_{\mathscr{B}} = \omega_y^{\mathscr{B}}$, with $\omega_x^{\mathscr{B}}(\cdot) = \langle x,\pi_{\mathscr{B}}(\cdot)x \rangle$, $\omega_y^{\mathscr{B}}(\cdot) = \langle y,\pi_{\mathscr{B}}(\cdot)y \rangle$. Hence,

$$\rho(v_{\mathscr{B}},w_{\mathscr{B}}) \geqslant |\langle x,y \rangle| = \rho(v,w). \blacksquare$$

Finally we consider the case of a coupled system. Suppose that $\mathfrak{A}_1 \subseteq B(H_1)$ and $\mathfrak{A}_2 \subseteq B(H_2)$ are von Neumann algebras on H_1 and H_2. Let $\mathfrak{A} = \mathfrak{A}_1 \otimes \mathfrak{A}_2$ be the tensor product von Neumann algebra on $H = H_1 \otimes H_2$, $v,w \in N(\mathfrak{A})$ normal states on $\mathfrak{A}$ such that $w \leqslant \lambda v$, $\lambda \geqslant 1$, and let $H_v(w)$ be the v entropy of w. Moreover, let $\mathscr{B}_1$ and $\mathscr{B}_2$ be the von Neumann subalgebras $\mathfrak{A}_1 \otimes I$ and $I \otimes \mathfrak{A}_2$, respectively. We now define on $\mathfrak{A}_1,\mathfrak{A}_2$ the factor states $v_1(A_1) = v_{\mathscr{B}_1}(A_1 \otimes I)$, $v_2(A_2) = v_{\mathscr{B}_2}(I \otimes A_2)$, $w_1(A_1) = w_{\mathscr{B}_1}(A_1 \otimes I)$, $w_2(A_2) = w_{\mathscr{B}_2}(I \otimes A_2)$. It is easily seen that $v_i,w_i \in N(\mathfrak{A}_i)$, $w_i \leqslant \lambda v_i$, $i = 1,2$. Let $H_{v_i}(w_i)$ be the corresponding entropies. Finally we define $v_1 \otimes v_2$, $w_1 \otimes w_2 \in N(\mathfrak{A})$ by linear extension from $v_1 \otimes v_2(A_1 \otimes A_2) = v_1(A_1)v_2(A_2)$; $w_i \otimes w_2(A_1 \otimes A_2) = w_1(A_1)w_2(A_2)$. Since $w_1 \otimes w_2 \leqslant \lambda^2 v_1 \otimes v_2$, the entropy $H_{v_1 \otimes v_2}(w_1 \otimes w_2)$ exists.

Theorem 4.3: Let $v,w,v_1 \otimes v_2$, $w_1 \otimes w_2$ be defined as above. Then

$$H_{v_1 \otimes v_2}(w_1 \otimes w_2) \geqslant [H_v(w)]^2. \tag{4.3}$$

Proof: From the definition of v_i and w_i it follows that $H_{v_i}(w_i) = H_{v_{\mathscr{B}_i}}(w_{\mathscr{B}_i})$, and from Theorem 4.2 that $H_{v_{\mathscr{B}_i}}(w_{\mathscr{B}_i}) \geqslant H_v(w)$. Moreover, we have, with T_i the Sakai operators satisfying $w_i(\cdot) = v_i(T_i \cdot T_i)$,

$$\begin{aligned}
H_{v_1}(w_1)\cdot H_{v_2}(w_2) &= v_1(T_1)\cdot v_2(T_2) = v_1 \otimes v_2(T_1 \otimes T_2)\\
&= H_{v_1 \otimes v_2}(w_1 \otimes w_2).
\end{aligned}$$

We therefore obtain

$$\begin{aligned}
H_{v_1 \otimes v_2}(w_1 \otimes w_2) &= H_{v_1}(w_1)\cdot H_{v_2}(w_2)\\
&= H_{v_{\mathscr{B}_1}}(w_{\mathscr{B}_1})H_{v_{\mathscr{B}_2}}(w_{\mathscr{B}_2}) \geqslant [H_v(w)]^2 \blacksquare
\end{aligned}$$

For methodological reasons it is useful to consider also the following *alternative proof.*

According to Araki's Lemma there exists a representation $\pi(\mathfrak{A})$ on H_π and unit vectors $x,y\in H_\pi$ such that $v(\cdot) = \langle x,\pi(\cdot)x\rangle$, $w(\cdot) = \langle y,\pi(\cdot)y\rangle$, $H_v(w) = |\langle x,y\rangle|$. Let π_1, π_2 be the representations of $\mathfrak{A}_1$, $\mathfrak{A}_2$ given by $\pi_1(A_1) = \pi(A_1\otimes I)$, $\pi_2(A_2) = \pi(I\otimes A_2)$, and define on $H_\pi\otimes H_\pi$ the representation $\pi_1\otimes\pi_2$ by extension of $\pi_1\otimes\pi_2(A_1\otimes A_2)z_1\otimes z_2 = \pi_1(A_1)z_1\otimes\pi_2(A_2)z_2$; $z_1,z_2\in H_\pi$. From

$$\langle\langle x\otimes x,\pi_1\otimes\pi_2(A_1\otimes A_2)x\otimes x\rangle\rangle = v_1(A_1)\cdot v_2(A_2)$$
$$= v_1\otimes v_2(A_1\otimes A_2),$$
$$\langle\langle y\otimes y,\pi_1\otimes\pi_2(A_1\otimes A_2)y\otimes y\rangle\rangle = w_1(A_1)\cdot w_2(A_2)$$
$$= w_1\otimes w_2(A_1\otimes A_2),$$

it then follows that

$$H_{v_1\otimes v_2}(w_1\otimes w_2)\geqslant|\langle\langle x\otimes x,y\otimes y\rangle\rangle| = |\langle x,y\rangle|^2 = [H_v(w)]^2.$$

5. CONDITIONING

In this Section we consider the effect upon $H_v(w)$ of conditioning w relative to a subalgebra $\mathscr{B}\subset\mathfrak{A}$.[3]

Let $\mathfrak{A}$ be a von Neumann algebra and $v\in N(\mathfrak{A})$ a faithful state. We define on $\mathfrak{A}$ the inner product $\langle A_1,A_2\rangle = v(A_1^\dagger A_2)$ and denote by $L^2(\mathfrak{A},v)$ the Hilbert space completion of this inner product space. Let π be the faithful *-representation of $\mathfrak{A}$ in $L^2(\mathfrak{A},v)$ defined by $\pi(A_1)A_2 = A_1A_2$. For any $w\in N(\mathfrak{A})$, $w\leqslant\lambda v$, with Sakai operator $T\in\mathfrak{A}_+$ and v entropy $H_v(w) = v(T)$, we have

$$w(\cdot) = \langle T,\pi(\cdot)T\rangle,\quad \langle T,T\rangle = 1,\quad H_v(w) = \langle I,T\rangle.$$

We extend w to a state on $B[L^2(\mathfrak{A},v)]$ by defining $w(A) = \langle T,AT\rangle$ for all $A\in B[L^2(\mathfrak{A},v)]$. Moreover, since the states $w\in N(\mathfrak{A})$ correspond to certain pure states on $B[L^2(\mathfrak{A},v)]$, we can extend the definition of H_v to all pure states of the form $\omega_z(A) = \langle z,Az\rangle$, $A\in B[L^2(\mathfrak{A},v)]$, by setting

$$H_v(\omega_z) = |\langle I,z\rangle|. \tag{5.1}$$

If now $\mathscr{B}\subset\mathfrak{A}$ is a von Neumann subalgebra, then $L^2(\mathscr{B},v_{\mathscr{B}})$ is a closed subspace of $L^2(\mathfrak{A},v)$. Let $P\in B[L^2(\mathfrak{A},v)]$ be the projection onto $L^2(\mathscr{B},v_{\mathscr{B}})$.

Definition 5.1: Let $w\in N(\mathfrak{A})$, $w\leqslant\lambda v$, and $T\in\mathfrak{A}_+$ the corresponding Sakai operator. If $PT\neq 0$, then

$$(w|\mathscr{B})(\cdot) = \frac{\langle PT,\pi(\cdot)PT\rangle}{\|PT\|^2} \tag{5.2}$$

is a state in $N(\mathfrak{A})$ called the $(\mathscr{B},v)$ *conditioning of* w. On the other hand the conditional expectation of $A\in\mathfrak{A}$ has been defined as follows[3]:

Definition 5.2: Let $A\in\mathfrak{A}$, then the $(\mathscr{B},v)$ expectation of A is an operator $\in(A)\in B[L^2(\mathfrak{A},v)]$ defined by

$$\in(A) = P\pi(A)P. \tag{5.3}$$

Combining the definition 5.1 and 5.2 we notice that

$$(w|\mathscr{B})(A) = \frac{w[\in(A)]}{\|PT\|^2};\quad A\in\mathfrak{A}. \tag{5.4}$$

If w_0 is the $(\mathscr{B},v)$ inference from $w_{\mathscr{B}}$ we have

$w_0(A) = \langle T_0,\pi(A)T_0\rangle$, $T_0\in L^2(\mathscr{B},v)$, and from $PT_0 = T_0$ follows that

$$(w_0|\mathscr{B})(\cdot) = w_0(\cdot);\quad w_0[\in(\cdot)] = w_0(\cdot). \tag{5.5}$$

The $(\mathscr{B},v)$ *inference* w_0 *is therefore unchanged by* $(\mathscr{B},v)$ *conditioning and does not distinguish between* $A\in\mathfrak{A}$ *and its* $(\mathscr{B},v)$ *expectation* $\in(A)$. This typical property is an indication for the coherence of our concepts of inference and conditioning.

Let us now show that conditioning does not decrease the v entropy $H_v(w)$.

Theorem 5.1: Let $\mathfrak{A}$, $\mathscr{B}$, v, w, $w|\mathscr{B}$ be as above. Then

$$H_v(w|\mathscr{B})\geqslant H_v(w). \tag{5.6}$$

Proof: From (5.2) it follows that

$$(w|\mathscr{B})(\cdot) = \left\langle\frac{PT}{\|PT\|},\pi(\cdot)\frac{PT}{\|PT\|}\right\rangle,$$

and (5.1 therefore implies

$$H_v(w|\mathscr{B}) = \left|\left\langle I,\frac{PT}{\|PT\|}\right\rangle\right| = \frac{|\langle I,PT\rangle|}{\|PT\|}$$
$$= \frac{|\langle PI,T\rangle|}{\|PT\|} = \frac{|\langle I,T\rangle|}{\|PT\|}\geqslant|\langle I,T\rangle| = H_v(w) \;\blacksquare$$

6. REMARKS

In recent years much effort has gone into generalizing von Neumann's entropy

$$S(W) = -\operatorname{Tr}(W\log W). \tag{6.1}$$

It can be shown that, in the special case of the trace, (6.1) and (6.2) lead to the same maximum entropy states. But for general v this is not true since (according to the last remark of Sec. 2) H_v is essentially the only entropy which characterizes statistical inference. (For the computation of the maximum Lindblad and Naudts entropy states, cf., Ref. 15.)

$$H(W) = \operatorname{Tr}(W^{1/2}). \tag{6.2}$$

It can be shown that, in the special case of the trace, (6.1) and (6.2) lead to the same maximum entropy states. But for general v this is not true since (according to the last remark of Sec. 2) H_v is essentially the only entropy which characterizes statistical inference. (For the computation of the maximum Lindblad and Naudts entropy states, cf., Ref. 15.)

For the classical probability measures $\mu < \nu$ with Radon–Nikodym derivative $d\mu/d\nu$, the analogues of (6.1) and (6.2) read

$$S_\nu(\mu) = -\int\frac{d\mu}{d\nu}\log\frac{d\mu}{d\nu}\,d\nu, \tag{6.3}$$

$$H_\nu(\mu) = \int\left(\frac{d\mu}{d\nu}\right)^{1/2}d\nu. \tag{6.4}$$

The possibility of (6.4) as an alternative to (6.3) in the classical case has been considered by Kagan, Linnik, and Rao.[16]

From the proof of theorem 4.3 follows that

$$H_{v_1\otimes v_2}(w_1\otimes w_2) = H_{v_1}(w_1)H_{v_2}(w_2)$$

is multiplicative for product states, whereas the generalizations of $S(W)$ are additive. This difference is, however, of a cosmetic nature and could be avoided by replacing the definition (2.2) by

$$\widetilde{H}_v(w) = \log v(T).$$

In the case where v is tracial, Araki and Lieb[17] have proved weak subadditivity of the entropy for coupled systems. The analogous statement for our multiplicative entropy would read

$$H_{v_1 \otimes v_2}(w_1 \otimes w_2) \geqslant H_v(w). \tag{6.5}$$

Instead of (6.5) we have only proved (4.3) which is a weaker result, since $0 \leqslant H_v(w) \leqslant 1$ implies $H_v(w) \geqslant [H_v(w)]^2$. Whether or not the inequality (6.5) holds remains an open problem which appears to be difficult even when v is tracial.

We conclude our series of remarks by communicating the following suggestion of the referee: The relation (3.4) could serve as a new definition for $H_v(w)$. The limitation to relatively bounded states v and w could thereby be avoided.

[1]J.-P. Marchand, Found. Phys. **7**, 35 (1977).
[2]J.-P. Marchand and W. Wyss, J. Stat. Phys. **16**, 349 (1977).
[3]S. Gudder and J.-P. Marchand, Rep. Math. Phys. **12**, 317 (1977).
[4]S. Sakai, Bull. Am. Math. Soc. **71**, 149 (1965).
[5]D.J.C. Bures, Trans. Am. Math. Soc. **135**, 199 (1969).
[6]A. Uhlmann, Rep. Math. Phys. **9**, 273 (1976).
[7]H. Araki, Publ. RIMS, Kyoto Univ. **8**, 335 (1972).
[8]G. Pedersen and M. Takesaki, Proc. Am. Math. Soc. **36**, 311 (1972).
[9]I.E. Segal, J. Math. Mech. **9**, 623 (1960).
[10]H. Umegaki, Kodai Math. Sem. Rep. **14**, 59 (1962).
[11]B. Ruskai, Ann. Inst. H. Poincaré, **19**, 357 (1973).
[12]W. Ochs and H. Spohn: "A Characterization of the Segal Entropy," preprint Univ. of Munich.
[13]J. Naudts, Commun. Math. Phys. **37**, 175 (1974).
[14]G. Lindblad, Commun. Math. Phys. **33**, 305 (1973).
[15]R. Benoist, J.-P. Marchand, and W. Wyss, Lett. Math. Phys. 1979 (to appear).
[16]A. Kagan, I. Linnik, and C. Rao, *Characterization Problems in Mathematical Statistics* (Wiley, New York, 1973).
[17]H. Araki and E.H. Lieb, Commun. Math. Phys. **18**, 160 (1970).

Singular potentials and analytic regularization in classical Yang–Mills equations

C. G. Bollini and J. J. Giambiagi
Centro Brasileiro de Pesquisas Físicas, Av. Wenceslau Braz, 71 - Rio de Janeiro, Brasil

J. Tiomno
Departamento de Física - Pontifícia Universidade Católica CxP 38071, Rio de Janeiro, Brasil

(Received 19 March 1979; accepted for publication 4 May 1979)

The class of instaton solutions with "extension" parameter λ^2 positive is extended to λ^2 negative. The nature of the singular sphere of radius $|\lambda|$ is analyzed in the light of the analytical regularization method. This leads to well defined solutions of the Yang–Mills equations. Some of them are sourceless ("$\pm$i o" and "*Vp*"), others correspond to currents concentrated on the sphere of singularity (" + " and " − "). Although the equations are nonlinear, the "*Vp*" solutions turn out to be the real part of the " $\pm$ i o" solutions. The ansatz of t'Hooft for the superposition of instantons is used to sum the contributions corresponding to λ^2 with positive and negative signs. A subsequent limiting process allows then the construction of solutions of the "multipole" type. The general situation of potentials having a denominator D, with a corresponding surface of singularity at $D = 0$, is also considered in the same light.

1. INTRODUCTION

Belavin *et al.*[1] and 't Hooft[2] have given an important solution of the homogeneous Yang–Mills equation (Euclidean metric):

$$A_\mu = -2i\frac{\sigma_{\mu\nu}x_\nu}{x^2+\lambda^2}, \quad F_{\mu\nu} = 4i\frac{\lambda^2\sigma_{\mu\nu}}{(x^2+\lambda^2)^2}, \tag{1.1}$$

$$\sigma_{ij} = \tfrac{1}{2}\epsilon_{ijk}\sigma_k, \quad \sigma_{i4} = \tfrac{1}{2}\sigma_i, \quad \sigma_{\mu\nu} = -\sigma_{\nu\mu}. \tag{1.2}$$

(1.1) is a solution for any real value of the constant λ. It is natural to consider also complex values of λ, in particular imaginary ones. In this case ($\lambda^2 = -|\lambda^2|$) we have a singular sphere of radius $|\lambda|$. Outside this sphere (1.1) continues to be a solution, but it is not clear what happens at the surface itself. [Equation (1.1) is not well defined there.] We must give additional rules in order to complete the definition of the potential.

In Sec. 2 we define the " $\pm io$" prescription. In Sec. 3 we study the "outside" and "inside" solutions which are implied by the " $\pm$ " potentials. In Sec. 4, the principal value potential ("Vp") is introduced via a combination of the " + " and " − " prescription. In Sec. 5 we find the topological numbers of those solutions. In Sec. 6 we briefly treat the linear denominator case. Once the use of negative values of λ^2 has been made clear and mathematically justified, it is then possible, following 't Hooft[3] method[4,5] to superimpose instantons with positive and negative λ^2, forming the analogue of what could be called a "dipole instanton." This is done in Sec. 7, where it is pointed out that the procedure can be extended to higher "multipoles."

Finally, in Sec. 8 we discuss the general case of potentials, having a denominator D, which are singular at a surface defined by the equation $D = 0$. All these singularities can be treated by following the method of analytic regularization, which goes back to Riesz[6] methods,[7,8] together with the distribution theory of Guelfand–Shilov.[9]

Appendix A contains some formulas from Ref. 9 which are extensively used in the text. Appendix B is an alternative (and equivalent) way to treat the singularities.

2. *io* POTENTIALS

When λ is pure imaginary in (1.1), we have a singularity at the sphere $x^2 = |\lambda^2|$. A natural way to attach a well-defined meaning to this singularity is to approximate the imaginary axis, from the right, for instance. For λ's with $\mathrm{Re}\lambda > 0$, (1.1) is a sourceless solution. Thus, we take

$$\lambda = \lambda_1 + i\lambda_2, \quad (\lambda_1 > 0),$$

$$\lambda^2 = \lambda_1^2 - \lambda_2^2 + 2i\lambda_1\lambda_2 \to -\lambda_2^2 \pm io, \quad (\text{for } \lambda_1 \to 0).$$

The $\pm$ sign corresponds to $\lambda_2 \gtrless 0$.

For simplicity we shall work with the positive sign. So, instead of (1.1), we shall write

$$A_\mu^{(+io)} = -2i\frac{\sigma_{\mu\nu}x_\nu}{(x^2-\lambda^2+io)}. \tag{2.1}$$

We know from the results obtained by Guelfand and Shilov,[9] that the distribution

$$(x^2-\lambda^2+io)^\alpha \tag{2.2}$$

is a well-defined functional analytic (entire) in the parameter α. All powers and derivatives are well defined. In particular, we can compute the field $F_{\mu\nu}$ and the current corresponding to (2.1).

$$F_{\mu\nu} = \partial_\mu A_\nu - \partial_\nu A_\mu + [A_\mu, A_\nu], \tag{2.3}$$

$$F_{\mu\nu}^{(+io)} = 4i\frac{\lambda^2\sigma_{\nu\mu}}{(x^2-\lambda^2+io)^2}, \tag{2.4}$$

$$J_\mu = \partial_\nu F_{\nu\mu} + [A_\nu, F_{\nu\mu}], \tag{2.5}$$

$$J_\mu^{(+io)} = 0. \tag{2.6}$$

Equation (2.6) follows simply from the self-duality of (2.4).

 © 1979 American Institute of Physics

3. " + " AND " − " POTENTIALS

Due to (2.6) the potential (2.1) is a complex solution of the Yang–Mills equation. Equation (A7) shows explicitly its real and imaginary parts.

In order to study the real part of (2.1), defined through (A8), (A7) and (A1), (A2), we shall consider the potential

$$A_\mu^{(\alpha)+} = -2i\sigma_{\mu\nu} x_\nu (x^2-\lambda^2)_+^\alpha, \tag{3.1}$$

which is identically zero for $x^2<\lambda^2$. [cf. (A1).]

Field and current corresponding to (3.1) can easily be computed:

$$\begin{aligned} F_{\nu\mu}^{(\alpha)+} &= 4i\sigma_{\mu\nu}[x^2-(x^2-\lambda^2)_+^{-\alpha}](x^2-\lambda^2)_+^{2\alpha} \\ &+ 4i(\sigma_{\nu\tau}x_\mu - \sigma_{\mu\tau}x_\nu)x_\tau(\alpha(x^2-\lambda^2)_+^{\alpha-1} \\ &+ (x^2-\lambda^2)_+^{2\alpha}). \end{aligned} \tag{3.2}$$

$$\begin{aligned} J_\mu^{(\alpha)+} &= -8i\sigma_{\mu\nu}x_\nu[3(x^2-\lambda^2)_+^{2\alpha} + 3\alpha(x^2-\lambda^2)_+^{\alpha-1} \\ &+ \alpha(\alpha-1)x^2(x^2-\lambda^2)_+^{\alpha-2} \\ &- 2x^2(x^2-\lambda^2)_+^{3\alpha}]. \end{aligned} \tag{3.3}$$

If we take naively the limit $\alpha\to -1$, the second member of (3.3) gives zero. However, one must be cautious in taking that limit, as the distribution $(x^2-\lambda^2)_+^\alpha$ has poles at negative integer values for α. When (A3) is taken into account in (3.1), (3.2), and (3.3), we can see that all three expressions give simple poles when $\alpha\to -1$. The residues at the poles are distributions concentrated on the surface of the sphere $x^2=\lambda^2$. Furthermore, the finite part of (3.3) (at $\alpha=-1$) is a derivative of a δ function. In fact, by taking

$$\text{P.f.}\Psi(\alpha)|_{\alpha=-n} = \frac{d}{d\alpha}(\alpha+n)\Psi(\alpha)|_{\alpha=-n}, \tag{3.4}$$

and using (A9) (for $n=1$), we get (see A5)

$$\text{P.f.}J_\mu^{(\alpha)+}|_{\alpha=-1} = 12i\sigma_{\mu\nu}x_\nu\lambda^2\delta''(x^2-\lambda^2). \tag{3.5}$$

We see then that, in the limit $\alpha\to -1$, the source of (3.1) is concentrated on the surface of the sphere $x^2=\lambda^2$.

A similar procedure can be followed with the potential [compare with Ref. 9]:

$$A_\mu^{(\alpha)-} = +2i\sigma_{\mu\nu}x_\nu(x^2-\lambda^2)_-^\alpha, \tag{3.6}$$

where $(x^2-\lambda^2)_-^\alpha$ is defined by (A2), and is zero outside the sphere of radius λ. The field and current for (3.6) are

$$\begin{aligned} F_{\mu\nu}^{(\alpha)-} &= 4i\sigma_{\mu\nu}[x^2+(x^2-\lambda^2)_-^{-\alpha}](x^2-\lambda^2)_-^{2\alpha} \\ &+ 4i(\sigma_{\nu\tau}x_\mu - \sigma_{\mu\tau}x_\nu)x_\tau[\alpha(x^2-\lambda^2)_-^{\alpha-1} \\ &+ (x^2-\lambda^2)_-^{2\alpha}]. \end{aligned} \tag{3.7}$$

$$\begin{aligned} J_\mu^{(\alpha)-} &= -8i\sigma_{\nu\mu}x_\nu[3\alpha(x^2-\lambda^2)_-^{\alpha-1} + 3(x^2-\lambda^2)_-^{2\alpha} \\ &- \alpha(\alpha-1)x^2(x^2-\lambda^2)_-^{\alpha-2} + 2x^2 \\ &\times(x^2-\lambda^2)_-^{3\alpha}]. \end{aligned} \tag{3.8}$$

Taking into account (A4), we again find that, in the limit $\alpha\to -1$, the source is concentrated on the sphere.

$$\text{P.f.}J_\mu^{(\alpha)-}|_{\alpha=-1} = -12i\lambda^2\sigma_{\mu\nu}x_\nu\delta''(x^2-\lambda^2). \tag{3.9}$$

It is worth mentioning that, in the limit $\alpha\to -1$, the source (3.9) is equal and opposite to (3.5).

4. Vp POTENTIAL

We introduce the definition

$$A_\mu^{(\alpha)} = A_\mu^{(\alpha)+} + A_\mu^{(\alpha)-}, \tag{4.1}$$

i.e.,

$$A_\mu^{(\alpha)} = 2i\sigma_{\mu\nu}x_\nu[(x^2-\lambda^2)_+^\alpha - (x^2-\lambda^2)_-^\alpha]. \tag{4.2}$$

For $\alpha\to -1$ the pole parts of the + and − distributions cancel each other, leaving only a finite result,

$$A_\mu^{\text{Vp}} = -2i\sigma_{\mu\nu}x_\nu[x^2-\lambda^2]^{-1}, \tag{4.3}$$

which is real solution where $[x^2-\lambda^2]^{-1}$ is given by (A8) and coincides with Cauchy's principal value.

It is easy to compute the field and current for the potential (4.2). The + and − distributions do not interfere with one another as one of them is zero when the other is not. For this reason the field corresponding to (4.2) is a superposition of (3.2) and (3.7).

$$F_{\nu\mu}^{(\alpha)} = F_{\nu\mu}^{(\alpha)+} + F_{\nu\mu}^{(\alpha)-}. \tag{4.4}$$

Analogously,

$$J_\mu^{(\alpha)} = J_\mu^{(\alpha)+} + J_\mu^{(\alpha)-}. \tag{4.5}$$

where $J_\mu^{(\alpha)+}$ and $J_\mu^{(\alpha)-}$ are given by (3.3) and (3.8). Taking now the limit $\alpha\to -1$, we find that the current is zero. The corresponding field is given by (4.4) [with (3.2) and (3.7)]. For $\alpha\to -1$ we get

$$F_{\nu\mu}^{\text{Vp}} = 4i\sigma_{\mu\nu}\lambda^2[x^2-\lambda^2]^{-2}, \tag{4.6}$$

where the finite part $[\ \ ]^{-n}$ is given by (A8). Note that (4.6) is self-dual, showing again that the field is sourceless.

According to (A7), the potential (4.3) and the field (4.6) are respectively the real parts of (2.1) and (2.4). We then see that the $\pm io$ potential (2.1) is a complex solution of the homogeneous Yang–Mills equation, whose real part (4.3) is also a solution (see also Ref. 10).

Note that in (4.3) and (4.6), the square brackets are in fact labels for the limiting process implied by the analytic regularization method. If one takes (4.3) directly in the Yang–Mills equation, one finds ambiguities due to the existence of products which are not well defined. (See also Ref. 11.)

5. TOPOLOGICAL NUMBER

The Gauge fields so far considered (2.4), (3.2), (3.7), and (4.6) are all singular on the sphere $x^2=\lambda^2$. For the calculation of the topological number of (2.4) one must take the integral over all space of

$$\text{Tr}\{F_{\mu\nu}^{(io)}\tilde F_{\mu\nu}^{(io)}\} = \frac{-96\lambda^4}{(x^2-\lambda^2+io)^4}. \tag{5.1}$$

One would think that the already mentioned singularity renders the integral divergent. However, it is well known, and easy to check, that when the denominator is $(x^2+\lambda^2)^{+4}$, the topological number comes out to be one independent of the value of λ^2. It is then natural to expect that the integration of (5.1) will give also the same result as is obtained by a continuation in λ to the imaginary axis. Nevertheless, to be consistent, one would like to obtain it by direct calculation using (A6),

$$(x^2 - \lambda^2 + io)^\alpha = (x^2 - \lambda^2)^\alpha_+ + e^{i\pi\alpha}(x^2 - \lambda^2)^\alpha_-,$$

as a functional analytic in α, and then take the limit $\alpha \to -4$.

To do that, we shall consider a "trial" function $\Psi(x^2)$, which is 1 for $x^2 < \Lambda^2$ and zero for $x^2 > \Lambda^2$. Of course this is not a proper trial function, but it can be approximated as much as we want by infinite differentiable functions. We then have

$$((x^2 - \lambda^2 + io)^\alpha, \Psi)$$
$$= \pi^2 \int_{\lambda^2}^{\Lambda^2} dx^2 x^2 (x^2 - \lambda^2)^\alpha + \pi^2 e^{i\pi\alpha} \int_0^{\lambda^2} dx^2 x^2 (\lambda^2 - x^2)^\alpha$$
$$= \pi^2 \left(\frac{(\Lambda^2 - \lambda^2)^{\alpha+2}}{\alpha+2} + \lambda^2 \frac{(\Lambda^2 - \lambda^2)^{\alpha+1}}{\alpha+1} \right)$$
$$+ \pi^2 e^{i\pi\alpha} \left(-\frac{(\lambda^2)^{\alpha+2}}{\alpha+2} + \lambda^2 \frac{(\lambda^2)^{\alpha+1}}{\alpha+1} \right),$$

where a factor $2\pi^2$ comes from angular integration. For $\alpha = -4$ we have

$$((x^2 - \lambda^2 + io)^{-4}, \Psi) = -\pi^2 \left(\frac{1}{2} (\Lambda^2 - \lambda^2)^{-2} + \frac{\lambda^2}{3} \times (\Lambda^2 - \lambda^2)^{-3} \right) + \frac{\pi^2}{6} \lambda^{-4}. \quad (5.2)$$

(5.2) gives the result of the integration over a sphere of radius Λ. If we now take limit $\Lambda \to \infty$ we get

$$\lim_{\Lambda \to \infty} ((x^2 - \lambda^2 + io)^{-4}, \Psi) = \int d^4x (x^2 - \lambda^2 + io)^{-4} = \frac{\pi^2}{6\lambda^4}. \quad (5.3)$$

So, for the topological number we obtain

$$-\frac{1}{16\pi^2} \int F^{(io)}_{\mu\nu} \tilde{F}^{(io)}_{\mu\nu} d^4x = \frac{96\lambda^4}{16\pi^2} \cdot \frac{\pi^2}{6\lambda^4} = 1. \quad (5.4)$$

One can also repeat the calculation using a Gaussian function $\exp(-x^2/a^2)$, which is a true trial function. The result, for $a \to \infty$, reproduces (5.4).

We would like to point out that it is also possible to directly use Guelfand–Shilov's definition of integrals for certain functionals (Ref. 9, p. 65), namely,

$$\int_0^\infty z^\alpha \, dz = 0 \quad \text{(any complex } \alpha). \quad (5.5)$$

With this definition

$$\int (x^2 - \lambda^2 + io)^\alpha d^4x$$
$$= \pi^2 \int_{\lambda^2}^\infty dx^2 x^2 (x^2 - \lambda^2)^\alpha + \pi^2 e^{i\pi\alpha} \int_0^{\lambda^2} dx^2 x^2 (\lambda^2 - x^2)^\alpha$$
$$= \pi^2 \int_0^\infty dz (z^{\alpha+1} + \lambda^2 z^\alpha) + \pi^2 e^{i\pi\alpha} \int_0^{\lambda^2} dz (-z^{\alpha+1} + \lambda^2 z^\alpha)$$
$$= \pi^2 e^{i\pi\alpha} (\lambda^2)^{\alpha+2} \left(\frac{1}{\alpha+1} - \frac{1}{\alpha+2} \right)$$
$$= \frac{\pi^2 (\lambda^2)^{\alpha+2} e^{i\pi\alpha}}{(\alpha+1)(\alpha+2)},$$

which, for $\alpha = 4$, again reproduces (5.3), (5.4). With the same procedure we can compute the topological numbers of the previously discussed solutions. The + field (3.2) turns out to have topological number zero. All the others (−, Vp, $\pm io$) have topological numbers equal to one.

Of course, if instead of using the self-dual matrix $\sigma_{\mu\nu}$, we use the anti-self-dual one $\tilde{\sigma}_{\mu\nu}$, the signs of the Chern number would be reversed.

6. SOLUTIONS WITH LINEAR DENOMINATORS

It is well known[11] that by choosing

$$\rho = \frac{1}{(x.n + \lambda)} \quad (6.1)$$

in the ansatz proposed by t'Hooft, Wilzcek, and Corrigan, Fairlie,[3,4,5] we get the self-dual solution:

$$A_\mu = -i \frac{\sigma_{\mu\nu} n_\nu}{(x.n + \lambda)}, \quad (6.2)$$

$$F_{\mu\nu} = -i \frac{\sigma_{\mu\nu}}{(x.n + \lambda)^2}. \quad (6.3)$$

It can also be directly verified that with

$$A_\mu = +i \frac{\sigma_{\mu\nu} n_\nu}{(x.n + \lambda)} \quad (6.4)$$

we get the anti-self-dual solution:

$$F_{\mu\nu} = \frac{i}{(x.n + \lambda)^2} (2\sigma_{\mu\rho} n_\rho n_\nu - 2\sigma_{\nu\rho} n_\rho n_\mu - \sigma_{\mu\nu} n^2). \quad (6.5)$$

These solutions are all singular at the plane

$$x.n + \lambda = 0. \quad (6.6)$$

In order to give a meaning to those singular expression, in a way similar to the cases already examined, we shall compute the fields and currents for the + and − potentials.

$$A^{(\alpha)+}_\mu = -is\sigma_{\mu\nu} n_\nu (x.n + \lambda)^\alpha_+, \quad (s^2 = 1), \quad (6.7)$$

$$A^{(\alpha)-}_\mu = +is\sigma_{\mu\nu} n_\nu (x.n + \lambda)^\alpha_-, \quad (s^2 = 1). \quad (6.8)$$

Which lead to

$$F^{(\alpha)\pm}_{\nu\mu} = i\sigma_{\mu\nu} n^2 (x.n + \lambda)^2_\pm + i(\sigma_{\nu\tau} n_\mu - \sigma_{\mu\tau} n_\nu) n_\tau \times ((x.n + \lambda)^{2\alpha}_\pm + s\alpha (x.n + \lambda)^{\alpha-1}_\pm) \quad (6.9)$$

$$J^{(\alpha)\pm}_\mu = isn^2 \sigma_{\mu\nu} n_\nu (2(x.n + \lambda)^{3\alpha}_\pm - \alpha(\alpha - 1) \times (x.n + \lambda)^{\alpha-2}_\pm). \quad (6.10)$$

It is easy to see that outside the singular plane the current is zero for $\alpha = -1$. So, the source (if it exists) is concentrated on the plane $x.n + \lambda = 0$.

As was previously done, we can now define the potentials with the labels $\pm io$, Vp and in the present case also $||$, which is defined by [compare with (4.1)]:

$$A^{(\alpha)||}_\mu = A^{(\alpha)+}_\mu - A^{(\alpha)-}_\mu. \quad (6.11)$$

The $\pm io$ and Vp potentials are sourceless, while $\pm$ and $||$

are solutions of the inhomogeneous equation concentrated on the plane $x.n + \lambda = 0$.

7. "MULTIPOLE" SOLUTIONS

To clarify the ideas we shall recall 't Hooft[3] method[4,5,11] for two instantons.

We chose

$$\rho = 1 + \frac{\lambda_1^2}{x^2} + \frac{\lambda_2^2}{(x - x_2)^2}, \tag{7.1}$$

from which we get the potential

$$A_\mu = i\sigma_{\mu\nu}\,\partial_\nu \ln\rho. \tag{7.2}$$

Now, once the idea of a negative λ^2 has been accepted, we can choose in (7.1), $\lambda_2^2 = -\lambda_1^2$, and take the limit $x_2 \to 0$, $\lambda_2^2 \to \infty$ in such a way that $\lambda_2^2 x_{2\mu} = P_\mu$ is a constant vector. It is then easily seen that

$$\rho \to 1 + P.x/x^4, \tag{7.3}$$

where P defines a "dipole moment" for the above mentioned two instantons case.

From (7.3) we obtain

$$A_\mu = i\frac{\sigma_{\mu\nu}Q_\nu}{D} \tag{7.4}$$

with

$$D = x^4 + x.P \tag{7.5}$$

and

$$Q_\nu = P_\nu - 4\frac{P.x}{x^2}x_\nu. \tag{7.6}$$

After a straightforward calculation we get

$$F_{\nu\mu} = i\sigma_{\mu\nu}\frac{Q^2}{D^2} + i\frac{\sigma_{\mu\tau}}{D^2}[DQ_{\tau,\nu} - Q_\tau(Q_\nu + D_\nu)]$$
$$- i\frac{\sigma_{\nu\tau}}{D^2}[DQ_{\tau,\mu} - Q_\tau(Q_\mu + D_\mu)]. \tag{7.7}$$

It is not difficult to check, using (7.5) and (7.6) in (7.7), that $F_{\nu\mu}$ is anti-self-dual, implying that we have a solution of the homogeneous equation, outside the surface $D = 0$.

It is easy to see that we can still have (7.4) as a solution if we change (7.5) to

$$D = \lambda x^4 + x.P, \tag{7.8}$$

for arbitrary λ, including $\lambda = 0$. The method can be generalized to higher "multipoles." For instance, with

$$\rho = 1 + \frac{x.Q.x}{x^6}, \tag{7.9}$$

we have again (43), but now

$$D = x^6 + x.Q.x, \tag{7.10}$$

$$Q_\nu = 2Q_{\nu\mu}x_\mu - 6x_\nu\frac{x.Q.x}{x^2}, \tag{7.11}$$

(7.4) and (7.7), [with (7.10) and (7.11)], provide us with a new solution of the sourceless equation, except perhaps at the surface $D = 0$.

8. REGULARIZED SINGULAR SOLUTIONS

The solutions found in Sec. 7, are singular at the surface defined by

$$D = 0, \tag{8.1}$$

where D is the denominator of the potential (7.4).

In order to give a well defined meaning to expressions such as (7.4) and (7.7), [containing a denominator D for which (8.1) has a real solution] we shall use the methods described previously.

We can give several prescriptions for the definition of the singularity. All of them can be built up from the + and − distributions which have the following definition:

$$D_+^\alpha = \begin{cases} = D^\alpha, & \text{for } D > 0, \\ = 0, & \text{for } D < 0. \end{cases} \tag{8.2}$$

$$D_- = \begin{cases} = 0, & \text{for } D > 0, \\ = |D|^\alpha, & \text{for } D < 0. \end{cases} \tag{8.3}$$

Both (8.2) and (8.3) are meromorphic functionals in the parameter (see Ref. 9, Chapter 3 sec. 4). We now follow the usual pattern:

$$A_\mu^{(\alpha)+} = i\sigma_{\mu\nu}Q_\nu D_+^\alpha. \tag{8.4}$$

$$A_\mu^{(\alpha)-} = -i\sigma_{\mu\nu}Q_\nu D_-^\alpha. \tag{8.5}$$

$$A_\mu^{(\alpha)\pm io} = A_\mu^{(\alpha)+} \pm e^{i\pi\alpha}A_\mu^{(\alpha)-}. \tag{8.6}$$

$$A_\mu^{(\alpha)\mathrm{Vp}} = A_\mu^{(\alpha)+} + A_\mu^{(\alpha)-}. \tag{8.7}$$

For $\alpha \to -1$, (8.6) and (8.7) are solutions of the homogeneous equation, while (8.4) and (8.5) have sources concentrated on the surface (8.1).

APPENDIX A

$$(x^2 - \lambda^2)_+^\alpha = \begin{cases} (x^2 - \lambda^2)^\alpha, & \text{if } x^2 > \lambda^2, \\ 0, & \text{if } x^2 < \lambda^2, \end{cases} \tag{A1}$$

$$(x^2 - \lambda^2)_-^\alpha = \begin{cases} (\lambda^2 - x^2)^\alpha, & \text{if } x^2 < \lambda^2, \\ 0, & \text{if } x^2 > \lambda^2. \end{cases} \tag{A2}$$

Equations (A1) and (A2) are well defined distributions, analytic in α with poles for $\alpha = -n$ (n, positive integer). Near these poles we have

$$(x^2 - \lambda^2)_+^\alpha = \frac{(-1)^{n-1}\delta^{(n-1)}(x^2 - \lambda^2)}{(n-1)!(\alpha + n)} + [x^2 - \lambda^2]_+^{-n} + 0(\alpha + n) \tag{A3}$$

$$(x^2 - \lambda^2)_-^\alpha = \frac{\delta^{n-1}(x^2 - \lambda^2)}{(n-1)!(\alpha + n)} + [x^2 - \lambda^2]_-^{-n} + 0(\alpha + n), \tag{A4}$$

where

$$[x^2 - \lambda^2]_\pm^{-n} = Pf(x^2 - \lambda^2)_\pm^\alpha|_{\alpha = -n},$$

$$\delta^k(u) = \frac{d^k\delta(u)}{du^k},$$

and

$$\delta^{(k-1)}(\lambda^2 - x^2) = \frac{(-1)^{k-1}}{2^k\lambda x^{k-1}}(\delta^{k-1}(x - \lambda) - \delta^{k-1}(x + \lambda)). \tag{A5}$$

We can now define

$$(x^2-\lambda^2 \pm io)^\alpha = (x^2-\lambda^2)_+ + e^{\pm i\pi\alpha}(x^2-\lambda^2)_-^\alpha \quad \text{(A6)}$$

It is easily seen from (A3) and (A4) that this new description is analytical everywhere. In particular,

$$(x^2-\lambda^2+io)^{-n} = [x^2-\lambda^2]^{-n} + i\pi \frac{(-1)^n}{(n-1)!} \times \delta^{(n-1)}(x^2-\lambda^2), \quad \text{(A7)}$$

where

$$[x^2-\lambda^2]^{-n} = [x^2-\lambda^2]_+^{-n} + (-1)^n [x^2-\lambda^2]_-^{-n} \quad \text{(A8)}$$

which, for $n=1$ gives

$$\mathrm{Vp}(x^2-\lambda^2)^{-1} = (x^2-\lambda^2)_+^{-1} - (x^2-\lambda^2)_-^{-1}$$

It is possible to show that (Ref. 9, p. 347)

$$(x^2-\lambda^2)\delta^{(n)}(x^2-\lambda^2) + n\delta^{(n-1)}(x^2-\lambda^2) = 0. \quad \text{(A9)}$$

Note also that if we multiply (A8) times a similar expression with m in place of n, the result is not well defined. However, multiplying $(x^2-\lambda^2+io)^{-n}$ times $(x^2-\lambda^2+io)^{-m}$, using (A7) and then taking the real part of the result, we get

$$[x^2-\lambda^2]^{-n-m} = \Big\{ [x^2-\lambda^2]^{-m} \cdot [x^2-\lambda^2]^{-n} - \frac{\pi^2(-1)^{n+m}}{(n-1)!(m-1)!}\delta^{(m-1)} \times (x^2-\lambda^2)\delta^{(n-1)}(x^2-\lambda^2)\Big\}. \quad \text{(A10)}$$

As a matter of fact, each of the terms on the right hand side is meaningless; however the complete combination is well defined (see Ref. 12).

APPENDIX B

We present here a perhaps more intuitive way to deal with the singularities produced by the zeros of the denominators. Let us introduce a regularizing function $\eta_\epsilon(x)$ with the following properties: $\eta_\epsilon(x)$ and all its derivatives are null at the origin

$$\eta_\epsilon(0), \frac{d^P\eta_\epsilon(x)}{dx^P}\Big|_{x=0} = 0, \quad (\text{all } p).$$

Furthermore

$$\eta_\epsilon(x) = 1, \quad \text{and also} \quad \frac{d^P\eta_\epsilon(x)}{dx^P}\Big|_{x=\epsilon} = 0, \quad (\text{all } p) \quad x \geqslant \epsilon.$$

The actual form of $\eta_\epsilon(x)$ is irrelevant, as long as it is differentiable any number of times. At the end of the calculations the limit $\epsilon \to 0$ is to be taken, eliminating all pole terms in ϵ (regularization), thus keeping only the finite part. We start with the regularized potential:

$$A_\mu^+ = -2i\sigma_{\mu\nu}x^\nu(x^2-\lambda^2)_\epsilon^{-1} = -2i\sigma_{\mu\nu}x^\nu \frac{\eta_\epsilon(z)}{Z}, \quad \text{(B1)}$$

$$Z = (x^2-\lambda^2). \quad \text{(B2)}$$

We define

$$V = \frac{\eta_\epsilon(Z)}{Z} \quad \text{(B3)}$$

From (B1) we get

$$F_{\mu\nu}^+ = -4i\sigma_{\mu\nu}([Z+\lambda^2]V^2 - V) + 4i(\sigma_{\mu\tau}x_\tau x_\nu - \sigma_{\nu\tau}x^\tau x_\mu)\Big[V^2 + \frac{d}{dZ}V\Big], \quad \text{(B4)}$$

and from Yang–Mills equations,

$$J_\mu^+ = -8i\sigma_{\mu\nu}x^\nu\Big\{\Big[(V^{-1}+\lambda^2)\Big(\frac{d^2}{dZ^2}V - 2V^3\Big)\Big] + 3\Big[V^2 + \frac{d}{dZ}V\Big]\Big\}, \quad \text{(B5)}$$

from which

$$J_\mu^+ = -8i\sigma_{\mu\nu}x^\nu\Big\{\lambda^2\Big[\frac{\eta''_\epsilon}{Z} - 2\frac{\eta'_\epsilon}{Z^2}\Big] + \Big[\frac{\eta'_\epsilon}{Z} + \eta''\Big]\Big\} \quad \text{(B6)}$$

In order to understand the distributions of the second term in the limit $\epsilon \to 0$, we apply, for instance, the square bracket to a function $\Phi(x)$.

$$\Big(\Big[\frac{\eta''_\epsilon}{Z} - 2\frac{\eta'_\epsilon}{Z^2}\Big], \Phi\Big) = \int_\epsilon^\infty \Big\{\frac{d^2}{dZ^2}\frac{\Phi(Z)}{Z} + 2\frac{d}{dZ}\frac{\Phi(Z)}{Z^2}\Big\}dZ$$

$$= -\frac{d}{dZ}\frac{\Phi(Z)}{Z}\Big|_{z=\epsilon}^\infty - 2\frac{\Phi(Z)}{Z^2}$$

$$= -\frac{\Phi'(\epsilon)}{\epsilon^2} + \frac{\Phi(\epsilon)}{\epsilon^2} - 2\frac{\Phi(\epsilon)}{\epsilon^2}$$

(where use has been made definition and properties of η_ϵ). Using $\Phi(\epsilon) = \Phi(0) + \epsilon\Phi'(0) + (\epsilon^2/2)\Phi''(0)$ and dropping the pole terms to get the finite part, one finally gets

$$\mathrm{Pf}\Big(\Big[\frac{\eta''_\epsilon}{Z} - 2\frac{\eta'_\epsilon}{Z^2}\Big], \Phi\Big)_{\epsilon\to 0} = -\frac{3}{2}\Phi''(0). \quad \text{(B7)}$$

with a similar procedure, it is shown that the finite part of the second bracket is zero. So, we get for the current the same result as given in (3.5).

[1] A.A. Belavin, A. Polgakov, A. Schwarz and Y. Tyupking, Phys. Lett. B **59**, 85 (1975).
[2] G. t' Hooft, Phys. Rev. Lett. **37**, 8 (1976).
[3] G. t' Hooft, unpublished.
[4] F. Wilczek, *Quark confinement and Field Theory* (Wiley, New York, 1977).
[5] E. Corrigan and D.B. Fairlie, Phys. Lett. B **67**, 69 (1977).
[6] M. Riesz, Soc. Math. France **66**, (1938).
[7] T. Gustavson, Fysiogr. Sällsk i lund Förh. **15**, No. 28 (1945); **16**, No. 2 (1946).
[8] E.R. Speer, J. Math. Phys. **9**, 404 (1968).
[9] I.M. Guelfand and G.E. Shilov, *Les Distributions* (Dunod, Paris, 1962).
[10] C. Rebbi, BNL 23004, July 1977.
[11] A. González Dominguez, *On some Heterodox Distributional Multiplicative Products* (Publicaciones del Instituto Argentino de Matemática, Buenos Aires, 1978).
[12] S. Sciuto, "Topics on Instantons," CERN preprint, 1977.

Solutions of the Rarita–Schwinger equation in the Kerr–Newman space

S. Einstein and R. Finkelstein
Department of Physics, University of California, Los Angeles, California 90024

(Received 9 February 1979; accepted for publication 4 May 1979)

Solutions of the massless Rarita–Schwinger equation are found in a Kerr–Newman background. These solutions have arbitrary energy and arbitrary angular momentum about the axis of rotation, and in the eikonal limit define orbits that coincide with the principal null congruences. They are simple generalizations of corresponding solutions of the Dirac equations.

1. INTRODUCTION

Supergravity may be characterized in terms of the covariant coupling of a helicity 3/2 field to the gravitational field. Exact solutions of the supergravity equations describe both fields mutually codetermined so as to satisfy the full supergravity field equations. It would be very interesting to find exact solutions of the full equations.

A less ambitious project, which is begun here, is to find exact solutions of the 3/2 field equations in a given gravitational background. This problem is very similar to that of finding exact solutions of the Dirac equation in a given gravitational field. Our more precise goal is to generalize our earlier solutions of the Dirac equation in a Kerr-Schild space[1] to the Rarita–Schwinger 3/2 field in the same space.

Our previous work was based on the observation that the Kerr–Schild metric defines a local rotation group so that it is possible to associate a Dirac spinor with the spin representation of this group. One could of course construct the 3/2 representation in the same way, but that is not the path followed here. Instead, we shall relate this simple procedure to the more systematic Newman–Penrose formalism.

2. KERR–NEWMAN BACKGROUND

The space surrounding a rotating charged mass may be described by the metric

$$g_{\mu\nu} = \eta_{\mu\nu} - 2v l_\mu l_\nu, \tag{2.1}$$

where $\eta_{\mu\nu}$ is the Minkowski metric, v is a scalar function, and l_μ is null and geodesic with respect to both $\eta_{\mu\nu}$ and $g_{\mu\nu}$. The vector l_μ describes the spiral motion of a massless particle moving in the field of the rotating charged mass. There are two classes of these spirals, corresponding to incoming and outgoing motion, and they define the principal null congruences. The corresponding energy–momentum vectors are also the repeated null directions of the Weyl tensor. In the present problem these spiral orbits provide an eikonal limit that is satisfied by an interesting class of quantum mechanical motions.

3. NEWMAN–PENROSE FORMALISM[2]

Let v_a^μ be a null tetrad satisfying the conditions

$$g_{\mu\lambda} v_a^\mu v_b^\lambda = \eta_{ab}, \tag{3.1a}$$

$$\eta_{ab} v_\mu^a v_\lambda^b = g_{\mu\lambda}, \tag{3.1b}$$

where

$$\eta_{ab} = \begin{pmatrix} 0 & -1 & 0 & 0 \\ -1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 \end{pmatrix} \tag{3.2}$$

and $g_{\mu\lambda}$ is the general relativistic metric. A more explicit notation is frequently used, namely

$$v_a^\mu = (m^\mu, \bar{m}^\mu, n^\mu, l^\mu). \tag{3.3}$$

Here l and n are real while m and $\bar{m}$ are complex conjugates. The only nonvanishing scalar products are

$$-m\bar{m} = nl = 1. \tag{3.4}$$

The Ricci coefficients of rotation may be expressed in terms of this null tetrad:

$$\omega_{cab} = v_{b\mu|\nu} v_a^\mu v_c^\nu. \tag{3.5}$$

In the present context these coefficients are referred to as the coefficients of rotation or the Newman–Penrose coefficients. If all tensors are replaced by the invariants that result from projection onto the tetrad, then all the field equations depend only on tetrad indices and become coordinate independent. The resulting equations are then covariant under the six parameter group of rotations of the tetrad. In our problem it is convenient to eliminate the freedom in orienting the tetrad by choosing the n and l directions of the tetrad along the motion of the incoming and outgoing photons. Then four of the NP coefficients vanish. In the usual notation these are[3]

$$\kappa = \sigma = \nu = \lambda = 0. \tag{3.6}$$

By choosing a suitable scale for the vector l one may also make another NP coefficient vanish, namely,

$$\epsilon = 0. \tag{3.7}$$

All of these simplifications are to be understood here.

Introduce the constant Dirac matrices satisfying the relations

$$(\gamma_a, \gamma_b)_+ = 2\eta_{ab} \tag{3.8}$$

and the position dependent matrices:

$$\gamma^\mu = v_a^\mu \gamma^a. \tag{3.9}$$

Then

$$(\gamma^\mu, \gamma^\nu)_+ = 2g^{\mu\nu} = -m^\mu \bar{m}^\nu - \bar{m}^\mu m^\nu + l^\mu n^\nu + n^\mu l^\nu. \tag{3.10}$$

Equations (3.10) are satisfied by the following representa-

 0022-2488/79/091972-05$01.00 © 1979 American Institute of Physics

tion for the γ^μ:

$$\gamma^\mu = \sqrt{2}\begin{bmatrix} 0 & 0 & l^\mu & -m^\mu \\ 0 & 0 & -\bar{m}^\mu & n^\mu \\ n^\mu & m^\mu & 0 & 0 \\ \bar{m}^\mu & l^\mu & 0 & 0 \end{bmatrix}. \qquad (3.11)$$

The charge conjugation matrix will also be required:

$$C = \begin{bmatrix} 0 & 1 & 0 & 0 \\ -1 & 0 & 0 & 0 \\ 0 & 0 & 0 & -1 \\ 0 & 0 & 1 & 0 \end{bmatrix}, \qquad (3.12)$$

with the property

$$C\gamma^\mu C^{-1} = -(\gamma^\mu)^T. \qquad (3.13)$$

Although η_{ab} is not diagonal, γ_5 still anticommutes with all γ_a:

$$(\gamma_5,\gamma_a)_+ = 0, \qquad (3.14)$$

where

$$\gamma_5 = (1/4!)\epsilon^{abcd}\gamma_a\gamma_b\gamma_c\gamma_d. \qquad (3.15)$$

Then also

$$(\gamma_5,\gamma_\alpha)_+ = 0, \qquad (3.16)$$

and

$$\gamma_5 = (1/4!)v\epsilon^{\alpha\beta\lambda\mu}\gamma_\alpha\gamma_\beta\gamma_\lambda\gamma_\mu,$$

where v is $\det(v_a^\alpha)$. The numerical form of γ_5 is

$$\gamma_5 = \begin{pmatrix} 1 & & & \\ & 1 & & \\ & & -1 & \\ & & & -1 \end{pmatrix}.$$

4. THE DIRAC WAVE FUNCTION

The discussion is here limited to massless or very high energy particles. Then the Dirac equation is

$$\gamma^\mu\nabla_\mu\psi = 0, \qquad (4.1)$$

where ∇_μ is the covariant derivative:

$$\nabla_\mu = \partial_\mu + \tfrac{1}{2}\omega_{\mu ab}\sigma^{ab} \qquad (4.2)$$

and $\omega_{\mu ab}$ are the coefficients of rotation. Here

$$\sigma^{ab} = \tfrac{1}{4}(\gamma^a,\gamma^b), \qquad (4.3)$$

so that (4.1) may also be written as

$$(\gamma^\mu\partial_\mu + \tfrac{1}{4}\omega_{\mu\alpha\beta}\gamma^\mu\gamma^\alpha\gamma^\beta)\psi = 0. \qquad (4.4)$$

If the Dirac equation in the Kerr–Newman space is written in the NP formalism, it is separable in Boyer–Lindquist coordinates, as shown by Teukolsky[3] for the massless case, and as shown by Chandrasekhar[4] for the massive case.

The explicit solutions of the massless Dirac equation that we found earlier correspond to those particular motions that approach the null geodesics in the eikonal limit, so that the corresponding fermions move along the same paths as photons while their spins are either parallel or antiparallel to their momentum, just as for photons. These particular solutions were selected by the condition

$$\tau\psi = 0, \qquad (4.5)$$

where

$$\tau = \gamma^\mu l_\mu. \qquad (4.6)$$

In these equations l_μ was the null vector directing the classical motion, and the γ_μ corresponded to the usual diagonal Minkowski metric.

We impose (4.5) again but γ^μ is now given by (3.11). If one writes

$$\psi = \begin{bmatrix} E \\ F \\ G \\ H \end{bmatrix}, \qquad (4.7)$$

then one finds from (4.5) and (3.11) that two components of ψ vanish:

$$E = H = 0. \qquad (4.8)$$

For the two remaining components of ψ one finds by (4.1) the four equations

$$(D + \bar{\epsilon} - \bar{\rho})G = 0, \qquad (4.9a)$$

$$(\bar{\delta} + \bar{\beta} - \bar{\tau})G = 0, \qquad (4.9b)$$

and

$$(D + \epsilon - \rho)F = 0, \qquad (4.10a)$$

$$(\delta + \beta - \tau)F = 0, \qquad (4.10b)$$

where

$$D = l^\mu\partial_\mu, \quad \delta = m^\mu\partial_\mu, \quad \Delta = \eta^\mu\partial_\mu, \quad \bar{\delta} = \bar{m}^\mu\partial_\mu, \qquad (4.11)$$

$$\omega_{124} = \bar{\rho}, \quad \omega_{234} = \alpha + \bar{\beta}, \quad \omega_{434} = \epsilon + \bar{\epsilon},$$
$$\omega_{324} = \bar{\tau}, \quad \omega_{212} = \alpha - \bar{\beta}, \quad \omega_{421} = \epsilon - \bar{\epsilon}. \qquad (4.12)$$

By (4.9) and (4.10) F and $\bar{G}$ satisfy the same differential equations, and it is therefore enough to consider (4.10).

In the Kerr–Newman space the following differential equations also hold if the tetrad is chosed in the way already described.

$$D\rho = \rho^2, \qquad (4.13)$$

$$\delta\rho = \tau\rho. \qquad (4.14)$$

Let

$$F = \rho f. \qquad (4.15)$$

Then with the help of (4.13) and (4.14) one finds

$$Df = 0, \qquad (4.16)$$

$$(\delta + \beta)f = 0. \qquad (4.17)$$

In Kerr coordinates[5]

$$(\partial_t + \partial_r)f = 0, \qquad (4.16a)$$

$$\left(ia\sin\theta\frac{\partial}{\partial t} + \frac{\partial}{\partial\theta} + \frac{i}{\sin\theta}\frac{\partial}{\partial\phi}\right)f + \tfrac{1}{2}\cot\theta f = 0 \qquad (4.17a)$$

leading to the partial wave solution

$$F(t,r,\theta,\phi) \sim e^{ikt}\frac{e^{-ik\omega}}{\omega}\left(\frac{1-\sigma/a}{1+\sigma/a}\right)^{(1/2)m} \times \frac{e^{im\phi}}{(1-\sigma^2/a^2)^{1/4}} \qquad (4.18)$$

and if constants, k and m are chosen the same for F and G, then

$$G = \bar{F}. \tag{4.19}$$

Here

$$\omega = r - i\sigma, \tag{4.20a}$$

$$\sigma = a\cos\theta, \tag{4.20b}$$

where ω is the complex distance[1] whose real and imaginary parts are radial and angular coordinates.

The complete solution is the usual sum of partial waves.

5. THE DIRAC CURRENT AND THE EIKONAL LIMIT

The equation that is charge conjugate to (4.1) may be written in the form

$$\tilde{\psi}\overleftarrow{\nabla}_a\gamma^a = 0, \tag{5.1}$$

where

$$\overleftarrow{\nabla}_a = \overleftarrow{\partial}_a - \tfrac{1}{2}\omega_{abc}\sigma^{bc} \tag{5.2}$$

$$\tilde{\psi} = \psi C. \tag{5.3}$$

Define the current

$$S^a = \tilde{\psi}\gamma^a\psi. \tag{5.4}$$

Then

$$\mathbf{D}_a S^a \equiv \partial_a S^a + \omega^a_{ab} S^b = 0, \tag{5.5}$$

where

$$\omega^a_{ab} = v^\mu_{b\,|\mu} = (1/\sqrt{-g})\partial_\mu(\sqrt{-g}v^\mu_b). \tag{5.6}$$

This is the usual conservation of current:

$$\mathbf{D}_a S^a = (1/\sqrt{-g})\partial_\mu(\sqrt{-g}S^\mu) = 0. \tag{5.7}$$

One then finds by (5.4) the current associated with one partial wave:

$$S^\mu = \sqrt{2}(2FG)l^\mu = 2\sqrt{2}|F|^2 l^\mu, \tag{5.8}$$

and by (4.18) and (4.19)

$$S^\circ = 2\sqrt{2}|F|^2 l^\circ = 2\sqrt{2}\,\frac{e^{-2k\sigma}}{|\omega|^2}\left(\frac{1-\sigma/a}{1+\sigma/a}\right)^m \times \frac{1}{(1-\sigma^2/a^2)^{1/2}}\,l^\circ. \tag{5.9}$$

The invariant volume element is determined by

$$\sqrt{-g} = |\omega|^2\sin\theta, \tag{5.10}$$

and therefore

$$S^\circ\sqrt{-g} \sim e^{-2k\sigma}\left(\frac{1-\sigma/a}{1+\sigma/a}\right)^m. \tag{5.11}$$

By maximizing this expression with respect to σ and passing to the limit of high angular momentum ($m\hbar$) one obtains the classical paths

$$L_z = -ak\sin^2\theta \tag{5.12}$$

as previously noted.[1,6]

6. RARITA–SCHWINGER EQUATION

This differential equation may be written in the following form[7]

$$\epsilon^{\lambda\mu\rho\sigma}\gamma_5\gamma_\mu\mathscr{D}_\rho\psi_\sigma = 0, \tag{6.1}$$

where

$$\mathscr{D}_\rho\psi_\sigma = (\partial_\rho + \tfrac{1}{2}\omega_{\rho mn}\sigma^{mn})\psi_\sigma \tag{6.2}$$

and γ_5 is still given by (3.15). Equation (6.1) may be rewritten in various forms depending on identities like

$$\epsilon^{\lambda\mu\rho\sigma}\gamma_\lambda\gamma_5 = (\sigma^{\mu\rho},\gamma^\sigma)_+ = \gamma^\mu g^{\rho\sigma} - \gamma^\rho g^{\mu\sigma} - \gamma^\sigma(g^{\mu\rho} - \gamma^\rho\gamma^\mu). \tag{6.3}$$

With the use of (6.3) one finds

$$(g^{\rho\sigma} - \gamma^\rho\gamma^\sigma)\mathscr{D}_\rho\psi_\sigma = 0 \tag{6.4}$$

and

$$\gamma^\rho(\mathscr{D}_\sigma\psi_\rho - \mathscr{D}_\rho\psi_\sigma) = 0. \tag{6.5}$$

It will be convenient to use this last form.[7]

7. SOLUTIONS OF THE VECTOR-SPINOR EQUATION

The procedure is the same as for the Dirac equation. One imposes the analogue of (4.5):

$$\tau\psi_k = 0, \tag{7.1}$$

where τ is given by (4.6). If one writes

$$\psi_k = \begin{pmatrix} E_k \\ F_k \\ G_k \\ H_k \end{pmatrix}, \tag{7.2}$$

then one finds

$$E_k = H_k = 0. \tag{7.3}$$

The differential equation (6.5) then becomes

$$2\gamma^a\partial_{[a}\psi_{b]} = [\omega_{abd} - \omega_{bad}]\gamma^a\psi^d + \tfrac{1}{4}\gamma^c\gamma^a\gamma^d \times(\omega_{cad}\psi_b - \omega_{bad}\psi_c) = 0. \tag{7.4}$$

In component form half of these equations may be expressed as follows:

$$DG_1 - \delta G_4 - \sigma G_2 - (\epsilon - 2\bar{\epsilon} + \bar{\rho})G_1 + \kappa G_3 + (\beta - \bar{\pi})G_4 = 0, \tag{7.4a}$$

$$\delta G_2 - \bar{\delta}G_1 + \beta G_2 + (\alpha - 2\bar{\beta} + \bar{\tau})G_1 - \rho G_3 - (\mu - \bar{\mu})G_4 = 0, \tag{7.4b}$$

$$-\tau F_1 + \sigma F_3 = 0, \tag{7.4c}$$

$$DF_1 - \delta F_4 + (\bar{\epsilon} - \rho - \bar{\rho})F_1 + \kappa F_3 + (\bar{\alpha} - \bar{\pi})F_4 = 0, \tag{7.4d}$$

$$DG_3 - \Delta G_4 - \pi G_1 - (\tau + \bar{\pi})G_2 + (2\bar{\epsilon} + \epsilon - \bar{\rho})G_3 + \gamma G_4 = 0, \tag{7.4e}$$

$$\Delta G_2 - \bar{\delta}G_3 + \lambda G_1 + (\bar{\mu} + \gamma)G_2 - (\alpha + 2\bar{\beta} - \bar{\tau})G_3 - \nu G_4 = 0, \tag{7.4f}$$

$$DG_2 - \bar{\delta}G_4 - \bar{\sigma}G_1 + (\epsilon - \rho)G_2 + (\alpha - \pi + \bar{\tau})G_4 = 0, \tag{7.4g}$$

$$\bar{\kappa}G_1 - \bar{\rho}G_4 = 0. \tag{7.4h}$$

The remaining equations are obtained by complex conjugation and the substitution $(G_1G_2G_3G_4) \leftrightarrows (\bar{F}_2\bar{F}_1\bar{F}_3\bar{F}_4)$.

We are everywhere assuming that l^μ is a geodesic vector field.

$$l_{\mu|\nu} l^{\nu} = 0, \quad \text{or} \quad \omega_{4a4} = 0. \tag{7.5}$$

Therefore,

$$\kappa = \epsilon + \bar{\epsilon} = 0. \tag{7.6}$$

Then

$$F_4 = G_4 = 0 \tag{7.7a}$$

by (7.4h) and its complex conjugate. This last equation may also be expressed as follows:

$$l_\alpha \psi^\alpha = 0. \tag{7.7b}$$

Therefore, ψ^α as a vector is orthogonal to l^α. This statement complements the original condition (4.5) of spinor transversality.

If one assumes that the background space has vanishing shear then

$$\sigma = 0. \tag{7.8}$$

It is also possible to scale the tetrad vector l^μ so that $\epsilon = 0$ as in (3.7). Then just by virtue of (7.6), (7.8), and (3.7), and without using (3.6), one finds first that[8]

$$F_1 = G_2 = 0. \tag{7.9}$$

Then (7.4) reduces to the following set:

$$DG_1 = \bar{\rho} G_1, \tag{7.10a}$$

$$\bar{\delta} G_1 = (\alpha - 2\bar{\beta} + \bar{\tau}) G_1 - \rho G_3, \tag{7.10b}$$

$$DG_3 = \pi G_1 + \bar{\rho} G_3, \tag{7.10c}$$

$$\bar{\delta} G_3 = \lambda G_1 - (\alpha + 2\bar{\beta} - \bar{\tau}) G_3. \tag{7.10d}$$

The four remaining equations satisfied by (F_2, F_3) are just the same as the four equations satisfied by $(\bar{G}_1, \bar{G}_3)$ and it is therefore enough to discuss the G equations.

When the background space is finally restricted to be Kerr–Newman there are additional simplifications, namely,

$$\lambda = 0, \tag{7.11}$$

$$\pi = \alpha + \bar{\beta}, \tag{7.12}$$

$$\rho \bar{\pi} = \bar{\rho} \tau. \tag{7.13}$$

The G equations then become

$$DG_1 = \bar{\rho} G_1, \tag{7.14a}$$

$$DG_3 = \pi G_1 + \bar{\rho} G_3, \tag{7.14b}$$

$$\bar{\delta} G_1 = (\pi + \bar{\tau} - 3\bar{\beta}) G_1 - \rho G_3, \tag{7.14c}$$

$$\bar{\delta} G_3 = -(\pi - \bar{\tau} + \bar{\beta}) G_3. \tag{7.14d}$$

In addition to (4.13) and (4.14) we shall make use of the following differential equations connecting the spin coefficients:

$$D\tau = \rho(\tau + \bar{\pi}) \tag{7.15}$$

$$\delta \tau = \tau(\tau + 2\beta - \bar{\pi}). \tag{7.16}$$

These relations depend of course on the special Kerr–Newman geometry. To solve (7.14) set

$$G_1 = \bar{\rho} g_1, \tag{7.17}$$

$$G_3 = \bar{\tau} g_3. \tag{7.18}$$

One then finds

$$Dg_1 = 0, \tag{7.19a}$$

$$Dg_3 = \rho(g_1 - g_3), \tag{7.19b}$$

$$\bar{\delta} g_1 = (\pi - 3\bar{\beta}) g_1 - \pi g_3, \tag{7.19c}$$

$$\bar{\delta} g_3 = -3\bar{\beta} g_3. \tag{7.19d}$$

Consider the special solutions

$$g_1 = g_3 = \bar{y}. \tag{7.20}$$

Then

$$Dy = 0, \tag{7.20a}$$

$$(\delta + 3\beta) y = 0. \tag{7.20b}$$

These equations are of the same form as (4.16) and (4.17) and therefore the solutions are also. One finds the partial wave solution

$$y = A e^{ik(t-r)} e^{im\phi} e^{-ak\cos\theta} \left(\frac{1 - \cos\theta}{1 + \cos\theta} \right)^{(1/2)m} \times \left(\frac{1}{\sin\theta} \right)^{3/2}. \tag{7.21}$$

Then

$$G_1 = \bar{\rho} \bar{y}, \tag{7.22}$$

$$G_3 = \bar{\tau} \bar{y}, \tag{7.23}$$

where

$$\rho = -\omega^{-1} = -(r - ia\cos\theta)^{-1}, \tag{7.24}$$

$$\tau = -\frac{ia\sin\theta}{\sqrt{2}\,|\omega|^2}. \tag{7.25}$$

In addition to the special solutions just found there are also solutions in which the functions g_1 and g_3 are different. Introduce

$$h = \rho(g_1 - g_3). \tag{7.26}$$

Then h satisfies the homogeneous equations

$$Dh = 0, \tag{7.27a}$$

$$(\bar{\delta} + 3\bar{\beta}) h = 0, \tag{7.27b}$$

while g_1 and g_3 are determined in terms of h with the aid of the following set:

$$Dg_1 = 0, \tag{7.28a}$$

$$(\bar{\delta} + 3\bar{\beta}) g_1 = (\pi/\rho) h, \tag{7.28b}$$

$$Dg_3 = h, \tag{7.29a}$$

$$(\bar{\delta} + 3\bar{\beta}) g_3 = 0. \tag{7.29b}$$

The solution of these equations may be expressed in the following form

$$g_1 = c_1 h + g_1^s, \tag{7.30}$$

$$g_3 = c_3 h + g_3^s, \tag{7.31}$$

where

$$g_1^s = ia[\cos\theta] h, \tag{7.30a}$$

$$g_3^s = rh. \tag{7.31b}$$

By (7.26), (7.30), and (7.31)

$$\rho^{-1} h = (c_1 - c_3) h + (ia\cos\theta - r) h. \tag{7.32}$$

This condition may be satisfied by

$$h = 0 \tag{7.33}$$

which is the case already considered, or by

$$\rho^{-1} = c_1 - c_3 + ia\cos\theta - r$$

which implies

$$c_1 = c_3. \tag{7.34}$$

Therefore, if $h \neq 0$, the solution is

$$g_1 = (c + ia\cos\theta)h, \tag{7.35}$$

$$g_3 = (c + r)h, \tag{7.36}$$

where

$$h = \bar{y} \tag{7.37}$$

as before.

An alternative form of y is

$$y = Ae^{ikt}e^{-ik\omega}\left(\frac{1-\sigma/a}{1+\sigma/a}\right)^{(1/2)m}\frac{e^{+im\phi}}{(1-\sigma^2/a^2)^{3/4}}. \tag{7.38}$$

Let

$$\mathscr{Y}_S = Ae^{ikt}e^{-ik\omega}\left(\frac{1-\sigma/a}{1+\sigma/a}\right)^{(1/2)m}\frac{e^{+im\phi}}{(1-\sigma^2/a^2)^{(1/2)S}}, \tag{7.39}$$

where S is the spin. Then the solution of the Dirac equation is a spinor with two nonvanishing components which depend on a single function:

$$F = \bar{G} \sim \omega^{-1}\mathscr{Y}_{1/2}$$
$$E = H = 0. \tag{7.40}$$

The corresponding solution of the Rarita–Schwinger equation is a vector spinor with four nonvanishing components which depend on two functions with different radial and angular dependence. We distinguish two cases:

$$h = 0,$$
$$F_2 = \bar{G}_1 = \rho\mathscr{Y}_{3/2} \sim -\omega^{-1}\mathscr{Y}_{3/2},$$
$$F_3 = \bar{G}_3 = \tau\mathscr{Y}_{3/2} \sim -\frac{ia\sin\theta}{\sqrt{2}\,|\omega|^2}\mathscr{Y}_{3/2}, \tag{7.41a}$$
$$E_k = H_k = F_1 = G_2 = F_4 = G_4 = 0,$$

$$h \neq 0,$$
$$F_2 = \bar{G}_1 = \rho[c - ia\cos\theta]\mathscr{Y}_{3/2},$$
$$F_3 = \bar{G}_3 = \tau[c + r]\mathscr{Y}_{3/2}, \tag{7.41b}$$
$$E_k = H_k = F_1 = F_2 = F_4 = G_4 = 0.$$

8. CURRENT AND EIKONAL LIMIT OF 3/2 SOLUTION

One cannot give a completely satisfactory physical discussion of these solutions since the 3/2 field does not in general propagate causally unless it is coupled to the gravitational field precisely as it is in supergravity.[9] Since we are not discussing exact solutions of the full supergravity equations, however, that case is excluded here. Nevertheless, one may define a formal current that is exactly conserved for the free 3/2 equation and approximately conserved in weak gravitational fields, namely,

$$S^\mu = \epsilon^{\mu\lambda\alpha\beta}\bar{\psi}_\lambda\gamma_5\gamma_\alpha\psi_\beta. \tag{8.1}$$

Then

$$\partial_\mu S^\mu = \omega_{\mu ab}\epsilon^{\mu\lambda b\rho}\bar{\psi}_\lambda\gamma_5\gamma^a\psi_\rho \tag{8.2}$$

and $\partial_\mu S^\mu$ vanishes at large distances since $\omega_{\mu ab} = 0(1/r)$. One then finds that the nonvanishing component of S^μ is

$$S^l \sim |G_1|^2 \sim |\rho|^2|y|^2 \sim \frac{e^{-2k\sigma}}{r^2 + a^2\cos^2\theta}\left(\frac{1-\sigma/a}{1+\sigma/a}\right)^m \times\left(\frac{1}{1-\sigma^2/a^2}\right)^{3/2}. \tag{8.3}$$

The current is directed along l and

$$S^l\sqrt{-g} \sim e^{-2k\sigma}\left(\frac{1-\sigma/a}{1+\sigma/a}\right)^m\frac{1}{(1-\sigma^2/a^2)}. \tag{8.4}$$

In the limit of high angular momentum and energy one obtains again the classical paths (5.12).

[1]S. Einstein and R. Finkelstein, J. Math. Phys. **18**, 664 (1977).
[2]E. Newman and R. Penrose, J. Math. Phys. **3**, 566 (1962).
[3]S. Teukolsky, Astrophys. J. **185**, 635 (1973).
[4]S. Chandrasekhar, Proc. R. Soc. London Ser. A **349**, 571 (1976).
[5]S.K. Bose, J. Math. Phys. **16**, 772 (1975).
[6]S. Einstein and R. Finkelstein, Phys. Rev. **15**, 2721 (1977).
[7]D. Freedman and P. van Nieuwenhuizen, Phys. Rev. **14**, 912 (1976).

[8]In view of (7.3) Eqs. (7.7a) and (7.9) may be summarized simply as $\gamma\cdot\psi = 0$.

[9]S. Deser and B. Zumino, Phys. Rev. Lett. **38**, 1433 (1977); G. Velo and D. Zwanziger, Phys. Rev. **186**, 1337 (1969). Solutions of the Rarita–Schwinger equation in an external electromagnetic potential may propagate faster than light. When the vector spinor is part of the supergravity field, however, there is no anomalous propagation.

Construction of Lie algebras and Lie superalgebras from ternary algebras[a)]

Itzhak Bars[b)]

Lyman Laboratory of Physics, Harvard University, Cambridge, Massachusetts 02138 and J. W. Gibbs Laboratory of Physics, Yale University, New Haven, Connecticut 06520

Murat Günaydin

Department of Physics, University of Bonn, Nussallee 12, D-53 Bonn, West Germany

(Received 6 December 1978)

Ternary algebras are algebras which close under suitable triple products. They have been shown to be building blocks of ordinary Lie algebras. They may acquire a deep physical meaning in fundamental theories given the important role played by Lie (super) algebras in theoretical physics. In this paper we introduce the concept of superternary algebras involving Bose and Fermi variables. Using them as building blocks, we give a unified construction of Lie algebras and superalgebras in terms of (super) ternary algebras. We prove theorems that must be satisfied for the validity of this construction, which is a generalization of Kantor's results. A large number of examples and explicit constructions of the Lie algebras A_n, B_n, C_n, D_n, F_4, E_6, E_7, E_8, and Lie superalgebras $A(m,n)$, $B(m,n)$, $D(m,n)$, $P(n)$, $Q(n)$ are given. We speculate on possible physical applications of (super) ternary algebras.

I. INTRODUCTION

Lie groups have found extensive applications in physics and play a basic role in describing the symmetries of nature. They are essential in the construction of fundamental theories of particle interactions and via the principle of local gauge invariance they lead to unified gauge theories of elementary particles.[1] The recent development of supersymmetry and supergravity[2] based on Lie superalgebras[3] may eventually lead to a unified theory of all interactions including gravity. Despite the partial success of these programs there remain problems of fundamental nature whose resolution may require new mathematical avenues which have not yet found their way into physics.

Ternary algebras is one of the more recent mathematical constructs which have not yet found applications in physics. These are algebras that close under triple products. They have been shown to be building blocks of ordinary Lie algebras. In this paper our aim is twofold: first, to familiarize physicists with the concept of ternary algebras and their role in the construction of ordinary Lie algebras; second, to introduce the concept of superternary algebras and use them as building blocks of Lie superalgebras. To our knowledge the concept of superternary algebras and our construction are new in the mathematical literature. Given the importance of Lie algebras in physics and the fundamental role played by the ternary algebras, as explained in this paper, it is not inconceivable that they may acquire a deep physical meaning in future theories and help resolve some of the unsolved problems of theoretical physics.

As mentioned above a ternary algebra is an algebra that closes under triple products. Of course, all algebras that close under double products will close trivially under triple products. In addition there are algebras that close only under triple products. For example, pure imaginary numbers, fermionic variables, column and row matrices, etc., form such algebras. More generally one can consider the triple product as a mapping of a vector space V in the form

$$V\otimes V\otimes V\rightarrow V$$

Essentially all of the familiar mathematical structures that have found applications in physics can be given the structure of a ternary algebra under a suitable triple product. Here we give a few examples: Real and complex numbers, quaternions, octonions, vectors, tensors, spinors, fermions, gamma matrices, discrete groups, rectangular and square matrices over real, complex and quaternionic commuting or anticommuting fields, Lie algebras, Jordan algebras, direct products of any of these systems with each other, etc., etc. In this paper we give examples that involve most of these.

The general outline of the paper is as follows: In Sec. II we give a *unified* construction of Lie algebras and Lie superalgebras by using ternary algebras as building blocks. Our treatment is a generalization to superalgebras of the methods used by Kantor[4,5] in constructing ordinary Lie algebras. His methods are generalizations of the Tits, Koecher, and Faulkner constructions[6] of Lie algebras. We prove two theorems. In the first theorem we establish the conditions imposed by the structure of the Lie superalgebra on the underlying ternary algebra. In the second theorem we prove that the re-

a)Research supported in part by the U.S. Dept. of Energy under Contract No. EY-76-C-02-3075, and the National Science Foundation under Grant No. PHY77-22864.

b)A.P. Sloan Foundation Fellow. On leave from Yale University on a Junior Faculty Fellowship.

 © 1979 American Institute of Physics

quirements of the first theorem are satisfied by a very large class of ternary algebras which we call "associative" (defined later).

In Sec. III we give the construction of all ordinary Lie algebras, including the exceptional ones, by the method of ternary algebras. This includes the famous magic square as well. In Sec. IV we consider ternary algebras with purely fermionic parameters and build the classical superalgebras of type $A(m,n)$, $B(m,n)$, $D(m,n)$. In Sec. V we study ternary algebras which include both Fermi and Bose parameters and construct the Lie superalgebras $P(n)$. In Sec. VI we treat ternary algebras defined via Jordan multiplication (anticommutators). The elements of these ternary algebras belong to Jordan superalgebras.[7] We give explicit constructions for $A(m,n)$, $D(2m,n)$, $P(2n-1)$, $Q(2n-1)$. The construction of the remaining exceptional superalgebras $D(2,1;\alpha)$, $G(3)$, and $F(4)$ as well as the Cartan superalgebras W_n, S_n, $\widetilde{S}_n$, H_n will be given elsewhere.[8]

We conclude with a discussion of our results, open problems and possible physical applications. The Appendix contains the detailed proofs of Theorems 1 and 2.

Throughout the paper we use the notation of Cartan and Kac to denote Lie algebras or superalgebras, and we do not distinguish between compact and noncompact groups.

II. GENERAL FORMALISM

We consider a Lie algebra or Lie superalgebra. We will always combine infinitesimal parameters with generators. A Lie algebra has only commuting (Bose) parameters, while a Lie superalgebra has both commuting (Bose) and anticommuting (Fermi) parameters. In such a formulation we only need to consider commutators. Anticommutators of generators arise when we take the commutator of two fermionic generators G_α and G_β, together with their anticommuting parameters $\theta_{1\alpha}$ and $\theta_{2\beta}$, in the form $L_1 = \theta_{1\alpha}G_\alpha$ and $L_2 = \theta_{2\beta}G_\beta$.

$$\begin{aligned}[L_1,L_2] &= \theta_{1\alpha}\theta_{2\beta}G_\alpha G_\beta - \theta_{2\beta}\theta_{1\alpha}G_\beta G_\alpha \\ &= \theta_{1\alpha}\theta_{2\beta}(G_\alpha G_\beta + G_\beta G_\alpha) = \theta_{1\alpha}\theta_{2\beta}\{G_\alpha,G_\beta\}.\end{aligned} \tag{2.1}$$

We will denote bosonic parameters by ω_a, bosonic generators by G_a and the combination by $L = \omega_a G_a$ just as for the fermionic counterparts. Then both Lie algebras and Lie superalgebras are defined only through commutators.

$$\begin{gathered}[L_i,L_j] = C_{ij}^{\,k} L_k\,, \\ [[L_i,L_j],L_k] + [[L_j,L_k],L_i] + [[L_k,L_i],L_j] = 0\,,\end{gathered} \tag{2.2}$$

where $C_{ij}^{\,k}$ depends on the parameters and is always antisymmetric. Note that these equations reduce to the usual definition[3] of a Lie algebra or Lie superalgebra in terms of generators if the parameters are factored out in a specified order.

Grading: Everys simple Lie algebra or Lie superalgebra can be given a graded structure. This can be done by choosing an appropriate U(1) subgroup with respect to which the generators can be assigned definite quantum numbers. The U(1) quantum number is additive modulo the dimension of the grading. Thus, denoting the set of elements with quantum number k by $\mathscr{U}_k$ we can group the generators of the algebra $\mathscr{L} = \{L\}$ in the form

$$\mathscr{U}_{-N} \oplus \cdots \oplus \mathscr{U}_{-1} \oplus \mathscr{U}_0 \oplus \mathscr{U}_1 \oplus \cdots \oplus \mathscr{U}_N\,. \tag{2.3}$$

Because of the additive quantum number the commutators should close as

$$[\mathscr{U}_m, \mathscr{U}_n] \subset \mathscr{U}_{m+n}\,, \tag{2.4}$$

where $-N \leqslant m+n \leqslant N$ and N is the largest U(1) quantum number allowed in the algebra. For example, consider Sl(2) and its subalgebra U(1). Take the 2×2 representation of Sl(2) in the form

$$\begin{pmatrix} \frac{1}{2}a & b \\ c & -\frac{1}{2}a \end{pmatrix} = aG_0 + bG_+ + cG_-\,, \tag{2.5}$$

where

$$G_0 = \begin{pmatrix} \frac{1}{2} & \\ & -\frac{1}{2} \end{pmatrix},\quad G_+ = \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix},\quad G_- = \begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix}. \tag{2.6}$$

Then, as is well known $G_\pm$ have quantum numbers ± 1 with respect to G_0:

$$[G_0, G_\pm] = \pm G_\pm\,. \tag{2.7}$$

Thus, we identify $\mathscr{U}_{-1} = cG_-$, $\mathscr{U}_0 = aG_0$, $\mathscr{U}_1 = bG_+$.

As a second example, consider the Lie algebra SU(3) which can be expanded in terms of Gell-Mann's λ-matrices. Here, there are two choices for the U(1) subalgebra. If one takes twice the isospin generator $\lambda_3 = (^1\ {}_{-1}\ {}_0)$. Then one gets a five-dimensional grading with respect to the third component of isospin. Thus we can identify

$$\mathscr{U}_{-2} = \tfrac{1}{2}(\lambda_1 - i\lambda_2)\pi^- = \begin{pmatrix} 0 & 0 & 0 \\ \pi^- & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix},$$

$$\mathscr{U}_{-1} = \tfrac{1}{2}(\lambda_6 + i\lambda_7)K^0 + \tfrac{1}{2}(\lambda_4 - i\lambda_5)K^- = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & K^0 \\ K^- & 0 & 0 \end{pmatrix},$$

$$\mathscr{U}_0 = \lambda_3\pi^0 + \lambda_8\eta = \begin{bmatrix} \pi^0 + \frac{1}{(3)^{1/2}}\eta & 0 & 0 \\ 0 & -\pi^0 + \frac{1}{(3)^{1/2}}\eta & 0 \\ 0 & 0 & -\frac{2}{(3)^{1/2}}\eta \end{bmatrix},$$

$$\mathscr{U}_1 = \tfrac{1}{2}(\lambda_4 + i\lambda_5)K^+ + \tfrac{1}{2}(\lambda_6 - i\lambda_7)\bar{K}^0 = \begin{pmatrix} 0 & 0 & K^+ \\ 0 & 0 & 0 \\ 0 & \bar{K}^0 & 0 \end{pmatrix},$$

$$\mathscr{U}_2 = \tfrac{1}{2}(\lambda_1 + i\lambda_2)\pi^+ = \begin{pmatrix} 0 & \pi^+ & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}. \tag{2.8}$$

If one takes for the U(1) subalgebra the hypercharge $(1/3^{1/2})\lambda_8$, one gets a 3 dimensional grading. Thus we can identify

$$\mathscr{U}_{-1} = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ K^- & \bar{K}^0 & 0 \end{pmatrix},$$

$$\mathscr{U}_0 = \begin{bmatrix} \pi^0 + \frac{\eta}{(3)^{1/2}} & \pi^+ & 0 \\ \pi^- & -\pi^0 + \frac{\eta}{(3)^{1/2}} & 0 \\ 0 & 0 & -\frac{2\eta}{(3)^{1/2}} \end{bmatrix}, \tag{2.9}$$

$$\mathscr{U}_1 = \begin{pmatrix} 0 & 0 & K^+ \\ 0 & 0 & K^0 \\ 0 & 0 & 0 \end{pmatrix},$$

and there is no $\mathscr{U}_{\pm 2}$ space. In the above, the matrices multiplying the parameters π, η, K represent generators with definite hypercharge quantum numbers.

Similar decompositions are clearly possible for superalgebras. For example, consider the superalgebra OSp(1,2). In the three-dimensional representation it can be written

$$M = \begin{pmatrix} a & b & \theta \\ c & -a & \xi \\ -\xi & \theta & 0 \end{pmatrix}, \tag{2.10}$$

where θ, ξ are anticommuting parameters (Fermi) and a, b, c are commuting parameters (Bose). Then the matrices M close under *commutation* (no anticommutators). This algebra can be given a graded structure by choosing as the U(1) subalgebra

$$G_0 = \begin{pmatrix} 1 & & \\ & -1 & \\ & & 0 \end{pmatrix}. \tag{2.11}$$

Then we can identify

$$\mathscr{U}_{-2} = \begin{pmatrix} 0 & 0 & 0 \\ c & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}, \quad \mathscr{U}_2 = \begin{pmatrix} 0 & b & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix},$$

$$\mathscr{U}_{-1} = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & \xi \\ -\xi & 0 & 0 \end{pmatrix}, \quad \mathscr{U}_1 = \begin{pmatrix} 0 & 0 & \theta \\ 0 & 0 & 0 \\ 0 & \theta & 0 \end{pmatrix}, \tag{2.12}$$

$$\mathscr{U}_0 = \begin{pmatrix} a & 0 & 0 \\ 0 & -a & 0 \\ 0 & 0 & 0 \end{pmatrix}.$$

If the parameters are *factored out* in a definite order, then the generators in $\mathscr{U}_{-1}$ and $\mathscr{U}_1$ close into the others under *anticommutation*.

The last example illustrates the usefulness of keeping the anticommuting parameters always combined with the generators so that we can give a unified treatment only in terms of commutators.

It is clear that because of the grading, the complete algebra is generated by commuting the elements $\mathscr{U}_1(a) = a^i G^i_{+1}$ and $\mathscr{U}_{-1}(b) = b^j G^j_{-1}$. Thus,

$$\begin{aligned} &\mathscr{U}_0(a,b) \supset [\mathscr{U}_1(a), \mathscr{U}_{-1}(b)], \\ &\mathscr{U}_2(a,b) \supset [\mathscr{U}_1(a), \mathscr{U}_1(b)], \\ &\mathscr{U}_3(a,b,c) \supset [\mathscr{U}_1(a), [\mathscr{U}_1(b), \mathscr{U}_1(c)]], \quad \text{etc.} \end{aligned} \tag{2.13}$$

In the above set of parameters a^i (similarly b^i, c^i) can be purely bosonic or purely fermionic or some bosonic and some fermionic. For all cases $\{\mathscr{U}_{\pm k}\}$ close under commutators only.

Conjugation: In such a graded algebra one can introduce a conjugation which maps the $\mathscr{U}_{+1}$ to the $\mathscr{U}_{-1}$ space. The conjugation is defined on the generators which act on some representation space:

$$\widetilde{G}^i_{+1} = G^i_{-1}, \quad \widetilde{G}^i_{-1} = G^i_{+1}. \tag{2.14}$$

We will distinguish between two possible conjugations. We shall call a conjugation of first kind if it changes the order of operators in a product.

$$(G_i \cdots G_j G_k)^{\sim} = \widetilde{G}_k \widetilde{G}_j \cdots \widetilde{G}_i \quad \text{(first kind)}. \tag{2.15}$$

Hermitian conjugation of matrices is such an example. We shall call a conjugation of second kind if it preserves the order of operators (e.g., complex conjugation):

$$(G_i G_j \cdots G_k)^{\sim} = \widetilde{G}_i \widetilde{G}_j \cdots \widetilde{G}_k \quad \text{(second kind)}. \tag{2.16}$$

The parameters a_i, b_j, c_k, etc. multiplying the generators, whether they are bosonic or fermionic, are left unchanged by either one of these conjugations. This implies that under a conjuation of the second kind we have, e.g.,

$$(\mathscr{U}_1(a)\mathscr{U}_{-1}(b))^{\sim} = \mathscr{U}_{-1}(a)\mathscr{U}_1(b), \quad \text{(second kind)}. \tag{2.17}$$

But, under a conjugation of first kind when the parameters are reordered to be assembled back with the generators a phase (± 1) will arise from permuting Fermi parameters. For example

$$\begin{aligned} (\mathscr{U}_1(a)\mathscr{U}_{-1}(b))^{\sim} &= a_i b_j (G^i_{+1} G^j_{-1})^{\sim} = a_i b_j G^j_{+1} G^i_{-1} \\ &= \pm b_j a_i G^j_{+1} G^i_{-1} \\ &= \pm \mathscr{U}_1(b)\mathscr{U}_{-1}(a), \quad \text{(first kind)}, \end{aligned} \tag{2.18}$$

where the (-1) is present only when *both* a_i and b_j are anticommuting parameters. We see that, given the grading, the complete (super) algebra is generated only from $\mathscr{U}_1(a)$ and the conjugation operation.

Five-dimensional grading and construction of Lie (super) algebras: In this article we will consider five-dimensional gradings of Lie algebras and Lie superalgebras. As a special case, we will obtain the three-dimensional gradings. Kantor gave a formulation of ordinary Lie algebras with such gradings. The main tool in his formulation is the use of ternary algebras. These are algebras that close under a triple product. Here we shall present a systematic approach more familiar to physicists, by which we obtain his results and simultaneously give new constructions of Lie superalgebras. In our construction of Lie superalgebras we introduce super ternary algebras in terms of super triple products which to our knowledge is new in the literature.

We argued above that it is always possible to construct the full graded (super) algebra from the $\mathscr{U}_{+1}$ space. The construction proceeds by commuting the $\mathscr{U}_{\pm 1}$ spaces and imposing Jacobi identities at each step. This is an abstract construction in which no explicit representations of $\mathscr{U}_{\pm 1}$ are used. The net result is a derivation of the structure constants of the Lie (super)algebra in terms of triple products. The concept of triple products arises naturally in this process. The Jacobi identities for $\mathscr{L}$ will be automatically satisfied if the ternary algebra satisfies just two conditions. We will present this result in the form of two theorems. The first theorem establishes the restrictions on the ternary algebras so that the resulting algebra $\mathscr{L}$ is a Lie (super)algebra satisfying Jacobi identities. This is a generalization of Kantor's results to include Lie superalgebras. Our second theorem specifies an infinite class of ternary algebras that satisfy the conditions of the first theorem.

From now on we will denote the elements of $\mathscr{L}$ as follows:

$$\begin{aligned}&\mathscr{U}_{-2}(a,b)=\widetilde{K}_{ab},\quad \mathscr{U}_{-1}(a)=\widetilde{\mathscr{U}}_a,\\&\mathscr{U}_0(a,b)=S_{ab},\\&\mathscr{U}_1(a)=\mathscr{U}_a,\quad \mathscr{U}_2(a,b)=K_{ab}.\end{aligned}\tag{2.19}$$

So that we can write $\mathscr{L}=\widetilde{K}\oplus\widetilde{\mathscr{U}}\oplus S\oplus\mathscr{U}\oplus K$. Where S_{ab}, K_{ab},$\widetilde{K}_{ab}$ are obtained by commuting $\mathscr{U}_a$, $\widetilde{\mathscr{U}}_b$:

$$S_{ab}=[\mathscr{U}_a,\widetilde{\mathscr{U}}_b],\quad K_{ab}=[\mathscr{U}_a,\mathscr{U}_b],\quad \widetilde{K}_{ab}=[\widetilde{\mathscr{U}}_a,\widetilde{\mathscr{U}}_b].\tag{2.20}$$

Henceforth, we shall *define* the sets $\mathscr{U}_{0,\pm 2}$ only by elements obtained by commuting $\mathscr{U}_{\pm 1}$. Our explicit constructions will show that all simple Lie (super) algebras can be given this structure. Next, we commute these with the $\mathscr{U}$ and $\widetilde{\mathscr{U}}$ spaces and use the grading structure to write

$$[S_{ab},\mathscr{U}_c]=\widetilde{\mathscr{U}}_{s_{ab}(c)},\tag{2.21a}$$

$$[S_{ab},\widetilde{\mathscr{U}}_c]=\widetilde{\mathscr{U}}_{\tilde{s}_{ab}(c)},\tag{2.21b}$$

$$[K_{ab},\widetilde{\mathscr{U}}_c]=\mathscr{U}_{k_{ab}(c)},\tag{2.21c}$$

$$[\widetilde{K}_{ab},\mathscr{U}_c]=\widetilde{\mathscr{U}}_{\tilde{k}_{ab}(c)}.\tag{2.21d}$$

We recall that a stands for a set of parameters. Clearly for the algebra to close $s_{ab}(c)$ must belong to the same set of parameters. Similarly $\tilde{s}_{ab}(c)$, $k_{ab}(c)$, and $\tilde{k}_{ab}(c)$ must belong to the set. Let us denote

$$s_{ab}(c)=(abc).\tag{2.22}$$

The triple product (abc) is a mapping from the tensor product of three copies of the space of parameters into the space itself. We will denote the parameter space by V. Thus (abc) corresponds to

$$V\otimes V\otimes V\to V.\tag{2.23}$$

Thus, the triple product (abc) defines a ternary algebra over V.

Let us express $\tilde{s}_{ab}(c)$, $k_{ab}(c)$, and $\tilde{k}_{ab}(c)$ also in terms of triple products. we make use of Jacobi identities which must be imposed on $\mathscr{L}$:

$$\begin{aligned}\mathscr{U}_{k_{ab}(c)}&=[K_{ab},\widetilde{\mathscr{U}}_c]=[[\mathscr{U}_a,\mathscr{U}_b],\widetilde{\mathscr{U}}_c]\\&=[\mathscr{U}_a,[\mathscr{U}_b,\widetilde{\mathscr{U}}_c]]-[\mathscr{U}_b,[\mathscr{U}_a,\widetilde{\mathscr{U}}_c]]\\&=[\mathscr{U}_a,S_{bc}]-[\mathscr{U}_b,S_{ac}]\\&=\mathscr{U}_{s_{ac}(b)}-\mathscr{U}_{s_{bc}(a)}.\end{aligned}\tag{2.24}$$

Since $\mathscr{U}_x$ is linear in x (x is an infinitesimal parameter) we obtain

$$k_{ab}(c)=(acb)-(bca).\tag{2.25}$$

Note the antisymmetry of $k_{ab}(c)$ in $a\leftrightarrow b$, as expected from Eq. (2.21).

We next use the conjugation operation to obtain $\tilde{s}_{ab}(c)$ and $\tilde{k}_{ab}(c)$. First, we treat the pure bosonic case. From conjugating Eq. (2.21) we obtain for the conjugation of the first kind (which reverses the orders of operators)

$$\begin{aligned}\widetilde{\mathscr{U}}_{s_{ab}(c)}&=([[\mathscr{U}_a,\widetilde{\mathscr{U}}_b],\mathscr{U}_c])^{\sim}\\&=a_i\,b_j\,c_k\,[\widetilde{G}_k,[G_j,\widetilde{G}_i]]\quad\text{(first kind)}\\&=[\widetilde{\mathscr{U}}_c,[\mathscr{U}_b,\widetilde{\mathscr{U}}_a]]=-[S_{ba},\widetilde{\mathscr{U}}_c]=-\widetilde{\mathscr{U}}_{\tilde{s}_{ba}(c)},\end{aligned}\tag{2.26}$$

where the last line follows from the definition Eq. (2.21b). Similarly, for the conjugation of second kind (which maintains the order of operators) one gets

$$\begin{aligned}\widetilde{\mathscr{U}}_{s_{ab}(c)}&=([[\mathscr{U}_a,\widetilde{\mathscr{U}}_b],\mathscr{U}_c])^{\sim}\\&=a_ib_jc_k[[\widetilde{G}_i,G_j],\widetilde{G}_k]\quad\text{(second kind)}\\&=[[\widetilde{\mathscr{U}}_a,\mathscr{U}_b],\widetilde{\mathscr{U}}_c]=-[S_{ba},\widetilde{\mathscr{U}}_c]\\&=-\widetilde{\mathscr{U}}_{\tilde{s}_{ba}(c)}.\end{aligned}\tag{2.27}$$

Comparing the left and right hand side of these equations we learn that in the pure Bose case both kinds of conjugations yield

$$\tilde{s}_{ab}(c)=-s_{ba}(c)=-(bac)\quad\text{(Bose)}\tag{2.28}$$

Similarly, by conjugating Eq. (2.21c) and using Eq. (2.21d) we arrive at

$$\tilde{k}_{ab}(c)=k_{ab}(c)=(acb)-(bca)\quad\text{(Bose).}\tag{2.29}$$

We now return to the pure Fermi or mixed Bose–Fermi cases. There is a difference between the first and second conjugation. If the conjugation is of second kind, the result is identical to Eqs. (2.28)–(2.29). If the conjugation is of first kind, we first specialize to the pure Fermi case. Proceeding as above by factoring out a_i,b_j,c_k, we find that the anticommutator $\{\widetilde{G}_i,G_j\}$ appears instead of the commutator and, furthermore, from reordering $a_ib_jc_k=-c_kb_ja_i$ an extra minus sign arises giving (for pure Fermi and first kind conjugation)

$$\begin{aligned}&\tilde{s}_{ab}(c)=+s_{ba}(c)=(bac),\\&\tilde{k}_{ab}(c)=-k_{ab}(c)=-(acb)+(bca)\quad\text{(Fermi).}\end{aligned}\tag{2.30}$$

The mixed case of some parameters bosonic, others fermionic will give mixed phases (± 1) with the result that when the number of Fermi parameters exceeds the number of Bose parameters (among a,b,c) there is a (-1) otherwise a ($+1$)

Summarizing these results, we can write for all cases

$$\tilde{s}_{ab}(c)=-e^{i\phi}s_{ba}(c),\quad \tilde{k}_{ab}(c)=+e^{i\phi}k_{ab}(c),\tag{2.31}$$

where $\phi(a,b,c)=0$ or π:

$$\begin{aligned}&\phi(a,b,c)=\begin{pmatrix}0 & \text{Bose}>\text{Fermi}\\ \pi & \text{Bose}<\text{Fermi}\end{pmatrix}\quad\text{first conjugation,}\\&\phi(a,b,c)=0\quad\text{Bose or Fermi,}\quad\text{second conjugation.}\end{aligned}\tag{2.32}$$

Next, we calculate the commutators $[S_{ab},S_{cd}]$, $[S_{ab},K_{cd}]$, $[S_{ab},\widetilde{K}_{cd}]$, $[K_{ab},\widetilde{K}_{cd}]$ which completes the list of nonvanishing commutators in $\mathscr{L}$. All of these are derived by imposing Jacobi identities as follows:

$$[S_{ab},S_{cd}]=[[\mathscr{U}_a,\widetilde{\mathscr{U}}_b],S_{cd}]$$

$$= [\mathscr{U}_a,[\widetilde{\mathscr{U}}_b,S_{cd}]] - [\widetilde{\mathscr{U}}_b,[\mathscr{U}_a,S_{cd}]]$$
$$= -[\mathscr{U}_a,\widetilde{\mathscr{U}}_{\tilde{s}_{cd}(b)}] + [\widetilde{\mathscr{U}}_b,\mathscr{U}_{s_{cd}(a)}]$$
$$= -S_{a,\tilde{s}_{cd}(b)} - S_{s_{cd}(a),b}, \qquad (2.33)$$

$$[S_{ab},K_{cd}] = [[\mathscr{U}_a,\widetilde{\mathscr{U}}_b],K_{cd}]$$
$$= [\mathscr{U}_a,[\widetilde{\mathscr{U}}_b,K_{cd}]] - [\widetilde{\mathscr{U}}_b,[\mathscr{U}_a,K_{cd}]]$$
$$= -[\mathscr{U}_a,\mathscr{U}_{k_{cd}(b)}] = -K_{a,k_{cd}(b)}. \qquad (2.34)$$

Another expression for this commutator can be obtained

$$[S_{ab},K_{cd}] = [S_{ab},[\mathscr{U}_c,\mathscr{U}_d]]$$
$$= [[S_{ab},\mathscr{U}_c],\mathscr{U}_d] - [[S_{ab},\mathscr{U}_d],\mathscr{U}_c]$$
$$= [\mathscr{U}_{s_{ab}(c)},\mathscr{U}_d] - [\mathscr{U}_{s_{ab}(d)},\mathscr{U}_c]$$
$$= K_{s_{ab}(c),d} + K_{c,s_{ab}(d)}. \qquad (2.35)$$

Similarly we arrive at

$$[S_{ab},\widetilde{K}_{cd}] = -\widetilde{K}_{\tilde{k}_{cd}(a),b} = \widetilde{K}_{\tilde{s}_{ab}(c),d} + \widetilde{K}_{c,\tilde{s}_{ab}(d)} \qquad (2.36)$$

and

$$[K_{ab},\widetilde{K}_{cd}] = S_{k_{ab}(c),d} - S_{k_{ab}(d),c} = -S_{\widetilde{k}_{cd}(a),b} + S_{\widetilde{k}_{cd}(b),a}. \qquad (2.37)$$

There remains to impose all other Jacobi identities. This is considered in detail in the Appendix. Here we give the result in a theorem.

Theorem 1: The algebra $\mathscr{L}$ constructed as above, is a Lie algebra or superalgebra, obeying all Jacobi identities if the triple product satisfies the following two (super) generalizations of Jacobson's[9] and Kantor's[4] conditions which we call (JSl) and (KSl), respectively:

$$\text{(JSl):}\quad (ab(cdx)) - (cd(abx)) + (a\tilde{s}_{cd}(b)x) + ((cda)bx) = 0, \qquad (2.38)$$

$$\text{(KSl):}\quad \{(ax(cbd)) - ((cbd)xa) + (ab(cxd)) - (c(\tilde{s}_{ab}(x))d)\} - \{c\leftrightarrow d\} = 0, \qquad (2.39)$$

where $\tilde{s}$ is given in Eq. (2.31) and switches signs depending on the Bose or Fermi nature of the parameters. These follow from commuting Eqs. (2.33)–(2.34) with $\mathscr{U}_x$ and using Jacobi identities as shown in the Appendix.

Corollary: The three-dimensional grading $\mathscr{L} = \mathscr{U}_{-1} \oplus \mathscr{U}_0 \oplus \mathscr{U}_1$ is a special case of the five-dimensional grading when the space $\mathscr{U}_{\pm 2}$ vanishes. In this case $k_{cd}(x) = 0$ and the second condition (KSl) of Eq. (2.39) is automatically satisfied. But now $k_{cd}(x) = 0$ implies the supergeneralization of Jacobson's condition[9] (JS2)

$$\text{(JS2):}\quad (cxd) = (dxc), \qquad (2.40)$$

that is, the triple product must be symmetric under the interchange of the first and last set of parameters. Note that in the pure bosonic case (JSl) together with (JS2) become the defining relations for the usual Jordan triple system.[9] Our result is a generalization that includes anticommuting variables. We shall call such systems Jordan superternary algebras. Clearly these automatically lead to a three-dimensional grading for the Lie algebra or superalgebra. Examples of these will be given in Sec. VI.

To summarize, we list all the nonzero commutation rules and comment on the general structure.

$$[\mathscr{U}_a,\widetilde{\mathscr{U}}_b] = S_{ab},\quad [\mathscr{U}_a,\mathscr{U}_b] = K_{ab},\quad [\widetilde{\mathscr{U}}_a,\widetilde{\mathscr{U}}_b] = \widetilde{K}_{ab},$$
$$[S_{ab},\mathscr{U}_c] = \mathscr{U}_{s_{ab}(c)},\quad [S_{ab},\widetilde{\mathscr{U}}_c] = \widetilde{\mathscr{U}}_{\tilde{s}_{ab}(c)},$$
$$[K_{ab},\widetilde{\mathscr{U}}_c] = \mathscr{U}_{k_{ab}(c)},\quad [\widetilde{K}_{ab},\mathscr{U}_c] = \widetilde{\mathscr{U}}_{\widetilde{k}_{ab}(c)},$$
$$[S_{ab},S_{cd}] = S_{s_{ab}(c),d} + S_{c,\tilde{s}_{ab}(d)} = -S_{s_{cd}(a),b} - S_{a,\tilde{s}_{cd}(b)} \qquad (2.41)$$
$$[S_{ab},K_{cd}] = -K_{a,k_{cd}(b)} = K_{s_{ab}(c),d} + K_{c,s_{ab}(d)},$$
$$[S_{ab},\widetilde{K}_{cd}] = -\widetilde{K}_{\widetilde{k}_{cd}(a),b} = \widetilde{K}_{\tilde{s}_{ab}(c),d} + \widetilde{K}_{c,\tilde{s}_{ab}(d)},$$
$$[K_{ab},\widetilde{K}_{cd}] = S_{k_{ab}(c),d} - S_{k_{ab}(d),c} = -S_{\widetilde{k}_{cd}(a),b} + S_{\widetilde{k}_{cd}(b),a},$$

where s, $\tilde{s}$, k, and $\widetilde{k}$ are given in terms of the triple product in Eqs. (2.22), (2.25), (2.31), and (2.32). We see that the following sets of generators close among themselves to form subalgebras:

$$\mathscr{S} = \{S\},\quad \mathscr{M} = \{\widetilde{K} \oplus S \oplus K\},$$
$$\mathscr{L} = \{\widetilde{K} \oplus \widetilde{\mathscr{U}} \oplus S \oplus \mathscr{U} \oplus K\},\quad \Delta = \{S \oplus \mathscr{U} \oplus K\}$$
$$\text{or } \{S \oplus \widetilde{\mathscr{U}} \oplus \widetilde{K}\}, \qquad (2.42)$$
$$T = \{S \oplus K\} \text{ or } \{S \oplus \widetilde{K}\},\quad \mathscr{N} = \{\mathscr{U} \oplus K\}$$
$$\text{or } \{\widetilde{\mathscr{U}} \oplus \widetilde{K}\},$$

We further point out the generators that correspond to the coset spaces

$$\mathscr{L}/\mathscr{M} = \{\mathscr{U} \oplus \widetilde{\mathscr{U}}\},\quad \mathscr{L}/\mathscr{S} = \{\widetilde{K} \oplus \widetilde{\mathscr{U}} \oplus \mathscr{U} \oplus K\}, \qquad (2.43)$$

etc. Of these cosets spaces $\mathscr{L}/\mathscr{M}$ is a symmetric space.[10] Note from the structure of the commutation rules that $\mathscr{U} \oplus \widetilde{\mathscr{U}}$ fall into a representation of $\mathscr{M}$ while each of $\mathscr{U}$, $\widetilde{\mathscr{U}}$, K, $\widetilde{K}$ fall into a representation of S.

A class of triple products: Explicit construction of Lie algebras or superalgebras require explicit forms for the triple product (abc). Here we consider a very general class of triple products which are generalizations of Kantor's products to include the super case.

We define the triple product

$$(abc) = a\cdot(\bar{b}\cdot c) + c\cdot(\bar{b}\cdot a) - b\cdot(\bar{a}\cdot c)e^{i\phi(abc)}. \qquad (2.44)$$

It is remarkable that all Lie algebras and most superalgebras (probably all) can be constructed by means of this triple product. Here a, b, c belong to a vector space V with a multiplication ($\cdot$) and a conjugation ($-$) which closes under this triple product $V \times V \times V \rightarrow V$. This conjugation which acts on V is distinguished from the one discussed above that acted on operators. The closure property includes the requirements that the properties of (abc) under the conjugation be identical to those of its individual arguments, a, b, or c. In the last term, the phase $\phi(abc)$ takes the values $\phi = 0$ or π for the same cases described in Eq. (2.32).

A special case of the above algebra is obtained by taking the product $\{\cdot\}$ associative. We shall call a ternary algebra an "associative" ternary algebra if it can be defined by an associative multiplication of its arguments.

Thus, a class of associative ternary algebras is defined by taking the product ($\cdot$) in Eq. (2.44) associative. When the product is associative, we adopt the convention of omitting the parentheses. Thus, Eq. (2.44) reduces to

$$(abc) = a\bar{b}c + c\bar{b}a - e^{i\phi}b\bar{a}c. \tag{2.45}$$

We define another class of associative ternary algebras by

$$(abc) = a\bar{b}c + c\bar{b}a. \tag{2.46}$$

It can be checked that when the multiplication in Eq. (2.44) is taken as the Jordan multiplication $a \cdot b = \frac{1}{2}(ab + ba)$ and the conjugation is trivial $\bar{a} = a$, then the two triple products Eqs. (2.44) and (2.46) coincide. However, if the conjugation is nontrivial, this is a different triple product. Note that the triple product (2.46) is a super Jordan triple system by Eqs. (2.38) and (2.40) and therefore it leads to a three-dimensional grading for $\mathscr{L}$.

We can now state a theorem whose proof is given in the Appendix.

Theorem 2: The associative ternary algebras defined by Eqs. (2.45)–(2.46) satisfy automatically the conditions (JSI) and (KSI) of Theorem 1.

Theorem (2) is of great generality, because the only condition that is now required for the construction of a Lie algebra or superalgebra is merely the closure property $V \otimes V \otimes V \to V$ under the above associative triple products. This property is easily satisfied by many algebraic systems and gives a lot of freedom for possible physical applications. Examples of such systems are real numbers, complex numbers, imaginary numbers, quaternions, fermions, square or rectangular matrices, Dirac γ matrices, discrete groups, spinors, tensors, tensor products of any of the above, etc., etc. It is remarkable that any of these associative algebraic systems can serve as building blocks of Lie algebras or Lie superalgebras, by the construction of Eqs. (2.41), (2.45), (2.46) given in this paper.

III. LIE ALGEBRAS FROM TERNARY ALGEBRAS

Using the results of Sec. II, we now construct Lie algebras and Lie superalgebras. For completeness and clarity we shall begin with the purely bosonic case, recovering the results of Kantor. We then present our results for the superalgebras in Secs. IV, V, and VI.

To illustrate the method, we begin from the simplest ternary algebras consisting of the following vector spaces V:

1. Real number $\mathbb{R}$, complex numbers $\mathbb{C}$, quaternions $\mathbb{H}$ (Pauli matrices) and octonions $\mathbb{O}$ (collectively referred to as composition algebras) and their direct products: $\mathbb{C}\otimes\mathbb{C}$, $\mathbb{C}\otimes\mathbb{H}$, $\mathbb{C}\otimes\mathbb{O}$, $\mathbb{H}\otimes\mathbb{H}$, $\mathbb{H}\otimes\mathbb{O}$, $\mathbb{O}\otimes\mathbb{O}$. The vector spaces will lead to the Lie algebras appearing in the famous magic square.[16] This approach gives a simple construction of exceptional Lie algebras and their subalgebras.

1a. When $V = \mathbb{R}$, a, b, c are just real numbers and the triple product Eq. (2.45) reduces to the ordinary multiplication of these three real numbers.

$$s_{ab}(c) = (abc) = abc, \quad k_{ab}(c) = (acb) - (bca) = 0.$$

The number of generators are the same as the number of independent real numbers associated with a transformation. Thus, we have

$$\mathscr{U}_a = \{1 \text{ generator}\}, \quad \widetilde{\mathscr{U}}_b = \{1 \text{ generator}\},$$
$$S_{ab} = \{1 \text{ generator}\}. \tag{3.1}$$

According to Eqs. (2.41) these generators close to form the Lie algebra B_1. For clarity, we give a representation of these operators. This can be done in any representation of this algebra we can write

$$\mathscr{U}_a = \begin{pmatrix} 0 & a \\ 0 & 0 \end{pmatrix}, \quad \widetilde{\mathscr{U}}_b = \begin{pmatrix} 0 & 0 \\ b & 0 \end{pmatrix}, S_{ab} = \begin{pmatrix} ab & 0 \\ 0 & -ab \end{pmatrix}. \tag{3.2}$$

By explicitly commuting these matrices, one can verify our assertions.

1b. When $V = \mathbb{C}$, a, b, c are simply complex numbers and $\bar{a} = a^*$ is a complex conjugation. The triple product (2.45) gives

$$s_{ab}(c) = (abc) = 2ab^*c - ba^*c,$$
$$k_{ab}(c) = (acb) - (bca) = c(b^*a - a^*b). \tag{3.3}$$

The number of generators are equal to the number of real parameters occurrring in $\mathscr{U}_a$, $\widetilde{\mathscr{U}}_b$, $\widetilde{K}_{ab}$, K_{ab}, and S_{ab}. Thus, we find

$$\mathscr{U}_a = \{2 \text{ generators}\}, \quad \widetilde{\mathscr{U}}_b = \{2 \text{ generators}\},$$
$$K_{ab} = \{1 \text{ generator}\}, \quad \widetilde{K}_{cd} = \{1 \text{ generator}\}, \tag{3.4}$$
$$S_{ab} = \{2 \text{ generators}\}.$$

Thus, recalling the subgroup structures discussed in Eq. (2.42) $S = E \oplus E$ is the Lie algebra of the group $U(1)\times U(1)$, $M = \widetilde{K} \oplus S \oplus K$ is the algebra $A_1 \oplus E$, the space $\mathscr{U} \oplus \widetilde{\mathscr{U}}$ form two doublets under commutation with $A_1 \oplus E$. The full algebra has eight generators corresponding to A_2. the structure constants can be constructed directly from the functions $s_{ab}(c)$ and $k_{ab}(c)$. In the fundamental 3×3 representation of A_2 we may write the matrix representation

$$\mathscr{U}_a = \begin{pmatrix} 0 & 0 & a \\ 0 & 0 & 0 \\ 0 & a^* & 0 \end{pmatrix}, \quad \widetilde{\mathscr{U}}_b = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & b \\ b^* & 0 & 0 \end{pmatrix}, \tag{3.5}$$

$$K_{ab} = \begin{pmatrix} 0 & ab^* - ba^* & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}, \quad \widetilde{K}_{cd} = \begin{pmatrix} 0 & 0 & 0 \\ cd^* - dc^* & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix},$$

$$S_{ab} = \begin{pmatrix} +ab^* & 0 & 0 \\ 0 & -ba^* & 0 \\ 0 & & a^*b - b^*a \end{pmatrix}.$$

1c. When $V = \mathbb{H}$; a, b, c are quaternions, which can be represented in terms of Pauli matrices

$$a = a_0 - i\vec{\sigma}\cdot\vec{a}, \quad \bar{a} = a_0 + i\vec{\sigma}\cdot\vec{a}, \tag{3.6}$$

where $\bar{a}$ corresponds to quaternion conjugation which is equivalent to Hermitian conjugation in this representation.

In this case we have

$$s_{ab}(c) = (abc) = (a\bar{b} - b\bar{a})c + c\bar{b}a,$$
$$k_{ab}(c) = c(\bar{b}a - \bar{a}b). \tag{3.7}$$

Clearly $s_{ab}(c)$ and $k_{ab}(c)$ are quaternions showing the closure property. $s_{ab}(c)$ corresponds to a transformation on c generated by a left multiplication with an imaginary quaternion (traceless 2×2 anti-Hermitian matrix) and a right multiplication with a full quaternion. Thus, they generate the transformation

$SU(2)\otimes SU(2)\otimes U(1)$ on c.

Counting the dimensions of real parameters, we have

$$\mathscr{U}_a=\{4 \text{ generators}\},\quad \widetilde{\mathscr{U}}_b=\{4 \text{ generators}\},$$
$$K_{ab}=\{3 \text{ generators}\},\quad \widetilde{K}_{cd}=\{3 \text{ generators}\},\quad (3.8)$$
$$S_{ab}=\{7 \text{ generators}\}.$$

The Lie algebras they generate are

$$S=A_1\oplus A_1\oplus E,\quad \mathscr{M}=\widetilde{K}\oplus S\oplus K=C_2\oplus C_1,$$
$$\mathscr{L}=\widetilde{K}\oplus\widetilde{\mathscr{U}}\oplus S\oplus\mathscr{U}\oplus K=C_3. \quad (3.9)$$

1d. $V=\mathbb{O}$, which means a, b, c are octonions. For the definition and properties of octonions we refer the reader to Refs. 11 and 12. Even though the octonions form a nonassociative algebra, the ternary algebra they generate under the triple product of Eq. (2.44),

$$s_{ab}(c)=(abc)=a(\bar{b}c)+c(\bar{b}a)-b(\bar{a}c), \quad (3.10)$$

satisfies the conditions (JS1) and (KSl) of Theorem 1.[13]

We have further,

$$k_{ab}(c)=c(\bar{b}a-\bar{a}b). \quad (3.11)$$

In this case the $s_{ab}(c)$ considered as a transformation on c corresponds to automorphic rotations of octonions [generators of the exceptional group $G(2)$] and multiplication by octonions,[14] thus generating the Lie algebra of $SO(7)\times U(1)$.

Counting the number of independent parameters is much more complicated in this case due to the nonassociativity of octonions. We find

$$\mathscr{U}_a=\{8 \text{ generators}\},\quad \widetilde{\mathscr{U}}_b=\{8 \text{ generators}\},$$
$$K_{ab}=\{7 \text{ generators}\},\quad \widetilde{K}_{ab}=\{7 \text{ generators}\},$$
$$S_{ab}=\{22 \text{ generators}\}.$$

The Lie algebras they generate are

$$S=B_3\oplus E,\quad \mathscr{M}=\widetilde{K}\oplus S\oplus K=B_4,$$
$$\mathscr{L}=\widetilde{K}\oplus\widetilde{U}\oplus S\oplus U\oplus K=F_4, \quad (3.13)$$

where F_4 is the exceptional group of Cartan with 52 generators. It is known to be the automorphism group of the exceptional Jordan algebra of Jordan, von Neumann, and Wigner.[15]

2. *Tensor products of composition algebras*: Let us now consider the case when the vector space V corresponds to the tensor product of two composition algebras $\mathbb{A}=(\mathbb{C},\mathbb{H},\mathbb{O})$ considered above. That is, $V=\mathbb{A}_1\otimes\mathbb{A}_2=\{\mathbb{C}\otimes\mathbb{C},\ \mathbb{C}\otimes\mathbb{H},\ \mathbb{C}\otimes\mathbb{O},\ \mathbb{H}\otimes\mathbb{H},\ \mathbb{H}\otimes\mathbb{O},\ \mathbb{O}\otimes\mathbb{O}\}$. The conjugation in V will be defined as the tensor product of conjugation in $\mathbb{A}_1$ and $\mathbb{A}_2$, i.e.,

$$\overline{a_1\otimes a_2}=\bar{a}_1\otimes\bar{a}_2. \quad (3.14)$$

We shall also adopt the convention of denoting the elements of $\mathbb{A}_1$ by small letters, the elements of $\mathbb{A}_2$ by capital letters and dropping the tensor product sign $\otimes$, i.e., $a_1\otimes a_2\to aA$. The triple product in the tensor product space $V=\mathbb{A}_1\otimes\mathbb{A}_2$ becomes

$$\begin{aligned}s_{aAbB}(cC)&=(aAbBcC)\\&=(aA)[(\overline{bB})(cC)]+(cC)[(\overline{bB})(aA)]\\&\quad-(bB)[(\overline{aA})(cC)]\\&=a(\bar{b}c)A(\bar{B}C)+c(\bar{b}a)C(\bar{B}A)-b(\bar{a}c)B(\bar{A}C).\end{aligned} \quad (3.15)$$

This can be written in the form

$$\begin{aligned}s_{aAbB}(cC)&=(a,b)c[A(\bar{B}C)-B(\bar{A}C)+C(\bar{B}A)]\\&\quad+[a(\bar{b}c)-b(\bar{a}c)+c(\bar{b}a)](A,B)C\\&\quad-(a,b)c(A,B)C+\tfrac{1}{4}c(\bar{b}a-\bar{a}b)C(\bar{B}A-\bar{A}B)\\&=(a,b)cs_{AB}(C)+s_{ab}(c)(A,B)C-(a,b)c(A,B)C\\&\quad+\tfrac{1}{4}c(\bar{b}a-\bar{a}b)C(\bar{B}A-\bar{A}B),\end{aligned} \quad (3.16)$$

where we defined the scalars

$$(a,b)\equiv\tfrac{1}{2}(a\bar{b}+b\bar{a})\quad (A,B)\equiv\tfrac{1}{2}(A\bar{B}+B\bar{A}). \quad (3.17)$$

Similarly,

$$\begin{aligned}k_{aAbB}(cC)&=-cC[(\bar{a}\bar{A})(bB)-(\bar{b}\bar{B})(aA)]\\&=-c(a,b)C\tfrac{1}{2}(\bar{A}B-\bar{B}A)\\&\quad-c\tfrac{1}{2}(\bar{a}b-\bar{b}a)C(A,B).\end{aligned} \quad (3.18)$$

As an illustration we present the case of tensor products of two quarternions a and A $aA\in\mathbb{H}\otimes\mathbb{H}$. Referring to the equations above we see that $S_{aA,bB}$ generate the following transformations on cC [see Eqs. (3.7) and (3.9)]:

$$s_{AB}(C)\to A_1+A_1+E,\quad s_{ab}(c)\to A_1+A_1+E. \quad (3.19)$$

In addition we have the right multiplication of both quaternions cC by the product of imaginary quaternions $(\bar{b}a-\bar{a}b)\otimes(\bar{B}A-\bar{A}B)(3\times 3=9$ parameters). These latter transformations combine with the right handed A_1's of $s_{AB}(C)$ and $s_{ab}(c)$ to form D_3 leaving us with left-handed A_1+A_1+E. Therefore, the total transformation corresponds to the Lie algebra

$$S=A_1+A_1+E+D_3=\{22 \text{ generators}\}. \quad (3.20)$$

Examining $k_{aA,bB}(cC)$ we see that it corresponds to right multiplication with imaginary quaternions ($3+3=6$ parameters). Thus, we find

$$K=\{6 \text{ generators}\},\quad \widetilde{K}=\{6 \text{ generators}\}. \quad (3.21)$$

Noting the combination $\widetilde{K}+K+D_3+E=D_4$, we arrive at

$$\mathscr{M}=\widetilde{K}+S+K=D_4+A_1+A_1=\{34 \text{ generators}\}. \quad (3.22)$$

Finally combining these with $\mathscr{U}=\{16 \text{ generators}\}$ and $\widetilde{\mathscr{U}}=\{16 \text{ generators}\}$ we arrive at the full Lie algebra

$$\begin{aligned}\mathscr{L}&=\widetilde{\mathscr{U}}+\widetilde{K}+S+\mathscr{U}+K=D_6\\&=\{66 \text{ generators}\}.\end{aligned} \quad (3.23)$$

We now list the results for the remaining tensor product cases

2a. $V=\mathbb{C}\otimes\mathbb{C}$,

$$S=E+E+E+E,$$
$$\mathscr{M}=A_1+A_1+E+E,\quad \mathscr{L}=A_2+A_2, \quad (3.24)$$

2b. $V=\mathbb{C}\otimes\mathbb{H}$,

$$S=D_2+A_1+E+E,\quad \mathscr{M}=A_3+A_1+E,$$
$$\mathscr{L}=A_5, \quad (3.25)$$

2c. $V=\mathbb{C}\otimes\mathbb{O}$,

$$S=D_4+E+E,\quad \mathscr{M}=D_5+E,\quad \mathscr{L}=E_6, \quad (3.26)$$

TABLE I. Magic square. In each entry the first line refers to S, the second to $\mathscr{M} = \widetilde{K} + S + K$ and the third to $\mathscr{L} = \widetilde{K} + \widetilde{\mathscr{U}} + S + \mathscr{U} + K$.

	$\mathbb{R}$	$\mathbb{C}$	$\mathbb{H}$	$\mathbb{O}$
$\mathbb{R}$	E E B_1	$E+E$ A_1+E A_2	A_1+A_1+E C_2+C_1 C_3	B_3+E B_4 F_4
$\mathbb{C}$	$E+E$ A_1+E A_2	$E+E+E+E$ A_1+A_1+E+E A_2+A_2	D_2+A_1+E+E A_3+A_1+E A_5	D_4+E+E D_5+E E_6
$\mathbb{H}$	A_1+A_1+E C_2+C_1 C_3	D_2+A_1+E+E A_3+A_1+E A_5	$D_3+A_1+A_1+E$ $D_4+A_1+A_1$ D_6	D_5+A_1+E D_6+A_1 E_7
$\mathbb{O}$	B_3+E B_4 F_4	D_4+E+E D_5+E E_6	D_5+A_1+E D_6+A_1 E_7	D_7+E D_8 E_8

2d. $V = \mathbb{H} \otimes \mathbb{H}$, as given above explicitly,

$$S = D_3 + A_1 + A_1 + E,$$
$$\mathscr{M} = D_4 + A_1 + A_1, \quad \mathscr{L} = D_6, \tag{3.27}$$

2e. $V = \mathbb{H} \otimes \mathbb{O}$,

$$S = D_5 + A_1 + E, \quad \mathscr{M} = D_6 + A_1, \quad \mathscr{L} = E_7, \tag{3.28}$$

2f. $V = \mathbb{O} \otimes \mathbb{O}$,

$$S = D_7 + E, \quad \mathscr{M} = D_8, \quad \mathscr{L} = E_8. \tag{3.29}$$

Recapitulating, the results of 1(a)–1(d) and (2a)–(2f) are summarized by the so called magic square[16] given in Table I. In each entry the first line refers to S, the second to $\mathscr{M} = \widetilde{K} + S + K$ and the third to $\mathscr{L} = \widetilde{K} + \widetilde{\mathscr{U}} + S + \mathscr{U} + K$.

Note that the magical property of this table, namely its symmetry under "transposition," is not magical in this construction, as it follows automatically from the symmety of the tensor product.

This table is useful in analyzing the subgroup structure of the groups it contains, in particular the exceptional groups: Any set of algebras $(S, \mathscr{M}, \mathscr{L})$ appearing above or to the left of any entry is a set of subalgebras of the algebras appearing in that entry. Of course within each block the usual subalgebra structure $S \subset \mathscr{M} \subset \mathscr{L}$ appears.

3. *Higher tensor products of composition algebras*: As a generalization of the previous structures, one can consider higher tensor products of composition algebras $A = (\mathbb{R}, \mathbb{C}, \mathbb{H}, \mathbb{O})$ in the form

$$\mathbb{A}_1 \otimes \mathbb{A}_2 \otimes \cdots \otimes \mathbb{A}_n. \tag{3.30}$$

When the $\mathbb{A}_i$ are associative (i.e., not octonion) Theorems 1 and 2 are satisfied, leading automatically to a Lie algebra.

As an example, we develop the case of n copies of the quaternion algebra

$$a \to \mathbb{H}_1 \otimes \mathbb{H}_2 \otimes \cdots H_n. \tag{3.31}$$

Here we find

$$\mathscr{U}_a = \{2^{2n} \text{ generators}\}, \quad \widetilde{\mathscr{U}}_a = \{2^{2n} \text{ generators}\},$$
$$K_{ab} = \{2^{n-1}[2^n - (-1)^n] \text{generators}\},$$
$$\widetilde{K}_{ab} = \{2^{n-1}[2^n - (-1)^n] \text{generators}\}, \tag{3.32}$$
$$S_{ab} = \{[(2^{2n} + 2^{2n-1}) - (-1)^n 2^{n-1}] \text{ generators}\}.$$

Thus, we arrive at the algebras

$$n = \text{even}, S = D_{(2^{n-1})} + A_{(2^n - 1)} + E,$$
$$\mathscr{M} = D_{(2^{n-1})} + D_{(2^n)}, \quad \mathscr{L} = D_{(3\times 2^{n-1})}. \tag{3.33}$$
$$n = \text{odd}, S = C_{(2^{n-1})} + A_{(2^n - 1)} + E,$$
$$\mathscr{M} = C_{(2^{n-1})} + C_{(2^n)}, \quad \mathscr{L} = C_{(3\times 2^{n-1})}. \tag{3.34}$$

4. *Rectangular matrices over* $\mathbb{R}, \mathbb{C}, \mathbb{H}$: We consider the ternary algebra of $n \times m$ rectangular matrices

$$a = \begin{pmatrix} a_{11} & \cdots & a_{1m} \\ \vdots & & \vdots \\ \cdot & & \cdot \\ a_{n1} & \cdots & a_{nm} \end{pmatrix}, \tag{3.35}$$

where the entries a_{ij} are taken over $\mathbb{R}, \mathbb{C}, \mathbb{H}$. Conjugation is defined by taking the transpose and conjugation in the composition algebra,

$$\bar{a} = \begin{pmatrix} \bar{a}_{11} & \cdots & \bar{a}_{n1} \\ \vdots & & \vdots \\ \bar{a}_{1m} & \cdots & \bar{a}_{nm} \end{pmatrix}. \tag{3.36}$$

Clearly the triple product

$$s_{ab}(c) = (abc) = a\bar{b}c + c\bar{b}a - b\bar{a}c \tag{3.37}$$

closes for such a space. Similarly,

$$k_{ab}(c) = c(\bar{b}a - \bar{a}b). \tag{3.38}$$

S_{ab} consists of a transformation on c in the form $(a\bar{b} - b\bar{a})$ acting from the left and $\bar{b}a$ acting from the right. Clearly, $(a\bar{b} - b\bar{a})$ is an $n \times n$ anti-Hermitian matrix over the given composition algebra, while $\bar{b}a$ is an arbitrary $m \times m$ matrix over the same algebra $\mathbb{R}$, $\mathbb{C}$, or $\mathbb{H}$. Of course, the left and right multiplications commute with each other. From $k_{ab}(c)$ we learn that there are as many K_{ab} generators as there are real parameters in the $m \times m$ anti-Hermitian matrix $(\bar{b}a - \bar{a}b)$. If we denote the dimensions of the composition algebras $\mathbb{R}$, $\mathbb{C}$, $\mathbb{H}$ by $d = 1,2,4$, respectively, we find the following counting of the generators:

$$\mathscr{U}_a = \{dmn \text{ generators}\}, \quad \widetilde{\mathscr{U}}_a = \{dmn \text{ generators}\},$$
$$K_{ab} = \{\tfrac{1}{2}dm(m-1) + m(d-1) \text{ generators}\},$$
$$\widetilde{K}_{ab} = \{\tfrac{1}{2}dm(m-1) + m(d-1) \text{ generators}\}, \quad (3.39)$$
$$S_{ab} = \{\tfrac{1}{2}dn(n-1) + n(d-1) + dm^2 \text{ generators}\}.$$

These combine to give the following Lie algebras:

$$\mathbb{R}: \quad n = \text{even} = 2k,$$
$$S = D_k + A_{m-1} + E, \quad \mathscr{M} = D_k + D_m,$$
$$\mathscr{L} = D_{k+m}, \quad (3.40)$$
$$n = \text{odd} = 2k+1,$$
$$S = B_k + A_{m-1} + E, \quad \mathscr{M} = B_k + D_m,$$
$$\mathscr{L} = B_{k+m}, \quad (3.41)$$
$$\mathbb{C}: \quad S = A_{n-1} + A_{m-1} + A_{m-1} + E + E,$$
$$\mathscr{M} = A_{n-1} + A_{2m-1} + E, \quad \mathscr{L} = A_{n+2m-1}, \quad (3.42)$$
$$\mathbb{H}: \quad S = C_n + A_{2m-1} + E,$$
$$\mathscr{M} = C_n + C_{2m}, \quad \mathscr{L} = C_{n+2m}, \quad (3.43)$$

Therefore, we see that the construction on these rectangular matrices over associative composition algebras cover *all* "classical" Lie algebras. Including 1×1 matrices over octonions and tensor products of octonions with other composition algebras (as we have done for the magic square) we get *all* Lie algebras in a unified approach!! (except G_2)[4].

IV. LIE SUPERALGEBRAS FROM SUPERTERNARY ALGEBRAS—PURE FERMION CASE

In this section we consider constructions of Lie superalgebras based on ternary algebras whose elements are purely fermionic. Therefore, we shall use the triple product with a sign change from the bosonic case as indicated in Eq. (2.44)

$$(abc) = a\cdot(\bar{b}\cdot c) + c\cdot(\bar{b}\cdot a) + b\cdot(\bar{a}\cdot c). \quad (4.1)$$

The same sign changes occur in Eqs. (2.38)–(2.39) JSI and KSI, as given by (2.32).

We will specialize to the associative cases: 1) $V = \mathbb{R}, \mathbb{C}, \mathbb{H}$, 2) $V = \mathbb{C}\otimes\mathbb{C}$, $\mathbb{C}\otimes\mathbb{H}$, $\mathbb{H}\otimes\mathbb{H}$, 3) $V = \mathbb{H}\otimes\mathbb{H}\otimes\cdots\otimes\mathbb{H}$ 4), V = rectangular matrices over $\mathbb{R}$, $\mathbb{C}$, $\mathbb{H}$. These are parallel to the Bose case considered above except for octonions $\mathbb{O}$, but generate Lie superalgebras instead of Lie algebras, as we show here.

1a: When $V = \mathbb{R}$, a,b,c are real anticommuting variables, with the trivial conjugation $\bar{a} = a$. Taking into account the anticommuation property $ab = -ba$, the triple product Eq. (4.1) becomes

$$s_{ab}(c) = -abc, \quad k_{ab}(c) = c(ab - ba) = 2cab. \quad (4.2)$$

Therefore, we obtain the number of generators by counting the number of real Fermi or Bose parameters in each class of generators

$$\mathscr{U}_a = \{1 \text{ Fermi generator}\}, \quad \widetilde{\mathscr{U}}_a = \{1 \text{ Fermi generator}\},$$
$$K_{ab} = \{1 \text{ Bose generator}\}, \quad \widetilde{K}_{ab} = \{1 \text{ Bose generator}\},$$
$$S_{ab} = \{1 \text{ Bose generator}\}.$$

Then according to the commutation rules Eq. (2.41) we find the algebras

$$S = E, \quad \mathscr{M} = C_1, \quad \mathscr{L} = B(0,1). \quad (4.4)$$

In denoting the superalgebras we are following the notation of Kac,[3] and we are not distinguishing between compact and noncompact cases.

1b: When $V = \mathbb{C}$, the parameters a, b, c are complex anticommuting numbers with complex conjugation $\bar{a} = a^*$. The triple product (abc) reduces to

$$s_{ab}(c) = -a^*bc, \quad k_{ab}(c) = c(a^*b - b^*a). \quad (4.5)$$

Note that for anticommuting numbers $a^*b - b^*a$ is real and *not* imaginary.

Counting the parameters as usual we arrive at

$$\mathscr{U}_a = \{2 \text{ Fermi generators}\},$$
$$\widetilde{\mathscr{U}}_a = \{2 \text{ Fermi generators}\},$$
$$K_{ab} = \{1 \text{ Bose generators}\},$$
$$\widetilde{K}_{ab} = \{1 \text{ Bose generator}\}, \quad (4.6)$$
$$S_{ab} = \{2 \text{ Bose generators}\}.$$

These combine to give the algebras

$$S = E + E, \quad \mathscr{M} = A_1 + E, \quad \mathscr{L} = A(1,0). \quad (4.7)$$

1c: When $V = \mathbb{H}$, a, b, c are quaternions with anticommuting parameters, which may be written in terms of Pauli matrices $a = a_0 - i\vec{\sigma}\cdot\vec{a}$. The conjugation is just quaternionic conjugation, however now we must watch the *orders* of the Fermions when conjugating products. That is,

$$\overline{ab} = -\bar{b}\bar{a},$$
$$\overline{abc} = -\bar{c}\bar{b}\bar{a}, \quad (4.8)$$
$$\overline{abcd} = +\bar{d}\bar{c}\bar{b}\bar{a},$$

etc.,

where the $(-)$ signs arise from reordering the anticommuting parameters a_i, b_i, etc. The triple product takes the form

$$s_{ab}(c) = a\bar{b}c + c\bar{b}a + b\bar{a}c, \quad k_{ab}(c) = c(\bar{a}b - \bar{b}a). \quad (4.9)$$

Here we see that S_{ab} generates on the quaternion c the left-handed transformation $(a\bar{b} + b\bar{a})c$ and the right-hand transformation $c(\bar{b}a)$. These transformations commute with each other. Using the anticommutation property of a, b, etc., we can write

$$a\bar{b} + b\bar{a} = a\bar{b} - (\overline{a\bar{b}})] \quad (4.10)$$

showing that it is an imaginary quaternion with three Bose parameters. Thus, this is just an SU(2) transformation on c from the left. The right-hand transformation $c(\bar{b}a)$ contains four Bose parameters corresponding to $SU(2)\times U(1)$. By similar arguments $k_{ab}(c)$ is a one-parameter transformation as seen from

$$\bar{a}b - \bar{b}a = \bar{a}b + (\overline{\bar{a}b}) = 2a(a,b), \quad (4.11)$$

where $(a,b) = a_ib_i$ is the scalar product of two quaternions. Summarizing, we have the following numbers for the generators:

$$\mathscr{U}_a = \{4 \text{ Fermi generators}\},$$
$$\widetilde{\mathscr{U}}_a = \{4 \text{ Fermi generators}\},$$
$$K_{ab} = \{1 \text{ Bose generator}\},$$
$$\widetilde{K}_{ab} = \{1 \text{ Bose generator}\}, \quad (4.12)$$
$$S_{ab} = \{7 \text{ Bose generators}\}.$$

These form the algebras

$$S = A_1 + A_1 + E, \quad \mathscr{M} = A_1 + A_1 + A_1, \quad \mathscr{L} = D(2,1). \tag{4.13}$$

2. *Tensor products* $\mathbb{C}\otimes\mathbb{C}$, $\mathbb{C}\otimes\mathbb{H}$, $\mathbb{H}\otimes\mathbb{H}$: The tensor products defined here must be overall anticommuting numbers. This can be achieved by taking one factor bosonic and the other fermionic. Since the factors commute with each other, it is irrelevant which is the fermionic one. We shall use the convention aA where the first factor a is fermionic the second factor A is bosonic. We shall treat the $\mathbb{H}\otimes\mathbb{H}$ case. The others are analogous, and we will give the results at the end. The triple product takes the form

$$s_{aA,bB}(cC) = [a\bar{b}c][A\bar{B}C] + [c\bar{b}a][C\bar{B}A] + [b\bar{a}c][B\bar{A}C],$$
$$k_{aA,bB}(cC) = [c\bar{a}b][C\bar{A}B] - [c\bar{b}a][C\bar{B}A]. \tag{4.14}$$

This can be rewritten as

$$S_{aA,bB}(cC) = [a\bar{b}c + c\bar{b}a + b\bar{a}c][(A,B)C] + [(a,b)c] \times [A\bar{B}C - C\bar{B}A - B\bar{A}C] + [(a,b)c][(A,B)C] + [c(\bar{b}a + \bar{a}b)/2][C(\bar{B}A - \bar{A}B)/2], \tag{4.15}$$

$$k_{aA,bB}(cC) = 2[c(a,b)][C(A,B)] + 2[c(\bar{a}b + \bar{b}a)/2][C(\bar{A}B - \bar{B}A)/2], \tag{4.16}$$

where as usual, we are denoting by (a,b) and (A,B) the scalar product between two quaternions. By taking into account the anticommuting property, the reader can convince himself that the transformations generated by $S_{aA,bB}$ on (cC) are identical to those of the corresponding pure bosonic case. In fact, this is in general true for any associative V with the triple products that we have given for these pure Bose and Fermi cases, Eq. (2.45) with $\phi = 0$ and $\phi = \pi$, respectively. Thus, the results for the algebra S can be obtained from the corresponding Bose case. Here it is given by

$$S = \{22 \text{ Bose generators}\} = A_1 + A_1 + E + D_3. \tag{4.17}$$

On the other hand, with anticommuting parameters the space K generates a transformation different from the corresponding pure Bose case. The number of parameters are obtained from Eq. (4.16)

$(a,b)\otimes(A,B) \to 1$ parameter,

$((\bar{a}b + \bar{b}a)/2)\otimes((\bar{A}B - \bar{B}A)/2) \to 3\times 3 = 9$ parameters, (4.18)

giving a total of 10 parameters. Summarizing, we have

$$\begin{aligned} \mathscr{U}_a &= \{16 \text{ Fermi generators}\},\\ \tilde{\mathscr{U}}_a &= \{16 \text{ Fermi generators}\},\\ K_{ab} &= \{10 \text{ Bose generators}\},\\ \tilde{K}_{ab} &= \{10 \text{ Bose generators}\},\\ S_{ab} &= \{22 \text{ Bose generators}\}. \end{aligned} \tag{4.19}$$

These form the algebras [noting $A_1 + A_1 \approx D(2)$]:

$$S = D_2 + E + D_3, \quad \mathscr{M} = D_2 + C_4, \quad \mathscr{L} = D(2,4). \tag{4.20}$$

We now summarize the results for the $\mathbb{C}\otimes\mathbb{C}$ and $\mathbb{C}\otimes\mathbb{H}$ cases which are analogous to the $\mathbb{H}\otimes\mathbb{H}$ case treated above. The resulting algebras are subalgebras of the $\mathbb{H}\otimes\mathbb{H}$ case by virtue of the fact that the ternary algebras on $\mathbb{C}\otimes\mathbb{C}$ and $\mathbb{C}\otimes\mathbb{H}$ are subcases of $\mathbb{H}\otimes\mathbb{H}$. The results are:

$\mathbb{C}\otimes\mathbb{C}$:

$$\begin{aligned} \mathscr{U}_a &= \{4 \text{ Fermi generators}\},\\ \tilde{\mathscr{U}}_a &= \{4 \text{ Fermi generators}\},\\ K_{ab} &= \{2 \text{ Bose generators}\},\\ \tilde{K}_{ab} &= \{2 \text{ Bose generators}\},\\ S_{ab} &= \{4 \text{ Bose generators}\}, \end{aligned} \tag{4.21}$$

giving the algebras

$$\begin{aligned} S &= E + E + E + E,\\ \mathscr{M} &= A_1 + A_1 + E + E,\\ \mathscr{L} &= A(1,0) + A(1,0). \end{aligned} \tag{4.22}$$

$\mathbb{C}\otimes\mathbb{H}$:

$$\begin{aligned} \mathscr{U}_a &= \{8 \text{ Fermi generators}\},\\ \tilde{\mathscr{U}}_a &= \{8 \text{ Fermi generators}\},\\ K_{ab} &= \{4 \text{ Bose generators}\},\\ \tilde{K}_{ab} &= \{4 \text{ Bose generators}\},\\ S_{ab} &= \{11 \text{ Bose generators}\}, \end{aligned} \tag{4.23}$$

They lead to the algebras

$$S = A_1 + A_1 + E + E, \quad \mathscr{M} = A_3 + A_1 + E, \quad \mathscr{L} = A(3,1). \tag{4.24}$$

We are now ready to write down the analog of the magic square for superalgebras. The results are given in Table II. The octonionic row and column have not been completed at this stage of our work, although for $\mathbb{R}\otimes\mathbb{O}$ or $\mathbb{O}\otimes\mathbb{R}$ we believe we should have the superalgebra $F(4)$ with $S = B_3 + E$, $\mathscr{M} = B_3 + A_1$ and $\mathscr{L} = F(4)$. The properties of the super magic square are clearly analogous to the Bose case.

3. *Higher tensor products of associative composition algebras*: These tensor products are defined by taking an odd number of Fermi factors with any number of Bose factors. However, in our construction an even number of Fermi factors lead to identical results as the same even number of Bose factors. Therefore, it is suficient to specialize to a single Fermi factor with $(n-1)$ Bose factors. As an example, we will work out the case when all factors (Bose or Fermi) are quaternionic. Other cases can be treated similarly. Thus, we take the tensor product (we denote $[a_1\otimes a_2\cdots\otimes a_n]$ by $[a_1,a_2\cdots a_n]$)

$$a = a_1 a_2 \cdots a_n, \tag{4.25}$$

where a_1 is a quaternion with anticommuting parameters and a_i $i = 2,\ldots,n$ are qaternions with Bose parameters. The triple product Eq. (4.1) takes the form

$$s_{ab}(c) = (a_1a_2\cdots a_n)(\bar{b}_1\bar{b}_2\cdots\bar{b}_n)(c_1c_2\cdots c_n) + (c_1c_2\cdots c_n)(\bar{b}_1\bar{b}_2\cdots\bar{b}_n)(a_1a_2\cdots a_n) + (b_1b_2\cdots b_n)(\bar{a}_1\bar{a}_2\cdots\bar{a}_n)(c_1c_2\cdots c_n), \tag{4.26}$$

$$k_{ab}(c) = (c_1c_2\cdots c_n)[(\bar{a}_1\bar{a}_2\cdots\bar{a}_n)(b_1b_2\cdots b_n) - (\bar{b}_1\bar{b}_2\cdots\bar{b}_n)(a_1a_2\cdots a_n)]. \tag{4.27}$$

The transformation generated by S_{ab} is the same as the corresponding Bose case treated in Sec. III, and corresponds to

TABLE II. The super magic square. In each entry the first line refers to S the second to $\mathscr{M} = \widetilde{K} + S + K$ and the third $\mathscr{L} = \widetilde{K} + \widetilde{\mathscr{U}} + S + \mathscr{U} + K$.

	$\mathbb{R}$	$\mathbb{C}$	$\mathbb{H}$	$\mathbb{O}$
$\mathbb{R}$	E C_1 $B(0,1)$	$E+E$ A_1+E $A(1,0)$	A_1+A_1+E $A_1+A_1+A_1$ $D(2,1)$	
$\mathbb{C}$	$E+E$ A_1+E $A(1,0)$	$E+E+E+E$ A_1+A_1+E+E $A(1,0)+A(1,0)$	$A_1+A_1+A_1+E+E$ A_3+A_1+E $A(3,1)$	
$\mathbb{H}$	A_1+A_1+E $A_1+A_1+A_1$ $D(2,1)$	$A_1+A_1+A_1+E+E$ A_3+A_1+E $A(3,1)$	D_2+E+D_3 D_2+C_4 $D(2,4)$	
$\mathbb{O}$				

the algebras

$$n = \text{even}, \quad S = D_{(2^{n-1})} + A_{(2^n-1)} + E,$$
$$n = \text{odd}, \quad S = C_{(2^{n-1})} + A_{(2^n-1)} + E. \tag{4.28}$$

The K space differs from the corresponding Bose case. Counting the number of real parameters in

$$(\bar{a}_1\bar{a}_2\cdots\bar{a}_n)(b_1b_2\cdots b_n) - (\bar{b}_1\bar{b}_2\cdots\bar{b}_n)(a_1a_2\cdots a_n), \tag{4.29}$$

we find $2^{n-1}[2^n + (-1)^n]$ parameters. Therefore, we have

$$\mathscr{U}_a = \{4^n \text{ Fermi generators}\},$$
$$\widetilde{\mathscr{U}}_a = \{4^n \text{ Fermi generators}\},$$
$$K_{ab} = \{2^{n-1}[2^n + (-1)^n] \text{ Bose generators}\}, \tag{4.30}$$
$$\widetilde{K}_{ab} = \{2^{n-1}[2^n + (-1)^n] \text{ Bose generators}\},$$
$$S_{ab} = \{(2^{2n} + 2^{2n-1} - (-1)^n 2^{n-1}) \text{ Bose generators}\},$$

These lead to the following superalgebras:

$$n = \text{even} \quad S = D_{(2^{n-1})} + A_{(2^n-1)} + E,$$
$$\mathscr{M} = D_{(2^{n-1})} + C_{(2^n)}, \quad \mathscr{L} = D(2^{n-1},2^n). \tag{4.31}$$
$$n = \text{odd} \quad S = C_{(2^{n-1})} + A_{(2^n-1)} + E$$
$$\mathscr{M} = C_{(2^{n-1})} + 2_{(2^n)}, \quad \mathscr{L} = D(2^n,2^{n-1}). \tag{4.32}$$

4. *Rectangular matrices over* $\mathbb{R},\mathbb{C},\mathbb{H}$ *with anticommuting parameters*: We consider the ternary algebra of $n\times m$ rectangular matrices

$$a = \begin{pmatrix} a_{11} & \cdots\cdots & a_{1m} \\ \vdots & & \\ a_{n1} & \cdots\cdots & a_{nm} \end{pmatrix}, \tag{4.33}$$

where the entries a_{ij} are taken over $\mathbb{R}$, $\mathbb{C}$, $\mathbb{H}$ with totally anticommuting parameters. Conjugation is defined by taking the transpose and conjugation in the composition algebra:

$$\bar{a} = \begin{bmatrix} \bar{a}_{11} & \cdots & \bar{a}_{1m} \\ \vdots & & \\ \bar{a}_{1m} & \cdots & \bar{a}_{nm} \end{bmatrix}, \tag{4.34}$$

The triple product

$$s_{ab}(c) = (abc) = a\bar{b}c + c\bar{b}a + b\bar{a}c \tag{4.35}$$

closes for such a space. Similarly,

$$k_{ab}(c) = c(\bar{a}b - \bar{b}a). \tag{4.36}$$

S_{ab} generates a transformation on c which corresponds to a left multiplication by $(a\bar{b} + b\bar{a})$ and a right multiplication by $(\bar{b}a)$. Since the parameters are anticommuting, these transformations correspond to a left multiplication by a $n\times n$ anti-Hermitian matrix over the given composition algebra and to a right multiplication by an arbitrary $(m\times m)$ matrix over the same algebra $\mathbb{R}$, $\mathbb{C}$ or $\mathbb{H}$. Clearly, the left and right multiplications commute with each other. To find the number of generators in K_{ab}, we note that $(a\bar{b} - \bar{b}a)$ appearing in $k_{ab}(c)$ is a Hermitian $(m\times m)$ matrix due to the anticommutativity of its parameters. Denoting the dimensions of the composition algebras $\mathbb{R}$, $\mathbb{C}$, $\mathbb{H}$ by $d = 1,2,4$ respectively, we find the following counting of generators:

$$\mathscr{U}_a = \{dmn \text{ Fermi generators}\},$$
$$\widetilde{\mathscr{U}}_a = \{dmn \text{ Fermi generators}\},$$
$$K_{ab} = \left\{\left(d\ \frac{m(m-1)}{2} + m\right) \text{ Bose generators}\right\},$$
$$\widetilde{K}_{ab} = \left\{\left(d\ \frac{m(m-1)}{2} + m\right) \text{ Bose generators}\right\},$$
$$S_{ab} = \left\{\left(d\ \frac{n(n-1)}{2} + (d-1)n + dm^2\right) \text{ Bose generators}\right\}, \tag{4.37}$$

These lead to the following Lie superalgebras:

$\mathbb{R}$: $n = \text{even} = 2k$,

$$S = D_k + A_{m-1} + E, \quad \mathscr{M} = D_k + C_m,$$
$$\mathscr{L} = D(k,m), \tag{4.38}$$
$$n = \text{odd} = 2k+1, \quad S = B_k + A_{m-1} + E,$$
$$\mathscr{M} = B_k + C_m, \quad \mathscr{L} = B(k,m), \tag{4.39}$$

$$\mathbb{C}: S = A_{n-1} + A_{m-1} + A_{m-1} + E + E,$$
$$\mathscr{M} = A_{n-1} + A_{2m-1} + E,$$
$$\mathscr{L} = A(n-1, 2m-1), \tag{4.40}$$
$$\mathbb{H}: S = C_n + A_{2m-1} + E,$$
$$\mathscr{M} = C_n + D_{2m}, \quad \mathscr{L} = D(2m,n). \tag{4.41}$$

This construction of Lie superalgebras can be directly generalized to the case of rectangular matrices taken over tensor products of division algebras.

V. TERNARY ALGEBRAS WITH MIXED BOSE AND FERMI VARIABLES

In this section we consider constructions of Lie superalgebras based on ternary algebras which contain both Bose and Fermi components. Furthermore, we take the conjugation of the generators of type II as explained in Sec. 2. Thus for the triple product we take:

$$(abc) = a\cdot(\overline{b}\cdot c) + c\cdot(\overline{b}\cdot a) - b\cdot(\overline{a}\cdot c). \tag{5.1}$$

Furthermore, we recall that with a conjugation of type II the conditions (JSl) and (KSl) have the same sign patterns as the pure bosonic case as given by Eq. (2.32).

First we shall consider the case of rectangular matrices taken over R which have the form

$$a = \begin{matrix} & k \quad k \\ l & \\ l & \end{matrix}\left[\begin{array}{c|c} A & \xi \\ \hline \eta & B \end{array}\right] \tag{5.2}$$

where the A and B are bosonic $(l)\times(k)$ matrices and the η and ξ are fermionic $(l)\times(k)$ matrices. The conjugation in this space will be defined as

$$\bar{a} = \begin{matrix} & l \quad l \\ k & \\ k & \end{matrix}\left[\begin{array}{c|c} -B^T & \xi^T \\ \hline -\eta^T & -A^T \end{array}\right]. \tag{5.3}$$

With the product $(\cdot)$ being the ordinary matrix multiplication the triple product defined above gives

$$s_{a_1a_2}(a_3) = (a_1\bar{a}_2 - a_2\bar{a}_1)a_3 + a_3(\bar{a}_2a_1), \tag{5.4}$$

$$k_{a_1a_2}(a_3) = a_3(\bar{a}_2a_1 - \bar{a}_1a_2). \tag{5.5}$$

Now, the left multiplication by the $2l\times 2l$ matrix $(a_1\bar{a}_2 - a_2\bar{a}_1)$ has the form

$$\begin{pmatrix} C & \beta \\ \gamma & -C^T \end{pmatrix}, \tag{5.6}$$

where β is an $l\times l$ symmetric matrix $\beta^T = \beta$, and γ is an $l\times l$ antisymmetric matrix $\gamma^T = -\gamma$, and both β and γ are fermionic. C is a general bosonic matrix. Such $2l\times 2l$ matrices (with traceless C) close under commutation to form the Lie superalgebra P_{l-1}. The right multiplication by the $2k\times 2k$ matrix $(\bar{a}_2a_1)$ generates the Lie superalgebra $A(k-1,k-1)+E$.

Counting the number of independent parameters, we find (note that the supertrace of $a_1\bar{a}_2 - a_2\bar{a}_1$ is equal to two times the supertrace of $\bar{a}_2a_1$, hence 1 less independent parameter in S_{ab})

$$\begin{aligned}
\mathscr{U}_a &= \{2lk \text{ Bose generators}, \ 2lk \text{ Fermi generators}\},\\
\widetilde{\mathscr{U}}_b &= \{2lk \text{ Bose}, 2lk \text{ Fermi generators}\},\\
K_{ab} &= \{k^2 \text{ Bose}, k^2 \text{ Fermi generators}\},\\
\widetilde{K}_{ab} &= \{k^2 \text{ Bose}, \ k^2 \text{ Fermi generators}\}, \qquad (5.7)\\
S_{ab} &= \{l^2+2k^2-1) \text{ Bose},(l^2+2k^2) \text{ Fermi generators}\},
\end{aligned}$$

These generators form the following superalgebras

$$\begin{aligned}
\mathbb{R}: S &= P(l-1) + A(k-1,k-1) + E,\\
\mathscr{M} &= P(l-1) + P(2k-1) + E, \qquad (5.8)\\
\mathscr{L} &= P(l+2k-1).
\end{aligned}$$

The above procedure can be repeated when the elements of the ternary algebra Eq. (5.2) are taken over the complex commuting and anticommuting numbers $\mathbb{C}$. In this case the counting of the generators gives

$$\begin{aligned}
\mathscr{U}_a &= \{4lk \text{ Bose generators}, 4lk \text{ Fermi generators}\},\\
\widetilde{\mathscr{U}}_a &= \{4lk \text{ Bose}, \ 4lk \text{ Fermi generators}\},\\
K_{ab} &= \{2k^2 \text{ Bose}, \ 2k^2 \text{ Fermi generators}\},\\
\widetilde{K}_{ab} &= \{2k^2 \text{ Bose}, 2k^2 \text{ Fermi generators}\}, \qquad (5.9)\\
S_{ab} &= \{(2l^2+4k^2-1)\text{Bose},(2l^2+4k^2)\text{Fermigenerators}\},
\end{aligned}$$

The resulting superalgebras are

$$\begin{aligned}
\mathbb{C}: S &= A(l-1,l-1) + A(k-1,k-1)\\
&\quad + A(k-1,k-1) + E + E,\\
\mathscr{M} &= A(l-1,l-1) + A(2k-1,2k-1) + E, \qquad (5.10)\\
\mathscr{L} &= A(l+2k-1,l+2k-1).
\end{aligned}$$

Finally we take the ternary algebra over the quaternions $\mathbb{H}$. We find that the number of generators in $\mathscr{U}_a$, $\widetilde{\mathscr{U}}_a$, K_{ab}, $\widetilde{K}_{ab}$ are quadrupled compared to the real case $\mathbb{R}$. The resulting superalgebras are

$$\begin{aligned}
\mathbb{H}: S &= P(2l-1) + A(2k-1,2k-1) + E,\\
\mathscr{M} &= P(2l-1) + P(4k-1) + E, \qquad (5.11)\\
\mathscr{L} &= P(2l+4k-1).
\end{aligned}$$

VI. JORDAN SUPER TERNARY ALGEBRAS

In this section we shall consider the construction of Lie superalgebras over Jordan super ternary algebras. We recall that by definition a Jordan super ternary algebra has a triple product (abc) which is symmetric under $a\leftrightarrow c$ interchange [as in Eq. (2.40)].

$$(abc) = (cba), \tag{6.1}$$

and satisfies the identity JSl of Theorem 1. An example of such triple products was given in Eq. (2.46). Since, as mentioned before, a class of such algebras can be constructed via Jordan products $a\cdot b = \frac{1}{2}(ab+ba)$, we may use Jordan superalgebras to construct Lie superalgebras. All Jordan superalgebras have been classified by Kac.[7]

We find that every Jordan superalgebra under the triple product of Eq. (2.44) with the trivial conjugation, defines a Jordan superternary algebra:

$$(abc) = a\cdot(b\cdot c) + c\cdot(b\cdot a) - b\cdot(a\cdot c). \tag{6.2}$$

We specialize to special Jordan algebras $a\cdot b = \frac{1}{2}(ab+ba)$ with a,b associative matrices for which this triple product reduces to

$$(abc) = \tfrac{1}{2}(abc+cba), \tag{6.3}$$

and is a special case of Eq. (2.46). From this form it is evident that the K_{ab} space vanishes and the grading is three-dimensional. Following the classification of Kac[7] for the Jordan superalgebras, we construct the corresponding super ternary algebras and Lie superalgebras.

1: Consider the Jordan superalgebras defined by matrices of the form [$(m+n)\times(m+n)$ matrices]

$$a=\begin{matrix} & m & n \\ m & A & \xi \\ n & \eta & B \end{matrix} \qquad (6.4)$$

where A, B are bosonic and ξ and η are fermionic. These matrices close under multiplication. The triple product gives

$$(abc)=\tfrac{1}{2}(abc+cba). \qquad (6.5)$$

This has the form of a left multiplication by (ab) and right multiplication by (ba) which are independent. Thus we have

$$\mathscr{U}_a=\{m^2+n^2 \text{ Bose generators}, 2mn \text{ Fermi generators}\},$$
$$\widetilde{\mathscr{U}}_a=\{m^2+n^2 \text{ Bose generators}, 2mn \text{ Fermi generators}\}, \qquad (6.6)$$
$$S_{ab}=\{2(m^2+n^2) \text{ Bose}, 4mn \text{ Fermi generators}\},$$

The resulting algebras are

$$S=A(m-1,n-1)+A(m-1,n-1)+E,$$
$$\mathscr{L}=A(2m-1,2n-1). \qquad (6.7)$$

2: Consider the Jordan superalgebras generated by matrices of the form [$(2n+m)\times(2n+m)$ matrices]

$$a=\begin{matrix} & m & n & n \\ m & A & \xi & \eta \\ n & \eta^T & B & C \\ n & -\xi^T & D & B^T \end{matrix}, \qquad (6.8)$$

where A,B,C,D are bosonic and ξ and η are fermionic, and A is symmetric, C and D are antisymmetric.

$$A^T=A, \quad C^T=-C, \quad D^T=-D. \qquad (6.9)$$

These matrices close under anticommutation. The triple product

$$(abc)=(abc+cba), \qquad (6.10)$$

can be regarded as simultaneous left and right multiplication by (ab) and (ba), respectively. Note that just left or just right multiplication does not maintain the form of the matrix c. The number of parameters in the transformation is obtained either from ab or ba (not the sum). To see the independent transformations clearly, one can rewrite the above triple product in the form

$$(abc)=\tfrac{1}{2}[[a,b],c]+\tfrac{1}{2}\{\{a,b\},c\}, \qquad (6.11)$$

where the first term is interpreted as derivations and the second term as multiplication by the elements of the algebra, generating the so called structure algebra.[17] Each term corresponds to independent transformations maintaining the form of the matrix c. The first term generates a subalgebra; in this case it is the $B((m-1)/2,n)$ or $D(m/2,n)$ for $m=$ odd integer or $m=$ even integer respectively. Counting the parameters in (ab), we find

$$\mathscr{U}_a=\left\{\left[\frac{m(m+1)}{2}+n(2n-1)\right] \text{ Bose}, 2mn \text{ Fermi generators}\right\},$$
$$\widetilde{\mathscr{U}}_a=\left\{\left[\frac{m(m+1)}{2}+n(2n-1)\right] \text{ Bose}, 2mn \text{ Fermi generators}\right\}, \qquad (6.12)$$
$$S_{ab}=\{[m^2+4n^2] \text{ Bose}, 4mn \text{ Fermi generators}\},$$

The superalgebras they generate are

$$S=A(m-1,2n-1)+E, \quad \mathscr{L}=D(2n,m). \qquad (6.13)$$

3: We now consider the Jordan superalgebra defined by the matrices of the form

$$a=\begin{matrix} & n & n \\ n & A & \xi \\ n & \eta & A^T \end{matrix}, \qquad (6.14)$$

where A is an $n\times n$ bosonic matrix ξ is a symmetric $n\times n$ fermionic matrix and η is an $n\times n$ antisymmetric fermionic matrix. Such matrices close under anticommutation. Then the triple product reduces to

$$(abc)=abc+cba=\tfrac{1}{2}[[a,b],c]+\tfrac{1}{2}\{\{a,b\},c\}. \qquad (6.15)$$

From the second form we see that the derivations generated by the commutator term $[a,b]$ correspond to a $P(n-1)$ transformation on c. The remaining anticommutator term $\{a,b\}$ has n^2 Bose and n^2 Fermi generators. Thus, we find

$$\mathscr{U}_a=\{n^2 \text{ Bose}, n^2 \text{ Fermi generators}\},$$
$$\widetilde{\mathscr{U}}_a=\{n^2 \text{ Bose}, n^2 \text{ Fermi generators}\}, \qquad (6.16)$$
$$S_{ab}=\{2n^2-1 \text{ Bose}, 2n^2 \text{ Fermi generators}\},$$

The superalgebras they generate are

$$S=A(n-1,n-1), \quad \mathscr{L}=P(2n-1). \qquad (6.17)$$

4: Consider the Jordan superalgebra defined by matrices of the form

$$a=\begin{matrix} & n & n \\ n & A & \xi \\ n & \xi & A \end{matrix}. \qquad (6.18)$$

These matrices close under ordinary multiplication and of course under Jordan product. The triple product can be written as in the previous case. The derivation algebra $[a,b]$ on c has (n^2-1) Bose and (n^2-1) Fermi parameters corresponding to $Q(n-1)$. The remaining anticommutator algebra has n^2 Bose and n^2 Fermi parameters. Thus, we obtain

$$\mathscr{U}_a=\{n^2 \text{ Bose}, n^2 \text{ Fermi generators}\},$$
$$\widetilde{\mathscr{U}}_a=\{n^2 \text{ Bose}, n^2 \text{ Fermi generators}\}, \qquad (6.19)$$
$$S_{ab}=\{(2n^2-1) \text{ Bose}, (2n^2-1) \text{ Fermi generators}\},$$

The superalgebras they generate are

$$S=Q(n-1)+Q(n-1)+E+F$$
$$\mathscr{L}=Q(2n-1) \qquad (6.20)$$

Here we have denoted the 1 Fermi parameter Lie superalgebra [analogous to U(1) group] as F.

This list of Jordan type constructions is not complete. The remaining Jordan algebras in Kac's classification including the exceptional ones (D, D_t, E, F, K) can be used in conjunction with our construction to form Lie superalgebras. These will be discussed elsewhere.[8]

5: As mentioned previously, the triple product

$$(abc) = a\bar{b}c + c\bar{b}a, \tag{6.21}$$

with a nontrivial conjugation $\bar{b} \neq b$ is also a Jordan super ternary algebra, however it cannot be written in terms of Jordan products. Here we give one example of this case. Consider rectangular matrices of the form

$$\begin{pmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & & \\ \cdot & & \\ \cdot & & \\ a_{m1} & \cdots & a_{mn} \end{pmatrix}. \tag{6.22}$$

where a_{ij} are real Fermi parameters. According to the triple product above, they generate left and right transformations on c by the $n \times n$ matrix $a\bar{b}$ and $m \times m$ matrix $\bar{b}a$, respectively. Thus we have the counting

$$\begin{aligned} \mathscr{U}_a &= \{mn \text{ Fermi generators}\}, \\ \bar{\mathscr{U}}_a &= \{mn \text{ Fermi generators}\}, \\ S_{ab} &= \{(n^2 + m^2 - 1) \text{ Bose generators}\}, \end{aligned} \tag{6.23}$$

They form the algebras

$$S = A_{n-1} + A_{m-1} + E, \mathscr{L} = A(n-1, m-1). \tag{6.24}$$

Clearly, the above algebra may be taken over complex and quaternionic parameters leading to further generalizations.

VII. OPEN PROBLEMS AND POSSIBLE PHYSICAL APPLICATIONS

In this paper we did not discuss the superalgebras $D(2,1;\alpha)$, $G(3)$ and $F(4)$. These are associated with octonions, and we will present them in a separate publication.[8] The Cartan superalgebras[3] are also under investigation[8] by the method of ternary algebras.

The incomplete super magic square of Table II presents an interesting mathematical question: Is there a set of superalgebras that correspond to a complete magic square? The triple product of Eq. (2.44) fails to satisfy the conditions of Theorem 1 when the elements of the ternary algebra cover the simplest possibilities, namely direct products of octonions with other composition algebras. On the other hand, Kantor[4] has shown, in the bosonic case, that exceptional Lie groups can be constructed from a variety of ternary algebras. We conjecture that a complete super magic square exists and can be constructed by an appropriate superternary algebra. Further investigation of this point is in progress.

It would be interesting to embark on a complete classification of ternary algebras and superternary algebras and provide a list of all possible constructions of a given Lie (super) algebra from (super) ternary algebras. A partial list will be presented by us in a future publication.

An important question for physicists is whether ternary algebras play a role in physical theories. Modern physics relies heavily on symmetries, as illustrated most dramatically in "gauge theories" of particle physics,[1,2] as well as many other disciplines in physics. In this paper we have illustrated the mathematically fundamental nature of (super) ternary algebras in relation to Lie (super) algebras. The fact that Lie (super) algebras are completely characterized by the much smaller ternary algebras is of special significance. It is then natural to ask whether (super) ternary algebras may play a corresponding fundamental role in the formulation of physical theories and whether they may acquire a fundamental physical meaning? Although speculative, we will suggest some areas for investigation:

(1) Unified gauge theories[1] introduce a variety of degrees of freedom. A large number of these degrees of freedom are not observable either because they correspond to superheavy particles, or because they carry "color." The number of "observed" quarks is growing, and there is no convincing theoretical argument that limits the number of fundamental constituents. If more and more quark "flavors" are discovered in experiments, it will be very doubtful that unified gauge theories in their present form are fundamental theories. What, then is the alternative? It may be possible that theories with *fewer degrees of freedom* can simulate the successful aspects of gauge theories, thus giving a more fundamental view of Nature. The partially successful picture provided by gauge theories relies heavily on gauge symmetry; therefore, a new theory should reproduce such features. Since (super) ternary algebras characterize Lie (super) algebras with *fewer* parameters, it is natural to consider them as possible building blocks in more fundamental physical theories.

(2) Nambu[18] has generalized Hamiltonian dynamics to systems involving two Hamiltonians H_1 and H_2 in such a way that the time development of an operator $F(t)$ is given by $(dF/dt) = (H_1, H_2, F)$. The triple product (H_1, H_2, F) that occurs in this formulation can be viewed as forming a ternary algebra. Thus, in this framework our (super) ternary algebra formalism may be used to provide examples and further generalize Nambu's mechanics by relating it to the theory of Lie (super) algebras.

(3) If quantum mechanics is cast into a density matrix formalism, it can be shown how some of the well known structures can be reformulated[19,20] in terms of Jordan triple products which form ternary algebras. Thus, the method of ternary algebras is a natural path to consider generalizations of quantum mechanics such as octonionic quantum mechanics.[15,20] In fact, some recent attempts to try to incorporate the color degrees of freedom through octonionic quantum mechanics,[21,22] suggest some algebraic structures which are just special cases of Kantor's ternary algebras.[4] Thus, our supergeneralization of ternary algebras defined by the conditions JS1 and KS1 of Theorem 1, provide a natural framework to investigate generalizations of quantum mechanics.

(4) It is both of mathematical[7] and physical interest to consider gradings larger than 5. For example, O(8) supergravity has a nine-dimensional grading structure. This may or may not be associated with the grading of a super group. Mathematically, ternary algebras are not in general sufficient to describe such gradings for all groups. It would be interesting to study larger grading structures in general. Some work in this direction can be found in Kantor's paper.[4] To illustrate a possibly physically interesting case, consider

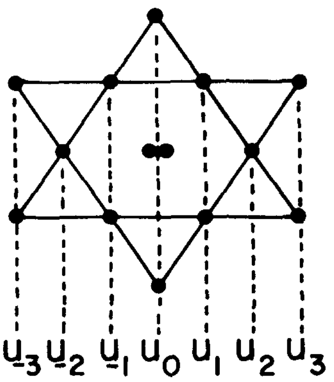


FIG. 1.

the seven-dimensional grading of G_2 with respect to the third component of an isospin subgroup (formed with short roots) as illustrated by the Cartan root diagram in Fig. 1. $\mathscr{U}_0$ forms a 4-parameter SU(2)×U(1) subgroup, $\mathscr{U}_{-3}+\mathscr{U}_0+\mathscr{U}_3$ forms an 8-parameter SU(3) subgroup, $\mathscr{U}_{-2}+\mathscr{U}_0+\mathscr{U}_2$ forms a 6-parameter SU(2)×SU(2) subgroup. Similar grading patterns can be envisaged for many Lie groups, including exceptional groups as well as supergroups. We shall call this type of seven-dimensional structure G(2) structure whenever it applies to a given (super) algebra. Since $\mathscr{U}_{\pm 1}$ generally play a more fundamental role, we may try to give them a fundamental physical meaning. Thus, let us associate $\mathscr{U}_{\pm 1}$ with some fundamental constituents. We will try to assign to these constituents a role similar to the more conventional quarks (or perhaps subquarks??). Then it is natural to think of the remaining generators as follows:

$\mathscr{U}_1=$ "quarks", $\mathscr{U}_{-1}=$ "antiquark",

$\mathscr{U}_2=$ "diquark, $\mathscr{U}_{-2}=$ "anti-diquark",

$\mathscr{U}_3=$ "baryon", $\mathscr{U}_{-3}=$ "antibaryon",

$\mathscr{U}_0=$ "meson".

If this is done for a superalgebra where $\mathscr{U}_{\pm 1}$(= "quarks") are purely fermionic, then $\mathscr{U}_{\pm 3}$(= "baryons") will also be fermionic. Furthermore, $\mathscr{U}_{-3}+\mathscr{U}_0+\mathscr{U}_3$(= "Baryons-" + "Mesons") will form a closed (super) algebra. Although at present we do not have a realistic scheme to compare to Nature, we find this structure amusing and worth of further study and speculation.

It is our belief that the (super) ternary algebraic approach is useful not only for formulating the known physics in a more general, perhaps more fundamental framework, but also for discovering new relevant structures and studying their symmetry properties.

ACKNOWLEDGMENTS

We had the pleasure of discussing our work with many physicists and mathematicians. Our special thanks are due to Professors Feza Gürsey, Murray Gell-Mann, Victor Kac, Howard Garland, George Seligman, Henri Ruegg, Richard Slansky, Gabor Domokos, S. Kövesi-Domokos and Jonathan Rosner. This work was initiated in the stimulating atmosphere of the Aspen Center for Physics, we thank the center for its hospitality. We also thank Harvard University where this work was completed, for the hospitality shown to both of us. One of us (M.G.) gratefully acknowledges the hospitality extended to him by Yale University and the Los Alamos Scientific Laboratory. Finally, I.B. would like to thank the A.P. Sloan Foundation for their fellowship and M.G. would like to thank the Deutsche Forschungsgemeinschaft for a travel grant.

APPENDIX

1: In this part we prove Theorem 1 which specifies the conditions imposed by the Jacobi identities on the ternary algebra.

Let us denote the Jacobi identity satisfied by three generators L_1,L_2,L_3 by $J(L_1L_2L_3)=0$,

$$J(L_1L_2L_3)\equiv[[L_1,L_2],L_3]+[[L_2,L_3],L_1]+[[L_3,L_1],L_2]=0.$$

The complete list of Jacobi identities to be satisfied by the operators in $\mathscr{L}=\tilde{K}\oplus\tilde{\mathscr{U}}\oplus S\oplus\mathscr{U}\oplus K$ are

$$J(\mathscr{U}_1\mathscr{U}_2\mathscr{U}_3)=0, \tag{A1}$$
$$J(\tilde{\mathscr{U}}_1\tilde{\mathscr{U}}_2\tilde{\mathscr{U}}_3)=0, \tag{A2}$$
$$J(K_1K_2K_3)=0, \tag{A3}$$
$$J(\tilde{K}_1\tilde{K}_2\tilde{K}_3)=0, \tag{A4}$$
$$J(\tilde{K}_1\tilde{K}_2\tilde{\mathscr{U}}_3)=0, \tag{A5}$$
$$J(K_1K_2\mathscr{U}_3)=0, \tag{A6}$$
$$J(K_1K_2\tilde{\mathscr{U}}_3)=0, \tag{A7}$$
$$J(\tilde{K}_1\tilde{K}_2\mathscr{U}_3)=0, \tag{A8}$$
$$J(K_1K_2S_3)=0, \tag{A9}$$
$$J(\tilde{K}_1\tilde{K}_2S_3)=0, \tag{A10}$$
$$J(\mathscr{U}_1\mathscr{U}_2K_3)=0, \tag{A11}$$
$$J(\tilde{\mathscr{U}}_1\tilde{\mathscr{U}}_2\tilde{K}_3)=0, \tag{A12}$$
$$J(K_1\mathscr{U}_2S_3)=0, \tag{A13}$$
$$J(\tilde{K}_1\tilde{\mathscr{U}}_2S_3)=0, \tag{A14}$$
$$J(\mathscr{U}_1\mathscr{U}_2\tilde{\mathscr{U}}_3)=0, \tag{A15}$$
$$J(\tilde{\mathscr{U}}_1\tilde{\mathscr{U}}_2\mathscr{U}_3)=0, \tag{A16}$$
$$J(\mathscr{U}_1\mathscr{U}_2\tilde{K}_3)=0, \tag{A17}$$
$$J(\tilde{\mathscr{U}}_1\tilde{\mathscr{U}}_2K_3)=0, \tag{A18}$$
$$J(\mathscr{U}_1\tilde{\mathscr{U}}_2S_3)=0, \tag{A19}$$
$$J(\mathscr{U}_1\tilde{\mathscr{U}}_2K_3)=0, \tag{A20}$$
$$J(\mathscr{U}_1\mathscr{U}_2S_3)=0, \tag{A21}$$
$$J(\mathscr{U}_1\tilde{\mathscr{U}}_2\tilde{K}_3)=0, \tag{A22}$$
$$J(\tilde{\mathscr{U}}_1\tilde{\mathscr{U}}_2K_3)=0, \tag{A23}$$
$$J(S_1S_2\mathscr{U}_3)=0, \tag{A24}$$
$$J(S_1S_2\tilde{\mathscr{U}}_3)=0, \tag{A25}$$
$$J(S_1K_2\tilde{\mathscr{U}}_3)=0, \tag{A26}$$
$$J(S_1\tilde{K}_2\mathscr{U}_3)=0, \tag{A27}$$
$$J(K_1\tilde{K}_2\mathscr{U}_3)=0, \tag{A28}$$
$$J(K_1\tilde{K}_2\tilde{\mathscr{U}}_3)=0, \tag{A29}$$
$$J(S_1S_2S_3)=0, \tag{A30}$$

$$J(S_1 S_2 K_3) = 0, \tag{A31}$$
$$J(S_1 S_2 \tilde{K}_3) = 0, \tag{A32}$$
$$J(K_1 \tilde{K}_2 S_3) = 0, \tag{A33}$$
$$J(K_1 K_2 \tilde{K}_3) = 0, \tag{A34}$$
$$J(\tilde{K}_1 \tilde{K}_2 K_3) = 0. \tag{A35}$$

Because of the grading structure identities (A1)–(A14) are trivially satisfied. Identities (A15)–(A23) were already imposed in deriving the commutation rules Eqs. (2.41) among the generators. So far no conditions are required on the triple product (abc). The remaining identities can no longer be true with an arbitrary triple product. From imposing Eq. (A24) and Eq. (A26), the conditions JSl and KSl of Theorem 1, respectively, can be derived as shown below. It can be proven that the remaining Jacobi identities are automatically satisfied provided (JSl) and (KSl) are true. Furthermore, in Eq. (2.41) it is seen that $[S,S]$, $[S,K]$, $[S,\tilde{K}]$ and $[K,\tilde{K}]$ can be expressed in two different ways; the equivalence of these expressions can be demonstrated if (JSl) and (KSl) are satisfied. Now, we show how (JSl) and (KSl) are derived: Let us write out the Jacobi identities

$$J(S_1 S_2 \mathscr{U}_3) = 0 = J(S_1 K_2 \tilde{\mathscr{U}}_3):$$
$$[[S_{ab},S_{cd}],\mathscr{U}_x] + [[S_{cd},\mathscr{U}_x],S_{ab}] + [[\mathscr{U}_x,S_{ab}],S_{cd}] = 0,$$
$$[[S_{ab},K_{cd}],\tilde{\mathscr{U}}_x] + [[K_{cd},\tilde{\mathscr{U}}_x],S_{ab}] + [[\tilde{\mathscr{U}}_x,S_{ab}],K_{cd}] = 0.$$

Now we substitute Eqs. (2.41) in them and evaluate the commutators. We find that each of these equations reduces to an expression of the form $\mathscr{U}_\alpha = 0$ which implies that the parameter $\alpha = 0$. These are the conditions JSl and KSl

$$(ab(cdx)) - (cd(abx)) + ((a\tilde{s}_{cd}(b)x) + ((cda)bx) = 0,$$
$$\{(ax(cbd)) - ((cbd)xa) + (ab(cxd))$$
$$- ((c\tilde{s}_{ab}(x)d)\} - \{c \leftrightarrow d\} = 0.$$

This proves Theorem 1.

2: Let us now prove that these identities are satisfied by the ternary algebras considered in this paper. We will first consider the (super) triple product of Eq. (2.44) which is in general nonassociative. We will omit the product symbol (·) for ease of writing, but will keep the parentheses. Note that in calculating these nested triple products, in some terms we need to evaluate $(\overline{cda})$, $(\overline{dcb})$, etc. We remind the reader that the conjugation in the algebra changes the order of matrices, quaternions, octonions, etc., and that whenever Fermi parameters need to be permuted, some (-1) factors must be taken into account. For the super ternary algebras considered in this paper the result of these considerations is as follows: If the triple product is

$$(abc) = a(\overline{bc}) + c(\overline{ba}) - e^{i\phi} b(\bar{a}c),$$

then the conjuation gives (with $e^{i2\phi} = 1$)

$$(\overline{abc}) = e^{i\phi}(\bar{c}b)\bar{a} + e^{i\phi}(\bar{a}b)\bar{c} - (\bar{c}a)\bar{b},$$

where the angle $\phi(a,b,c)$ is given in Eq. (2.32).

Inserting this triple product in JSl and rearranging terms, we obtain the expression

$$\begin{aligned}(\mathrm{JSl}) = {} & a\{\bar{b}[c(\bar{d}x)]\} - a\{[(\overline{bc})\bar{d}]x\} - c\{\bar{d}[a(\overline{bx})]\} \\ & + [c(\overline{da})](\overline{bx}) - x\{[(\overline{dc})\bar{b}]a\} + [x(\overline{dc})](\overline{ba}) \\ & + x\{\bar{b}[a(\overline{dc})]\} - [x(\overline{ba})](\overline{dc}) + a\{\bar{b}[x(\overline{dc})]\} \\ & - [a(\overline{bx})](\overline{dc}) - c\{\bar{d}[x(\overline{ba})]\} + [c(\overline{dx})](\overline{ba}) \\ & - a\{[(\overline{dc})\bar{b}]x\} + [a(\overline{dc})](\overline{bx}) + x\{\bar{b}[c(\overline{da})] \\ & - x\{[(\overline{bc})\bar{d}]a\} + b\{\bar{a}[d(\bar{c}x)]\} - b\{[(\bar{a}d)\bar{c}]x\} \\ & - d\{\bar{c}[b(\bar{a}x)]\} + [d(\bar{c}b)](\bar{a}x) - b\{[(\bar{c}d)\bar{a}]x\} \\ & + [b(\bar{c}d)](\bar{a}x) - e^{i\phi}\big(a\{\bar{b}[d(\bar{c}x)]\} - a\{[(\overline{bd})\bar{c}]x\} \\ & - d\{\bar{c}[a(\overline{bx})]\} + [d(\bar{c}a)](\overline{bx}) + x\{\bar{b}[d(\bar{c}a)]\} \\ & - x\{[(\overline{bd})\bar{c}]a\} - d\{\bar{c}[x(\overline{ba})]\} + [d(\bar{c}x)](\overline{ba}) \\ & + b\{\bar{a}[c(\bar{d}x)]\} - b\{[(\bar{a}c)\bar{d}]x\} - c\{\bar{d}[b(\bar{a}x)]\} \\ & + [c(\bar{d}b)](\bar{a}x) + b\{\bar{a}[x(\overline{dc})]\} - [b(\bar{a}x)](\overline{dc})\big),\end{aligned}$$

where we have used $e^{i2\phi} = 1$ for the pure Bose and Fermi cases. For the mixed cases considered in this paper this relation is still satisfied despite the fact that various angles $\phi(a,b,c), \phi(c,d,x), \phi(a,b,(cdx))$, etc. could occur. This results from the conjugation properties of the super ternary algebras we considered e.g. in Eqs. (5.2) and (5.3).

From the expression above we immediately conclude that if the multiplication is *associative* (i.e., can remove all parentheses) neighboring terms cancel each other so that (JSl) = 0 [Eq. (2.38)] is automatically satisfied, while nonassociative cases have to be considered one by one. We will not discuss the nonassociative super ternary algebras in this paper, but for the ordinary Bose case we refer the reader to Kantor's paper.[4]

Similar considerations apply to the condition (KSl) in Eq. (2.39). This equation acquires a simpler form because of the antisymmetry under the interchange of $c \leftrightarrow d$. Noting [for the triple product (2.44)]

$$(cxd) - (dxc) = e^{i\phi} x(\overline{dc} - \bar{c}d),$$

and inserting in (KSl) we find, after using $e^{i2\phi} = 1$ as in the previous case,

$$\begin{aligned}(\mathrm{KSl}) = {} & x\{\bar{a}[b(\bar{c}d - \overline{dc})]\} - [x(\bar{a}b)](\bar{c}d - \overline{dc}) \\ & + b\{\bar{a}[x(\bar{c}d - \overline{dc})]\} - [b(\bar{a}x)](\bar{c}d - \overline{dc}) \\ & + e^{i\phi}\big(x\{[(\bar{c}d - \overline{dc})\bar{b}]a\} - [x(\bar{c}d - \overline{dc})](\overline{ba}) \\ & + [a(\overline{bx})](\bar{c}d - \overline{dc}) - a\{\bar{b}[x(\bar{c}d - \overline{dc})]\}\big).\end{aligned}$$

Again, if the product is associative, this expression is automatically zero, as desired.

Similar considerations apply to the triple product in Eq. (2.46), which is associative. In this case (KSl) is automatically satisfied, since this triple product is symmetric, i.e., $(cxd) = (dxc)$, while (JSl) is also satisfied by explicit calculation.

Thus, we have proven Theorem 2.

[1]For reviews, see E.S. Abers and B.W. Lee, Phys. Rep. **9**, 1 (1973); S. Weinberg, Rev. Mod. Phys. **46**, 255 (1974); H. Pagels and W. Marciano, Phys. Reports C **36**, 137 (1978).
[2]See, e.g., S. Ferrara, "Supersymmetric theories of fundamental interac-

tions (CERN preprint) TH. 2514-CERN, and extensive references therein.
[3a]V.G. Kac, Funct. Anal. Appl. (1975); Djokovic and Hocshild, J. Pure Appl. Math. (1976); A. Pais and V. Rittenberg, J. Math. Phys. **16**, 2062 (1975); M. Scheunert, W. Nahm, and V. Rittenberg, J. Math. Phys. **17**, 1626 (1976).
[3b]P.G.O. Freund and I. Kaplansky, J. Math. Phys. **17**, 228 (1976); V.G. Kac, Commun. Math. Phys. **53**, 31 (1977); V.G. Kac, Adv. Math. **26**, 8 (1977). The last references contain the complete list of Lie superalgebras.
[4]I.L. Kantor, Sov. Math. Dokl. **14**, 254 (1973), and Trudy Sem. Vector. Anal. **16**, 407 (1972) (Russian).
[5]A subclass of ternary algebras with unital element, called "structurable algebras" are treated by B.N. Allison, Am. J. Math. **98**, 285 (1976); Trans. Am. Math. Soc. **114**, 75 (1976); "A class of nonassociative algebras with involution containing the class of Jordan algebras" (preprint, unpublished).
[6]J. Tits, Nederl. Akad. Wetensch., Proc. Ser. A **65**, 530 (1962); M. Koecher, Am. J. Math. **89**, 787 (1967); J.R. Faulkner, Trans. Am. Math.Soc. **155**, 397 (1971).
[7]V.G. Kac, Commun. Alg. **5**, 1375 (1977).
[8]I. Bars and M. Günaydin, in preparation.
[9]N. Jacobson, "Structure and Representation of Jordan Algebras," Am. Math. Soc. Colloq. Publ. **39** (1968).
[10]S. Helgason, "Differential Geometry and Symmetric Spaces" (Academic, New York, 1962).
[11]M. Günaydin and F. Gürsey, J. Math. Phys. **14**, 1651 (1973).
[12]R.D. Schafer, *An Introduction to Non-Associative Algebras* (Academic, New York, 1966).
[13]See e.g., Refs. 4, 5, 9, 12.
[14]See e.g., Ref. 5.
[15]P. Jordan, J. von Neumann, and E. Wigner, Ann. Math. **36**, 29 (1934).
[16]H. Freudenthal, Adv. Math. **1**, 145 (1965).
[17]For the corresponding Bose case, the term "structure algebra" is used in the mathematical literature, see, e.g., Ref. 5.
[18]Y. Nambu, Phys. Rev. D **7**, 2405 (1973).
[19]M. Günaydin, in "Second Workshop on Current Problems in High Energy Particle Theory, 1978," edited by G. Domokos and S. Kövesi Domokos (John Hopkins U., Baltimore, Maryland, 1978).
[20]F. Gürsey, in "Second Workshop on Current Problems in High Energy Particle Theory, 1978."
[21]M. Günaydin and F. Gürsey, to appear.
[22]M. Gell-Mann and J. Rosner, to appear.