Quasigroup construction and first class constraints

I. A. Batalin
Lebedev Physical Institute, Moscow, U.S.S.R.

(Received 25 March 1980; accepted for publication 20 June 1980)

A generalization of the Lie group construction is proposed wherein the composition law depends, apart from the parameters of the transformations composed, also on the transformed variables. This construction is met, in particular, on the hypersurfaces specified by the first class constraints in phase space.

PACS numbers: 02.20.Qs

INTRODUCTION

A generalization of the Lie group construction henceforth referred to as a quasigroup is proposed in the present paper. At the infinitesimal level the quasigroup is given by a set of generators that act as differential operators on functions of some initial variables. These generators obey Lie algebra commutation relations with the difference that the structural coefficients now depend, generally, on the initial variables. At the level of finite transformations the main difference between the Lie group and quasigroup is in the modification of the composition law which, in the quasigroup case, depends not only on the parameters of the transformations under the composition but also on the variables transformed.

As will be seen, the quasigroup construction is realized, in particular, on the hypersurfaces of the first class constraints in a phase space. An example of a gauge theory with the gauge transformations forming a quasisupergroup (see below) is given by supergravity with auxiliary fields meant to close the set of generators.

Let us sketch the contents of the paper. In Sec. 1. functional equations of the quasigroup are formulated and some of their consequences are studied. Here differential equations are obtained which are quasigroup counterparts of the Lie and Maurer–Cartan equations and their transformational properties are considered.

In Sec. 2 the formal integrability of the quasigroup differential equations is checked and the problem of reconstruction of the quasigroup, if the structural functions are given, is solved. (Solution of this problem for the Lie group is given in Ref. 1).

In Sec. 3 the quasigroup differential equations are extended to the case when initial variables and parameters are elements of a graded algebra (bosons and fermions).

In Sec. 4 a realization of the quasigroup is considered in a phase space on hypersurfaces of the first class constraints.

In Sec. 5, basing ourselves on the contents of the previous sections, we investigate the general structure of the quantum transition amplitude (the S-matrix) in dynamical systems subject to the first class constraints.

In the Appendix we introduce the left and right measures on a quasigroup and study their transformational properties.

1. QUASIGROUP

Let $g^a(1\leqslant a\leqslant n)$ be real variables for which the continuous law of transformation is given by

$$g^a\to\bar{g}^a = f^a(g,\theta)\,, \tag{1.1}$$

which depends actually on the set of real parameters $\theta^\alpha(1\leqslant\alpha\leqslant r)$. Assume that transformations (1.1) possess the following properties:

1) there exists a unit element which is common for all g^a and corresponds to

$$\theta^\alpha=0:\quad f^a(g,\theta=0)=g^a\,; \tag{1.2}$$

2) the modified compositions law holds:

$$f^a(f(g,\theta),\theta')=f^a(g,\varphi(\theta,\theta';g))\,; \tag{1.3}$$

3) the left and right units coincide:

$$\varphi^\alpha(\theta,0;g)=\theta^\alpha\,, \tag{1.4}$$

$$\varphi^\alpha(0,\theta';g)=\theta'^\alpha\,; \tag{1.5}$$

4) the modified law of associativity is satisfied:

$$\varphi^\alpha(\varphi(\theta,\theta';g),\theta'';g)=\varphi^\alpha(\theta,\varphi(\theta',\theta'';f(g,\theta));g)\,; \tag{1.6}$$

5) the transformation inverse to (1.1) exists and may be represented as

$$g^a=f^{-1a}(\bar{g},\theta)=f^a(\bar{g},\tilde{\theta}(\theta;\bar{g}))\,, \tag{1.7}$$

where the function $\tilde{\theta}$ satisfies the equations

$$\varphi^\alpha(\tilde{\theta}(\theta;\bar{g}),\theta;\bar{g})=0, \tag{1.8}$$

$$\varphi^\alpha(\theta,\tilde{\theta}(\theta;f(g,\theta));g)=0\,, \tag{1.9}$$

which give the left and right inverse elements.

The set of functional equations (1.1)–(1.9) defines a formal construction henceforth called quasigroup. In case the compositional function $\varphi^\alpha(\theta,\theta';g)$ does not depend upon g^a, these equations express ordinary group properties.

Taking the consistency of conditions 1)–5) for granted we shall obtain some of their consequences. The infinitesimal transformations follow from (1.1) when $\theta^\alpha\to 0$:

$$\delta g^a=R^a_\alpha(g)\,\theta^\alpha\,, \tag{1.10}$$

where

$$R^a_\alpha(g)\equiv\left.\frac{\partial f^a(g,\theta)}{\partial\theta^\alpha}\right|_{\theta=0}\,. \tag{1.11}$$

 0022-2488/81/091837-14$01.00

Using (1.10) in the relation,

$$W(\bar{g}) = W(g) \tag{1.12}$$

which expresses the invariance of a function $W(g)$, we come to the identity

$$R^a_\alpha(g)\, W_{,a}(g) \equiv 0, \tag{1.13}$$

where

$$W_{,a}(g) \equiv \frac{\partial W(g)}{\partial g^a}. \tag{1.14}$$

Expose the composition law (1.3) to the action of the operator

$$\left(\frac{\partial^2}{\partial\theta^\alpha \partial\theta'^\beta} - (\alpha \longleftrightarrow \beta)\right)\cdot(1.3)\Big|_{\theta=\theta'=0}. \tag{1.15}$$

With the use of (1.2), (1.4), (1.5), and (1.11) we obtain

$$R^a_{\alpha,b} R^b_\beta - R^a_{\beta,b} R^b_\alpha = -t^\gamma_{\alpha\beta} R^a_\gamma, \tag{1.16}$$

where

$$R^a_{\alpha,b} \equiv \frac{\partial R^a_\alpha(g)}{\partial g^b}, \tag{1.17}$$

$$t^\gamma_{\alpha\beta}(g) \equiv \left(\frac{\partial^2}{\partial\theta^\alpha \partial\theta'^\beta} - (\alpha \longleftrightarrow \beta)\right)\varphi^\gamma(\theta,\theta';g)\Big|_{\theta=\theta'=0}. \tag{1.18}$$

Expose the law of associativity to the operation

$$\left[\left(\frac{\partial^3}{\partial\theta^\alpha \partial\theta'^\beta \partial\theta''^\delta} - (\beta \longleftrightarrow \delta)\right) + \text{cyclic permutations of } \times (\alpha,\beta,\delta)\right]\cdot(1.6)\Big|_{\theta=\theta'=\theta''=0}. \tag{1.19}$$

Using (1.2), (1.4), (1.5), and (1.11) one obtains the modified Jacobi relations for the structure coefficients (1.18):

$$t^\mu_{\alpha\beta,a} R^a_\delta + t^\mu_{\delta\alpha,a} R^a_\beta + t^\mu_{\beta\delta,a} R^a_\alpha + t^\mu_{\alpha\gamma} t^\gamma_{\beta\delta} + t^\mu_{\delta\gamma} t^\gamma_{\alpha\beta} + t^\mu_{\beta\gamma} t^\gamma_{\delta\alpha} = 0, \tag{1.20}$$

where

$$t^\mu_{\alpha\beta,a} \equiv \frac{\partial t^\mu_{\alpha\beta}(g)}{\partial g^a}. \tag{1.21}$$

Owing to (1.16) the generators

$$\Gamma_\alpha = R^a_\alpha(g)\frac{\partial}{\partial g^a} \tag{1.22}$$

obey the commutation relations of quasialgebra

$$[\Gamma_\alpha, \Gamma_\beta] = t^\gamma_{\alpha\beta}(g)\Gamma_\gamma. \tag{1.23}$$

Now expose (1.3) to the action of

$$\frac{\partial}{\partial\theta'}\cdot(1.3)\Big|_{\theta'=0}. \tag{1.24}$$

With the aid of (1.4) and (1.11) we obtain an analog of the Lie equation,

$$\frac{\partial\bar{g}^a}{\partial\theta^\beta} = R^a_\alpha(\bar{g})\lambda^\alpha_\beta(\theta;g), \tag{1.25}$$

where

$$\lambda^\alpha_\beta \mu^\beta_\gamma = \delta^\alpha_\gamma, \tag{1.26}$$

$$\mu^\alpha_\beta(\theta;g) \equiv \frac{\partial\varphi^\alpha(\theta,\theta';g)}{\partial\theta'^\beta}\Big|_{\theta'=0}, \tag{1.27}$$

and it is assumed that

$$\operatorname{Det}\{\mu\} \neq 0. \tag{1.28}$$

Apply the operation

$$\left(\frac{\partial^2}{\partial\theta'^\gamma \partial\theta''^\delta} - (\gamma \longleftrightarrow \delta)\right)\cdot(1.6)\Big|_{\theta'=\theta''=0} \tag{1.29}$$

to (1.6). With the use of (1.2), (1.4), (1.5), and (1.18) we obtain the equation for $\mu(\theta;g)$,

$$\frac{\partial\mu^\alpha_\delta}{\partial\theta^\beta}\mu^\beta_\gamma - \frac{\partial\mu^\alpha_\gamma}{\partial\theta^\beta}\mu^\beta_\delta = -t^\beta_{\delta\gamma}(\bar{g})\mu^\alpha_\beta, \tag{1.30}$$

whence an analog of the Maurer–Cartan equation follows for $\lambda(\theta;g)$,

$$\frac{\partial\lambda^\alpha_\gamma}{\partial\theta^\beta} - \frac{\partial\lambda^\alpha_\beta}{\partial\theta^\gamma} + t^\alpha_{\mu\nu}(\bar{g})\lambda^\mu_\beta\lambda^\nu_\gamma = 0, \tag{1.31}$$

wherein the property

$$\lambda^\alpha_\beta(\theta=0;g) = \delta^\alpha_\beta \tag{1.32}$$

holds in virtue of (1.5). Apply the operation

$$\frac{\partial}{\partial\theta^\gamma}\cdot(1.3)\Big|_{\theta=0} \tag{1.33}$$

to (1.3). Using (1.2), (1.5), (1.11), and (1.25) we obtain the transformation property

$$S^a_b(g,\theta)R^b_\gamma(g) = R^a_\alpha(\bar{g})\,U_\gamma^{-1\alpha}(\theta;g), \tag{1.34}$$

where we write θ instead of θ' and, besides,

$$S^a_b(g,\theta) \equiv \frac{\partial\bar{g}^a}{\partial g^b}, \tag{1.35}$$

$$U_\gamma^{-1\alpha} \equiv \lambda^\alpha_\beta\tilde{\mu}^\beta_\gamma, \tag{1.36}$$

$$\tilde{\mu}^\alpha_\beta(\theta;g) \equiv \frac{\partial\varphi^\alpha(\theta',\theta;g)}{\partial\theta'^\beta}\Big|_{\theta'=0}. \tag{1.37}$$

It is assumed that

$$\operatorname{Det}\{S\} \neq 0, \tag{1.38}$$

$$\operatorname{Det}\{\tilde{\mu}\} \neq 0. \tag{1.39}$$

As combined with the transformation property of the functions (1.14),

$$W_{,a}(g) = W_{,b}(\bar{g})S^b_a(g,\theta), \tag{1.40}$$

which follows from (1.12), relation (1.34) provides the covariant nature of the identities (1.13). Using (1.25) and (1.34) we obtain an equation for the inverse transformation (1.7)

$$\frac{\partial g^a}{\partial\theta^\beta} = -R^a_\alpha(g)\tilde{\lambda}^\alpha_\beta(\theta;g), \tag{1.41}$$

where

$$\tilde{\lambda}^\alpha_\beta\tilde{\mu}^\beta_\gamma = \delta^\alpha_\gamma. \tag{1.42}$$

Let us apply the operation

$$\left(\frac{\partial^2}{\partial\theta^\delta \partial\theta'^\gamma} - (\delta \longleftrightarrow \gamma)\right)\cdot(1.6)\Big|_{\theta=\theta'=0} \tag{1.43}$$

to (1.6). With the help of (1.2), (1.4), (1.5), (1.11), (1.18), (1.37), and (1.42) we obtain the equation for $\widetilde{\mu}(\theta; g)$

$$(D_\beta \widetilde{\mu}^\alpha_\delta)\widetilde{\mu}^\beta_\gamma - (D_\beta \widetilde{\mu}^\alpha_\gamma)\widetilde{\mu}^\beta_\delta = t^\beta_{\delta\gamma}(g)\widetilde{\mu}^\alpha_\beta\,, \tag{1.44}$$

where we again write θ for θ'' and, besides,

$$D_\beta \equiv \frac{\partial}{\partial\theta^\beta} - R^a_\alpha(g)\tilde{\lambda}^\alpha_\beta \frac{\partial}{\partial g^a}\,. \tag{1.45}$$

Equation (1.44) leads to the following equation for $\tilde{\lambda}(\theta; g)$

$$D_\beta\tilde{\lambda}^\alpha_\gamma - D_\gamma\tilde{\lambda}^\alpha_\beta - t^\alpha_{\mu\nu}(g)\tilde{\lambda}^\mu_\beta\tilde{\lambda}^\nu_\gamma = 0\,. \tag{1.46}$$

Now expose (1.6) to the operation

$$\frac{\partial^2}{\partial\theta^\gamma\partial\theta''^\delta}\cdot(1.6)\Big|_{\theta=\theta''=0}\,. \tag{1.47}$$

Using (1.2), (1.4), (1.5), (1.11), (1.27), (1.37), and (1.42) we obtain the relation

$$(D_\beta\,\mu^\alpha_\delta)\widetilde{\mu}^\beta_\gamma - \frac{\partial\widetilde{\mu}^\alpha_\gamma}{\partial\theta^\beta}\mu^\beta_\delta = 0\,, \tag{1.48}$$

with θ standing for θ'. An equivalent form of (1.48) is

$$(D_\beta\,\mu^\alpha_\delta)\lambda^\delta_\gamma - \frac{\partial\widetilde{\mu}^\alpha_\delta}{\partial\theta^\gamma}\tilde{\lambda}^\delta_\beta = 0\,. \tag{1.49}$$

It follows from (1.16) and (1.46) that operators (1.45) commute with one another

$$[D_\beta, D_\gamma] = 0\,. \tag{1.50}$$

We have, besides, due to (1.25) and (1.34),

$$D_\beta\,\bar{g}^a = 0\,. \tag{1.51}$$

Components g^a are independent variables in Eqs. (1.44), (1.46), and (1.48). Substitute now the inverse transformation (1.7) into these equations. With $\bar{g}^a$ as independent variables we obtain, by using (1.34) and (1.41),

$$\frac{\partial\widetilde{\mu}^\alpha_\delta}{\partial\theta^\beta}\widetilde{\mu}^\beta_\gamma - \frac{\partial\widetilde{\mu}^\alpha_\gamma}{\partial\theta^\beta}\widetilde{\mu}^\beta_\delta = t^\beta_{\delta\gamma}(g)\widetilde{\mu}^\alpha_\beta\,, \tag{1.52}$$

$$\frac{\partial\tilde{\lambda}^\alpha_\gamma}{\partial\theta^\beta} - \frac{\partial\tilde{\lambda}^\alpha_\beta}{\partial\theta^\gamma} - t^\alpha_{\mu\nu}(g)\tilde{\lambda}^\mu_\beta\tilde{\lambda}^\nu_\gamma = 0\,, \tag{1.53}$$

$$\frac{\partial\mu^\alpha_\delta}{\partial\theta^\beta}\widetilde{\mu}^\beta_\gamma - (\bar{D}_\beta\,\widetilde{\mu}^\alpha_\gamma)\mu^\beta_\delta = 0\,, \tag{1.54}$$

$$\frac{\partial\mu^\alpha_\delta}{\partial\theta^\beta}\lambda^\delta_\gamma - (\bar{D}_\gamma\,\widetilde{\mu}^\alpha_\delta)\tilde{\lambda}^\delta_\beta = 0\,, \tag{1.55}$$

where

$$\bar{D}_\beta \equiv \frac{\partial}{\partial\theta^\beta} + R^a_\alpha(\bar{g})\lambda^\alpha_\beta\frac{\partial}{\partial\bar{g}^a}\,. \tag{1.56}$$

The θ-derivatives are to be calculated in (1.52)–(1.56) taking into account the extra dependence on θ, coming from the substitution of g^a from (1.7). Operators (1.56) commute with one another,

$$[\bar{D}_\beta, \bar{D}_\gamma] = 0\,. \tag{1.57}$$

This follows from (1.16) and Eq. (1.31) after the latter is, using (1.34) and (1.41), written with $\bar{g}^a$ as independent variables. We have, besides, due to (1.34) and (1.41),

$$\bar{D}_\beta\,g^a = 0\,. \tag{1.58}$$

Equations (1.41), (1.52), and (1.53) with the independent variables.

Let us now obtain analogs of the Lie equations for the composition function $\varphi^\alpha(\theta,\theta'; g)$. Apply the operation

$$\frac{\partial}{\partial\theta''}\cdot(1.6)\Big|_{\theta''=0} \tag{1.59}$$

to (1.6). Taking into account (1.4), (1.26), and (1.27) we obtain

$$\frac{\partial\varphi^\alpha(\theta,\theta'; g)}{\partial\theta'^\gamma} = \mu^\alpha_\beta(\varphi(\theta,\theta'; g); g)\,\lambda^\beta_\gamma(\theta'; \bar{g}(\theta))\,. \tag{1.60}$$

Expose (1.6) to the operation

$$\frac{\partial}{\partial\theta}\cdot(1.6)\Big|_{\theta=0}\,. \tag{1.61}$$

Equations (1.2), (1.5), (1.11), (1.37), and (1.42) give

$$D_\gamma\varphi^\alpha(\theta,\theta'; g) = \widetilde{\mu}^\alpha_\beta(\varphi(\theta,\theta'; g); g)\,\tilde{\lambda}^\beta_\gamma(\theta; g)\,, \tag{1.62}$$

where we write θ,θ' instead of θ',θ''. Operators D_γ are defined as (1.45).

When written with $\bar{g}^a$ as independent variables [Eq. (1.7) must be used to this end], Eq. (1.62) takes the form

$$\frac{\partial\varphi^\alpha(\theta,\theta'; g)}{\partial\theta^\gamma} = \widetilde{\mu}^\alpha_\beta(\varphi(\theta,\theta'; g); g)\,\tilde{\lambda}^\beta_\gamma(\theta; g)\,, \tag{1.63}$$

where the extra dependence on θ resulting from the substitution of g^a from (1.7) should be taken into account when fulfilling the θ^γ-differentiation shown in the left-hand side of (1.63).

Now expose (1.6) to the operation

$$\frac{\partial}{\partial\theta'^\gamma}(1.6)\Big|_{\theta'=0}\,. \tag{1.64}$$

Using (1.4), (1.5), and (1.37) we come to the important relation between derivatives of the composition function

$$\frac{\partial\varphi^\alpha(\theta,\theta'; g)}{\partial\theta^\beta}\mu^\beta_\gamma(\theta; g) = \frac{\partial\varphi^\alpha(\theta,\theta'; g)}{\partial\theta'^\beta}\widetilde{\mu}^\beta_\gamma(\theta'; \bar{g}(\theta))\,, \tag{1.65}$$

where again we use θ' instead of θ''. Differentiating (1.8) with respect to θ and using (1.65) we obtain the equation (with $\bar{g}^a$ as independent variables)

$$\frac{\partial\tilde{\theta}^\beta(\theta; g)}{\partial\theta^\gamma} = -\mu^\beta_\alpha(\tilde{\theta}(\theta; \bar{g}); \bar{g})\,\tilde{\lambda}^\alpha_\gamma(\theta; f^{-1}(\bar{g},\theta))\,, \tag{1.66}$$

where

$$\tilde{\theta}^\alpha(\theta = 0; \bar{g}) = 0\,, \tag{1.67}$$

as a consequence of (1.4) and (1.8). With the use of (1.61) and (1.63) we obtain from (1.65) the following relation for the matrix (1.36):

$$U^{-1\alpha}_\gamma(\varphi(\theta,\theta'; g); g) - U^{-1\alpha}_\beta(\theta'; \bar{g}(\theta))\,U^{-1\beta}_\gamma(\theta; g) = -\lambda^\alpha_\beta(\varphi(\theta,\theta'; g); g)\frac{\partial\varphi^\alpha(\theta,\theta'; g)}{\partial g^a}R^a_\alpha(g)\,. \tag{1.68}$$

Putting

$$\theta^\alpha = \tilde{\theta}^\alpha(\theta'; g)\,, \tag{1.69}$$

in (1.68) one obtains

$$\delta^\alpha_\gamma + \frac{\partial \varphi^\alpha(\tilde\theta(\theta;g),\theta;\overleftarrow{g})}{\partial g^a} R^a_\alpha(g)$$
$$= U_\beta^{-1\alpha}(\theta; f^{-1}(g,\theta))\, U_\gamma^{-1\beta}(\tilde\theta(\theta;g);g)\,, \qquad (1.70)$$

where the arrow indicates the argument subject to the differentiation. Relations (1.68) and (1.79) show the deflection of the compositional properties of the matrix (1.36) from the multiplicativity characteristic of the group case. Deflection from multiplicativity appears also in the compositional properties of the matrix (1.35). Differentiating the law of composition (1.3) with respect to g^c and using (1.25) and (1.34) we obtain the corresponding relation

$$S^a_b(\bar g(\theta),\theta')\, S^b_c(g,\theta)$$
$$= S^a_b(g,\varphi(\theta,\theta';g))$$
$$\times\left[\delta^b_c + R^a_\alpha(g)\tilde\lambda^\alpha_\beta(\varphi(\theta,\theta';g);g)\,\frac{\partial\varphi^\alpha(\theta,\theta';g)}{\partial g^c}\right]. \qquad (1.71)$$

Now we are going to derive a formula for the determinant of the matrix (1.35). From (1.25) we have

$$\frac{\partial}{\partial\theta^\beta} Sp \ln\{S(g,\theta)\}$$
$$= R^a_{\alpha,a}(\bar g)\lambda^\alpha_\beta + S_a^{-1b} R^a_\alpha(\bar g)\cdot\frac{\partial\lambda^\alpha_\beta}{\partial g^b}$$
$$= R^a_{\alpha,a}(\bar g)\lambda^\alpha_\beta - \frac{\partial\mu^\alpha_\delta}{\partial g^a} R^a_\gamma(g)\tilde\lambda^\gamma_\alpha\lambda^\delta_\beta\,, \qquad (1.72)$$

where (1.34) is taken into account in the second equality. Let us write, next, (1.49) in a more explicit form

$$\frac{\partial\mu^\alpha_\delta}{\partial\theta^\beta}\lambda^\delta_\gamma - \frac{\partial\widetilde{\mu^\alpha_\delta}}{\partial\theta^\gamma}\tilde\lambda^\delta_\beta = \frac{\partial\mu^\alpha_\delta}{\partial g^a} R^a_\epsilon(g)\tilde\lambda^\epsilon_\beta\lambda^\delta_\gamma\,. \qquad (1.73)$$

From (1.73) we obtain subsequently

$$\frac{\partial\mu^\alpha_\delta}{\partial g^a} R^a_\gamma(g)\tilde\lambda^\gamma_\alpha\lambda^\delta_\beta$$
$$= \frac{\partial\mu^\alpha_\delta}{\partial\theta^\alpha}\lambda^\delta_\beta - \frac{\partial\widetilde{\mu^\alpha_\delta}}{\partial\theta^\beta}\tilde\lambda^\delta_\alpha$$
$$= -\mu^\alpha_\delta\frac{\partial\lambda^\delta_\beta}{\partial\theta^\alpha} - \frac{\partial}{\partial\theta^\beta} Sp \ln\tilde\mu$$
$$= -\mu^\alpha_\delta\left(\frac{\partial\lambda^\delta_\alpha}{\partial\theta^\beta} - t^\delta_{\mu\nu}(\bar g)\lambda^\mu_\alpha\lambda^\nu_\beta\right) + \frac{\partial}{\partial\theta^\beta} Sp \ln\tilde\lambda$$
$$= -\frac{\partial}{\partial\theta^\beta} Sp\ln\lambda + t^\mu_{\mu\nu}(\bar g)\lambda^\nu_\beta + \frac{\partial}{\partial\theta^\beta} Sp\ln\tilde\lambda\,, \qquad (1.74)$$

where (1.31) has been taken into account. The substitution of (1.74) into (1.72) gives

$$\frac{\partial}{\partial\theta^\beta} Sp\ln\{S(g,\theta)\}$$
$$= A_\alpha(\bar g)\lambda^\alpha_\beta + \frac{\partial}{\partial\theta^\beta}(Sp\ln\lambda - Sp\ln\tilde\lambda)\,, \qquad (1.75)$$

where

$$A_\alpha(g) \equiv R^a_{\alpha,a}(g) + t^\beta_{\alpha\beta}(g)\,. \qquad (1.76)$$

It follows from (1.75), in particular, that

$$A_\alpha(\bar g)\lambda^\alpha_\beta = \frac{\partial \ln E}{\partial\theta^\beta}\,. \qquad (1.77)$$

Equation (1.77) is at the same time the definition of the function $E(g,\theta)$. With the boundary condition

$$E(g,\theta=0) = 1, \qquad (1.78)$$

we obtain from (1.75) and (1.77)

$$\mathrm{Det}\{S(g,\theta)\} = E(g,\theta)\frac{\mathrm{Det}\{\lambda(\theta;g)\}}{\mathrm{Det}\{\tilde\lambda(\theta;g)\}}\,, \qquad (1.79)$$

where

$$E(g,\theta) = \exp\left\{\int_0^\theta A_\alpha(\bar g(\theta'))\,\lambda^\alpha_\beta(\theta';g)\,d\theta'^\beta\right\}. \qquad (1.80)$$

The integral in (1.80) does not depend upon the form of the path of integration. Keeping this in mind and using (1.3), (1.4), and (1.60) it is easy to obtain for the function E that

$$E(g,\theta)\,E(\bar g(\theta),\theta') = E(g,\varphi(\theta,\theta';g))\,. \qquad (1.81)$$

Making the formal change

$$\theta \to \tilde\theta(\theta;g)\,;\quad \theta'\to\theta \qquad (1.82)$$

in (1.81) and taking (1.7), (1.8) and (1.78) into account one obtains

$$E(g,\tilde\theta(\theta;g))\,E(f^{-1}(g,\theta),\theta) = 1\,. \qquad (1.83)$$

To conclude this section consider transformation properties of the quasigroup under the transformation of parameters

$$\theta^\alpha \to \psi^\alpha(\theta;g)\,, \qquad (1.84)$$

where

$$\psi^\alpha(\theta=0;g) = 0\,, \qquad (1.85)$$

$$\mathrm{Det}\left\{\frac{\partial\psi}{\partial\theta}\right\} \neq 0\,. \qquad (1.86)$$

The change (1.84) introduces the new law of transformation of variables g^a:

$$\bar g^a = f^a_1(g,\theta) \equiv f^a(g,\psi(\theta;g))\,, \qquad (1.87)$$

with

$$f^a_1(g,\theta=0) = g^a \qquad (1.88)$$

due to (1.2) and (1.85). Making the formal change

$$\theta^\alpha \to \psi^\alpha(\theta;g)\,, \qquad (1.89)$$

$$\theta'^\alpha \to \psi^\alpha(\theta';f_1(g,\theta)) \qquad (1.90)$$

in (1.3), we come to the composition law for transformations (1.87)

$$f^a_1(f_1(g,\theta),\theta') = f^a_1(g,\varphi_1(\theta,\theta';g))\,, \qquad (1.91)$$

where the new composition function is determined by the equation

$$\varphi^\alpha(\psi(\theta;g),\psi(\theta';f_1(g,\theta));g)$$
$$= \psi^\alpha(\varphi_1(\theta,\theta';g);g)\,. \qquad (1.92)$$

Taking into account (1.4), (1.5), (1.85), and (1.86) it follows from (1.92) that

$$\varphi^\alpha_1(\theta,0;g) = \theta^\alpha\,, \qquad (1.93)$$

$$\varphi^\alpha_1(0,\theta';g) = \theta'^\alpha\,. \qquad (1.94)$$

Making in (1.6) the formal change (1.89) and (1.90) along with

$$\theta''^\alpha \to \psi^\alpha(\theta'';f_1(f_1(g,\theta),\theta'))\,, \qquad (1.95)$$

one obtains, using (1.92), the law of associativity

$$\varphi_1^\alpha(\varphi_1(\theta,\theta';g),\theta'';g) = \varphi_1^\alpha(\theta,\varphi_1(\theta',\theta'';f_1(g,\theta));g). \quad (1.96)$$

The transformation inverse to (1.87) is written in the form

$$g^\alpha = f_1^{-1\alpha}(\bar{g},\theta) = f_1^\alpha(\bar{g},\tilde{\theta}_1(\theta;\bar{g})), \quad (1.97)$$

where the functions $\tilde{\theta}_1^\alpha$ satisfy the equations

$$\varphi_1^\alpha(\tilde{\theta}_1(\theta;\bar{g}),\theta;\bar{g}) = 0, \quad (1.98)$$

$$\varphi_1^\alpha(\theta,\tilde{\theta}_1(\theta;f_1(g,\theta));g) = 0, \quad (1.99)$$

and are connected with the functions $\tilde{\theta}^\alpha$ by the relations

$$\tilde{\theta}^\alpha(\psi(\theta;f_1^{-1}(\bar{g},\theta));\bar{g}) = \psi^\alpha(\tilde{\theta}_1(\theta;\bar{g});\bar{g}). \quad (1.100)$$

Introduce, further, the functions

$$R_{1\beta}^\alpha(g) \equiv \left.\frac{\partial f_1^\alpha(g,\theta)}{\partial\theta^\beta}\right|_{\theta=0} = R_\alpha^\alpha(g)\,\Omega_\beta^\alpha(g) \quad (1.101)$$

$$\Omega_\beta^\alpha(g) \equiv \left.\frac{\partial\psi^\alpha(\theta;g)}{\partial\theta^\beta}\right|_{\theta=0}. \quad (1.102)$$

Differentiating (1.92) with respect to θ' and setting $\theta'=0$ one has

$$\frac{\partial\psi^\alpha(\theta;g)}{\partial\theta^\beta}\mu_{1\gamma}^\beta(\theta;g) = \mu_\beta^\alpha(\psi(\theta;g);g)\,\Omega_\gamma^\beta(f_1(g,\theta)), \quad (1.103)$$

where

$$\mu_1^\alpha(\theta;g) \equiv \left.\frac{\partial\varphi_1^\alpha(\theta,\theta';g)}{\partial\theta'^\beta}\right|_{\theta'=0}. \quad (1.104)$$

The matrix inverse to (1.104) is defined by the equation

$$\lambda_{1\beta}^\alpha\,\mu_{1\gamma}^\beta = \delta_\gamma^\alpha. \quad (1.105)$$

Differentiating (1.92) with respect to θ and setting $\theta=0$ one has

$$(D_{1\beta}\psi^\alpha(\theta;g))\,\tilde{\mu}_{1\gamma}^\beta(\theta;g) = \tilde{\mu}_\beta^\alpha(\psi(\theta;g);g)\,\Omega_\gamma^\beta(g), \quad (1.106)$$

where we again write θ instead of θ' and, besides,

$$\tilde{\mu}_{1\beta}^\alpha(\theta;g) \equiv \left.\frac{\partial\varphi_1^\alpha(\theta',\theta;g)}{\partial\theta'^\beta}\right|_{\theta'=0}, \quad (1.107)$$

$$\tilde{\lambda}_{1\beta}^\alpha\,\tilde{\mu}_{1\gamma}^\beta = \delta_\gamma^\alpha, \quad (1.108)$$

$$D_{1\beta} \equiv \frac{\partial}{\partial\theta^\beta} - R_{1\alpha}^\alpha(g)\,\tilde{\lambda}_{1\beta}^\alpha\frac{\partial}{\partial g^\alpha}. \quad (1.109)$$

Now apply the operation

$$\left.\left(\frac{\partial^2}{\partial\theta^\gamma\partial\theta'^\delta} - (\gamma\longleftrightarrow\delta)\right)\cdot(1.92)\right|_{\theta=\theta'=0} \quad (1.110)$$

to (1.92). This gives

$$\Omega_\beta^\alpha(g)\,t_{1\gamma\delta}^\beta(g) = t_{\mu\nu}^\alpha(g)\,\Omega_\gamma^\mu(g)\,\Omega_\delta^\nu(g) + R_{1\gamma}^\alpha(g)\frac{\partial\Omega_\delta^\alpha(g)}{\partial g^\alpha} - R_{1\delta}^\alpha(g)\frac{\partial\Omega_\gamma^\alpha(g)}{\partial g^\alpha}, \quad (1.111)$$

where

$$t_{1\alpha\beta}^\gamma(g) \equiv \left.\left(\frac{\partial^2}{\partial\theta^\alpha\partial\theta'^\beta} - (\alpha\longleftrightarrow\beta)\right)\varphi_1^\gamma(\theta,\theta';g)\right|_{\theta=\theta'=0} \quad (1.112)$$

are new structure coefficients.

Thus, the new functions $f_1^\alpha(g,\theta)$, $\varphi_1^\alpha(\theta,\theta';g)$, $\tilde{\theta}_1^\alpha(\theta;g)$ satisfy the same functional equations as the initial ones $f^\alpha(g,\theta)$, $\varphi^\alpha(\theta,\theta';g)$, $\tilde{\theta}^\alpha(\theta;g)$.

To finish the consideration of the transformation properties of the quasigroup we note that the inverse transformation (1.7) is also a particular case of the change of parameters (1.84). Indeed, let us write $\bar{g}^\alpha$ instead of g^α in (1.84) and set

$$\psi^\alpha(\theta;\bar{g}) = \tilde{\theta}^\alpha(\theta;\bar{g}). \quad (1.113)$$

Then

$$f_1^\alpha(\bar{g},\theta) = f^{-1\alpha}(\bar{g},\theta). \quad (1.114)$$

Equation (1.91) leads to

$$f^{-1\alpha}(f^{-1}(\bar{g},\theta),\theta') = f^{-1\alpha}(\bar{g},\tilde{\varphi}(\theta,\theta';\bar{g})), \quad (1.115)$$

where, according to (1.92)

$$\varphi^\alpha(\tilde{\theta}(\theta;\bar{g}),\tilde{\theta}(\theta';f^{-1}(\bar{g},\theta));\bar{g}) = \tilde{\theta}^\alpha(\tilde{\varphi}(\theta,\theta';\bar{g});\bar{g}). \quad (1.116)$$

From (1.96) it follows that

$$\tilde{\varphi}^\alpha(\tilde{\varphi}(\theta,\theta';\bar{g}),\theta'';\bar{g}) = \tilde{\varphi}^\alpha(\theta,\tilde{\varphi}(\theta',\theta'';f^{-1}(\bar{g},\theta));\bar{g}). \quad (1.117)$$

It may be shown, that the function $\tilde{\varphi}^\alpha$ is related to the function φ^α by the following formula

$$\tilde{\varphi}^\alpha(\theta',\theta;f(g,\theta')) = \varphi^\alpha(\theta,\theta';f^{-1}(g,\theta)). \quad (1.118)$$

2. INTEGRABILITY OF THE EQUATIONS OF QUASIGROUP. RECONSTRUCTION OF QUASIGROUP FROM THE STRUCTURE FUNCTIONS

In the previous section we obtained some consequences of the functional equations (1.1)–(1.9). In particular, the differential description of the quasigroup is contained in the following set of equations:

$$\frac{\partial\bar{g}^\alpha}{\partial\theta^\beta} = R_\alpha^\alpha(\bar{g})\,\lambda_\beta^\alpha, \quad (1.25')$$

$$\bar{g}^\alpha(\theta=0) = g^\alpha, \quad (1.2')$$

$$\frac{\partial\lambda_\gamma^\alpha}{\partial\theta^\beta} - \frac{\partial\lambda_\beta^\alpha}{\partial\theta^\gamma} + t_{\mu\nu}^\alpha(\bar{g})\,\lambda_\beta^\mu\,\lambda_\gamma^\nu = 0, \quad (1.31')$$

$$\lambda_\beta^\alpha(\theta=0) = \delta_\beta^\alpha, \quad (1.32')$$

which we complement by the equation for the compositional function

$$\frac{\partial\varphi^\alpha(\theta,\theta';g)}{\partial\theta'^\gamma} = \mu_\beta^\alpha(\varphi(\theta,\theta';g);g)\,\lambda_\gamma^\beta(\theta';\bar{g}(\theta)), \quad (1.60')$$

$$\lambda_\beta^\alpha\,\mu_\gamma^\beta = \delta_\gamma^\alpha, \quad (1.26')$$

$$\varphi^\alpha(\theta,\theta'=0;g) = \theta^\alpha. \quad (1.4')$$

We formulate now the following problem. Let there be given structural functions $R_\alpha^\alpha(g)$, $t_{\alpha\beta}^\gamma(g)$ that obey the Eqs. (1.16) and (1.20). Let us try, using (1.25')–(1.32'), to reconstruct the quasigroup law of transformation, at least locally, i.e., in a sufficiently small, but finite environment of the zero values of the parameters.

Let us, first of all check the fulfillment of the formal conditions of integrability. Differentiating (1.25′) with respect to θ^γ and performing the antisymmetrization with respect to the indices (γ,β) we obtain zero on the left-hand side identically, while the vanishing of the right-hand side is provided by Eqs. (1.16), (1.25′), and (1.31′). Differentiate (1.31′) with respect to θ^δ and add next the cyclic permutations of the indices (γ,β,δ). The second derivatives of λ cancel while the other terms disappear owing to (1.20), (1.25′), and (1.31′). Thus, the formal conditions of integrability of Eqs. (1.25′) and (1.31′) are fulfilled, which provides existence of a solution. This solution, however, is determined with the arbitrariness of the transformations of parameters under which the system (1.25′)–(1.32′) is covariant with the given functions R^a_α, $t^\gamma_{\alpha\beta}$. These transformations have the form (1.84) with $\Omega^\alpha_\beta(g)=\delta^\alpha_\beta$, where Ω^α_β is defined as (1.102). To fix a unique solution of the set (1.25′)–(1.32′) it is necessary to impose auxiliary conditions on the functions λ^α_β thus making a special choice of parametrization of the quasigroup (see below).

Assume now that functions $\bar g^a(\theta)$, $\lambda^\alpha_\beta(\theta)$ which satisfy Eqs. (1.25′)–(1.32′) in a special parametrization, are found. Let us check the integrability of Eq. (1.60′). Differentiating (1.60′) with respect to θ'^δ and performing the antisymmetrization with respect to the indices (δ,γ) we get identically zero on the left-hand side, while on the right-hand side we obtain via (1.31′), (1.60′), and (1.26′)

$$\mu^\alpha_\beta(\varphi(\theta,\theta';g);g)(t^\beta_{\mu\nu}(\bar g_1) - t^\beta_{\mu\nu}(\bar g_2))\lambda^\mu_\gamma(\theta';\bar g(\theta))\lambda^\nu_\delta(\theta';\bar g(\theta)), \tag{2.1}$$

where

$$\bar g^a_1 \equiv f^a(f(g,\theta),\theta'), \tag{2.2}$$

$$\bar g^a_2 \equiv f^a(g,\varphi(\theta,\theta';g)), \tag{2.3}$$

$$f^a(g,\theta)\equiv \bar g^a(\theta). \tag{2.4}$$

Differentiation of (2.2) leads to

$$\frac{\partial \bar g^a_1}{\partial\theta'^\beta} = R^a_\alpha(\bar g_1)\lambda^\alpha_\beta(\theta';\bar g(\theta)), \tag{2.5}$$

with

$$\bar g^a_1(\theta'=0)=\bar g^a(\theta). \tag{2.6}$$

Differentiating (2.3) with respect to θ' we have

$$\begin{aligned}\frac{\partial \bar g^a_2}{\partial\theta'^\beta} &= R^a_\alpha(\bar g_2)\lambda^\alpha_\gamma(\varphi(\theta,\theta';g);g)\frac{\partial\varphi^\gamma(\theta,\theta';g)}{\partial\theta'^\beta}\\ &= R^a_\alpha(\bar g_2)\lambda^\alpha_\gamma(\varphi(\theta,\theta';g);g)\mu^\gamma_\delta(\varphi(\theta,\theta';g);g)\\ &\quad\times\lambda^\delta_\beta(\theta';\bar g(\theta))\\ &= R^a_\alpha(\bar g_2)\lambda^\alpha_\beta(\theta';\bar g(\theta)),\end{aligned} \tag{2.7}$$

with

$$\bar g_2(\theta'=0)=\bar g^a(\theta). \tag{2.8}$$

Comparison of (2.5) and (2.6) with (2.7) and (2.8) shows that expressions (2.2) and (2.3), as functions of θ', obey one and the same differential equation (whose integrability has been already established) and the same boundary conditions at $\theta'=0$. Hence we conclude that

$$\bar g^a_1(\theta')=\bar g^a_2(\theta'), \tag{2.9}$$

i.e.,

$$f^a(f(g,\theta),\theta')=f^a(g,\varphi(\theta,\theta';g)). \tag{1.3′}$$

By virtue of (1.3′) expression (2.1) vanishes. Thus confirming the integrability of (1.60′). Simultaneously the law of composition (1.3′) is obtained as a consequence of the quasigroup differential equations.

By analogy with (2.2) and (2.3) let us introduce the functions

$$\Phi^\alpha_1\equiv\varphi^\alpha(\varphi(\theta,\theta';g),\theta'',g), \tag{2.10}$$

$$\Phi^\alpha_2\equiv\varphi^\alpha(\theta,\varphi(\theta',\theta'';f(g,\theta));g) \tag{2.11}$$

into consideration. Taking (1.60′) into account one has for (2.10)

$$\frac{\partial\Phi^\alpha_1}{\partial\theta''^\gamma}=\mu^\alpha_\beta(\Phi_1;g)\lambda^\beta_\gamma(\theta'';\bar g_2), \tag{2.12}$$

while

$$\Phi^\alpha_1(\theta''=0)=\varphi^\alpha(\theta,\theta';g). \tag{2.13}$$

Analogously one has for (2.11)

$$\begin{aligned}\frac{\partial\Phi^\alpha_2}{\partial\theta''^\gamma} &= \mu^\alpha_\beta(\Phi_2;g)\lambda^\beta_\delta(\varphi(\theta',\theta'';\bar g(\theta));\bar g(\theta))\\ &\quad\times\frac{\partial\varphi^\delta(\theta',\theta'';\bar g(\theta))}{\partial\theta''^\gamma}\\ &= \mu^\alpha_\beta(\Phi_2;g)\lambda^\beta_\delta(\varphi(\theta'\,\theta'';\bar g(\theta));\bar g(\theta))\\ &\quad\times\mu^\delta_\epsilon(\varphi(\theta',\theta'';\bar g(\theta));\bar g(\theta))\lambda^\epsilon_\gamma(\theta'';\bar g_1)\\ &= \mu^\alpha_\beta(\Phi_2;g)\lambda^\beta_\gamma(\theta'';\bar g_2),\end{aligned} \tag{2.14}$$

while

$$\Phi^\alpha_2(\theta''=0)=\varphi^\alpha(\theta,\theta';g). \tag{2.15}$$

By comparing (2.12) and (2.13) with (2.19) and (2.15) we conclude that

$$\Phi^\alpha_1(\theta'')=\Phi^\alpha_2(\theta''), \tag{2.16}$$

i.e.,

$$\varphi^\alpha(\varphi(\theta,\theta';g),\theta'';g)=\varphi^\alpha(\theta,\varphi(\theta',\theta'';f(g,\theta));g). \tag{1.6′}$$

Therefore, the modified law of associativity is obtained as a consequence of the differential equations of quasigroup.

Further on, we find from (1.60′)–(1.4′)

$$\varphi^\alpha(\theta=0,\theta';g)=\theta'^\alpha. \tag{1.5′}$$

Define now a function $\tilde\theta^\alpha(\theta;\bar g)$ of the independent variables $\theta;\bar g$ as a solution of the equation

$$\varphi^\alpha(\tilde\theta(\theta;\bar g),\theta;\bar g)=0. \tag{1.8′}$$

Then, taking (1.2′) and (1.3′) into account we obtain

$$\begin{aligned}&f^a(f(\bar g,\tilde\theta(\theta;\bar g)),\theta)\\ &\quad=f^a(\bar g,\varphi(\tilde\theta(\theta;\bar g),\theta;\bar g))=f^a(\bar g,0)=\bar g^a,\end{aligned} \tag{2.17}$$

whence

$$f^{-1a}(\bar g,\theta)=f^a(\bar g,\tilde\theta(\theta;\bar g)). \tag{1.7′}$$

Substitution of (2.4) into (1.8′) gives

$$\varphi^\alpha(\tilde\theta(\theta;f(g,\theta)),\theta;f(g,\theta))=0. \tag{2.18}$$

It follows from (1.4′) and (2.18) that

$$\varphi^\alpha(\theta,\varphi(\tilde\theta(\theta;f(g,\theta)),\theta;f(g,\theta));g)=\theta^\alpha. \tag{2.19}$$

With the help of (1.6′) we may represent (2.19) in the form

$$\varphi^{\alpha}(\varphi(\theta,\tilde{\theta}(\theta;f(g,\theta));g),\theta;g)=\theta^{\alpha}. \tag{2.20}$$

In virtue of (1.5′) this equation has the solution

$$\varphi^{\alpha}(\theta,\tilde{\theta}(\theta;f(g,\theta));g)=0, \tag{1.9′}$$

which is unique in a sufficiently small neighborhood of the point $\theta=0$.

Thus, we have not only confirmed the formal integrability of the set of differential equations (1.25′)–(1.4′) but also derived the functional equations of the quasigroup from it.

Let us now proceed directly to reconstructing quasigroup when the structure functions are given. For the special parametrization we choose the canonical one with the functions satisfying the condition

$$\lambda^{\alpha}_{\beta}\,\theta^{\beta}=\theta^{\alpha}, \tag{2.21}$$

or, what is the same,

$$\mu^{\alpha}_{\beta}\,\theta^{\beta}=\theta^{\alpha}. \tag{2.22}$$

Multiplying (1.25′) and (1.31′) by θ^{β} and θ^{γ}, respectively, and using (2.21) we obtain for the functions

$$G^{a}(x)\equiv \bar{g}^{a}(x\theta), \tag{2.23}$$

$$\Lambda^{\alpha}_{\beta}(x)\equiv x\lambda^{\alpha}_{\beta}(x\theta), \tag{2.24}$$

and the set of equations

$$\frac{dG^{a}}{dx}=R^{a}_{\alpha}(G)\,\theta^{\alpha}, \tag{2.25}$$

$$G^{a}(x=0)=g^{a}, \tag{2.26}$$

$$\frac{d\Lambda^{\alpha}_{\beta}}{dx}=\delta^{\alpha}_{\beta}+t^{\alpha}_{\mu\nu}(G)\,\theta^{\nu}\Lambda^{\mu}_{\beta}, \tag{2.27}$$

$$\Lambda^{\alpha}_{\beta}(x=0)=0. \tag{2.28}$$

The function (2.23) may be found from (2.25) and (2.26) and then susbtituted into (2.27). The latter equation, along with (2.28), determines the function (2.24) and (2.28). Then we have

$$\bar{g}^{a}(\theta)=G^{a}(x=1), \tag{2.29}$$

$$\lambda^{\alpha}_{\beta}(\theta)=\Lambda^{\alpha}_{\beta}(x=1). \tag{2.30}$$

Next one must show that the functions (2.29) and (2.30) do satisfy Eqs. (1.25′)–(1.32′). To this end we must study some consequences of (2.25)–(2.28). Introduce the function

$$h^{\alpha}(x)\equiv\Lambda^{\alpha}_{\beta}(x)\,\theta^{\beta}-x\theta^{\alpha}. \tag{2.31}$$

With the use of (2.27) we have

$$\frac{dh^{\alpha}}{dx}=t^{\alpha}_{\mu\nu}(G)\,\theta^{\nu}h^{\mu}, \tag{2.32}$$

with the property

$$h^{\alpha}(x=0)=0, \tag{2.33}$$

which follows from (2.28). From (2.32) and (2.33) it follows that

$$h^{\alpha}(x)=0, \tag{2.34}$$

so that for all x the relation holds

$$\Lambda^{\alpha}_{\beta}(x)\,\theta^{\beta}=x\theta^{\alpha} \tag{2.35}$$

as a consequence of (2.27) and (2.28). Introduce now the function

$$h^{a}_{\beta}(x)\equiv\frac{\partial G^{a}}{\partial\theta^{\beta}}-R^{a}_{\alpha}(G)\,\Lambda^{\alpha}_{\beta}. \tag{2.36}$$

Differentiating it with respect to x we have

$$\begin{aligned}\frac{dh^{a}_{\beta}}{dx}&=\frac{d}{dx}\frac{dG^{a}}{\partial\theta^{\beta}}\\&\quad-R^{a}_{\alpha,b}(G)\frac{dG^{b}}{dx}\Lambda^{\alpha}_{\beta}-R^{a}_{\alpha}(G)\frac{d\Lambda^{\alpha}_{\beta}}{dx}\\&=\frac{\partial}{\partial\theta^{\beta}}\frac{dG^{a}}{dx}-R^{a}_{\alpha,b}(G)\,R^{b}_{\gamma}(G)\,\theta^{\gamma}\Lambda^{\alpha}_{\beta}\\&\quad-R^{a}_{\alpha}(G)(\delta^{\alpha}_{\beta}+t^{\alpha}_{\mu\nu}(G)\,\theta^{\nu}\Lambda^{\mu}_{\beta}).\end{aligned} \tag{2.37}$$

The differentiation of (2.25) with respect to θ^{β} gives

$$\begin{aligned}\frac{\partial}{\partial\theta^{\beta}}\frac{dG^{a}}{dx}&=R^{a}_{\alpha,b}(G)\frac{\partial G^{b}}{\partial\theta^{\beta}}\theta^{\alpha}+R^{a}_{\beta}(G)\\&=R^{a}_{\alpha,b}(G)(h^{b}_{\beta}+R^{b}_{\gamma}(G)\,\Lambda^{\gamma}_{\beta})\,\theta^{\alpha}+R^{a}_{\beta}(G).\end{aligned} \tag{2.38}$$

Substituting (2.38) into (2.37) and taking (1.16) into account we obtain

$$\frac{dh^{a}_{\beta}}{dx}=R^{a}_{\alpha,b}(G)\,\theta^{\alpha}h^{b}_{\beta}, \tag{2.39}$$

with condition

$$h^{a}_{\beta}(x=0)=0, \tag{2.40}$$

which holds owing to (2.26), (2.28), and (2.36). Equations (2.39) and (2.40) lead to

$$h^{a}_{\beta}(x)=0, \tag{2.41}$$

i.e. for all x the relation

$$\frac{\partial G^{a}}{\partial\theta^{\beta}}=R^{a}_{\alpha}(G)\,\Lambda^{\alpha}_{\beta} \tag{2.42}$$

holds as a consequence of (2.25)–(2.28). Besides, from (2.25) and (2.26) it follows that

$$G^{a}|_{\theta=0}=g^{a}. \tag{2.43}$$

Introduce next the function

$$h^{\alpha}_{\beta\gamma}(x)\equiv\frac{\partial\Lambda^{\alpha}_{\gamma}}{\partial\theta^{\beta}}-\frac{\partial\Lambda^{\alpha}_{\beta}}{\partial\theta^{\gamma}}+t^{\alpha}_{\mu\nu}(G)\,\Lambda^{\mu}_{\beta}\Lambda^{\nu}_{\gamma}. \tag{2.44}$$

Its differentiation with respect to x and the use of (2.25) and (2.27) result in

$$\begin{aligned}\frac{dh^{\alpha}_{\beta\gamma}}{dx}&=\frac{\partial}{\partial\theta^{\beta}}\frac{d\Lambda^{\alpha}_{\beta}}{dx}-\frac{\partial}{\partial\theta^{\gamma}}\frac{d\Lambda^{\alpha}_{\beta}}{dx}\\&\quad+t^{\alpha}_{\mu\nu,a}(G)\,R^{a}_{\delta}(G)\,\theta^{\delta}\Lambda^{\mu}_{\beta}\Lambda^{\nu}_{\gamma}\\&\quad+t^{\alpha}_{\mu\nu}(G)(\delta^{\mu}_{\beta}+t^{\mu}_{\sigma\tau}(G)\,\theta^{\tau}\Lambda^{\sigma}_{\beta})\Lambda^{\nu}_{\gamma}\\&\quad+t^{\alpha}_{\mu\nu}(G)\,\Lambda^{\mu}_{\beta}(\delta^{\nu}_{\gamma}+t^{\nu}_{\sigma\tau}(G)\,\theta^{\tau}\Lambda^{\sigma}_{\gamma}).\end{aligned} \tag{2.45}$$

Differentiation of (2.27) with respect to θ^{γ} with the use of (2.42) gives

$$\begin{aligned}&\frac{\partial}{\partial\theta^{\gamma}}\frac{d\Lambda^{\alpha}_{\beta}}{dx}\\&\quad=t^{\alpha}_{\mu\nu,a}(G)\,\theta^{\nu}R^{a}_{\delta}(G)\,\Lambda^{\delta}_{\gamma}\Lambda^{\mu}_{\beta}\end{aligned}$$

$$+ t^{\alpha}_{\mu\gamma}(G)\Lambda^{\mu}_{\beta} + t^{\alpha}_{\mu\nu}(G)\theta^{\nu}\frac{\partial\Lambda^{\mu}_{\beta}}{\partial\theta^{\gamma}}. \tag{2.46}$$

The part of (2.46) antisymmetric with respect to the indices (β,γ) is

$$\begin{aligned}\frac{\partial}{\partial\theta^{\beta}}\frac{d\Lambda^{\alpha}_{\gamma}}{dx} - \frac{\partial}{\partial\theta^{\gamma}}\frac{d\Lambda^{\alpha}_{\beta}}{dx} &= t^{\alpha}_{\mu\nu,a}(G)\theta^{\nu}R^{a}_{\delta}(G)(\Lambda^{\delta}_{\beta}\Lambda^{\mu}_{\gamma} - \Lambda^{\delta}_{\gamma}\Lambda^{\mu}_{\beta}) \\ &\quad + t^{\alpha}_{\mu\beta}(G)\Lambda^{\mu}_{\gamma} - t^{\alpha}_{\mu\gamma}(G)\Lambda^{\mu}_{\beta} + t^{\alpha}_{\mu\nu}(G)\theta^{\nu}h^{\mu}_{\beta\gamma} \\ &\quad - t^{\alpha}_{\mu\nu}(G)\theta^{\nu}t^{\mu}_{\sigma\tau}(G)\Lambda^{\sigma}_{\beta}\Lambda^{\tau}_{\gamma}, \end{aligned} \tag{2.47}$$

where the definition (2.44) has been taken into account. Substitution of (2.47) into (2.45) and the use of (1.20) result in

$$\frac{dh^{\alpha}_{\beta\gamma}}{dx} = t^{\alpha}_{\mu\nu}(G)\theta^{\nu}h^{\mu}_{\beta\gamma}, \tag{2.48}$$

with the condition

$$h^{\alpha}_{\beta\gamma}(x=0) = 0 \tag{2.49}$$

that follows from (2.28). Equations (2.48) and (2.49) provide

$$h^{\alpha}_{\beta\gamma}(x) = 0, \tag{2.50}$$

for all x. Thus the relation

$$\frac{\partial\Lambda^{\alpha}_{\gamma}}{\partial\theta^{\beta}} - \frac{\partial\Lambda^{\alpha}_{\beta}}{\partial\theta^{\gamma}} + t^{\alpha}_{\mu\beta\nu}(G)\Lambda^{\mu}_{\beta}\Lambda^{\nu}_{\gamma} = 0 \tag{2.51}$$

holds as a consequence of (2.25)–(2.28). Besides, it follows from (2.27) and (2.28) that

$$\Lambda^{\alpha}_{\beta}\big|_{\theta=0} = x\delta^{\alpha}_{\beta}. \tag{2.52}$$

Setting $x = 1$ in (2.35), (2.42), (2.43), (2.51), and (2.52) we see that the functions (2.29) and (2.30) do realize solution of the set of equations (1.25′)–(1.32′) with canonical variables (2.21).

Now consider Eq. (1.60′) for the compositional function in the canonical coordinates. Multiplying it by θ'^{γ} and using (2.21) we obtain for the function

$$\Phi^{\alpha}(x) \equiv \varphi^{\alpha}(\theta, x\theta'; g) \tag{2.53}$$

and the following equation

$$\frac{d\Phi^{\alpha}}{dx} = \mu^{\alpha}_{\beta}(\Phi; g)\theta'^{\beta}, \tag{2.54}$$

$$\Phi^{\alpha}(x=0) = \theta^{\alpha}, \tag{2.55}$$

where $\mu^{\alpha}_{\beta}(\theta; g)$ is the matrix inverse to (2.30). After the solution of Eq. (2.54) subject to the condition (2.55) is found one has

$$\varphi^{\alpha}(\theta, \theta'; g) = \Phi^{\alpha}(x=1). \tag{2.56}$$

We are going to show that the function (2.56) found in this way satisfies (1.60′), indeed. To this end we introduce the function

$$h^{\alpha}_{\gamma}(x) \equiv \frac{\partial\Phi^{\alpha}}{\partial\theta'^{\gamma}} - \mu^{\alpha}_{\beta}(\Phi)\overset{\bullet}{\Lambda}{}^{\beta}_{\gamma}, \tag{2.57}$$

where

$$\mu^{\alpha}_{\beta}(\Phi) \equiv \mu^{\alpha}_{\beta}(\Phi; g), \tag{2.58}$$

$$\overset{\bullet}{\Lambda}{}^{\alpha}_{\beta} \equiv x\lambda^{\alpha}_{\beta}(x\theta'; \bar{g}(\theta)), \tag{2.59}$$

so that the equations hold

$$\frac{d\overset{\bullet}{\Lambda}{}^{\alpha}_{\beta}}{dx} = \delta^{\alpha}_{\beta} + t^{\alpha}_{\mu\nu}(\overset{\bullet}{G})\theta'^{\nu}\overset{\bullet}{\Lambda}{}^{\mu}_{\beta}, \tag{2.60}$$

$$\overset{\bullet}{\Lambda}{}^{\alpha}_{\beta}(x=0) = 0, \tag{2.61}$$

$$\frac{d\overset{\bullet}{G}{}^{a}}{dx} = R^{a}_{\alpha}(\overset{\bullet}{G})\theta'^{\alpha}, \tag{2.62}$$

$$\overset{\bullet}{G}(x=0) = \bar{g}^{a}(\theta), \tag{2.63}$$

with $\bar{g}^{a}(\theta)$ in (2.63) defined as (2.29). The function $\overset{\bullet}{G}{}^{a}$, $\overset{\bullet}{\Lambda}{}^{\alpha}_{\beta}$, in accordance with the above said, satisfy the equations

$$\frac{\partial\overset{\bullet}{G}{}^{a}}{\partial\theta'^{\beta}} = R^{a}_{\alpha}(\overset{\bullet}{G})\overset{\bullet}{\Lambda}{}^{\alpha}_{\beta}, \tag{2.64}$$

$$\overset{\bullet}{G}{}^{a}\Big|_{\theta'=0} = \bar{g}^{a}(\theta), \tag{2.65}$$

$$\frac{\partial\overset{\bullet}{\Lambda}{}^{\alpha}_{\gamma}}{\partial\theta'^{\beta}} - \frac{\partial\overset{\bullet}{\Lambda}{}^{\alpha}_{\beta}}{\partial\theta'^{\gamma}} + t^{\alpha}_{\mu\nu}(\overset{\bullet}{G})\overset{\bullet}{\Lambda}{}^{\mu}_{\beta}\overset{\bullet}{\Lambda}{}^{\nu}_{\gamma} = 0, \tag{2.66}$$

$$\overset{\bullet}{\Lambda}{}^{\alpha}_{\beta}\Big|_{\theta'=0} = x\delta^{\alpha}_{\beta}. \tag{2.67}$$

From (2.57) we obtain taking (2.54) and (2.60) into account

$$\begin{aligned}\frac{dh^{\alpha}_{\gamma}}{dx} &= \frac{\partial}{\partial\theta'^{\gamma}}\frac{d\Phi^{\alpha}}{dx} \\ &\quad - \frac{\partial\mu^{\alpha}_{\beta}(\Phi)}{\partial\Phi^{\delta}}\mu^{\delta}_{\epsilon}(\Phi)\theta'^{\epsilon}\overset{\bullet}{\Lambda}{}^{\beta}_{\gamma} \\ &\quad - \mu^{\alpha}_{\beta}(\Phi)(\delta^{\beta}_{\gamma} + t^{\beta}_{\mu\nu}(\overset{\bullet}{G})\theta'^{\nu}\overset{\bullet}{\Lambda}{}^{\mu}_{\gamma}). \end{aligned} \tag{2.68}$$

Differentiating (2.54) with respect to θ'^{γ} we have

$$\frac{\partial}{\partial\theta'^{\gamma}}\frac{d\Phi^{\alpha}}{dx} = \frac{\partial\mu^{\alpha}_{\beta}(\Phi)}{\partial\Phi^{\delta}}\theta'^{\beta}(h^{\delta}_{\gamma} + h^{\delta}_{\epsilon}(\Phi)\overset{\bullet}{\Lambda}{}^{\epsilon}_{\gamma}) + \mu^{\alpha}_{\gamma}(\Phi). \tag{2.69}$$

Substitution of (2.69) into (2.68) leads to

$$\begin{aligned}\frac{dh^{\alpha}_{\gamma}}{dx} &= \frac{\partial\mu^{\alpha}_{\beta}(\Phi)}{\partial\Phi^{\delta}}\theta'^{\beta}h^{\delta}_{\gamma} \\ &\quad + \mu^{\alpha}_{\beta}(\Phi)(t^{\beta}_{\mu\nu}(\bar{g}(\Phi)) - t^{\beta}_{\mu\nu}(\overset{\bullet}{G}))\theta'^{\nu}\overset{\bullet}{\Lambda}{}^{\mu}_{\gamma}, \end{aligned} \tag{2.70}$$

where the effect the Eq. (1.31′) has on the functions $\mu^{\alpha}_{\beta}(\theta; g)$ has been taken into account

$$\frac{\partial\mu^{\alpha}_{\delta}}{\partial\theta^{\beta}}\mu^{\beta}_{\gamma} - \frac{\partial\mu^{\alpha}_{\gamma}}{\partial\theta^{\beta}}\mu^{\beta}_{\delta} = -t^{\beta}_{\delta\gamma}(\bar{g}(\theta))\mu^{\alpha}_{\beta}. \tag{1.30'}$$

The use of (1.25′) and (2.54) leads to

$$\begin{aligned}\frac{d\bar{g}^{a}(\Phi)}{dx} &= \frac{\partial\bar{g}^{a}(\Phi)}{\partial\Phi^{\beta}}\frac{d\Phi^{\beta}}{dx} \\ &= R^{a}_{\alpha}(\bar{g}(\Phi))\lambda^{\alpha}_{\beta}(\Phi)\mu^{\beta}_{\gamma}(\Phi)\theta'^{\gamma} \\ &= R^{a}_{\alpha}(\bar{g}(\Phi))\theta'^{\alpha}, \end{aligned} \tag{2.71}$$

where

$$\bar{g}^a(\Phi(x=0)) = \bar{g}^a(\theta). \tag{2.72}$$

By comparing (2.62) and (2.63) with (2.71) and (2.72) we conclude that

$$\bar{g}^a(\Phi(x)) = \overset{\bullet}{G}{}^a(x). \tag{2.73}$$

In virtue of (2.73) one has from (2.70)

$$\frac{dh^\alpha_\gamma}{dx} = \frac{\partial\mu^\alpha_\beta(\Phi)}{\partial\Phi^\delta}\theta'^\beta h^\delta_\gamma, \tag{2.74}$$

with the equality

$$h^\alpha_\gamma(x=0)=0 \tag{2.75}$$

following from (2.55) and (2.61). The only solution of Eq. (2.74) with the boundary condition (2.75) is

$$h^\alpha_\gamma(x)=0. \tag{2.76}$$

Therefore, at every x the relation

$$\frac{\partial\Phi^\alpha}{\partial\theta'^\gamma}=\mu^\alpha_\beta(\Phi)\overset{\bullet}{\Lambda}{}^\beta_\gamma \tag{2.77}$$

holds a consequence of (2.54) and (2.55). Besides, it follows from (2.54) and (2.55) that

$$\Phi^\alpha|_{\theta'=0}=\theta^\alpha. \tag{2.78}$$

Setting $x=1$ in (2.59), (2.77), and (2.78) one sees that the function (2.56) satisfies (1.60′), indeed.

Thus, the reconstruction of the local quasigroup, with the structure function $R^a_\alpha(g)$, $t^\gamma_{\alpha\beta}(g)$ subject to (1.16) and (1.20) given reduces, if canonical parameter are used, to the following steps: (2.25), (2.26), and (2.29) lead to the transformation law $\bar{g}^a(\theta)$; the auxiliary functions $\lambda^\alpha_\beta(\theta;g)$ are found from (2.27), (2.28), and (2.30); the composition law $\varphi^\alpha(\theta,\theta';g)$ follows from (2.54), (2.55), and (2.56).

It remains to confirm that the canonical parametrization is admissible at least locally. Let $\bar{g}^a(\theta)$, $\lambda^\alpha_\beta(\theta)$ be a solution of the set (1.25′)–(1.32′). This set, with the structure functions $R^a_\alpha(g)$, $t^\gamma_{\alpha\beta}(g)$ given, was already mentioned to be covariant under the transformations (1.84) if $\Omega^\alpha_\beta(g)=\delta^\alpha_\beta$, where Ω^α_β is defined as (1.102). Write in this case the transformation law (1.103) as

$$\frac{\partial\psi^\alpha(\theta;g)}{\partial\theta^\gamma}=\mu^\alpha_\beta(\psi(\theta;g);g)\,\lambda^\beta_{1\gamma}(\theta;g), \tag{2.79}$$

where

$$\psi^\alpha(\theta=0;g)=0. \tag{2.80}$$

The function $\lambda^\alpha_{1\beta}(\theta;g)$ along with the corresponding function $\bar{g}^a_1(\theta;g)$ satisfies a set of equations which coincides with (1.25′)–(1.32′). To confirm that the canonical parametrization is admissible suffices it to point a function $\psi^\alpha(\theta;g)$ that satisfies (2.79) and (2.80) with the function $\lambda^\alpha_{1\beta}(\theta;g)$ obeying the condition of canonicity

$$\lambda^\alpha_{1\beta}\,\theta^\beta=\theta^\alpha. \tag{2.81}$$

Using the method that was developed above as applied to (2.29), (2.30), and (2.56), it is easy to show that the function

$$\psi^\alpha(\theta;g)\equiv\Phi^\alpha_0(x=1), \tag{2.82}$$

where $\Phi^\alpha_0(x)$ is a solution of the equation

$$\frac{d\Phi^\alpha_0}{dx}=\mu^\alpha_\beta(\Phi_0;g)\,\theta^\beta \tag{2.83}$$

satisfying the condition

$$\Phi^\alpha_0(x=0)=0 \tag{2.84}$$

meets the requirement formulated as (2.79), (2.80), and (2.81). Therefore, canonical parameterization is admissible, at least locally.

We conclude this section by noting that for the canonical parameterization the relations are true:

$$\tilde{\theta}^\alpha=-\theta^\alpha, \tag{2.85}$$

$$\tilde{\lambda}^\alpha_\beta(\theta;\bar{g}(-\theta))=\lambda^\alpha_\beta(-\theta;g). \tag{2.86}$$

Note also that the function (2.29) may be formally presented as

$$\bar{g}^a(\theta)=\exp\{\theta^\alpha\Gamma_\alpha\}\,g^a, \tag{2.87}$$

where Γ_α are the generators (1.22).

3. QUASISUPERGROUP

In the previous sections the variables g^a and parameters θ^α were assumed to be commuting (Bose) quantitites. Now we are going to extend the quasigroup construction to the case when the variables g^a and parameters θ^α are elements of a graded algebra:

$$g^ag^b=(-1)^{n_an_b}g^bg^a, \tag{3.1}$$

$$\theta^\alpha\theta^\beta=(-1)^{n_\alpha n_\beta}\theta^\beta\theta^\alpha, \tag{3.2}$$

$$g^a\theta^\alpha=(-1)^{n_an_\alpha}\theta^\alpha g^a, \tag{3.3}$$

where n_a, n_α are Grassmann parities of the variables g^a and parameters θ^α, respectively. Each of the parities n_a, n_α takes the values of 0 or 1 depending on whether the corresponding quantity is an even (bosonic) or odd (fermionic) element.

The functional equations (1.1)–(1.9) may be formally extended to the case (3.1)–(3.3) to now define a construction called quasisupergroup.

Infinitesimal transformations are defined by the formula (1.10), where

$$R^a_\alpha(\,)\equiv\left.\frac{\partial_rf^a(g,\theta)}{\partial\theta^\alpha}\right|_{\theta=0}. \tag{3.4}$$

Here ∂_r is the right derivative. These functions satisfy the relations

$$\frac{\partial_rR^a_\alpha}{\partial g^b}R^b_\beta-(-1)^{n_\alpha n_\beta}\frac{\partial_rR^a_\beta}{\partial g^b}R^b_\alpha = -R^a_\gamma t^\gamma_{\alpha\beta}, \tag{3.5}$$

where

$$t^\gamma_{\alpha\beta}(g)\equiv\left[(-1)^{n_\alpha n_\beta}\frac{\partial^2_r}{\partial\theta^\alpha\partial\theta'^\beta}-\frac{\partial^2_r}{\partial\theta^\beta\partial\theta'^\alpha}\right]\varphi^\gamma(\theta,\theta';g)\Big|_{\theta=\theta'} \tag{3.6}$$

are the structure coefficients subject to the relations

$$t^\gamma_{\alpha\beta}=-(-1)^{n_\alpha n_\beta}t^\gamma_{\beta\alpha}, \tag{3.7}$$

$$(-1)^{n_\alpha n_\delta}\frac{\partial_r t^\mu_{\alpha\beta}}{\partial g^a}R^a_\delta+(-1)^{n_\delta n_\beta}\frac{\partial_r t^\mu_{\delta\alpha}}{\partial g^a}R^a_\beta$$
$$+(-1)^{n_\beta n_\alpha}\frac{\partial_r t^\mu_{\beta\delta}}{\partial g^a}R^a_\alpha+(-1)^{n_\alpha n_\delta}t^\mu_{\alpha\gamma}\,t^\gamma_{\beta\delta}$$
$$+(-1)^{n_\delta n_\beta}t^\mu_{\delta\gamma}\,t^\gamma_{\alpha\beta}+(-1)^{n_\beta n_\alpha}t^\mu_{\beta\gamma}\,t^\gamma_{\delta\alpha}=0. \tag{3.8}$$

The analog of Eq. (1.25) has the form

$$\frac{\partial_r\bar g^a}{\partial\theta^\beta}=R^a_\alpha(\bar g)\lambda^\alpha_\beta, \tag{3.9}$$

where

$$\bar g^a(\theta=0)=g^a, \tag{3.10}$$

$$\lambda^\alpha_\beta\mu^\beta_\gamma=\delta^\alpha_\gamma, \tag{3.11}$$

$$\mu^\alpha_\beta(\theta;g)\equiv\left.\frac{\partial_r\varphi^\alpha(\theta,\theta';g)}{\partial\theta'^\beta}\right|_{\theta'=0}. \tag{3.12}$$

Functions (3.12) obey the equation

$$\frac{\partial_r\mu^\alpha_\beta}{\partial\theta^\gamma}\mu^\gamma_\delta-(-1)^{n_\beta n_\delta}\frac{\partial_r\mu^\alpha_\delta}{\partial\theta^\gamma}\mu^\gamma_\beta$$
$$=-\mu^\alpha_\gamma\,t^\gamma_{\beta\delta}(\bar g) \tag{3.13}$$

and the condition

$$\mu^\alpha_\beta(\theta=0;g)=\delta^\alpha_\beta. \tag{3.13a}$$

For the functions $\lambda^\alpha_\beta(\theta;g)$, one has the equation

$$\frac{\partial_r\lambda^\alpha_\gamma}{\partial\theta^\beta}-(-1)^{n_\gamma n_\beta}\frac{\partial_r\lambda^\alpha_\beta}{\partial\theta^\gamma}$$
$$+(-1)^{n_\gamma n_\mu}t^\alpha_{\mu\nu}(\bar g)\lambda^\nu_\gamma\lambda^\mu_\beta=0 \tag{3.14}$$

and the boundary condition

$$\lambda^\alpha_\beta(\theta=0;g)=\delta^\alpha_\beta. \tag{3.15}$$

The equation for the composition function $\varphi^\alpha(\theta,\theta';g)$ has the form

$$\frac{\partial_r\varphi^\alpha(\theta,\theta';g)}{\partial\theta'^\gamma}$$
$$=\mu^\alpha_\beta(\varphi(\theta,\theta';g);g)\,\lambda^\beta_\gamma(\theta';\bar g(\theta)), \tag{3.16}$$

with the boundary condition

$$\varphi^\alpha(\theta,\theta'=0;g)=\theta^\alpha. \tag{3.17}$$

Within Eqs. (3.9), (3.10), and (3.14)–(3.17) one may put forward a problem of reconstructing the quasisupergroup if the structure functions $R^a_\alpha(g)$, $t^\gamma_{\alpha\beta}(g)$ are given satisfying (3.5), (3.7), and (3.8). This problem can be solved for canonical parametrization (2.21) where the order of factors is fixed. The final result is as follows: the function $\bar g^a(\theta)$ is given by (2.29), (2.25), and (2.26), the auxiliary function $\lambda^\alpha_\beta(\theta;g)$ is given by (2.30), (2.27), and (2.28), the compositional function $\varphi^\alpha(\theta,\theta';g)$ is given by (2.56), (2.54), and (2.55), the order of factors in (2.25), (2.27), and (2.54) being kept fixed.

We present now the formal expression for $\bar g^a(\theta)$ in the canonical parametrization

$$\bar g^a(\theta)=g^a\exp\{\Gamma_\alpha\theta^\alpha\}, \tag{3.18}$$

where

$$\Gamma_\alpha=\frac{\overleftarrow{\partial}_r}{\partial g^a}R^a_\alpha(g) \tag{3.19}$$

are generators acting to the left. In virtue of (3.5) the generators (3.19) satisfy the relations

$$\Gamma_\alpha\Gamma_\beta-(-1)^{n_\alpha n_\beta}\Gamma_\beta\Gamma_\alpha=-\Gamma_\gamma\,t^\gamma_{\alpha\beta}. \tag{3.20}$$

Thus, we see that inclusion of fermions does not cause any essential alterations in the formal construction of the quasigroup.

4. FIRST CLASS CONSTRAINTS AS A REALIZATION OF THE QUASIGROUP

Let $q^i, p_i (i=1,\dots,N)$ be canonical coordinates and moments of a dynamical system with the initial Hamiltonian $H(q,p)$ and the first class constraints $T_\alpha(q,p)$ $(\alpha=1,\dots,r)$. The constraints T_α are in involution among themselves and with the Hamiltonian:

$$\{T_\alpha,T_\beta\}=U^\gamma_{\alpha\beta}T_\gamma, \tag{4.1}$$

$$\{H,T_\alpha\}=V^\beta_\alpha T_\beta. \tag{4.2}$$

It is assumed for simplicity, that the dynamical variables are bosonic and the second class constraints are absent. Then $\{\cdots\}$ are ordinary Poisson brackets

$$\{A,B\}\equiv\frac{\partial A}{\partial q^i}\frac{\partial B}{\partial p_i}-\frac{\partial A}{\partial p_i}\frac{\partial B}{\partial q^i}. \tag{4.3}$$

It is convenient to combine canonical coordinates and moments in a column

$$g^a=\begin{pmatrix}q^i\\p_i\end{pmatrix} \tag{4.4}$$

and refer in what follows to the lower case latin indices as running through the values $(1,2,\dots,2N)$. Then (4.3) takes the form

$$\{A,B\}=\frac{\partial A}{\partial g^a}\epsilon^{ab}\frac{\partial B}{\partial g^b}, \tag{4.5}$$

where $\epsilon^{ab}=-\epsilon^{ba}$ is the corresponding simplectic matrix.

Introduce transformations of the variables (4.4) which are generated by the constraints

$$\delta g^a=R^a_\alpha(g)\,\theta^\alpha, \tag{4.6}$$

where

$$R^a_\alpha(g)\equiv\{g^a,T_\alpha\}=\epsilon^{ab}\frac{\partial T_\alpha}{\partial g^b} \tag{4.7}$$

and $\theta^\alpha\to 0$. Let us calculate now the Lie bracket of the functions (4.7). The use of (4.1) gives

$$\frac{\partial R^a_\alpha}{\partial g^b}R^b_\beta-(\alpha\longleftrightarrow\beta)$$
$$=U^\gamma_{\alpha\beta}R^a_\gamma+\epsilon^{ab}\frac{\partial U^\gamma_{\alpha\beta}}{\partial g^b}T_\gamma, \tag{4.8}$$

so that

$$\left.\left(\frac{\partial R^a_\alpha}{\partial g^b}R^b_\beta-(\alpha\longleftrightarrow\beta)+t^\gamma_{\alpha\beta}R^a_\gamma\right)\right|_{T=0}=0, \tag{4.9}$$

where the new designation of the structure functions of the involution is introduced as

$$t^\gamma_{\alpha\beta}(g)\equiv-U^\gamma_{\alpha\beta}(g). \tag{4.10}$$

Equation (4.1), along with the completeness and irreducibility of the constraints, allows us to state that

$$\left.\left(\frac{\partial t^{\mu}_{\alpha\beta}}{\partial g^{a}} R^{a}_{\delta} + t^{\mu}_{\alpha\gamma}\, t^{\gamma}_{\beta\delta}\right)\right|_{T=0} + \text{cycl. perm. } (\alpha,\beta,\delta) = 0\,. \tag{4.11}$$

Let us define finite transformations $g^a \to \bar{g}^a(\theta)$ of the variables (4.4) by means of the equations

$$\frac{\partial \bar{g}^a}{\partial \theta^{\beta}} = R^{a}_{\alpha}(\bar{g})\,\lambda^{\alpha}_{\beta}\,, \tag{4.12}$$

$$\bar{g}^a(\theta=0) = g^a\,, \tag{4.13}$$

$$\frac{\partial \lambda^{\alpha}_{\gamma}}{\partial \theta^{\beta}} - \frac{\partial \lambda^{\alpha}_{\beta}}{\partial \theta^{\gamma}} + t^{\alpha}_{\mu\nu}(\bar{g})\,\lambda^{\mu}_{\beta}\,\lambda^{\nu}_{\gamma} = 0\,, \tag{4.14}$$

$$\lambda^{\alpha}_{\beta}(\theta=0) = \delta^{\alpha}_{\beta}\,, \tag{4.15}$$

where the structure functions are defined by (4.7) and (4.10). Using (4.12) one has

$$\frac{\partial T_{\alpha}(\bar{g})}{\partial \theta^{\delta}} = -\, t^{\gamma}_{\alpha\beta}(\bar{g})\,\lambda^{\beta}_{\delta}\,T_{\gamma}(\bar{g})\,, \tag{4.16}$$

where

$$T_{\alpha}(\bar{g})\big|_{\theta=0} = T_{\alpha}(g) \tag{4.17}$$

due to (4.13). If the initial data is localized on the hypersurface specialized by the constraints

$$T_{\alpha}(g) = 0 \tag{4.18}$$

one has in virtue of (4.16) and (4.17)

$$T_{\alpha}(\bar{g}) = 0\,. \tag{4.19}$$

This means that the transformation given by (4.12)–(4.15) leaves the variables (4.4) on the hypersurface (4.18). Appealing to (4.9), (4.11) and (4.19) one can confirm the integrability of equations (4.12)–(4.15) under the condition that the initial data is localized on the hypersurface (4.18).

Equations (4.12)–(4.15) define naturally an action of the (local) quasigroup on the hypersurface of constraints. This quasigroup is meant to be reconstructed based on the structure functions (4.7) and (4.10) as it was described in Sec. 2. The classical Hamiltonian action

$$W(g) = \int dt\, \left[\tfrac{1}{2} g^a \epsilon_{ab}\, \dot{g}^b - H(g)\right], \tag{4.20}$$

with the relation $\epsilon_{ab}\,\epsilon^{bc} = \delta^{c}_{a}$ holding, on the hypersurface of constraints (4.18) is invariant under the transformations given by (4.12)–(4.15).

Up to this stage we were dealing with the complete set of the primary variables (4.4). One may, however, confine oneself explicitly to the hypersurface of constraints (4.18) by eliminating r variables. To this end let us divide the variables (4.4) as follows

$$g^a = \begin{pmatrix} g'^{\alpha} \\ g''^{A} \end{pmatrix}, \tag{4.21}$$

where $\alpha = 1,\dots,r$; $A = r+1,\dots,2N-r$. Assume that the equation of constraints (4.18) when written in the form

$$T_{\alpha}(g', g'') = 0 \tag{4.22}$$

may be solved with respect to the variables g'^{α}, i.e.,

$$\operatorname{Det}\left\{\frac{\partial T(g', g'')}{\partial g'}\right\} \neq 0\,. \tag{4.23}$$

Then it follows from (4.22) that

$$g'^{\alpha} = \tau^{\alpha}(g'') \tag{4.24}$$

and also

$$\frac{\partial T_{\alpha}}{\partial g'^{\mu}}\,\frac{\partial g'^{\mu}}{\partial g''^{A}} + \frac{\partial T_{\alpha}}{\partial g''^{A}} = 0\,. \tag{4.25}$$

From (4.1) on the hypersurface of constraints and (4.25) it follows that

$$R^{\mu}_{\alpha} = \frac{\partial g'^{\mu}}{\partial g''^{A}}\, R^{A}_{\alpha}\,. \tag{4.26}$$

Transformations of variables g''^{A} on the hypersurface of constraints are

$$\delta g''^{A} = R^{A}_{\alpha}(g'')\,\theta^{\alpha}\,, \tag{4.27}$$

where

$$R^{A}_{\alpha}(g'') \equiv R^{A}_{\alpha}(g', g'')\big|_{g'=\tau(g'')}\,. \tag{4.28}$$

For the function (4.28) from (4.9) and (4.26) we derive the relations

$$\frac{\partial R^{A}_{\alpha}}{\partial g''^{B}}\, R^{B}_{\beta} - (\alpha \longleftrightarrow \beta) = -\, t^{\gamma}_{\alpha\beta}(g'')\, R^{A}_{\gamma}\,, \tag{4.29}$$

where

$$t^{\gamma}_{\alpha\beta}(g'') \equiv t^{\gamma}_{\alpha\beta}(g', g'')\big|_{g'=\tau(g'')}\,. \tag{4.30}$$

$$\left(\frac{\partial t^{\mu}_{\alpha\beta}}{\partial g''^{A}} R^{A}_{\delta} + t^{\mu}_{\alpha\gamma}\, t^{\gamma}_{\beta\delta}\right) + \text{cycl. perm. } (\alpha,\beta,\delta) = 0\,. \tag{4.31}$$

The division (4.21) of the primary variables leads naturally to analogous division in the functions

$$\bar{g}^a = \begin{pmatrix} \bar{g}'^{\alpha} \\ \bar{g}''^{A} \end{pmatrix} \tag{4.32}$$

so that

$$T_{\alpha}(\bar{g}, \bar{g}'') = 0 \tag{4.33}$$

simultaneously with (4.22).

Equations (4.12)–(4.15) may be represented on the hypersurface of constraints as

$$\frac{\partial \bar{g}''^{A}}{\partial \theta^{\beta}} = R^{A}_{\alpha}(\bar{g}'')\,\lambda^{\alpha}_{\beta}\,, \tag{4.34}$$

$$\bar{g}''^{A}(\theta=0) = g''^{A}\,, \tag{4.35}$$

$$\frac{\partial \lambda^{\alpha}_{\gamma}}{\partial \theta^{\beta}} - \frac{\partial \lambda^{\alpha}_{\beta}}{\partial \theta^{\gamma}} + t^{\alpha}_{\mu\nu}(\bar{g}'')\,\lambda^{\mu}_{\beta}\,\lambda^{\nu}_{\gamma} = 0\,, \tag{4.36}$$

$$\lambda^{\alpha}_{\beta}(\theta=0) = \delta^{\alpha}_{\beta}\,, \tag{4.37}$$

where the structure functions are defined by (4.28) and (4.30). Equations (4.34)–(4.37) contain only independent variables of the hypersurface of constraints.

5. GAUGE INVARIANCE

Let g^a be real variables and $W(g)$ be a function invariant under the infinitesimal transformations

$$\delta g^a = R^{a}_{\alpha}(g)\,\theta^{\alpha}\,, \tag{5.1}$$

where the functions $R^{a}_{\alpha}(g)$ obey (1.16) and (1.20).

Consider the formal integral over all g^a:

$$Z_\Psi = \int \exp\{iW(g)\}\, \delta(\Psi(g))\, \Delta_\Psi(g)\, d\mu(g)\,, \quad (5.2)$$

where $\Psi^\alpha(g)$ is the "gauge" function. [Its derivatives are denoted for brevity by subscripts after a comma, like: $\Psi^\alpha_{,a}(g) \equiv \partial\Psi^\alpha(g)/\partial g^a$]

$$\Delta_\Psi(g) \equiv \mathrm{Det}\,\{\mathscr{D}^{-1}\}\,, \quad (5.3)$$

$$\mathscr{D}_\beta^{-1\alpha}(g) = \Psi^\alpha_{,a}(g)\, R^a_\beta(g)\,, \quad (5.4)$$

$$d\mu(g) = M(g) \prod_a dg^a\,. \quad (5.5)$$

The function $M(g)$ satisfies the equation

$$\Gamma^T_\alpha M = t^\beta_{\alpha\beta} M\,, \quad (5.6)$$

where

$$\Gamma^T_\alpha = -\frac{\partial}{\partial g^a} R^a_\alpha(g) \quad (5.7)$$

are the transposed generators and $t^\gamma_{\alpha\beta}(g)$ are the structure coefficients in (1.16). In virtue of (1.16) Γ^T_α obey the commutation relations

$$[\Gamma^T_\alpha, \Gamma^T_\beta] = -\Gamma^T_\gamma t^\gamma_{\alpha\beta} = t^\gamma_{\alpha\beta,a} R^a_\gamma - t^\gamma_{\alpha\beta}\Gamma^T_\gamma\,. \quad (5.8)$$

The compatibility of Eqs. (5.6) is provided by (5.8) and the relation

$$t^\mu_{\alpha\beta,a} R^a_\mu - t^\gamma_{\alpha\beta} t^\mu_{\gamma\mu} - (t^\mu_{\alpha\mu,a} R^a_\beta - (\alpha \longleftrightarrow \beta)) = 0\,, \quad (5.9)$$

which follows from (1.20).

The law of finite quasigroup transformation of the function M follows from (1.25), (1.77), and (5.6):

$$M(\bar g) = M(g)(E(g,\theta))^{-1}\,, \quad (5.10)$$

where $E(g,\theta)$ is defined as (1.80). Equations (1.79) and (5.10) determine the transformation of the integration measure

$$d\mu(\bar g) = \gamma(\theta; g)\, d\mu(g)\,, \quad (5.11)$$

where

$$\gamma(\theta; g) = \frac{\mathrm{Det}\{\lambda(\theta; g)\}}{\mathrm{Det}\{\tilde\lambda(\theta; g)\}}\,. \quad (5.12)$$

An important property of the integral (5.2) is its independence of a choice of the "gauge" function $\Psi^\alpha(g)$ (the gauge invariance). To see this let us perform the variation of the integration variables in (5.2)

$$g^a \to g^a + \delta g^a\,, \quad (5.13)$$

where

$$\delta g^a = R^a_\alpha \mathscr{D}^\alpha_\beta \delta\Psi^\beta \quad (5.14)$$

and $\mathscr{D}^\alpha_\beta$ is the matrix inverse to (5.4), $\delta\Psi^\alpha(g)$ is an arbitrary infinitesimal function. The variation (5.13) and (5.14) induces, up to the first order in $\delta\Psi^\alpha$ the change of the gauge function

$$\Psi^\alpha(g) \to \Psi^\alpha(g) + \delta\Psi^\alpha(g)\,. \quad (5.15)$$

Using (1.16) and (5.6) and the invariance of $W(g)$ one can easily see that within the first order in $\delta\Psi^\alpha$ the relation holds,

$$Z_\Psi = Z_{\Psi+\delta\Psi}\,. \quad (5.16)$$

Now we are in a position to return to dynamical systems subject to the first class constraints. We shall see that the quantum transition amplitude exactly reproduces the structure of the integral (5.2).

Let g^a be again the variables (4.4) taken at a time moment t. Within canonical gauges $\Psi^\alpha(g)$ the quantum transition amplitude for the systems given by the action (4.20) and the first class constraints $T_\alpha(g)$ has the form[2,3]:

$$Z_\Psi = \int \exp\{iW(g)\} \prod_t \delta(\Psi(g))\, \Delta_\Psi(g)\, \delta(T(g)) \prod_a dg^a\,, \quad (5.17)$$

where $\Delta_\Psi(g)$ is defined by (5.3) and (5.4).

The factor $\delta(T(g))$ acts in (5.17) in two ways: in the first place it localizes the integrand in (5.17) onto the hypersurface of constraints, in the second place it is an analog of $M(g)$ in (5.5), as we are going now to see. Let us substitute the functions (4.7) for R^a_α into (5.7). It follows from (4.7) that

$$R^a_{\alpha,a} \equiv 0 \quad (5.18)$$

and hence

$$\Gamma^T_\alpha = -\Gamma_\alpha = -R^a_\alpha \frac{\partial}{\partial g^a}\,. \quad (5.19)$$

Further on, we get in succession

$$\begin{aligned}\Gamma^T_\alpha \delta(T(g)) &= -R^a_\alpha \frac{\partial}{\partial g^a}\delta(T) \\ &= -R^a_\alpha \frac{\partial T_\beta}{\partial g^a}\frac{\partial}{\partial T_\beta}\delta(T) \\ &= -\{T_\beta, T_\alpha\}\frac{\partial}{\partial T_\beta}\delta(T) \\ &= -U^\gamma_{\beta\alpha} T_\gamma \frac{\partial}{\partial T_\beta}\delta(T) \\ &= -t^\gamma_{\alpha\beta} T_\gamma \frac{\partial}{\partial T_\beta}\delta(T) \\ &= -t^\gamma_{\alpha\beta}\left[\frac{\partial}{\partial T_\beta} T_\gamma\, \delta(T) - \delta^\beta_\gamma\, \delta(T)\right] \\ &= t^\beta_{\alpha\beta}(g)\, \delta(T(g))\,. \end{aligned} \quad (5.20)$$

In other words, the function $\delta(T(g))$ satisfies an equation analogous to (5.6).

The functional integration in (5.17) is carried out over the complete set of variables (4.4). One can, however by making the division (4.21) integrate over the variables g'^α thus removing the δ-function of the constraints. In this way we obtain

$$Z_\Psi = \int \exp\{iW(g'')\} \times \prod_t \delta(\Psi(g''))\, \Delta_\Psi(g'')\, d\mu(g'')\,, \quad (5.21)$$

where

$$W(g'') \equiv W(g', g'')|_{g' = \tau(g'')}\,, \quad (5.22)$$

$$\Psi^\alpha(g'') \equiv \Psi^\alpha(g', g'')|_{g' = \tau(g'')}\,, \quad (5.23)$$

$$\Delta_\Psi(g'') \equiv \mathrm{Det}\{\mathscr{D}^{-1}(g'')\}\,, \quad (5.24)$$

$$\mathscr{D}_{\beta}^{-1\alpha}(g'') = \Psi^{\alpha}_{,A}(g'')\, R^{A}_{\beta}(g'') . \qquad (5.25)$$

The functions $R^{A}_{\alpha}(g)$ are defined by (4.28).

The integration measure $d\mu(g'')$ has the form

$$d\mu(g'') \equiv M(g'') \prod_{A} dg''^{A} , \qquad (5.26)$$

where

$$M(g'') = \left(\mathrm{Det}\left\{\frac{\partial T(g', g'')}{\partial g'}\right\}\right)^{-1}\bigg|_{g' = \tau(g'')} . \qquad (5.27)$$

One can show that the function (5.27) satisfies the equation

$$\Gamma^{T}_{\alpha}(g'')\, M(g'') = t^{\beta}_{\alpha\beta}(g'')\, M(g'') , \qquad (5.28)$$

where

$$\Gamma^{T}_{\alpha}(g'') \equiv -\frac{\partial}{\partial g''^{A}} R^{A}_{\alpha}(g'') \qquad (5.29)$$

and the functions $t^{\gamma}_{\alpha\beta}(g'')$ are defined by (4.30).

Note that the action (5.22) is invariant under the transformations (4.27).

Thus, we see that expressions (5.17) and (5.21) have the same general structure as the integral (5.2). Their gauge invariance is provided via the same mechanism that leads to (5.16).

ACKNOWLEDGMENTS

In conclusion the author thanks Professor F.A. Berezin and Professor E.S. Fradkin for useful discussions.

APPENDIX

Here we introduce "left" and "right" measures on the quasigroup

$$dG(\theta; g) \equiv \mathrm{Det}\{\lambda(\theta; g)\}\,[d\theta] , \qquad (A1)$$

$$d\tilde{G}(\theta; g) \equiv \mathrm{Det}\{\tilde{\lambda}(\theta; f^{-1}(g,\theta))\}\,[d\theta] , \qquad (A2)$$

where

$$[d\theta] = \prod_{\alpha} d\theta^{\alpha} . \qquad (A3)$$

Let us change to the new variables θ^{α}_{L},

$$\theta^{\alpha}_{L} \equiv \varphi^{\alpha}(\theta', \theta; f^{-1}(g, \theta')) . \qquad (A4)$$

Using the relation

$$\frac{\partial \theta^{\alpha}_{L}}{\partial \theta^{\gamma}} = \mu^{\alpha}_{\beta}(\theta_{L}; f^{-1}(g,\theta'))\, \lambda^{\beta}_{\gamma}(\theta; g) \qquad (A5)$$

which follows from (1.60) we obtain

$$dG(\theta; g) = dG(\theta_{L}; f^{-1}(g,\theta')) . \qquad (A6)$$

Let us change now to the new variables θ^{α}_{R} in (A2):

$$\theta^{\alpha}_{R} \equiv \varphi^{\alpha}(\theta, \theta'; f^{-1}(g, \theta)) . \qquad (A7)$$

Using the relations

$$\frac{\partial \theta^{\alpha}_{R}}{\partial \theta^{\gamma}} = \tilde{\mu}^{\alpha}_{\beta}(\theta_{R}; f^{-1}(g,\theta))\, \tilde{\lambda}^{\beta}_{\gamma}(\theta; f^{-1}(g, \theta)) \qquad (A8)$$

which follows from (1.63) and also the relation

$$f^{-1\alpha}(g, \theta) = f^{-1\alpha}(f(g, \theta'), \theta_{R}) , \qquad (A9)$$

which follows from (1.115) and (1.118), we obtain

$$d\tilde{G}(\theta; g) = d\tilde{G}(\theta_{R}; f(g,\theta')) . \qquad (A10)$$

The properties (A6) and (A8) are analogs of the left and right invariances of the corresponding group measures. Note that the inversion

$$\theta^{\alpha} \to \tilde{\theta}^{\alpha}(\theta; g) \qquad (A11)$$

turns (A1) into (A2) due to (1.7) and (1.66).

Let now $\Psi^{\alpha}(g)$ be an admissible "gauge" function. Represent the unity as

$$1 = \int \delta(\Psi(f(g,\theta)))\, \mathrm{Det}\left\{\frac{\partial \Psi(f(g,\theta))}{\partial \theta}\right\} [d\theta] , \qquad (A12)$$

where the integration is to be performed over the region which corresponds to the quasigroup as a whole.

Using (1.25) one has

$$\frac{\partial \Psi^{\alpha}(f(g,\theta))}{\partial \theta^{\gamma}} = \Psi^{\alpha}_{,a}(f(g,\theta))\, \frac{\partial f^{a}(g,\theta)}{\partial \theta^{\gamma}}$$
$$= \Psi^{\alpha}_{,a}(f(g,\theta))\, R^{a}_{\beta}(f(g,\theta))\, \lambda^{\beta}_{\gamma}(\theta; g) , \qquad (A13)$$

whence

$$\mathrm{Det}\left\{\frac{\partial \Psi(f(g,\theta))}{\partial \theta}\right\}$$
$$= \Delta_{\Psi}(f(g,\theta))\, \mathrm{Det}\{\lambda(\theta; g)\} , \qquad (A14)$$

where $\Delta_{\Psi}(g)$ is defined by (5.3) and (5.4). Using (A14) in (A12) we obtain

$$1 = \int \delta(\Psi(f(g,\theta))\, \Delta_{\Psi}(f(g,\theta))\, dG(\theta; g) , \qquad (A15)$$

where dG is the measure (A1). The inversion (A11) converts (A15) into

$$1 = \int \delta(\Psi(f^{-1}(g,\theta)))\, \Delta_{\Psi}(f^{-1}(g,\theta))\, d\tilde{G}(\theta; g) , \qquad (A16)$$

and $d\tilde{G}$ is the measure (A2).

We are now going to give an alternative proof to the gauge invariance of the integral (5.2) with the aid of (A15) and (A16). Introduce into the integrand in (5.2) the expansion of the unity (A16) where another admissible "gauge" function Ψ_1 is taken

$$Z_{\Psi} = \int\int d\tilde{G}(\theta; g)\, \Delta_{\Psi_1}(f^{-1}(g,\theta))\, \delta(\Psi_1(f^{-1}(g,\theta)))$$
$$\times \exp\{iW(g)\}\, \Delta_{\Psi}(g)\, \delta(\Psi(g))\, d\mu(g) . \qquad (A17)$$

Perform the change of variables

$$g^{a} \to f^{a}(g,\theta) \qquad (A18)$$

in (A17). Keeping in mind the invariance of $W(g)$ and using (5.11) and (5.12) we have

$$Z_{\Psi} = \int\int [d\theta]\, \mathrm{Det}\{\tilde{\lambda}(\theta; g)\}\, \Delta_{\Psi_1}(g)\, \delta(\Psi_1(g))$$
$$\times \exp\{iW(g)\}\, \Delta_{\Psi}(f((g,\theta))\, \delta(\Psi(f(g,\theta)))$$
$$\times \ \gamma(\theta; g)\, d\mu(g)$$
$$= \int\int dG(\theta; g)\, \Delta_{\Psi}(f(g,\theta))\, \delta(\Psi(f(g,\theta)))$$
$$\times \exp\{iW(g)\}\, \Delta_{\Psi_1}(g)\, \delta(\Psi_1(g))\, d\mu(g) = Z_{\Psi_1} , \qquad (A19)$$

where (A15) is taken into account in the final equality.

[1] L. S. Pontriagin, *Nepreryvnie Gruppy* (*Continuous Groups*), 3rd edition (Nauka, Moscow, 1973).
[2] L. D. Faddeev, Teor. Mat. Fiz. **1**, 3 (1969).
[3] E. S. Fradkin, Acta Univ. Wratislav. N 207, Proc. Xth Winter School of Theoretical Physics in Karpacz (1973).

A criterion for completeness of Casimir operators

F. Berdjis
Max-Planck Institut, 'D813 Starnberg, West Germany

(Received 6 February 1981; accepted for publication 10 April 1981)

We give a criterion for a set of Casimir operators of any semisimple Lie algebra to be an algebraically independent generating set of the algebra of the Casimir operators. The criterion is applied to supply complete sets for the Casimir operators of A_l, B_l, C_l, and D_l. With the aid of a method to construct some Casimir operators we then furnish complete sets for the Casimir operators of the exceptional Lie algebras using the criterion.

PACS numbers: 02.20.Sv

In the physics literature there are many papers which construct Casimir operators for the classical series A_l, B_l, C_l, and D_l and for the exceptional Lie algebra G_2[1–16] (see Refs. 17–23). But their proof of completeness is in no way satisfactory. In this paper I give a criterion for a set of Casimir operators of any finite-dimensional complex semisimple Lie algebra to be complete in the sense that it generates algebraically all Casimir operators and is algebraically independent. This is important because the simultaneous eigenvalues of such a set of operators characterizes the irreducible finite-dimensional representations of the Lie algebra in a one-to-one manner. I will apply this criterion then to give complete sets of Casimir operators for A_l, B_l, C_l, D_l, and G_2. I give also a method for construction of some Casimir operators of any semisimple Lie algebra which fulfill a part of the criterion. With the aid of this method I will give in a second paper complete sets of Casimir operators for the exceptional Lie algebras F_4, E_6, E_7, and E_8. In a third paper I will construct complete sets of Casimir operators for all the real restrictions and real forms of the above Lie algebras.

In the following L is a finite-dimensional complex semisimple Lie algebra, H a Cartan subalgebra of L, $U(L)$ the universal enveloping algebra of L, and $Z(L)$ the center of $U(L)$. Casimir operators are representations of the elements of $Z(L)$, which we call Casimir elements. In a more precise language the above statements on Casimir operators are intended for Casimir elements. To formulate our criterion we need the following definitions and assertions. Let $S(L)$ be the symmetric algebra over L and $S_{\mathrm{Int}}(L)$ the algebra of the invariants of the adjoint group $\mathrm{Int}(L)$ of L in $S(L)$; that means $S_{\mathrm{Int}}(L) = \{s \in S(L) / \tilde{g}(s) = s, g \in \mathrm{Int}(L)\}$, where $\tilde{g}$ is the unique extension of $g \in \mathrm{Int}(L)$ to an algebra automophism of $S(L)$. Then there exists a vector space isomorphism $\lambda : S(L) \rightarrow U(L)$, $\lambda(1_{S(L)}) := 1_{U(L)}$,

$$\lambda(y_1 y_2 \ldots y_p) := \frac{1}{p!} \sum_{\sigma \in S_p} y_{\sigma(1)} \cdot y_{\sigma(2)} \cdot \cdots \cdot y_{\sigma(p)}, \tag{1}$$

where $p \in \mathbb{N}$, $y_j \in L$, $1 \leqslant j \leqslant p$, and S_p is the symmetric group, with $\lambda(S_{\mathrm{Int}}(L)) = Z(L)$ (see Ref. 24, pp. 344–346) (by a dot $\cdot$ we denote the product in a noncommutative associative algebra, whereas for the product in a commutative associative algebra we denote nothing). To prove our criterion we need

Lemma 1: Let $\{I_1, \ldots, I_l\}$, $l \in \mathbb{N}$, be a subset of $S_{\mathrm{Int}}(L)$. Then $\{I_1, \ldots, I_l\}$ generates $S_{\mathrm{Int}}(L)$ iff $\{\lambda(I_1), \ldots, \lambda(I_l)\}$ generates $Z(L)$.

Proof: The right hand side of Eq. (1) can be brought into the form $y_1 \cdot y_2 \cdot \ldots \cdot y_p + r$ with grad $(r) < p$, where grad means the filtration of r (see Ref. 25, p. 75). Now let $\{x_1, \ldots, x_n\}$ be a basis of L and g_i the graduation of I_i, $1 \leqslant i \leqslant l$. Then I_i can be written in the form $I_i = \Sigma_{\mu_1 + \cdots \mu_n \leqslant g_i} a^i_{\mu_1 \cdots \mu_n} x_1^{\mu_1} \cdots x_n^{\mu_n}$, $a^i_{\mu_1 \cdots \mu_n} \in \mathbb{C}$ (complex numbers), $\mu_k \in \{0\} \cup \mathbb{N}$, $1 \leqslant k \leqslant n$. With the above we derive

$$\begin{aligned} \lambda(I_i I_j) &= \sum a^i_{\mu_1 \cdots \mu_n} a^j_{\nu_1 \cdots \nu_n} \lambda(x_1^{\mu_1} \cdots x_n^{\mu_n} x_1^{\nu_1} \cdots x_n^{\nu_n}) \\ &= \lambda(I_i) \cdot \lambda(I_j) + \lambda(r) \end{aligned} \tag{2}$$

with $r \in S_{\mathrm{Int}}(L)$ and grad (r) $< g_i + g_j$, where we set $r = 0$ for grad (r) < 0. Now let $z \in Z(L)$. Then there exists $s \in S_{\mathrm{Int}}(L)$ with $\lambda(s) = z$ and s has the form

$$s = \sum_{n_1, \ldots, n_l \in \{0\} \cup \mathbb{N}} b_{n_1 \cdots n_l} I_1^{n_1} \cdots I_l^{n_l}, \quad b_{n_1 \cdots n_l} \in \mathbb{C}.$$

It follows from Eq. (2) that

$$\lambda(s) - \sum_{n_1, \ldots, n_l \in \{0\} \cup \mathbb{N}} b_{n_1 \cdots n_l} \lambda(I_1)^{n_1} \cdot \cdots \cdot \lambda(I_l)^{n_l} = \lambda(r),$$

with grad $(r) <$ grad(s) and $\lambda(r) \in Z(L)$. With this it follows by induction on grad(z), $z \in Z(L)$, that $\{\lambda(I_1), \ldots, \lambda(I_l)\}$ generates $Z(L)$, because for grad $(z) = 0$, plainly, z is generated by $\lambda(1_{S(L)}) = 1_{U(L)}$ and $\{I_1, \ldots, I_l\}$ generates $1_{S(L)}$. The opposite conclusion is proved analogously by using the following form of Eq. (2):

$$I_i I_j - r = \lambda^{-1}(\lambda(I_i) \cdot \lambda(I_j)), \quad \mathrm{grad}(r) < g_i + g_j. \qquad \square$$

Using Eq. (2) one can easily show,[22,26] that if $\{I_1, \ldots, (I_l)\}$ is algebraically independent and consists of homogeneous elements, then $\{\lambda(I_1), \ldots, \lambda(I_l)\}$ is algebraically independent. We remark that the above assertions are valid for every real or complex (not necessarily semisimple) Lie algebra.

Now let L^* and H^* be the algebraic duals of L and H, respectively, W^* the set of nonvanishing roots of H in L, $\mathrm{Int}(W^*, H)$ the Weyl group of H in L (acting on H^*), $S_{\mathrm{Int}}(H^*)$ the invariants of $\mathrm{Int}(W^*, H)$ in $S(H^*)$ [$S(\cdot)$ means the symmetric algebra], $P(L)$ the algebra of polynomial functions over L [it is $P(L) = S(L^*)$], $P_{\mathrm{Int}}(L)$ the algebra of the invariants of $\mathrm{Int}(L)$ in $P(L)$ [$P_{\mathrm{Int}}(L) := \{p \in P(L) / p \circ g = p, g \in \mathrm{Int}(L)\}$, "$\circ$" means composition of mappings] and R: $P(L) \rightarrow P(H) = S(H^*)$, $R(p) := p|_H$ (the restriction of p on H). Then we have $R(P_{\mathrm{Int}}(L)) = S_{\mathrm{Int}}(H^*)$ and the mapping $\tilde{R}$:

 0022-2488/81/091851-06$01.00

$P_{\text{Int}}(L) \to S_{\text{Int}}(H^*)$, $\widetilde{R}(p) := R(p)$ is an algebra isomorphism (see Ref. 27, pp. 504, 521–523). The algebras $S_{\text{Int}}(L)$ and $P_{\text{Int}}(L)$ are isomorphic and the isomorphism can be constructed with the aid of the Killing form $K(.,.)$ of L: The mapping $\widetilde{\delta}: L \to L^*, \widetilde{\delta}(x) := K(x,.)[K(x,.): L \to \mathbb{C}, K(x,.)(y) := K(x,y)]$ is linear and bijective and its extension to δ: $S(L) \to P(L)$ has, due to the invariance of $K(.,.)$ under the action of $\text{Int}(L)$, the property: $\delta(S_{\text{Int}}(L)) = P_{\text{Int}}(L)$. Thus $S_{\text{Int}}(H^*)$ *and* $S_{\text{Int}}(L)$ are isomorphic. It is known by a theorem of Chevalley[28] that there exists an algebraically independent subset $\{I_1,...,I_l\}$ of $S_{\text{Int}}(H^*)$, where $l := \dim(H)$ is the rank of L and I_i *is* homogeneous, such that $\{1_{S(H^*)}, I_1,...,I_l\}$ generates $S_{\text{Int}}(H^*)$. It follows by Lemma 1 that $Z(L)$ is isomorphic to the ring $\mathbb{C}[t_1,...,t_l]$ of indeterminates t_i and that every subset $\{z_1,...,z_l\}$ of $Z(L)$ which generates together with $1_{U(L)}$ the algebra $Z(L)$ is algebraically independent. For the sake of brevity we define the algebra $U^+(L)$ which is generated by L in $U(L)$ [$U^+(L)$ is an ideal in $U(L)$] and $Z^+(L) := Z(L) \cap U^+(L)$.

Now we are in a position to formulate and to prove the criterion.

Theorem: Let $L = H \underset{\alpha \in W^*}{\oplus} L_\alpha$ be the decomposition of L into root spaces, $B := \{h_i, b_\alpha / h_i \in H, 1 \leqslant i \leqslant l, b_\alpha \in L_\alpha, \alpha \in W^*\}$ a basis of L, $\mathbb{H}$ the (commutative) subalgebra in $U(L)$ generated by H, $\{z_1,...,z_l\}$ a subset of $Z^+(L)$ such that $\text{grad}(z_1) \leqslant \text{grad}(z_2) \leqslant \cdots \leqslant \text{grad}(z_l)$, $z_i = h_{z_i} + b_{z_i}, 1 \leqslant i \leqslant l$, with $h_{z_i} \in \mathbb{H}$, $\text{grad}(h_{z_i}) = \text{grad}(z_i)$ and b_{z_i} is a sum of such monomials u in the elements of B so that $\text{grad}(u) \leqslant \text{grad}(z_i)$ and for $\text{grad}(u) = \text{grad}(z_i)$ at least one factor in u is an element of $\{b_\alpha / \alpha \in W^*\}$, and $\{(R \circ \delta \circ \lambda^{-1})(h_{z_i}) / 1 \leqslant i \leqslant l\}$, a subset of homogeneous elements of $S(H^*)$ which together with $1_{S(H^*)}$ generates $S_{\text{Int}}(H^*)$. Then $\{z_1,...,z_l\}$ is an algebraically independent set which generates $Z^+(L)$.

Proof: We set $\{x_1,...,x_n\} := B$, where $n := \dim(L)$. As in the proof of Lemma 1 it holds that: $\lambda^{-1}(x_{i_1} \cdot x_{i_2} \cdot \cdots \cdot x_{i_p}) = x_{i_1} x_{i_2} \ldots x_{i_p} + s, p \in \mathbb{N}, \text{grad}(s) < p, x_{i_1},...,x_{i_p} \in B$. Thus $\lambda^{-1}(b_{z_i}) = p_i(x_1,...,x_n) + p_{i0}(h_1,...,h_l)$, where $p_i \in \mathbb{C}[t_1,...,t_n]$, $p_{i0} \in \mathbb{C}[t_1,...,t_l]$, in every monomial of p_i there is at least one factor which is an element of $\{b_\alpha / \alpha \in W^*\}$ and $\text{grad}(p_{i0}) < \text{grad}(z_i)$. Now, as is well known, $K(b_\alpha, h) = 0, \alpha \in W^*, h \in H$ (see Ref. 29, p. 108). Therefore, we have $(R \circ \delta)(p_i(x_1,...,x_n)) = R[p_i(K(x_1,.),...,K(x_n,.))] = 0$, $1 \leqslant i \leqslant l$. With this it follows that $(R \circ \delta \circ \lambda^{-1})(z_i) = (R \circ \delta \circ \lambda^{-1})(h_{z_i}) + (R \circ \delta \circ \lambda^{-1})(b_{z_i}) = (R \circ \delta \circ \lambda^{-1})(h_{z_i}) + (R \circ \delta)(p_{i0}(h_1,...,h_l))$ and $\text{grad}[(R \circ \delta)(p_{i0}(h_1,...,h_l))] < \text{grad}(z_i)$. As we stated above, we have $(R \circ \delta \circ \lambda^{-1})(z_i) \in S_{\text{Int}}(H^*)$ and by assumption $(R \circ \delta \circ \lambda^{-1})(h_{z_i}) \in S_{\text{Int}}(H^*)$. Therefore $(R \circ \delta)(p_{i0}(h_1,...,h_l)) \in S_{\text{Int}}(H^*)$ and $\text{grad}[(R \circ \delta)(p_{i0}(h_1,...,h_l))] < \text{grad}(z_i)$. Now it follows by the generating property of the set $\{(R \circ \delta \circ \lambda^{-1})(h_{z_i}) / 1 \leqslant i \leqslant l\}$ and the homogeneity of its elements that $(R \circ \delta)(p_{10}(h_1,...,h_l)) = 0$, because it is not possible to generate algebraically by a set of homogeneous polynomials a polynomial the graduation of which is less than the minimum of the graduations of the polynomials of the set. Therefore, the following equations must hold:

$$
\begin{aligned}
(R \circ \delta \circ \lambda^{-1})(z_1) &= (R \circ \delta \circ \lambda^{-1})(h_{z_1}),\\
(R \circ \delta \circ \lambda^{-1})(z_2) &= (R \circ \delta \circ \lambda^{-1})(h_{z_2}) + (R \circ \delta)(p_{20}(h_1,...,h_l))\\
&= (R \circ \delta \circ \lambda^{-1})(h_{z_2}) + q_2((R \circ \delta \circ \lambda^{-1})(h_{z_1})),\\
(R \circ \delta \circ \lambda^{-1})(z_3) &= (R \circ \delta \circ \lambda^{-1})(h_{z_3}) + (R \circ \delta)(p_{30}(h_1,...,h_l))\\
&= (R \circ \delta \circ \lambda^{-1})(h_{z_3}) + q_3((R \circ \delta \circ \lambda^{-1})(h_{z_1}),\\
\vdots \qquad & \qquad (R \circ \delta \circ \lambda^{-1})(h_{z_2})),\\
(R \circ \delta \circ \lambda^{-1})(z_l) &= (R \circ \delta \circ \lambda^{-1})(h_{z_l}) + (R \circ \delta)(p_{l0}(h_1,...,h_l))\\
&= (R \circ \delta \circ \lambda^{-1})(h_{z_l})\\
&\quad + q_l((R \circ \delta \circ \lambda^{-1})(h_{z_1}),...,(R \circ \delta \circ \lambda^{-1})(h_{z_{l-1}})),
\end{aligned}
$$

where $q_j \in \mathbb{C}[t_1,...,t_{j-1}]$ and $\text{grad}[q_j((R \circ \delta \circ \lambda^{-1})(h_{z_1}),...,(R \circ \delta \circ \lambda^{-1})(h_{z_{j-1}}))] < \text{grad}(z_j)$, $2 \leqslant j \leqslant l$. These yield by iterative insertion that to every q_j there exist a $\hat{q}_j \in \mathbb{C}[t_1,...t_{j-1}]$, $2 \leqslant j \leqslant l$, with $q_j((R \circ \delta \circ \lambda^{-1})(h_{z_1}),...,(R \circ \delta \circ \lambda^{-1})(h_{z_{j-1}})) = \hat{q}_j((R \circ \delta \circ \lambda^{-1})(z_1),...,(R \circ \delta \circ \lambda^{-1})(z_{j-1}))$. Therefore, $\{(R \circ \delta \circ \lambda^{-1})(z_i) / 1 \leqslant i \leqslant l\}$ is, together with $1_{S(H^*)}$, a generating set of $S_{\text{Int}}(H^*)$. From this, Lemma 1, and what we said above about the isomorphism between $S_{\text{Int}}(H^*)$ and $S_{\text{Int}}(L)$ the theorem follows immediately. □

Remark 1: Let $\{x_1,...,x_n\}$ be any basis of L and $\{x_1^*,...,x_n^*\}$ its dual. Then $\delta(x_i) = \sum_{j=1}^{n} K(x_i,x_j) x_j^*$. Therefore, one can easily calculate, because of the commutativity of $\mathbb{H}$, the element $(\delta \circ \lambda^{-1})(x)$, $x \in \mathbb{H}$. This shows, as we shall see, how useful our criterion is in practice.

Remark 2: Let $(z_i) = \sum_{\nu_1,...,\nu_n \in \{0\} \cup \mathbb{N}} a^i_{\nu_1 \ldots \nu_n} x_1^{\nu_1} \cdot \ldots \cdot x_n^{\nu_n}, 1 \leqslant i \leqslant l$, $a^i_{\nu_1 \ldots \nu_n} \in \mathbb{C}$, from the theorem be such that s_i: $= \sum_{\nu_1,...,\nu_n \in \{0\} \cup \mathbb{N}} a^i_{\nu_1 \ldots \nu_n} x_1^{\nu_1} \ldots x_n^{\nu_n}$ is an element of $S_{\text{Int}}(L)$. Then it can be shown, analogously to our proof, that $\{s_1,...,s_l\}$ is algebraically independent and that $\{1_{S(L)}, s_i / 1 \leqslant i \leqslant l\}$ generates $S_{\text{Int}}(L)$. As this assumption is fulfilled by the Casimir elements which we will construct, we will have also constructed complete sets for the invariants of the adjoint groups of all simple Lie algebras.

Now we construct out of each class of simple Lie algebras A_l, B_l, C_l, D_l, and G_2 for one model a complete set of Casimir elements. Each time, we first present the model (see Ref. 29 for A_l, B_l, C_l, D_l, Ref. 30 for G_2).

A_l: Let M_{l+1} be the set of $(l+1) \times (l+1)$ matrices, $l \in \mathbb{N}$, $sl(l+1) := \{M \in M_{l+1} / \text{tr}(M) = 0\}$, $:= \delta_{ik} \delta_{jm}$, $1 \leqslant i,j,k,m \leqslant l+1$, $h_k := e_{kk} - e_{l+1,l+1}$, $1 \leqslant k \leqslant l+1$, and $B_{sl(l+1)} := \{h_k, e_{ij} / 1 \leqslant k \leqslant l, 1 \leqslant i,j \leqslant l+1, i \neq j\}$. Then $sl(l+1) \in A_l$, $B_{sl(l+1)}$ is a basis of $sl(l+1)$, the linear span H of $\{h_k / 1 \leqslant k \leqslant l\}$ is a Cartan subalgebra of $sl(l+1)$, $\{w_i - w_j, w_{l+1} - w_i, w_i - w_{l+1} / 1 \leqslant i \leqslant, j \leqslant l\}$ is the set of roots of H in $sl(l+1)$, where $\{w_i / 1 \leqslant i \leqslant l\}$ is the dual of $\{h_k / 1 \leqslant k \leqslant l\}$ and $w_{l+1} := -\sum_{i=1}^{l} w_i$, the root space of $w_i - w_j, i \neq j, w_{l+1} - w_i$ and $w_i - w_{l+1}$ is $\{ce_{ij} / c \in \mathbb{C}\}$, $\{ce_{l+1,i} / c \in \mathbb{C}\}$, and $\{ce_{i,l+1} / c \in \mathbb{C}\}$, respectively, and the Weyl group of H in $sl(l+1)$ is the symmetric group of $\{w_i / 1 \leqslant i \leqslant l+1\}$. We define now a new basis of $sl(l+1)$ which allows us to construct easily the Casimir elements. Let $x_{ij} := e_{ij}, i \neq j, 1 \leqslant i,j \leqslant l+1$, and $x_{kk} := h_k - \frac{1}{l+1} \sum_{j=1}^{l} h_j$, $1 \leqslant k \leqslant l+1$. Then $\{x_{ij}, x_{kk} / 1 \leqslant i,j \leqslant l+1, i \neq j, 1 \leqslant k \leqslant l\}$ is a basis of $sl(l+1)$ and it holds that $[x_{ij}, x_{km}] = \delta_{jk} x_{im} - \delta_{im} x_{kj}$, $1 \leqslant i, j, k, m \leqslant l+1$. Now let $(z_1)_{ij} := x_{ij}$, $(z_{n+1})_{ij} := \sum_{k=1}^{l+1} x_{ik} \cdot (z_n)_{kj}$, $n \in \mathbb{N}$, and $z_n := \sum_{i=1}^{l+1} (z_n)_{ii}$. Then it follows by using the equation $\widetilde{\mathrm{ad}}\,(x_{ij})((z_n)_{km}) = \delta_{jk}(z_n)_{im} - \delta_{im}(z_n)_{kj}$ ($\widetilde{\mathrm{ad}}$ is the extension of the adjoint representation to the universal enveloping algebra), that $z_n \in Z^+(sl(l+1))$. Now we establish

Lemma 2: $\{z_n / 2 \leqslant n \leqslant l+1\}$ is a complete set of $Z^+(sl(l+1))$.

Proof: Let $h_{z_n} := \sum_{i=1}^{l+1} (x_{ii})^n$, $n \in \mathbb{N}$. Then one obtains using $K(h_i, h_j) = 2(l+1)(1+\delta_{ij})$, $\delta(h_i) = 2(l+1)(w_i - w_{l+1})$ and $R(w_i) = w_i$: $(R \circ \delta \circ \lambda^{-1})(h_{z_n}) = (2(l+1))^n \sum_{i=1}^{l+1} w_i^n$. These power sums generate the elementary symmetric polynomials, which build an integrity basis for $S_{\mathrm{Int}}(H^*)$, and vice versa. Thus Lemma 2 follows because h_{z_n} and $b_{z_n} := z_n - h_{z_n}$, $2 \leqslant n \leqslant l+1$, fulfill all assumptions of the theorem. $\square$

B_l: Let

$$S_l := \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 1_l \\ 0 & 1_l & 0 \end{bmatrix},$$

$l \in \mathbb{N}$, $\hat{B}_l := \{M \in M_{2l+1} / M \cdot S_l = -S_l \cdot M^t\}$ (M^t is the transpose of M), $h_i := e_{i+1,i+1} - e_{i+l+1,i+l+1}$, $e_{w_i} := e_{i+1,1} - e_{1,i+l+1}$, $e_{-w_i} := e^t_{w_i}$, $e_{w_i - w_j} := e_{i+1,j+1} - e_{j+l+1,i+l+1}$, $e_{-w_i - w_j} := e_{i+l+1,j+1} - e_{j+l+1,i+1}$, $e_{w_i + w_j} := e_{j+1,i+l+1} - e_{i+1,j+l+1}$, $1 \leqslant i, j \leqslant l$, and $B := \{h_i, e_{w_i}, e_{-w_i}, e_{w_j - w_j}, e_{-w_k - w_m}, e_{w_k + w_m} / i \neq j, k < m, 1 \leqslant i, j, k, m \leqslant l\}$. Then $\hat{B}_l \in B_l$ $(l \geqslant 2)$, B is a basis of $\hat{B}_l$, the linear span H of $\{h_i / 1 \leqslant i \leqslant l\}$ is a Cartan subalgebra of $\hat{B}_l$, the linear functional w_i: $H \to \mathbb{C}$, $1 \leqslant i \leqslant l$, which maps every matrix $(h_{ij})_{1 \leqslant i, j \leqslant 2l+1}$ of H on the entry $h_{i+1,i+1}$, is a root of H in $\hat{B}_l$ ($\{w_i / 1 \leqslant i \leqslant l\}$ is the dual of $\{h_i / 1 \leqslant i \leqslant l\}$), $\{\pm w_i, w_i - w_j, \pm (w_k + w_m) / i \neq j, k < m, 1 \leqslant i, j, k, m \leqslant l\}$ is the set W^* of H in $\hat{B}_l$ and the element of B indexed by an element of W^* is a vector from the corresponding root space. Now Let $x_{i,-i} := h_i$, $x_{i,-j} := e_{w_j - w_i}$, $x_{i0} := e_{-w_i}$, $x_{0,-i} := e_{-w_i}$, $x_{km} := e_{-w_m - w_k}$, $x_{-k,-m} := e_{w_k + w_m}$, $i \neq j$, $k < m$, $1 \leqslant i, j, k, m \leqslant l$, and $x_{ij} = -x_{ji}$, $-l \leqslant i, j \leqslant l$. Then $\{x_{ij} / i > j, -l \leqslant i, j \leqslant l\}$ is a basis of $\hat{B}_l$ and it holds that $[x_{ij}, x_{km}] = \delta_{j,-k} x_{im} - \delta_{-i,m} x_{kj} + \delta_{j,-m} x_{ki} - \delta_{i,k} x_{jm}$. By these commutation relations it is easily proved that $z_n := \Sigma_{-l \leqslant i_1, i_2, \ldots, i_n \leqslant l}\, x_{-i_1, i_2} \cdot x_{-i_2, i_3} \cdot \ldots \cdot x_{-i_n, i_1}$, $n \in \mathbb{N}$, is an element of $Z^+(\hat{B}_l)$.

Lemma 3: $\{z_2, z_4, \ldots, z_{2l}\}$ is a complete set of $Z^+(\hat{B}_l)$.

Proof: Let $h_{z_n} := 2\Sigma_{i=1}^{l}(x_{i,-i})^n$. Then $(R \circ \delta \circ \lambda^{-1})(h_{z_n})$ is proportional to $\Sigma_{i=1}^{l}(w_i)^n$. As $S_{\mathrm{Int}}(H^*)$ is generated by $\{\Sigma_{i=1}^{l}(w_i)^n / n = 2, 4, \ldots, 2l\}$ together with $1_{S(H^*)}$, h_{z_n} and $b_{z_n} := z_n - h_{z_n}$ fulfill the assumptions of the theorem. $\square$

C_l: Let

$$J_l := \begin{bmatrix} 0 & 1_l \\ -1_l & 0 \end{bmatrix},$$

$l \in \mathbb{N}$, $\mathrm{sp}(2l, \mathbb{C}) := \{M \in M_{2l} / M^t \cdot J_l = -J_l \cdot M\}$, $e_{ij} \in M_{2l}$, $1 \leqslant i, j \leqslant 2l$, $h_i := e_{ii} - e_{i+l,i+l}$, $e_{w_j - w_i} := e_{ij} - e_{j+l,i+l}$, $e_{w_i + w_j} := e_{i+l,j} + e_{j+l,i}$, $e_{-w_i - w_j} := (e_{w_i} + w_j)^t$, $e_{2w_i} := e_{i+l,i}$, $e_{-2w_i} := (e_{2w_i})^t$, $1 \leqslant i, j \leqslant l$, and $B_{\mathrm{sp}(2l,\mathbb{C})} := \{h_i, e_{\pm 2w_i}, e_{w_j - w_i}, e_{\pm(w_k + w_m)} / i \neq j, k < m, 1 \leqslant i, j, k, m \leqslant l\}$. Then $\mathrm{sp}(2l, \mathbb{C}) \in C_l$ $(l \geqslant 3)$, $B_{\mathrm{sp}(2l,\mathbb{C})}$ is a basis of $\mathrm{sp}(2l, \mathbb{C})$, the linear span H of $\{h_i / 1 \leqslant i \leqslant l\}$ is a Cartan subalgebra of $\mathrm{sp}(2l, \mathbb{C})$, the linear function w_i: $H \to \mathbb{C}$, $1 \leqslant i \leqslant l$, which maps every matrix of H on its ith diagonal entry, is an element of the dual of $\{h_i / 1 \leqslant i \leqslant l\}$, $\{w_i - w_{i+1}, 2w_l / 1 \leqslant i \leqslant l-1\}$ is a simple system of roots of $\mathrm{sp}(2l, \mathbb{C})$ relative to H, $\{\pm 2w_i, w_j - w_i, \pm(w_k + w_m) / i \neq j, k < m, 1 \leqslant i, j, k, m \leqslant l\}$ the set W^* of H in $\mathrm{sp}(2l, \mathbb{C})$, and the element of $B_{\mathrm{sp}(2l,\mathbb{C})}$ indexed by an element of W^* is a vector from the corresponding root space. It is easily proved that the first l power sums in $w_1^2, \ldots, w_l^2$ generate $S_{\mathrm{Int}}(H^*)$. Now let $x_{i,-i} := h_i$, $x_{i,-j} := e_{w_j - w_i}$, $x_{ij} := e_{-w_i - w_j}$, $x_{-i,-j} := e_{w_i + w_j}$, $x_{ii} := -2e_{-2w_i}$, $x_{-i,-i} := 2e_{2w_i}$, $1 \leqslant i, j \geqslant l$, and $x_{ij} \equiv x_{ji}$, $i, j \in \{\pm 1, \pm 2, \ldots, \pm 1\}$. Then $\{x_{ij} / i \geqslant j, i, j \neq 0, -l \leqslant i, j \leqslant l\}$ is a basis of $\mathrm{sp}(2l, \mathbb{C})$ and it holds that $[x_{ij}, x_{km}] = \epsilon_k \delta_{-j,k} x_{im} + \epsilon_m \delta_{-j,m} x_{ik} + \epsilon_k \delta_{-i,k} x_{jm} + \epsilon_m \delta_{-i,m} x_{jk}$, where $\epsilon_i = 1$ for $1 \leqslant i \leqslant l$ and $\epsilon_i = -1$ for $-l \leqslant i \leqslant -1$. Now let $(z)_{1ij} := \epsilon_i x_{ij}$, $(z_{n+1})_{ij} := \epsilon_i \Sigma_{-l \leqslant i_k \leqslant l}\, x_{i,-i_k} \cdot (z_n)_{i_k j}$, $i, j \in \{\pm 1, \pm 2, \ldots, \pm l\}$, $n \in \mathbb{N}$, and $(z_n) := \Sigma_{-l \leqslant i \leqslant l}(z_n)_{i,-i}$. Then it follows with the aid of $\widetilde{\mathrm{ad}}\,(x_{ij})((z_n)_{km}) = \epsilon_i \delta_{-j,k}(z_n)_{im} + \epsilon_m \delta_{-j,m}(z_n)_{ki} + \epsilon_j \delta_{-i,k}(z_n)_{jm} + \epsilon_m \delta_{-i,m}(z_n)_{kj}$, $n \in \mathbb{N}$, that $z_n \in Z^+(\mathrm{sp}(2l, \mathbb{C}))$.

§ Lemma 4: $\{z_2, z_4, \ldots, z_{2l}\}$ is a complete set of $Z^+(\mathrm{sp}(2l, \mathbb{C}))$.

Proof: $z_n = \Sigma_{-l \leqslant i_1, \ldots, i_n \leqslant l}\, \epsilon_{i_1} \cdots \epsilon_{i_n} x_{i_1, -i_2} \cdot x_{i_2, -i_3} \cdots x_{i_n, -i_1}$. Now let $h_{z_n} := \Sigma_{-l \leqslant i \leqslant l}(\epsilon_i)^n (x_{i,-i})^n$. Then $h_{z_n} = 2\Sigma_{i=1}^{l}(x_{i,-i})^n$ for even n. With $K(x_{i,-i}, x_{j,-j}) = K(h_i, h_j) = 4(l+1)\delta_{ij}$, $1 \leqslant i, j \leqslant l$, it follows that $(R \circ \delta \circ \lambda^{-1})(h_{z_n})$ is proportional to $\Sigma_{i=1}^{l} w_i^n$, n even. Therefore, $\{z_n / \mathrm{n} = 2, 4, \ldots, 2\mathrm{l}\}$ fulfills all the assumptions of the theorem. $\square$

D_l: Let

$$S_l := \begin{bmatrix} 0 & 1_l \\ 1_l & 0 \end{bmatrix},$$

$l \in \mathbb{N}$, $l \neq 1$ (we exclude $l = 1$ because otherwise we would get a nonsemisimple Lie algebra), $\hat{D}_l := \{M \in M_{2l} / M \cdot S_l = -S_l \cdot M^t\}$, $e_{ij} \in M_{2l}$, $1 \leqslant i, j \leqslant 2l$, $h_i := e_{ii} - e_{i+l,i+l}$, $e_{w_i - w_j} := e_{ji} - e_{i+l,j+l}$, $e_{w_i + w_j} := e_{i+l,j} - e_{j+l,i}$, $e_{-w_i - w_j} := (e_{w_i + w_j})^t$, $1 \leqslant i, j \leqslant l$, $B := \{h_i, e_{w_j - w_j}, e_{w_k + w_m}, e_{-w_k - w_m} / i \neq j, k < m, 1 \leqslant i, j, k, m \leqslant l\}$. Then $\hat{D}_l \in D_l$ $(l \geqslant 4)$, B is a basis of $\hat{D}_l$, the linear span H *of* $\{h_i / 1 \leqslant i \leqslant l\}$ a Cartan subalgebra of $\hat{D}_l$, the linear functional w_i: $H \to \mathbb{C}$, $1 \leqslant i \leqslant l$, which maps every matrix of H on its ith diagonal entry, an element of the dual of $\{h_i / 1 \leqslant i \leqslant l\}$, $\{w_i - w_{i+l}, w_{l-1} + w_l / 1 \leqslant i \leqslant l-1\}$ is a simple system of roots of $\hat{D}_l$ relative to H, $\{w_i - w_j, \pm(w_k + w_m) / i \neq j, k < m, 1 \leqslant i, j, k, m, \leqslant l\}$ the set W^* of H in $\hat{D}_l$, and the element of B indexed by an element of W^* is a vector from the corresponding root space. Let s_i, $1 \leqslant i \leqslant l-1$, be the power sum in l indeterminates. Then one can prove[31] that $\{s_i(w_1^2, w_2^2, \ldots, w_l^2), w_1 w_2 \cdots w_l / 1 \leqslant i \leqslant l-1\}$ is an algebraically independent set which together with $1_{S(H^*)}$ generates $S_{\mathrm{Int}}(H^*)$. Now let $x_{i,-i} := h_i$, $x_{i,-j} := e_{w_j - w_i}$, $x_{mk} := e_{-w_k - w_m}$, $x_{-k,-m} := e_{w_k + w_m}$, $i \neq j$, $k < m$, $1 \leqslant i, j, k, m \leqslant l$, and $x_{ij} \equiv -x_{ji}$, $i, j \in \{\pm 1, \pm 2, \ldots, \pm l\}$. Then $\{x_{ij}$

$/i>j,i,j\neq 0,\ -l\leqslant i,j\leqslant l\}$ is a basis of $\hat{D}_l$ and it holds that $[x_{ij},x_{km}] = \delta_{j,-k}x_{im} - \delta_{-i,m}x_{kj} + \delta_{j,-m}x_{ki} - \delta_{i,-k}x_{jm}$. From these commutation relations follows that $z_n := \Sigma_{i_1,\dots,i_n\in\{\pm 1,\dots,\pm l\}}\, x_{-i_1 i_2}\cdot x_{-i_2 i_3}\cdot \dots \cdot x_{-i_n i_1}$, $n\in\mathbb{N}$, and $\hat{z}_l := \Sigma_{i_1 j_1,\dots,i_l j_l\in\{\pm 1,\dots,\pm l\}}\, \epsilon_{i_1 j_1\cdots i_l j_l}\, x_{-i_1,-j_1}\cdot \dots \cdot x_{-i_l,-j_l}$ are elements of $Z^+(\hat{D}_l)$, where ϵ is the totally antisymmetric unit tensor.

Lemma 5. $\{z_2, z_4,\dots,z_{2l-2}, \hat{z}_l\}$ is a complete set of $Z^+(\hat{D}_l)$.

Proof: With the aid of $K(h_j,h_k) = 4(l-1)\,\delta_{jk}$ the proof is analogous to those of Lemmas 2–4. □

Before proceeding with the construction of a complete set for G_2 we establish two lemmas which we need for the case of all exceptional Lie algebras. The first lemma is a generalization of a method of Racah[17] and is found in Refs. 16 and 32; in Ref. 16 it is formulated only for compact semisimple Lie algebras and in both cases the proof is group-theoretical, where our proof is purely algebraic. In the following $\mathbb{K}$ is either $\mathbb{R}$ (real numbers) or $\mathbb{C}$.

Lemma 6. Let L be over $\mathbb{K}$ $\{x_1,\dots,x_n\}$, $n\in\mathbb{N}$, a basis of L, $\tilde{K} := (K_{ij})_{1\leqslant i,j\leqslant n}$, where $K_{ij} := K(x_i,x_j)$, $K^{ij} := (\tilde{K}^{-1})_{ij}$, $x^j := \Sigma_{i=1}^n K^{ji}x_i$, D a finite-dimensional representation of L, and $I_p(D) := \Sigma_{1\leqslant i_1,\dots,i_p\leqslant n}\operatorname{tr}(D(x_{i_1})\cdot D(x_{i_2})\cdot \dots \cdot D(x_{i_p}))\, x^{i_1}\cdot x^{i_2}\cdot \dots \cdot x^{i_p}$, $p\in\mathbb{N}$. Then $I_p(D)\in Z^+[L]$.

Proof: Let $c^k{}_{ij}$, $1\leqslant i,j,k\leqslant n$, be the structure constants of L relative to $\{x_1,\dots,x_n\}$. Then

$$\begin{aligned}
&\tilde{\operatorname{ad}}(x_k)(I_p(D))\\
&\quad = \sum_{\substack{1\leqslant i_1,\dots,i_p\leqslant n\\ 1\leqslant m\leqslant p\\ 1\leqslant j_m\leqslant n}} (\operatorname{tr}(D(x_{i_1})\cdot \dots \cdot D(x_{i_p}))K^{i_m j_m}x^{i_1}\cdot \dots \cdot x^{i_{m-1}}\\
&\qquad\cdot [x_k,x_{j_m}]\cdot x^{i_{m+1}}\cdot \dots \cdot x^{i_p})\\
&\quad = \sum_{\substack{1\leqslant i_1,\dots,i_p,j_m,r_m\leqslant n\\ 1\leqslant m\leqslant p}} (\operatorname{tr}(D(x_{i_1})\cdot \dots \cdot D(x_{i_p}))K^{i_m j_m}c^{r_m}{}_{kj_m}\\
&\qquad\times x^{i_1}\cdot \dots \cdot x^{i_{m-1}}\cdot x_{r_m}\cdot x^{i_{m+1}}\cdot \dots \cdot x^{i_p}).
\end{aligned}$$

Now, c_{ijk} being totally antisymmetric, it holds that $\Sigma_{r_m=1}^n K^{i_m r_m}c^{j_m}{}_{kr_m} = c^{j_m}{}_k{}^{i_m} = \Sigma_{k=1}^n K^{rj_m}c^{i_m}{}_{rk}$. Thus

$$\begin{aligned}
&\tilde{\operatorname{ad}}(x_k)(I_p(D))\\
&\quad = \sum_{\substack{1\leqslant i_1,\dots,i_p,r\leqslant n\\ 1\leqslant m\leqslant p}} (\operatorname{tr}(D(x_{i_1})\cdot \dots \cdot D(x_{i_{m-1}})\cdot c^{i_m}{}_{rk}D(x_{i_m})\cdot \dots\\
&\qquad\cdot D(x_{i_p}))x^{i_1}\cdot \dots \cdot x^{i_{m-1}}\cdot x^r\cdot x^{i_{m+1}}\cdot \dots \cdot x^{i_p})\\
&= \sum_{1\leqslant i_1,\dots,i_p\leqslant n}\ \sum_{1\leqslant m\leqslant p} (\operatorname{tr}(D(x_{i_1})\cdot \dots \cdot D(x_{i_{m-1}})\cdot D([x_{i_m},x_k])\\
&\cdot D(x_{i_{m+1}})\cdot \dots \cdot D(x_{i_p}))x^{i_1}\cdot \dots \cdot x^{i_{m-1}}\cdot x^{i_m}\cdot x^{i_{m+1}}\cdot \dots \cdot x^{i_p})\\
&\quad = 0,
\end{aligned}$$

because

$$\begin{aligned}
&\sum_{1\leqslant m\leqslant p}\operatorname{tr}(D(x_{i_1})\cdot \dots \cdot D(x_{i_{m-1}})\cdot D([x_{i_m},x_k]\cdot D(x_{i_{m+1}})\cdot \dots D(x_{i_p}))]\\
&\quad = \operatorname{tr}D(x_{i_1})\cdot D(x_k)\cdot D(x_{i_2})\cdot \dots \cdot D(x_{i_p}))\\
&\qquad - \operatorname{tr}(D(x_k)\cdot D(x_{i_1})\cdot \dots \cdot D(x_{i_p}))\\
&\qquad + \operatorname{tr}(D(x_{i_1})\cdot D(x_{i_2})\cdot D(x_k)\cdot D(x_{i_3})\cdot \dots \cdot D(x_{i_p}))\\
&\qquad - \operatorname{tr}(D(x_{i_1})\cdot D(x_k)\cdot D(x_{i_2})\cdot \dots \cdot D(x_{i_p}))\\
&\qquad + \dots + \operatorname{tr}(D(x_{i_1})\cdot \dots \cdot D(x_{i_p})\cdot D(x_k))\\
&\qquad - \operatorname{tr}(D(x_{i_1})\cdot \dots \cdot D(x_{i_{p-1}})\cdot D(x_k)\cdot D(x_{i_p}))\\
&\quad = - \operatorname{tr}(D(x_k)\cdot D(x_{i_1})\cdot \dots \cdot D(x_{i_p}))\\
&\qquad + \operatorname{tr}(D(x_{i_1})\cdot \dots \cdot D(x_{i_p})\cdot D(x_k))\\
&\quad = 0.
\end{aligned}$$
□

Two equivalent representations D and $\tilde{D}$ supply the same Casimir elements: $I_p(D) = I_p(\tilde{D})$. Further, $I_p(D)$ is independent of a basis of L. Let $\{y_1,\dots,y_n\}$ be another basis of L and $x_i = \Sigma_{j=1}^n A_{ij}y_j$. Then $K^x_{ij} = K(x_i,x_j) = \Sigma_{k,m=1}^n A_{ik}A_{jm}K(y_k,y_m) \equiv \Sigma_{k,m=1}^n A_{ik}K^y_{km}A^t_{mj}$; that means $(A^t\cdot(K^x)^{-1}\cdot A)_{ij} = (K^y)^{ij}$. Thus $I_p(D) = \Sigma_{1\leqslant i_1,\dots,i_p\leqslant n}(\operatorname{tr}(D(y_{i_1})\cdot \dots \cdot D(y_{i_p}))y^{i_1}\cdot \dots \cdot y^{i_p})$. Let $\hat{D}_{jk} := \Sigma_{i=1}^n D(x_i)_{jk}x^i, 1\leqslant j,k\leqslant d$, where d is the dimension of the representation D, and $\hat{D} := (\hat{D}_{jk})_{1\leqslant j,k\leqslant d}$. Then $I_p(D) = \operatorname{tr}(\hat{D}^p) = \Sigma_{1\leqslant j_1,\dots,j_p\leqslant d}\hat{D}_{j_1 j_2}\cdot\hat{D}_{j_2 j_3}\cdot \dots \cdot \hat{D}_{j_p j_1}$. In this abbreviated notation we will state the Casimir elements of all exceptional Lie algebras. We denote that $I_p(D)$ interpreted as an element of $S(L)$ is an element of $S_{\text{Int}}(L)$ (cf. Remark 2 above).

Now we show that the Casimir elements constructed by Lemma 6 fulfill part of the assumptions of the above criterion.

Lemma 7: Let $\{h_i, b_\alpha/1\leqslant i\leqslant l, \alpha\in W^*\}$ be a basis of L as in the theorem, D a finite-dimensional representation of L, $z_p := I_p(D)$ for some $p\in\mathbb{N}$, $h_{z_p} := \operatorname{tr}((\Sigma_{i=1}^l D(h_i)h^i)^p)$ and $b_{z_p} := z_p - h_{z_p}$. Then $z_p = h_{z_p} + b_{z_p}$ fulfills the decomposition assumptions of the theorem and it holds that $(R\circ\delta\circ\lambda^{-1})(h_{z_p})\in S_{\text{Int}}(H^*)$.

Proof: In the above basis K_{ij} has the following form:

$$(K_{ij})_{1\leqslant i,j\leqslant n} = \begin{bmatrix} S & 0 & 0\\ 0 & 0 & D\\ 0 & D & 0\end{bmatrix},$$

where $S\in M_l$ is symmteric and $D\in M_{(n-l)/2}$ is diagonal. Therefore it follows by inspection that $z_p = h_{z_p} + b_{z_p}$ fulfills the decomposition assumptions. Now, as equivalent representations yield the same coefficients for the monomials of z_p, we assume that $D(h_i)$ is diagonal. Thus $D(h_i) = \operatorname{diag}\ (\underbrace{w_1(h_i),\dots,w_1(h_i)}_{d_1\text{ times}}, \underbrace{w_2(h_i),\dots,w_2(h_i)}_{d_2\text{ times}},\dots, \underbrace{w_m(h_i),\dots,w_m(h_i)}_{d_m\text{times}})$, where w_k, $1\leqslant k\leqslant m$, are the weights of H relative to D and d_k is the multiplicity of w_k $(d = \Sigma_{k=1}^m d_k)$. Now let $\alpha_i := \delta(h_i)$. Then there exist numbers a_{ki} such that $w_k = \Sigma_{q=1}^l a_{kq}\alpha_q$. Thus $(D(h_i))_{jj} = w_{k(j)}(h_i) = \Sigma_{q=1}^l a_{k(j)q}\alpha_q(h_i) = \Sigma_{q=1}^l a_{k(j)q}K(h_q,h_i)$ for some k depending on $j, 1\leqslant j\leqslant d$. From this follows, because of the form of K_{ij}, that $(\Sigma_{i=1}^l D(h_i)h^i)_{jj} = \Sigma_{i,r=1}^l (D(h_i))_{jj}K^{ir}h_r = \Sigma_{i,r,q=1}^l a_{k(j)q}K_{qi}K^{ir}h_r = \Sigma_{q=1}^l a_{k(j)q}h_q$. By $(R\circ\delta\circ\lambda^{-1})(h_q) = \alpha_q$ we have $(R\circ\delta\circ\lambda^{-1})(h_{z_p}) = \Sigma_{k=1}^m d_k w_k^p$. Now it is known that $g\in\operatorname{Int}(W^*,H)$ permutes the weights and that the multiplicity of the weights is invariant under g (see Ref. 29, p. 113). □

The degrees of the elements of a minimal set of homogeneous elements that together with $1_{S(H^*)}$ generates $S_{\text{Int}}(H^*)$ are known for all complex simple Lie algebras (see Ref. 27, pp. 508–509, 515–516, 518, and 520). By the abovecited theorem of Chevalley the cardinality of such a set is equal to the rank of the Lie algebra. By the knowledge of the degrees of the elements of such a set we construct minimal generating sets for the invariants of the Weyl groups of all exceptional Lie algebras. For G_2 the degrees are 2 and 6. Now we present a model for

G_2:

Let, for $a = (a_1, a_2, a_3) \in \mathbb{C}^3$, the matrix $M(a) \in M_3$ be defined as $(M(a))_{ij} := \Sigma_{k=1}^{3} \epsilon_{kji} a_k, 1 \leqslant i, j \leqslant 3$, where ϵ is the totally anti-symmetric unit tensor with $\epsilon_{123} = 1$. Further let $\widehat{G}_2 := \{ M \in M_7 /$

$$M = \begin{bmatrix} 0 & \sqrt{2}a & \sqrt{2}b \\ -\sqrt{2}b^t & S & M(a) \\ -\sqrt{2}a^t & M(b) & -S^t \end{bmatrix}, a, b \in \mathbb{C}^3, S \in sl(3) \},$$

$e_{ij} \in M_7$ as before, $h_1 := e_{22} - e_{55} - e_{33} + e_{66}$, $h_2 := e_{22} - e_{55} - e_{44} + e_{77}$, $g_{\alpha_{i,-j}} := e_{i+1,j+1} - e_{j+4,i+4}$, $1 \leqslant i, j \leqslant 3$, $g_{\alpha_i} := \sqrt{2}(e_{1,i+1} - e_{i+4,1}) + \epsilon_{ijk}(e_{k+1,j+4} - e_{j+1,k+4})$, $g_{\alpha_{-i}} := \sqrt{2}(e_{1,i+4} - e_{i+1,1}) + \epsilon_{ijk}(e_{k+4,j+1} - e_{j+4,k+1})$, $(ijk) \in \{(123), (213), (312)\}$, and $B := \{ h_1, h_2, g_{\alpha_{i,-j}}, g_{\alpha_i}, g_{\alpha_{-i}} / i \neq j, 1 \leqslant i, j \leqslant 3 \}$. Then $\widehat{G}_2 \in G_2$, B is a basis of $\widehat{G}_2$, the linear span of H of $\{h_1, h_2\}$ is a Cartan subalgebra of $\widehat{G}_2$, the linear functionals α_1 and $\alpha_{1,-3}$ on H defined by $\alpha_1(h_1) := -1, \alpha_1(h_2) := -1$, $\alpha_{1,-3}(h_1) := 1$, and $\alpha_{1,-3}(h_2) := 2$ are a simple system of roots of $\widehat{G}_2$ relative to H, $\{ \alpha_1, \alpha_{1,-3}, \alpha_{-1} := -\alpha_1, \alpha_{3,-1} := -\alpha_{1,-3}, \alpha_{\mp 2} := \pm(2\alpha_1 + \alpha_{1,-3}), \alpha_{\pm 3} := \pm(\alpha_1 + \alpha_{1,-3}), \alpha_{2,-1} := -\alpha_{1,-2} := 3\alpha_1 + \alpha_{1,-3}, \alpha_{2,-3} := -\alpha_{3,-2} := 3\alpha_1 + 2\alpha_{1,-3} \}$ is the set W^* of H in $\widehat{G}_2$, and the element of B indexed by an element of W^* is a vector from the corresponding root space.

Lemma 8: Let I_2 and I_6 be homogeneous elements of $S_{\rm Int}(H^*)$ with grad $(I_j) = j, j = 2,6$, and $\{I_2, I_6\}$ an algebraically independent set. Then the union of $\{I_2, I_6\}$ with $1_{S(H^*)}$ is a generating set of $S_{\rm Int}(H^*)$.

Proof: From the above there exist two homogeneous elements S_2 and S_6 with degrees 2 and 6, repsectively, which generate together with $1_{S(H^*)}$ the algebra $S_{\rm Int}(H^*)$. Thus, $I_2 = cS_2, c \in \mathbb{C}, c \neq 0$. Therefore, there is a $p \in \mathbb{C}[t_1, t_2]$ with $I_6 = p(I_2, S_6)$. From this follows, because of the degrees and homogeneity of I_2, I_6, and S_6, that $I_6 = a(I_2)^3 + bS_6, ab \in \mathbb{C}$, and $b \neq 0$, since $\{I_2, I_6\}$ is algebraically independent. Thus $\{S_2, S_6\}$ is generated by $\{I_2, I_6\}$. □

Using $K(B,B) := \{ K(a,b) / a,b \in B \}$ one obtains: $h^1 = \frac{1}{24}(2h_1 - h_2)$, $h^2 = \frac{1}{24}(-h_1 + 2h_2)$, $g^{\alpha_{i,-j}} = \frac{1}{8} g_{\alpha_{j,-i}}$, $g^{\alpha_i} = -\frac{1}{24} g_{\alpha_{-i}}$, $g^{\alpha_{-i}} = -\frac{1}{24} g_{\alpha_i}$, $i \neq j$, $1 \leqslant i, j \leqslant 3$. Now let D be the identity representation $x \to x$ of $\widehat{G}_2$, $x \in \widehat{G}_2$. Then $\widehat{D} = \frac{1}{8}G$ (cf. Lemma 6) by the above formula, where $G :=$

$$\begin{bmatrix} 0 & -(\sqrt{2}/3)g_{\alpha_{-1}} & -(\sqrt{2}/3)g_{\alpha_{-2}} & -(\sqrt{2}/3)g_{\alpha_{-3}} & -(\sqrt{2}/3)g_{\alpha_1} & -(\sqrt{2}/3)g_{\alpha_2} & -(\sqrt{2}/3)g_{\alpha_3} \\ (\sqrt{2}/3)g_{\alpha_1} & \frac{1}{3}(h_1 + h_2) & g_{\alpha_{2,-1}} & g_{\alpha_{3,-1}} & 0 & \frac{1}{3}g_{\alpha_{-3}} & -\frac{1}{3}g_{\alpha_{-2}} \\ (\sqrt{2}/3)g_{\alpha_2} & g_{\alpha_{1,-2}} & \frac{1}{3}(-2h_1 + h_2) & g_{\alpha_{3,-2}} & -\frac{1}{3}g_{\alpha_{-3}} & 0 & \frac{1}{3}g_{\alpha_{-1}} \\ (\sqrt{2}/3)g_{\alpha_3} & g_{\alpha_{1,-3}} & g_{\alpha_{2,-3}} & \frac{1}{3}(h_1 - 2h_2) & \frac{1}{3}g_{\alpha_{-2}} & \frac{1}{3}g_{\alpha_{-1}} & 0 \\ (\sqrt{2}/3)g_{\alpha_{-1}} & 0 & \frac{1}{3}g_{\alpha_3} & -\frac{1}{3}g_{\alpha_2} & -\frac{1}{3}(h_1 + h_2) & -g_{\alpha_{1,-2}} & -g_{\alpha_{1,-3}} \\ (\sqrt{2}/3)g_{\alpha_{-2}} & -\frac{1}{3}g_{\alpha_3} & 0 & \frac{1}{3}g_{\alpha_1} & -g_{\alpha_{2,-1}} & \frac{1}{3}(2h_1 - h_2) & -g_{\alpha_{2,-3}} \\ (\sqrt{2}/3)g_{\alpha_{-3}} & \frac{1}{3}g_{\alpha_2} & -\frac{1}{3}g_{\alpha_1} & 0 & -g_{\alpha_{3,-1}} & -g_{\alpha_{3,-2}} & \frac{1}{3}(-h_1 + 2h_2) \end{bmatrix}$$

We have by Lemma 6

$z_j := \mathrm{tr}(G^j) \in Z^+(\widehat{G}), j \in \mathbb{N}$.

Let

$$h_{z_j} := 2(\tfrac{1}{3})^j((h_1 + h_2)^j + (h_2 - 2h_1)^j + (h_1 - 2h_2)^j),$$

j even. Then $(R \circ \delta \circ \lambda^{-1})(h_{z_j}) \in S_{\rm Int}(H^*)$ by Lemma 7: It follows by calculation that $(R \circ \delta \circ \lambda^{-1})(h_{z_j})$

$$= 2(\tfrac{1}{3})^j 24^j((\alpha_1)^j + (\alpha_1 + \alpha_{1,-3})^j + (2\alpha_1 + \alpha_{1,-3})^j).$$

Now we can prove

Lemma 9: Let

$$I_j := (\alpha_1)^j + (\alpha_1 + \alpha_{1,-3})^j + (2\alpha_1 + \alpha_{1,-3})^j.$$

Then $\{I_2, I_6\}$ is algebraically independent and generates together with $1_{S(H^*)}$ the algebra $S_{\rm Int}(H^*)$.

Proof: Because of Lemma 8 we need only to prove the property of algebraic independence. Let

$$J := \det \begin{bmatrix} \partial I_2/\partial \alpha_1 & \partial I_2/\partial \alpha_{1,-3} \\ \partial I_6/\partial \alpha_1 & \partial I_6/\partial \alpha_{1,-3} \end{bmatrix}.$$

Then $J(2h_1 + h_2) = -1440 \neq 0$. □

Because of the above assertions we have proved by our criterion

Lemma 10: $\{z_2, z_6\}$ is a complete set in $Z^+(\widehat{G}_2)$.

Remark 3: Let $a := (g_{\alpha_1}, g_{\alpha_2}, g_{\alpha_3})$, $b := (g_{\alpha_{-1}}, g_{\alpha_{-2}}, g_{\alpha_{-3}})$, $c := (g_{\alpha_{1,-2}}, g_{\alpha_{1,-3}}, g_{\alpha_{3,-2}})$, and $d := (g_{\alpha_{2,-1}}, g_{\alpha_{3,-1}}, g_{\alpha_{2,-3}})$. Then $z_2 = 2[-\frac{1}{3}(a \cdot b + b \cdot a) + (c \cdot d + d \cdot c) + \frac{2}{3}((h_1)^2 + (h_2)^2 - h_1 \cdot h_2)]$.

ACKNOWLEDGMENT

I am indebted to Professor H. J. Meister for his interest in the work. I am also indebted to Professor C. F. von Weizsäcker, who made part of the work possible.

[1] M. Umezawa, Nucl. Phys. **48**, 111 (1963).
[2] W. J. Holman III and L. C. Biedenharn, in *Group Theory and its Application*, edited by E. M. Loebl (Academic, New York, 1971), Vol. II, p. 1–73.
[3] M. Micu, Nucl. Phys. **60**, 353 (1964).
[4] M. Umezawa, Nucl. Phys. **53**, 54 (1964).
[5] M. Umezawa, Nucl. Phys. **57**, 65 (1964).
[6] M. Umezawa, Proc. Neder. Akad. van Wet. Ser. B **69**, 579 (1966).
[7] M. Umezawa, Proc. Neder. Akad. van Wet. Ser. B **69**, 592 (1966).
[8] M. Umezawa, Proc. Neder. Akad. van Wet. Ser. B **69**, 607 (1966).
[9] M. Umezawa, Proc. Neder. Akad. van Wet. Ser. B **69**, 620 (1966).

[10]L. C. Biedenharn, J. Math. Phys. **4**, 436 (1963).
[11]A. M. Perelomov and V. S. Popov, Yad. Fiz. **3**, 924 (1966) [Sov. J. Nucl. Phys. **3**, 676 (1966)].
[12]A. M. Perelomov and V. S. Popov, Yad. Fiz. **3**, 1127 (1966) [Sov. J. Nucl. Phys. **3**, 819 (1966)].
[13]A. M. Perelomov and V. S. Popov, Yad. Fiz. **5**, 693 (1967) [Sov. J. Nucl. Phys. **5**, 489 (1967)].
[14]A. M. Perelomov and V. S. Popov, Yad. Fiz. **7**, 460 (1968) [Sov. J. Nucl. Phys. **7**, 290 (1968)].
[15]A. M. Perelomov and V. S. Popov, Izv. Akad. Nauk SSR Ser. Mat. **32**, 1368 (1968) (in Russian).
[16]B. Gruber and L. O'Raifeartaigh, J. Math. Phsy. **5**, 1796 (1964).
[17]G. Racah, *Group Theory and Spectroscopy*, in Erg. Exakt. Nat. **37**, 28 (Springer, Berlin, 1965).
[18]G. Racah, Rend. Lincei **8**, 108 (1950).
[19]I. M. Gel'fand, Mat. Sb. **26**, 103 (1950) (in Russian).
[20]E. G. Beltrametti and A. Blasi, Phys. Lett. **20**, 62 (1966).
[21]U. Mutze, Z. Phys. **229**, 224 (1969).
[22]L. Abellanas and L. M. Alonso, J. Math. Phys. **16**, 1580 (1975).
[23]M. A. Gauger, Pac. J. Math. **62**, 93 (1976).
[24]C. Chevalley, *Théorie des Groupes de Lie* (Hermann, Paris, 1968).
[25]J. Dixmier, *Enveloping Algebras* (Akademie, Berlin, 1977).
[26]F. Berdjis, "Untersuchung der unabhängigen Casimir Operatoren halbeinfacher Lie Algebren," Diplomarbeit, Universität München, 1972.
[27]H. Freudenthal and H. de Vries, *Linear Lie Groups* (Academic, New York, 1969).
[28]C. Chevalley, Am. J. Math. **77**, 778 (1955).
[29]N. Jacobson, *Lie Algebras* (Interscience, New York, 1966).
[30]J. E. Humphreys, *Introduction to Lie Algebras and Representation Theory* (Springer, New York, 1972).
[31]F. Berdjis, "Die Casimir Operatoren der halbeinfachen Lie Algebren," Doctoral dissertation, Universität München, 1977.
[32]A. O. Barut and R. Raczka, *Theory of Group Representations and Applications* (Polish Scientific Publishers, Warswa, 1977).

Casimir operators for F_4, E_6, E_7, and E_8

F. Berdjis
Max-Planck-Institut, D813 Starnberg, West Germany

E. Beslmüller
Sektion Physik der Universität München, D8000 München 2, Theresienstrasse 37, Federal Republic of Germany

(Received 6 February 1981; accepted for publication 10 April 1981)

We construct explicitly algebraically-independent generating sets for the Casimir operators and the invariants of the Weyl groups and adjoint groups of the exceptional Lie algebras F_4, E_6, E_7, and E_8.

PACS numbers: 02.20.Sv

This paper is an extension of the work done in Ref. 1. Up to now, sets generating the Casimir operators of F_4, E_6, E_7, and E_8, respectively, were not known. To find complete sets of Casimir operators for F_4, E_6, E_7, and E_8, we construct with the aid of the lowest-dimensional representations some elements of the center of the universal enveloping algebra. By a criterion given in Ref. 1, we show that the elements obtained in that way are a complete set of Casimir operators, in the sense that they generate algebraically all Casimir operators and are algebraically independent. In order to prove our assertions on Casimir operators we need complete sets of the invariants of the Weyl groups. Therefore, we constructed the Weyl invariants of the above exceptional Lie algebras. According to Remark 2 and the statements subsequent to Lemma 6 in Ref. 1, we have also constructed complete sets of the invariants of the adjoint groups of the Lie algebras under consideration. We emphasize that we have constructed complete sets for the above invariant algebras of the exceptional Lie algebras for the first time in the literature. The method of presentation is the same as that of G_2 in Ref. 1. The models of F_4, E_6, E_7, and E_8, are taken from Ref. 2. A mistake in representing E_8 in Ref. 2 is eliminated.

In order to compute our Casimir operators explicitly we have to know all the generators (e.g., a Cartan–Weyl basis) of the Lie algebras under consideration. But this knowledge is not necessary for our proof of the algebraic indepdence and the generating property of our sets of Casimir operators. For this we need only the explicit form of a basis of a Cartan subalgebra. Therefore, presenting the models of our Lie algebras, we restrict the explicit representation of a basis of the Lie algebra by furnishing a basis of a Cartan subalgebra as explicit matrices and indicating how to calculate all the other generators which are elements of rootspaces corresponding to nonzero roots.

Further we need the knowledge of the degrees of the elements of a minimal set of homogeneous elements that together with $1_{S(H^*)}$ generates the algebra $S_{\text{Int}}(H^*)$ (*we use the same symbols as in Ref.* 1). These are for[3]

$$\begin{aligned} &F_4\colon 2, 6, 8, 12;\\ &E_6\colon 2, 5, 6, 8, 9, 12;\\ &E_7\colon 2, 6, 8, 10, 12, 14, 18;\\ &E_8\colon 2, 8, 12, 14, 18, 20, 24, 30. \end{aligned} \tag{1}$$

Presenting our models of the above algebras we always give a list of objects first, followed by a list of propositions.

F_4

Let h_i, $e_i \in M_{26}$, $1 \leqslant i \leqslant 4$, as in Ref. 2, pp. 169 and 170; H be the linear span of $\{h_i/1\leqslant i\leqslant 4\}$, $A \equiv (A_{ij})_{1\leqslant i,j\leqslant 4} \in M_4$ as in Ref. 2, p. 200; α_i be linear functionals on H with $\alpha_i(h_j) := A_{ji}, 1\leqslant i, j \leqslant 4$, $I(4) := \{(1,0,0,0), (0,1,0,0), (0,0,1,0), (0,0,0,1), (1,1,0,0), (0,1,1,0), (0,0,1,1), (1,1,1,0), (0,1,1,1), (0,1,2,0), (1,1,1,1), (1,1,2,0), (0,1,2,1), (1,1,2,1), (1,2,2,0), (0,1,2,2), (1,1,2,2), (1,2,2,1), (1,2,2,2), (1,2,3,1), (1,2,3,2), (1,2,4,2), (1,3,4,2), (2,3,4,2)\}$, $W^+ := \{(\alpha_1, \alpha_2, \alpha_3, \alpha_4)\cdot a^t / a \in I(4)\}$ ($\cdot$ is matrix multiplication, t means transpose), $W^- := \{-\alpha/\alpha \in W^+\}$, $e_{\alpha_i} := e_i, 1\leqslant i\leqslant 4$,

$$e_{\alpha_{i_1}+\alpha_{i_2}+\alpha_{i_3}+\cdots+\alpha_{i_n}} := [\ldots[e_{\alpha_{i_1}}, e_{\alpha_{i_2}}], e_{\alpha_{i_3}}], \ldots, e_{\alpha_{i_n}}],$$
$$n, i_1, \ldots, i_n \in \mathbf{N}, \tag{2}$$

$e_{-\alpha} := -e_\alpha^t, \alpha \in W^+, B^+ := \{e_\alpha/\alpha \in W^+\}, B^- := \{e_\alpha/\alpha \in W^-\}$, and $F_4(\mathbf{C})$ be the linear span of $B^+ \cup B^- \cup H$. Then $F_4(\mathbf{C}) \in F_4$ with the Lie product $[x,y] := x\cdot y - y\cdot x$, $x, y \in F_4(\mathbf{C})$, H is a Cartan subalgebra of $F_4(\mathbf{C})$, A is the Cartan matrix of $F_4(\mathbf{C})$, $\{\alpha_i/1\leqslant i\leqslant 4\}$ is a set of simple roots of $F_4(\mathbf{C})$, W^+ and W^- are the set of positive and negative roots, respectively, and the elements of B^+ and B^- are elements of the rootspaces corresponding to positive and negative roots, respectively.

First we prove

Lemma 1: Let $\{I_j/j = 2, 6, 8, 12\}$ be a set of algebraically independent homogeneous elements of $S_{\text{Int}}(H^*)$ with grad $(I_j) = j, j = 2, 6, 8, 12$. Then the union of $\{I_j/j = 2, 6, 8, 12\}$ with $1_{S(H^*)}$ is a generating set of $S_{\text{Int}}(H^*)$.

Proof: Because of (1) there exist four homogeneous elements S_2, S_6, S_8, and S_{12} with degrees 2, 6, 8, and 12, respectively, which generate together with $1_{S(H^*)}$ the algebra $S_{\text{Int}}(H^*)$. Thus $I_2 = a_2 S_2$, $a_2 \in \mathbf{C}$, $a_2 \neq 0$. Therefore, there exists a polynomial $p \in \mathbf{C}[t_1, t_2]$ with $I_6 = p(I_2, S_6)$. Because of the degrees and homogeneity of I_2, I_6, and S_6 it follows that $I_6 = a_6(I_2)^3 + b_6 S_6$, $a_6, b_6 \in \mathbf{C}$, and $b_6 \neq 0$, since $\{I_2, I_6\}$ is algebraically independent. Analogously there exist polynomials $p_1 \in \mathbf{C}[t_1, t_2, t_3]$ and $p_2 \in \mathbf{C}[t_1, t_2, t_3, t_4]$ with $I_8 = p_1(I_2, I_6, S_8) = a_8(I_2)^4 + b_8 I_6 S_2 + c_8 S_8$, $a_8, b_8, c_8 \in \mathbf{C}$ and $I_{12} = p_2(I_2, I_6, I_8, S_{12}) = a_{12}(I_2)^6 + b_{12}(I_6)^2 + c_{12} I_6 (I_2)^3 + d_{12} I_8 (I_2)^2 + e_{12} S_{12}$, $a_{12}, b_{12}, \ldots, e_{12} \in \mathbf{C}$. Then $c_8 \neq 0$ and

$e_{12}\neq 0$, since $\{I_2, I_6, I_8\}$, and $\{I_2, I_6, I_8, I_{12}\}$ are algebraically independent. Thus $\{S_2, S_6, S_8, S_{12}\}$ is generated by $\{I_2, I_6, I_8, I_{12}\}$.□

Now we construct some Casimir operators with the aid of Lemma 6 in Ref. 1. For the Killing form restricted on $H\times H$ we compute

$$(K_{ij}/H\times H)_{1\leqslant i,j\leqslant 4} = 18\begin{bmatrix} 2 & -1 & 0 & 0\\ -1 & 2 & -2 & 0\\ 0 & -2 & 4 & -2\\ 0 & 0 & -2 & 4\end{bmatrix}$$

and

$$(K_{ij}^{-1}/H\times H)_{1\leqslant i,j\leqslant 4} = \frac{1}{36}\begin{bmatrix} 4 & 6 & 4 & 2\\ 6 & 12 & 8 & 4\\ 4 & 8 & 6 & 3\\ 2 & 4 & 3 & 2\end{bmatrix}$$

(see the proof of Lemma 7 in Ref. 1). Therefore, we have $h^1 = \frac{1}{36}(4h_1+6h_2+4h_3+2h_4)$, $h^2 = \frac{1}{36}(6h_1+12h_2+8h_3+4h_4)$,, $h^3=\frac{1}{36}(4h_1+8h_2+6h_3+3h_4)$, and $h^4=\frac{1}{36}(2h_1+4h_2+3h_3+2h_4)$. Let D be the identity representation of $F_4(\mathbf{C})$, $\{x_1,...,x_{52}\}:= B^+\cup B^-\cup H$, $\hat{D}_{jm} = \Sigma_{i=1}^{52} x^i D(x_i)_{jm}$, $1\leqslant j, m\leqslant 26$, and $\hat{D}:=(\hat{D}_{jm})_{1\leqslant j,m\leqslant 26}$. Then $z_k:=\operatorname{tr}(\hat{D}^k)\in Z^+(F_4(\mathbf{C}))$, $k\in\mathbf{N}$. Let $h_{z_k} := \operatorname{tr}[(\Sigma_{i=1}^4 D(h_i)h^i)^k]$. Then $h_{z_k} = (\frac{1}{36})^k((2h_1+4h_2+3h_3+2h_4)^k + (2h_1+4h_2+3h_3+h_4)^k + (2h_1+4h_2+2h_3+h_4)^k + (2h_1+2h_2+2h_3+h_4)^k + (2h_1+2h_2+h_3+h_4)^k + (2h_2+2h_3+h_4)^k + (2h_1+2h_2+h_3)^k + (2h_2+h_3+h_4)^k + (2h_2+h_3)^k + (h_3+h_4)^k + (h_3)^k + (h_4)^k)$, $k=2, 6, 8, 12$. Then $(R\circ\delta\circ\gamma^{-1})h_{z_k})\in S_{\text{Int}}(H^*)$ (see Lemma 7 in Ref. 1). It follows by calculation that

$$\begin{aligned} J_k &:= (R\circ\delta\circ\lambda^{-1})(h_{z_k})\\ &= (\tfrac{2}{36})^k((\alpha_1+2\alpha_2+3\alpha_3+2\alpha_4)^k + (\alpha_1+2\alpha_2+3\alpha_3+\alpha_4)^k\\ &\quad + (\alpha_1+2\alpha_2+2\alpha_3+\alpha_4)^k + (\alpha_1+\alpha_2+2\alpha_3+\alpha_4)^k\\ &\quad + (\alpha_1+\alpha_2+\alpha_3+\alpha_4)^k + (\alpha_2+2\alpha_3+\alpha_4)^k\\ &\quad + (\alpha_1+\alpha_2+\alpha_3)^k + (\alpha_2+\alpha_3+\alpha_4)^k + (\alpha_2+\alpha_3)^k\\ &\quad + (\alpha_3+\alpha_4)^k + (\alpha_3)^k + (\alpha_4)^k), k=2, 6, 8, 12.\end{aligned}$$

Now we can prove

Lemma 2: Let J_k be as in Eq. (3). Then $\{J_k/k=2, 6, 8, 12\}$ is algebraically independent and generates together with $1_{S(H^*)}$ the algebra $S_{\text{Int}}(H^*)$.

Proof: Because of Lemma 1 we have only to prove that $\{J_k/k=2, 6, 8, 12\}$ is algebraically independent. Let

$$J := \det\begin{bmatrix} \partial J_2/\partial\alpha_1 & \partial J_2/\partial\alpha_2 & \partial J_2/\partial\alpha_3 & \partial J_2/\partial\alpha_4\\ \partial J_6/\partial\alpha_1 & & & \cdot\\ \partial J_8/\partial\alpha_1 & & & \cdot\\ \partial J_{12}/\partial\alpha_1 & \cdot & \cdot & \partial J_{12}/\partial\alpha_4\end{bmatrix}.$$

Suppose

$$a_1\left(\frac{\partial J_2}{\partial\alpha_1}, \frac{\partial J_2}{\partial\alpha_2}, \frac{\partial J_2}{\partial\alpha_3}, \frac{\partial J_2}{\partial\alpha_4}\right) + \cdots + a_4\left(\frac{\partial J_{12}}{\partial\alpha_1}, \frac{\partial J_{12}}{\partial\alpha_2}, \frac{\partial J_{12}}{\partial\alpha_3}, \frac{\partial J_{12}}{\partial\alpha_4}\right)$$

$=0$, $a_1, a_2, a_3, a_4\in\mathbf{C}$. Then,

$$a_1\frac{\partial J_2}{\partial\alpha_1} + a_2\frac{\partial J_6}{\partial\alpha_1} + a_3\frac{\partial J_8}{\partial\alpha_1} + a_4\frac{\partial J_{12}}{\partial\alpha_1} = 0. \tag{4}$$

Since in (4) each homogeneous part has to be identically zero and since $\partial J_k/\partial\alpha_1\neq 0$, $k=2, 6, 8, 12$, it follows that $a_1=a_2=a_3=a_4=0$. Thus the rows of the above matrix are linearly independent and therefore $J\neq 0$.□

By the above we have proved

Lemma 3: $\{z_2, z_6, z_8, z_{12}\}$ is a complete set in $Z^+(F_4(\mathbf{C}))$.

Now we show in the same manner as for F_4 that for E_6 and E_7 there are, respectively, 6 and 7 Casimir operators which generate all Casimir operators and are algebraically independent. Since the methods and proofs are analogous to those for F_4, we state the lemmas without proofs.

E_6:

Let $h_i, e_i\in M_{27}$, $1\leqslant i\leqslant 6$ as in Ref. 2, pp. 170 and 171; H be the linear span of $\{h_i/1\leqslant i\leqslant 6\}$, $A\equiv(A_{ij})_{1\leqslant i,j\leqslant 6}\in M_6$ as in Ref. 2, p. 201; α_i be linear functionals on H with $\alpha_i(h_j) := A_{ji}$, $1\leqslant i, j\leqslant 6$,

$I(6.1) := \{(1,0,0,0,0,0), (1,1,0,0,0,0), (1,1,1,0,0,0), (1,1,1,0,0,1), (1,1,1,1,0,0), (1,1,1,1,0,1), (0,1,0,0,0,0), (0,1,1,0,0,0), (0,1,1,0,0,1), (0,1,1,1,0,0), (0,1,1,1,0,1), (0,0,1,0,0,0), (0,0,1,0,0,1), (0,0,1,1,0,0), (0,0,1,1,0,1), (0,0,0,1,0,0), (0,0,0,0,0,1), (0,1,2,1,0,1), (1,1,2,1,0,1), (0,1,2,1,1,1), (1,2,2,1,0,1), (0,1,2,2,1,1), (1,1,2,1,1,1), (1,2,2,1,1,1), (1,1,2,2,1,1), (1,2,2,2,1,1), (1,2,3,2,1,1), (1,2,3,2,1,2)\}$, (5)

$I(6.2) := \{(1,1,1,1,1,0), (1,1,1,1,1,1), (0,1,1,1,1,0), (0,1,1,1,1,1), (0,0,1,1,1,0), (0,0,1,1,1,1), ((0,0,0,1,1,0), (0,0,0,0,1,0)\}$, (6)

$W^+ := \{(\alpha_1,...,\alpha_6)\cdot a^t/a\in I(6.1)\cup I(6.2)\}$, $B := \{e_\alpha, e^t_{-\alpha}/\alpha\in W^+\}$ [e_α defined as in Eq. (2)], and $E_6(\mathbf{C})$ is the linear span of $B\cup H$. Then $E_6(\mathbf{C})\in E_6$ with the Lie product $[x,y] := x\cdot y - y\cdot x$, $x, y\in E_6(\mathbf{C})$, H is a Cartan subalgebra of $E_6(\mathbf{C})$, A is the Cartan matrix of $E_6(\mathbf{C})$, $\{\alpha_i/1\leqslant i\leqslant 6\}$ is a set of simple roots of $E_6(\mathbf{C})$, W^+ is the set of positive roots, and the elements of B are elements of the rootspaces corresponding to positive and negative roots, respectively.

Lemma 4: Let $\{I_j/j=2, 5, 6, 8, 9, 12\}$ be a set of algebraically independent homogeneous elements of $S_{\text{Int}}(H^*)$ with $\operatorname{grad}(I_j) = j$, $j=2, 5, 6, 8, 9, 12$. Then the union of $\{I_j/j=2, 5, 6, 8, 9, 12\}$ with $1_{S(H^*)}$ is a generating set of $S_{\text{Int}}(H^*)$.

Now we construct some Casimir operators. For the Killing form restricted on $H\times H$ we compute $(K_{ij}/H\times H)_{1\leqslant i,j\leqslant 6} = 12A_{ij}$, and $(K_{ij}^{-1}/H\times H)_{1\leqslant i,j\leqslant 6} = \frac{1}{12}(A_{ij}^{-1})_{1\leqslant i,j\leqslant 6}$, A^{-1} as in Ref. 2, p. 201. Therefore, we know that $h^i = \Sigma_{j=1}^6 K^{ij}/H\times H\ h_j$, $1\leqslant i\leqslant 6$. Let D be the identical representation of $E_6(\mathbf{C})$, $\{x_1,...,x_{78}\} := B\cup H$, $\hat{D}_{jm} = \Sigma_{i=1}^{78} D(x_i)_{jm} x^i$, $1\leqslant i, j\leqslant 27$, $\hat{D} := (\hat{D}_{jm})_{1\leqslant j,m\leqslant 27}$. Then $z_k = \operatorname{tr}(\hat{D}^k)\in Z^+(E_6(\mathbf{C}))$, $k\in\mathbf{N}$ by Lemma 6 in Ref. 1. Obviously $h_{z_k} = \operatorname{tr}[(\Sigma_{i=1}^6 D(h_i)h^i)^k]$. Since we know the matrix entries $D(h_i)$ we compute $h_{z_k} = \Sigma_{p\in P}(\frac{1}{36}p\cdot(h_1,...,h_6)^t)^k$, $k=2, 5, 6, 8, 9, 12$, where

$P := \{(-4,-5,-6,-4,-2,-3), (-1,-5,-6,-4,-2,-3), (-1,-2,-6,-4,-2,-3), (-1,-2,-3,-4,-2,-3), (-1,-2,-3,-1,-2,-3), (-1,-2,-3,-1,1,-3)$,

$(-1,-2,-3,-4,-2,0)$, $(-1,-2,-3,-1,-2,0)$, $(-1,-2,0,-1,-2,0)$, $(-1,1,0,-1,-2,0)$, $(-1,-2,-3,-1,1,0)$, $(-1,-2,0,-1,1,0)$, $(-1,-2,0,2,1,0)$, $(-1,1,0,-1,1,0)$, $(-2,-1,0,1,2,0)$, $(-1,1,0,2,1,0)$, $(-1,2,3,2,1,0)$, $(2,1,0,-1,1,0)$, $(2,1,0,2,1,0)$, $(2,1,3,2,1,0)$, $(2,4,3,2,1,0)$, $(-1,2,3,2,1,3)$, $(2,1,3,2,1,3)$, $(2,4,3,2,1,3)$, $(2,4,6,2,1,3)$, $(2,4,6,5,1,3)$, $(2,4,6,5,4,3)\}$. (7)

Since $\delta(h_i) = -12\alpha_i, 1\leqslant i\leqslant 6$, we have

$$J_k = (R\circ\delta\circ\lambda^{-1})(h_{z_k}) = \sum_{p\in P}(\tfrac{1}{3}p\cdot(\alpha_1,\alpha_2,\ldots,\alpha_6)^t)^k, \tag{8}$$

where P is as in (7), and $J_k\in S_{\rm Int}(H^*), k = 2, 5, 6, 8, 9, 12$ by Lemma 7 in Ref. 1. With the aid of $\partial J_k/\partial\alpha_1\neq 0$, $k\in\mathbf{N}$, $k\geqslant 2$ we can prove

Lemma 5: Let J_k be as in Eq. (8). Then $\{J_k/k = 2, 5, 6, 8, 9, 12\}$ is algebraically independent and generates together with $1_{S(H^*)}$ the algebra $S_{\rm Int}(H^*)$.

By the above we have proved

Lemma 6: $\{z_k/k = 2, 5, 6, 8, 9, 12\}$ is a complete set in $Z^+(E_6(\mathbf{C}))$.

E_7

Let $h_i, e_i\in M_{56}$, $1\leqslant i\leqslant 7$, as in Ref. 2, pp. 171, 172; H be the linear span of $\{h_i/1\leqslant i\leqslant 7\}$, $A\equiv(A_{ij})_{1\leqslant i,j\leqslant 7}\in M_7$ as in Ref. 2, p. 201; α_i be linear functionals on H with $\alpha_i(h_j) := A_{ji}, 1\leqslant i, j\leqslant 7$,

$I(7) := \{(0,1,2,1,1,1,1)$, $(1,1,2,1,1,1,1)$, $(0,1,2,2,1,1,1)$, $(1,2,2,1,1,1,1)$, $(1,1,2,2,1,1,1)$, $(0,1,2,2,2,1,1)$, $(1,2,2,2,1,1,1)$, $(1,1,2,2,2,1,1)$, $(1,2,3,2,1,1,1)$, $(1,2,2,2,2,1,1)$, $(1,2,3,2,1,1,2)$, $(1,2,3,2,2,1,1)$, $(1,2,3,2,2,1,2)$, $(1,2,3,3,2,1,1)$, $(1,2,3,3,2,1,2)$, $(1,2,4,3,2,1,2)$, $(1,3,4,3,2,1,2)$, $((2,3,4,3,2,1,2)\}$, (9)

$W^+ := \{(\alpha_1,\ldots,\alpha_5,\alpha_7)\cdot a^t/a\in I(6.1)\cup I(6.2)\}\cup\{(\alpha_1,\ldots,\alpha_5,\alpha_7)\cdot a^t + \alpha_6/a\in I(6.2)\}\cup\{\alpha_6\}\cup\{(\alpha_1,\ldots,\alpha_7)\cdot a^t/a\in I(7)\}$ [see Eqs. (5), (6), and (9)], $B := \{e_\alpha, -e_\alpha^t/a\in W^+\}$ [e_α defined as in Eq. (2)], and $E_7(\mathbf{C})$ be the linear span of $B\cup H$. Then $E_7(\mathbf{C})\in E_7$ with the Lie product $[x,y] := x\cdot y - y\cdot x$, $x, y\in E_7(\mathbf{C})$, H is a Cartan subalgebra of $E_7(\mathbf{C})$, A is the Cartan matrix of $E_7(\mathbf{C})$, $\{\alpha_i/1\leqslant i\leqslant 7\}$ is a set of simple roots of $E_7(\mathbf{C})$, W^+ is the set of positive roots, and the elements of B are elements of the rootspaces corresponding to positive and negative roots, respectively.

Lemma 7: Let $\{I_j/j = 2, 6, 8, 10, 12, 14, 18\}$ be a set of algebraically independent homogeneous elements of $S_{\rm Int}(H^*)$ with grad $(I_j) = j, j = 2, 6, 8, 10, 12, 14, 18$. Then the union of $\{I_j/j = 2, 6, 8, 10, 12, 14, 18\}$ with $1_{S(H^*)}$ is a generating set of $S_{\rm Int}(H^*)$.

Now we construct some Casimir operators. For the Killing form restricted on $H\times H$ we compute $(K_{ij}/H\times H)_{1\leqslant i,j\leqslant 7} = 24A_{ij}$. Therefore, we know $h^i = \Sigma_{j=1}^7 K^{ij}/H\times H h_j, 1\leqslant i\leqslant 7$. Let D be the identical representation of $E_7(\mathbf{C})$, $\{x_1,\ldots,x_{133}\} := B\cup H$, $\hat{D}_{jm} = \Sigma_{i=1}^{133} D(x_i)_{jm}x^i, 1\leqslant j, m\leqslant 56$, $\hat{D} := (\hat{D}_{jm})_{1\leqslant j,m\leqslant 56}$. Then $z_k = {\rm tr}(\hat{D}^k)\in Z^+(E_7(\mathbf{C})), k\in\mathbf{N}$, by Lemma 6 in Ref. 1. Obviously $h_{z_k} = {\rm tr}[(\Sigma_{i=1}^7 D(h_i)h^i)^k]$. Since we know the matrix entries $D(h_i)$ we compute $h_{z_k} = \Sigma_{p\in P}(\frac{1}{12}p\cdot(h_1,\ldots,h_7)^t)^k$, where

$P := \{(2,4,6,5,4,3,3)$, $(2,4,6,5,4,1,3)$, $(2,4,6,5,2,1,3)$, $(2,4,6,3,2,1,3)$, $(2,4,4,3,2,1,3)$, $(2,2,4,3,2,1,3)$, $(0,2,4,3,2,1,3)$, $(0,2,4,3,2,1,1)$, $(0,2,2,3,2,1,1)$, $(0,0,2,3,2,1,1)$, $(0,0,2,1,2,1,1)$, $(0,0,0,1,2,1,1)$, $(0,0,0,1,0,1,1)$, $(0,0,0,-1,0,1,1)$, $(0,0,0,-1,0,-1,1)$, $(0,0,0,-1,-2,-1,1)$, $(0,2,2,1,2,1,1)$, $(2,2,2,1,2,1,1)$, $(2,2,2,3,2,1,1)$, $(2,2,4,3,2,1,1)$, $(2,4,4,3,2,1,1)$, $(0,2,2,1,0,1,1)$, $(0,2,2,1,0,-1,1)$, $(0,0,2,1,0,-1,1)$, $(0,0,0,1,0,-1,1)$, $(0,0,2,1,0,1,1)$, $(2,2,2,1,0,1,1)$, $(2,2,2,1,0,-1,1)\}$. (10)

Since $\delta(h_i) = -24\alpha_i, 1\leqslant i\leqslant 7$, we have

$$J_k = (R\circ\delta\circ\lambda^{-1})(h_{z_k}) = \sum_{p\in P}(p\cdot(\alpha_1,\ldots,\alpha_7)^t)^k, \tag{11}$$

where P is as in (10) and $J_k\in S_{\rm Int}(H^*), k = 2, 6, 8, 10, 12, 14, 18$ by *Lemma 7* in Ref. 1. With the aid of $(\partial J_k/\partial\alpha_1)\neq 0$, $k\geqslant 2$, we can prove

Lemma 8: Let J_k be as in Eq. (11). Then $\{J_k/k = 2, 6, 8, 10, 12, 14, 18\}$ is algebraically independent and generates together with $1_{S(H^*)}$ the algebra $S_{\rm Int}(H^*)$.

By the above we have proved

Lemma 9: $\{z_k/k = 2, 6, 8, 10, 12, 14, 18\}$ is a complete set in $Z^+(E_7(\mathbf{C}))$.

$\mathbf{E_8}$:

Remark: There is a mistake in Ref. 2 representing the entries $(h_i)_{187,187}, 1\leqslant i\leqslant 8$, which would yield as corresponding root $\alpha := 2\alpha_1 + 5\alpha_2 + 8\alpha_3 + 7\alpha_4 + 5\alpha_5 + 3\alpha_6 + 2\alpha_7 + 4\alpha_8$. But α is not a root of E_8. In the following we furnish all the roots and a basis for a Cartan subalgebra of E_8. Let $A\equiv(A_{ij})_{1\leqslant i,j\leqslant 8}\in M_8$ as in Ref. 2, p. 202,

$I(8) := \{(0,1,2,1,1,1,1,1)$, $(1,1,2,1,1,1,1,1)$, $(0,1,2,2,1,1,1,1)$, $(1,2,2,1,1,1,1,1)$, $(1,1,2,2,1,1,1,1)$, $(0,1,2,2,2,1,1,1)$, $(1,2,2,2,1,1,1,1)$, $(1,1,2,2,2,1,1,1)$, $(0,1,2,2,2,2,1,1)$, $(1,2,3,2,1,1,1,1)$, $(1,2,2,2,2,1,1,1)$, $(1,1,2,2,2,2,1,1)$, $(1,2,3,2,1,1,1,2)$, $(1,2,3,2,2,1,1,1)$, $(1,2,2,2,2,2,1,1)$, $(1,2,3,2,2,1,1,2)$, $(1,2,3,3,2,1,1,1)$, $(1,2,3,2,2,2,1,1)$, $(1,2,3,3,2,1,1,2)$, $(1,2,3,2,2,2,1,2)$, $(1,2,3,3,2,2,1,1)$, $(1,2,4,3,2,1,1,2)$, $(1,2,3,3,2,2,1,2)$, $(1,2,3,3,3,2,1,1)$, $(1,3,4,3,2,1,1,2)$, $(1,2,4,3,2,2,1,2)$, $(1,2,3,3,3,2,1,2)$, $(2,3,4,3,2,1,1,2)$, $(1,3,4,3,2,2,1,2)$, $(1,2,4,3,3,2,1,2)$, $(2,3,4,3,2,2,1,2)$, $(1,3,4,3,3,2,1,2)$, $(1,2,4,4,3,2,1,2)$, $(2,3,4,3,3,2,1,2)$, $(1,3,4,4,3,2,1,2)$, $(1,3,5,4,3,2,1,2)$, $(2,3,4,4,3,2,1,2)$, $(1,3,5,4,3,2,1,3)$, $(2,3,5,4,3,2,1,2)$, $(2,3,5,4,3,2,1,3)$, $(2,4,5,4,3,2,1,2)$, $(2,4,5,4,3,2,1,3)$, $(2,4,6,4,3,2,1,3)$, $(2,4,6,5,3,2,1,3)$, $(2,4,6,5,4,2,1,3)$, $(2,4,6,5,4,3,1,3)$, $(2,4,6,5,4,3,2,3))\}$, (12)

$W^+ := \{(\alpha_1,\ldots,\alpha_5,\alpha_8)\cdot a^t/a\in I(6.1)\cup I(6.2)\}\cup\{(\alpha_1,\ldots,\alpha_5,\alpha_8)\cdot a^t + \alpha_6/a\in I(6.2)\}\cup\{\alpha_7\}\cup\{(\alpha_1,\ldots,\alpha_6,\alpha_8)\cdot a^t/a\in I(7)\}\cup\{(\alpha_1,\ldots,\alpha_5,\alpha_8)\cdot a^t + \alpha_6 + \alpha_7/a\in I(6.2)\}\cup\{\alpha_6,\alpha_6+\alpha_7\}\cup\{(\alpha_1,\ldots,\alpha_8)\cdot a^t/a\in I(8)\}$ [see Eqs. (5), (6), (9), and (12)], $\{\bar\beta_1,\ldots,\bar\beta_{120}\} := W^+$, $\{\beta_{129},\ldots,\beta_{248}\} := \{-\beta/\beta\in W^+\}$, $h_i\in M_{248}, 1\leqslant i\leqslant 8$, with $(h_i)_{jj} := \beta_j(h_i), 1\leqslant j\leqslant 120, 129\leqslant j\leqslant 248$, $(h_i)_{jj} := 0, 121\leqslant j\leqslant 128$, $(h_i)_{km} := 0$ for $k\neq m$, $1\leqslant k, m\leqslant 248$, and H be the linear span of $\{h_i/1\leqslant i\leqslant 8\}$. Then A is the Cartan matrix, W^+ the set of positive roots, $\{\alpha_1,\dots,\alpha_8\}$ a set of simple roots, and H a Cartan subalgebra of a model $E_8(\mathbf{C})\in E_8$. It is standard to construct by the above a canonical basis for $E_8(\mathbf{C})$ (see Ref. 4, p. 126; physicists call it Cartan–Weyl basis). As mentioned before we need only the basis of H explicitly.

Lemma 10: Let $\{I_j/j = 2, 8, 12, 14, 18, 20, 24, 30\}$ be a set of algebraically-independent homogeneous elements of $S_{\mathrm{Int}}(H^*)$ with $\mathrm{grad}(I_j) = j, j = 2, 8, 12, 14, 18, 20, 24, 30$. Then the union of $\{I_j/j = 2, 8, 12, 14, 18, 20, 24, 30\}$ with $1_{S(H^*)}$ is a generating set of $S_{\mathrm{Int}}(H^*)$.

Let $\{x_1,\dots,x_{248}\}$ be a canonical basis of $E_8(\mathbf{C})$ containing $\{h_i/1\leqslant i\leqslant 8\}$, D be the identical representation of $E_8(\mathbf{C})$ with $D(x_i) = x_i, 1\leqslant i\leqslant 248$, $\widehat{D}_{jm} = \Sigma_{i=1}^{248} D(x_i)_{jm} x^i, 1\leqslant j, m\leqslant 248$, and $\widehat{D} := (\widehat{D}_{jm})_{1\leqslant j,m\leqslant 248}$. Then $z_k = \mathrm{tr}(\widehat{D}^k)\in Z^+(E_8(\mathbf{C})), k\in\mathbf{N}$, by Lemma 6 in Ref. 1. Obviously $h_{z_k} = \mathrm{tr}\left[(\Sigma_{i=1}^{8} D(h_i)h^i)^k\right]$. Since the weights of D are the roots, we have

$$(R\circ\delta\circ\lambda^{-1})(h_{z_k}) = \Sigma_{\substack{1\leqslant i\leqslant 120\\ 129\leqslant i\leqslant 248}} (\beta_i)^k =: J_k \in S_{\mathrm{Int}}(H^*)$$

by Lemma 7 in Ref. 1. Therefore we can prove by the aid of $(\partial J_k/\partial\beta_7)\neq 0, k = 2, 8, 12, 14, 18, 20, 24, 30$

Lemma 11: Let $J_k := 2\Sigma_{i=1}^{120}(\beta_i)^k$. Then $\{J_k/k = 2, 8, 12, 14, 18, 20, 24, 30\}$ is algebraically independent and generates together with $1_{S(H^*)}$ the algebra $S_{\mathrm{Int}}(H^*)$,

Therefore, we have

Lemma 12: $\{z_k/k = 2, 8, 12, 14, 18, 20, 24, 30\}$ is a complete set in $Z^+(E_8(\mathbf{C}))$.

Remark: The lowest dimensional representation of E_8 is its adjoint representation with the aid of which we constructed a complete set for $Z^+(E_8)$. This is not always possible. Let L be a semisimple Lie algebra and W the set of its roots. Then one can construct Casimir operators to each degree $k\in\mathbf{N}$ by the method of Lemma 6 in Ref. 1 using the adjoint representation. It follows that $(R\circ\delta\circ\lambda^{-1})h_{z_k} = \Sigma_{\beta\in W}(\beta)^k$. This term vanishes for odd k. Therefore, it is obvious that for those Lie algebras which have generating elements of their Weyl-invariants $S_{\mathrm{Int}}(H^*)$ of odd degree one cannot hope to get a complete set of Casimir operators by the adjoint representation of L using the above method. For F_4 and E_7 one gets complete sets of Casimir operators using that method by the adjoint representation, but we reject its presentation. We declare further that we computed complete sets of Casimir operators in canonical bases. Due to the assertions subsequent to Lemma 6 in Ref. 1 one gets the same complete sets using other bases of the Lie algebras under consideration.

ACKNOWLEDGMENT

The authors are indebted to Professor C. F. von Weizsäcker and Professor G. Süssmann, who made this work possible.

[1] F. Berdjis, "A Criterion for Completeness of Casimir Operators", J. Math. Phys. **22**, 1851 (1981).

[2] J. M. Ekins and J. F. Cornwell, Rep. Math. Phys. **7**, 167 (1975).

[3] H. Freudenthal and H. de Vries, *Linear Lie Groups* (Academic, New York, 1969).

[4] N. Jacobson, *Lie Algebras* (Interscience, New York, 1966).

The analytical properties of the off-shell T matrix for infinite rank nonlocal separable potentials

C. Daskaloyannis
Department of Theoretical Physics, University of Thessaloniki, Thessaloniki, Greece

(Received 29 July 1980; accepted for publication 12 December 1980)

Analytical expression and uniform bounds for the l th partial wave off-shell T matrix are derived for infinite rank separable potentials. It is proved that Fredholm's alternative can be used to solve the Lipmann–Schwinger equation in some cases of noncompact nonlocal potentials in the strong L_p topology.

PACS numbers: 02.30.Rz, 03.80. + r

I. INTRODUCTION

The analytical properties of the two-body off-shell T matrix have been topics of intensive investigation and application in several branches of the nuclear physics.[1] As is well known the introduction of the nonlocal rank-N separable potentials simplifies considerably the calculations of the two- and three-body problem. The T matrix for rank-N separable potentials is studied by several authors.[2,3]

The problems related to the rank-N separable potentials can be solved with two methods. One is developed by K. Chadan[4,5] and the other is based on the manipulation of N-dimensional determinants.[2,3] Obviously these methods are not generalizable in the case of the nonlocal potentials expressed as infinite series of separable terms. In our work, these potentials are called infinite rank potentials.

In this paper a method of solution of the Lippmann–Schwinger (LS) equation is proposed. This method is applicable in the case of both finite- and infinite-rank separable potentials. The proposed solution does not require the compactness of the potential in the strong L_p-topology (the potential is not necessarily a Hilbert–Schmidt operator).

To preserve the analytical properties of the off-shell T matrix it has been necessary to study some properties of the intersection of the Lipschitz space[6–8] and the usual L_p-space. The necessary mathematical apparatus is given in the Appendix A. An extension of the notions of the product and the determinant introduced above as well as a generalization of Hadamard's lemma is introduced for these cases.

In Sec. II, the analytical expression of the l th wave off-shell T matrix is given and its uniform bounds are found. The restrictions imposed on the nonlocal potential are weaker than the compactness in the strong L_p-topology. This aim is obtained by the use of Fredholm's alternative[9,10] extended in some families of noncompact potentials. The proofs of the convergence of Fredholm's series in such cases are studied in Appendix B.

II. SOLUTION OF THE LS EQUATION

In this paper nonlocal potentials $\langle \mathbf{r}|V|\mathbf{r}'\rangle$ are used. The angle dependence of the potentials that are introduced here comes only through the angle between $\mathbf{r}$ and $\mathbf{r}'$. The coordinate representation of these potentials is given by the relation

$$\langle \mathbf{r}|V|\mathbf{r}'\rangle = \sum_{l=0}^{\infty} \sum_{|m|\leq l} (rr')^{-1} v_l(r,r') Y_l^m(\hat{\mathbf{r}}) Y_l^{m*}(\hat{\mathbf{r}}').$$

The free wavefunction is normalized as follows:

$$\int d\mathbf{r}\,\psi_o^*(\mathbf{k},\mathbf{r})\psi_o(\mathbf{k}',\mathbf{r}) = \delta(E-E')\delta_\Omega(\hat{\mathbf{k}},\hat{\mathbf{k}}'), \tag{1}$$

where $E = k^2$, δ_Ω is the solid angle δ function, $\hbar = 1$, and $2m = 1$.

The normalization (1) of the l th partial free wave function implies

$$\varphi_l(E,r) = r^{1/2} J_{l+1/2}(E^{1/2} r)/2^{1/2}.$$

The l th partial wave off-shell T matrix satisfies the off-shell LS equation[11]

$$T_l(E,E';z) = V_l(E,E') + \int_0^\infty dE''\, V_l(E,E'')T_l(E'',E';z)/(z-E''), \tag{2}$$

where

$$V_l(E,E') = \int_0^\infty dr \int_0^\infty dr'\, \varphi_l(E,r) v_l(r,r') \varphi_l(E',r).$$

The general form of the l th partial wave potential, in our normalization, is assumed to be expressed as a series of separable nonlocal potentials:

$$V(E,E') = (1/\pi) \sum_{n=1}^{\infty} f_n(E) g_n(E'). \tag{3}$$

In the last equation, as well as in the next, in our notation, the l index, which is related to the angular momentum, will be omitted for reasons of simplicity.

The functions f_n and g_n are assumed to satisfy the following restrictions:

$$\|f_n\| \leq F, \quad \text{and} \quad \sum_{n=1}^{\infty} \|g_n\|_{\mathrm{Lip}} = G < +\infty. \tag{4}$$

Here the norm $\|\cdot\|$ is defined as the sum of the norm $\|\cdot\|_{\mathrm{Lip}}$ of the Lipschitz space Lip (R_+, d^α)[6–8] and the usual norm $\|\cdot\|_p$ of the $L_p(R_+)$ space (R_+ is the positive real axis). Under the restrictions (4) the potential $V(E,E')$ given by Eq. (3) is not compact in the strong L_p-topology. The operator $V(E,E')$ given by Eq. (3) under the restrictions (4) is a nuclear operator[12] defined from the Banach space $L_p(R_+)$ into the Banach space Lip(R_+, d^α).

For a fixed E' the LS equation (2) can be studied as a Fredholm's second kind integral equation with kernel

$$N(E,E';z) = (1/\pi) \sum_{n=1}^{\infty} f_n(E) g_n(E')/(z-E'). \tag{5}$$

In Appendix B these integral equations, obeying the restrictions (4), are studied in detail; it is proved that the Fredholm's series converge. Therefore, the analytical solution of the LS equation is given by the equation

$$T(E,E',z) = V(E,E') + \frac{1}{\pi}\frac{\sum_{n,n'} f_n(E)\left[(1/\pi)\int_0^\infty ds T_{nn'}(s)/(z-s)\right]g_{n'}(E')}{1+(1/\pi)\int_0^\infty ds\rho(s)/(z-s)}, \quad (6)$$

where

$$\rho(s) = \sum_{m=1}^{\infty} \frac{(-1)^m}{m!} \sum_{n_1,n_2,\ldots,n_m} (\circ\text{-det}\{g_{n_i} f_{n_j}\})(s) \quad (7)$$

and

$$T_{nn'}(s) = \sum_{m=1}^{\infty} \frac{(-1)^m}{m!} \sum_{n_2,n_3,\ldots,n_m} \circ - \begin{Bmatrix} g_n g_{n'} & g_n f_{n_2} & g_n f_{n_3} & \cdots & g_n f_{n_m} \\ g_{n_2} f_{n'} & g_{n_2} f_{n_2} & g_{n_2} f_{n_3} & \cdots & g_{n_2} f_{n_m} \\ & & & & \\ g_{n_m} f_{f'} & g_{n_m} f_{n_2} & g_{n_m} f_{n_3} & \cdots & g_{n_m} f_{n_m} \end{Bmatrix}(s). \quad (8)$$

The $\circ$-determinant, which appears in the above equations, is a generalization of the usual determinant when the generalized $\circ$-product $g\circ f = \bar{g}f + g\bar{f}$ is used. The $\bar{g}$ and $\bar{f}$ are the Hilbert transforms[10,13] of the functions g and f respectively (for details see Appendices A and B). The functions $\rho(s)$ and $T_{nn'}(s)$ are well-defined functions in the space $X(R_+,\alpha,p)$.

The solution (6) of the LS equation can be formulated as follows:

$$T(E,E';z) = A(E,E';z)/D(z), \quad (9)$$

where

$$D(z) = 1 + (1/\pi)\int_0^\infty ds\, \rho(s)/(z-s). \quad (10)$$

The function $D(z)$ is a holomorphic function in the complex z plane with a cut on the positive real axis. The roots of this function correspond to the bound states of the problem. The function $A(E,E';z)$, for fixed E and E', is an holomorphic function with a cut on the real positive axis. So the off-shell T matrix is a meromorphic function with a cut on the real positive axis.

The uniform bounds of the function $D(z)$ are given in Appendix B:

$$||D(\cdot + i0) - 1|| < H_1(F{\cdot}G{\cdot}D(\alpha,p)), \quad (11)$$

$$||D(\cdot + iy) - 1|| < H_2(F{\cdot}G{\cdot}D(\alpha,p))/|y|, \quad (12)$$

where F and G are defined by (4), $D(\alpha,p)$ is a constant which depends only on α, and p and H_1, H_2 are entire functions defined by (B15) and (B16) in Appendix B.

The inequalities (11) and (12) involve the following asymptotic property:

$$\lim_{|z|\to\infty} D(z) = 1. \quad (13)$$

By the same way [see formula (B24) in Appendix B] the following bounds are found for the off-shell T matrix:

$$||D(\cdot + i0)[T(E,E';\cdot + i0) - V(E,E')]|| < F{\cdot}D^2(\alpha,p)\left(\sum_{n=1}^{\infty} f_n(E)|\,||g_n||_{\rm Lip}\right) \times\left(\sum_{n=1}^{\infty}|g_n(E')|\right)H_3(F{\cdot}G{\cdot}D(\alpha,p)), \quad (14)$$

where H_3 is an entire function defined by (B25) in Appendix B.

The inequality (A21), of Appendix A, imposes the bound

$$|D(x+iy)[T(E,E';z) - V(E,E')]| < \frac{F{\cdot}D^2(\alpha,p)}{|y|}\left(\sum_{n=1}^{\infty}|f_n(E)|{\cdot}||g_n||_{\rm Lip}\right) \times\left(\sum_{n=1}^{\infty}|g_n(E')|\right){\cdot}H_3(F{\cdot}G{\cdot}D(\alpha,p)). \quad (15)$$

The relations (14), (15), and (16) imply the following asymptotic property:

$$\lim_{|z|\to\infty} T_l(E,E';z) = V_l(E,E').$$

III. CONCLUSIONS

In this paper, Fredholm's alternative is used to solve the LS equation for infinite rank separable potentials. Also this method can be used to solve the LS equation when the potential is a finite rank separable one. In the last case all the determinants, where the rank is greater than the rank of the separable potential, are zero. Consequently the proposed method is a generalization of the known methods.

The analytical expression and the uniform bounds of the off-shell T matrix are given. The infinite rank T matrix has the same analytic properties as in the case of finite rank separable potentials.

Another result is that the infinite rank potential is not necessarily a compact operator in the strong L_p topology but it is a nuclear operator (therefore compact) from the L_p space into a Lipschitz space.

Osborn[14] and Levinger[15] have proved that for a local potential the off-shell T matrix is noncompact in the strong L_p topology. Therefore the convergence of any separable approximation for the T matrix from local potentials is impossible in this topology. In this paper the convergence of separable approximations for the T matrix for some families of noncompact potentials in the strong L_p topology is proved. This work is not sufficient to explain the sucessfulness of the separable approximations to the off-shell T matrix from a local potential because the imposed restrictions (4) are not satisfied for a Gaussian or a square-well potential, as it can be shown after extensive calculations.

APPENDIX A

In this appendix some properties of the functions in the space $X(R_+,\alpha,p)$ are studied. This space is defined as the following intersection:

$$X(R_+,\alpha,p) = {\rm Lip}(R_+,d^\alpha)\cap L_p(R_+). \quad (A1)$$

The space $\mathrm{Lip}(R_+,d^\alpha)$ is the well-known[6–8] Lipschitz space, d is the usual distance $d(x,y)=|x-y|$ and α is a positive number less than 1. This space is complete by the norm $||\cdot||_{\mathrm{Lip}}$ where

$$||f||_{\mathrm{Lip}}=||f||_\infty+||f||_{d^\alpha}, \tag{A2}$$

where

$$||f||_\infty=\max_{x\in R_+}(|f(x)|), \tag{A3}$$

$$||f||_{d^\alpha}=\sup_{x,y\in R_+}(|f(x)-f(y)|/d^\alpha(x,y)). \tag{A4}$$

The space $L_p(R_+)$, $p\geqslant 2$ is the space of real integrable functions with norm

$$||f||_p=\left[\int_0^\infty dx\,|f(x)|^p\right]^{1/p}. \tag{A5}$$

The space $X(R_+,\alpha,p)$ is a complete (Banach) space as the intersection of two complete spaces. In this space, the norm $||\cdot||$ is defined as follows:

$$||f||=||f||_{\mathrm{Lip}}+||f||_p. \tag{A6}$$

The following properties are necessary to study the space $X(R_+,\alpha,p)$.

Property 1: Let g be a function in $X(R_+,\alpha,p)$. Then

$$||g||_\infty\leqslant||g||_{d^\alpha}+||g||_p. \tag{A7}$$

Proof: Let g_m be the minimum of the function $|g(x)|$ when x belongs in the interval $I_m=[m,m+1]$, where $m=0,1,\cdots$. Then

$$g_m^p=g_m^p\int_m^{m+1}dx\leqslant\int_m^{m+1}dx\,|g(x)|^p\leqslant||g||_p.$$

Consequently

$$g_m\leqslant||g||_p.$$

For every $x\in I_m$ the following inequality holds:

$$|g(x)-g_m|\leqslant||g||_{d^\alpha},$$

as one can see from the definition (A4).

The above inequalities imply

$$|g(x)|\leqslant||g||_p+||g||_{d^\alpha}$$

for every m. Then the relation (A7) is true.

Property 2: Let g be a function which belongs in $\mathrm{Lip}(R_+,d^\alpha)$ and f a function in $X(R_+,\alpha,p)$. Then

$$||gf||\leqslant||g||_{\mathrm{Lip}}\cdot||f||<||g||\cdot||f||. \tag{A8}$$

Proof: The above inequality is a direct result of the inequalities

$$||gf||_p\leqslant||g||_\infty\,||f||_p,$$

and

$$||gf||_{\mathrm{Lip}}\leqslant||g||_{\mathrm{Lip}}\cdot||f||_{\mathrm{Lip}}.$$

In the next proposition we summarize some useful properties of the functions which belong in the $X(R_+,\alpha,p)$ space. The proofs are given by Titchmarch.[13]

Proposition 1: Let f be a function which belongs in the real $X(R_+,\alpha,p)$ space. For every $|y|\neq 0$ the function $\Phi(x+iy)$ is defined as follows:

$$\Phi(x+iy)=(1/\pi)\int_0^\infty dt f(t)/(z-t),$$

and

$$\Phi(x+i0)=\lim_{y\to 0^+}\Phi(x+iy).$$

Then

$$\text{(i)}\quad ||\Phi(\cdot+i0)||_p=\left[\int_{-\infty}^{+\infty}dx|\Phi(x+i0)|^p\right]^{1/p}<C_p||f||_p, \tag{A9}$$

where C_p is a constant which depends only on p (Ref. 13, Theorem 101, p. 132).

$$\text{(ii)}\quad ||\Phi(\cdot+i0)||_{d^\alpha}=\sup_{x,y\in R}(|\Phi(x+i0)-\Phi(y+i0)|/d^\alpha(x,y))<F_\alpha||f||_{d^\alpha}, \tag{A10}$$

where F_α is a constant which depends only on α (Ref. 13, Theorem 106, p. 145).

$$\text{(iii)}\quad ||\Phi(\cdot+iy)||_p<||f||_p/|y|. \tag{A11}$$

The last inequality is a result of the Hölder's inequality.[12] Property 1 [Eq. (A7)] is also valid for the complex $X(R_+,\alpha,p)$ space. Consequently the function $\Phi(x+i0)$ belongs in the complex $X(R_+,\alpha,p)$ space. The relations (A7), (A9), and (A10) imply the following inequality:

$$||f||<||\Phi(\cdot+i0)||<D(\alpha,p)||f||, \tag{A12}$$

where $D(\alpha,p)$ is a constant which depends only on α and p, and this constant is greater than 1.

The study of the integral equations is simplified considerably by the introduction of some kind of product.

Definition 1: Let φ_1 and φ_2 be functions in the real $X(R_+,\alpha,p)$ space, then the $\circ$-product of these functions is defined as follows:

$$(\varphi_1\circ\varphi_2)(x)=\varphi_1(x)\tilde{\varphi}_2(x)+\tilde{\varphi}_1(x)\varphi_2(x), \tag{A13}$$

where

$$\tilde{\varphi}_i(x)=\frac{1}{\pi}P\int_0^\infty dt\frac{\varphi_i(t)}{x-t},$$

is the Hilbert transform[10,13] of the function φ_i.

Property 2 [relation (A8)] implies that $\varphi_1\circ\varphi_2$ is a function in the $X(R_+,\alpha,p)$ space. By recurrence, the $\circ$-product of N functions can be defined

$$\circ\text{-}\prod_{i=1}^N\varphi_i=\left(\circ\text{-}\prod_{i=1}^{N-1}\varphi_i\right)\circ\varphi_N=\varphi_1\circ\varphi_2\circ\cdots\circ\varphi_N. \tag{A14}$$

The $\circ$-product is connected with the usual product by Proposition 2.

Proposition 2. If the functions $\varphi_1,\varphi_2,\ldots,\varphi_N$ belong in the real $X(R_+,\alpha,p)$ space, then:

$$\text{(i)}\quad \Phi(x+iy)=\prod_{k=1}^N\left(\frac{1}{\pi}\int_0^\infty dt\frac{\varphi_k(t)}{x+iy-t}\right)=(1/\pi)\int_0^\infty dt\left(\circ\text{-}\prod_{k=1}^N\varphi_k\right)(t)/(x+iy-t), \tag{A15}$$

and

$$\text{(ii)}\quad \|\circ\text{-}\prod_{k=1}^{N}\varphi_k\| < \|\Phi(\cdot + i0)\| < D^N(\alpha,p)\prod_{k=1}^{N}\|\varphi_k\|. \tag{A16}$$

The function $\Phi(z)$, $z = x + iy$, is holomorphic in the complex z plane with a cut on the positive real axis.

The proof of the relation (A15) is trivial in the case $N = 2$, if the imaginary parts of the successive members of (A15) are compared for $y \to 0^+$. The generalization by recurrence is obvious if the definition (A14) of the $\circ$-product is used. The relation (A16) is proved if the inequalities (A8) and (A12) are combined.

A natural generalization of Definition 1 is the introduction of the $\circ$-determinant.

Definition 2: If the functions φ_{ij}, where $1 \leqslant i \leqslant N$, $1 \leqslant j \leqslant N$, belong in the real $X(R_+,\alpha,p)$ space, then the $\circ$-determinant of these functions is defined as follows:

$$\circ\text{-det}\{\varphi_{ij}\} = \circ\text{-}\begin{Bmatrix}\varphi_{11}\varphi_{12}\cdots\varphi_{1N}\\ \varphi_{21}\varphi_{22}\cdots\varphi_{2N}\\ \vdots \qquad \vdots\\ \varphi_N\varphi_{N2}\cdots\varphi_{NN}\end{Bmatrix}$$

$$= \sum(-)^P\varphi_{k_1}\circ\varphi_{2k_2}\circ\cdots\circ\varphi_{k_N}; \tag{A17}$$

the sum is extended in all transpositions

$$\begin{pmatrix}1 & 2 & \cdots & N\\ k_1 & k_2 & \cdots & k_N\end{pmatrix}$$

and P is the parity of this transposition.

The $\circ$-determinant has all the linear properties of the ordinary determinants, that is to say, it is zero if some column (row) is the linear combination of others columns (rows). A direct result of the Definition 2 is the generalization of Proposition 2.

Proposition 3: If the functions φ_{ij}, where $1 \leqslant i \leqslant N$ and $1 \leqslant j \leqslant N$, belong in the real $X(R_+,\alpha,p)$ space, with $p \geqslant 2$, then

$$\text{(i)}\quad \Phi(x+iy) = \det\left((1/\pi)\int_0^\infty dt\,\varphi_{ij}(t)/(x+iy-t)\right)$$

$$= (1/\pi)\int_0^\infty dt\,(\circ\text{-det}\{\varphi_{ij}\})(t)/(x+iy-t), \tag{A18}$$

$$\text{(ii)}\quad \|\circ\text{-det}\{\varphi_{ij}\}\|_p < \|\Phi(\cdot + i0)\|_p < [\mathrm{D}(\alpha,p)]^N \prod_{i=1}^{N}\left(\sum_{j=1}^{N}\|\varphi_{ij}\|^2\right)^{1/2}, \tag{A19}$$

$$\text{(iii)}\quad \|\circ\text{-det}\{\varphi_{ij}\}\|_{d^\alpha} < \|\Phi(\cdot + i0)\|_{d^\alpha} < [\mathrm{D}(\alpha,p)]^N \prod_{i=1}^{N}\left(\sum_{j=1}^{N}\|\varphi_{ij}\|^2\right)^{1/2}, \tag{A20}$$

$$\text{(iv)}\quad \|\Phi(\cdot + iy)\|_\infty < \frac{[D(\alpha,p)]^N}{|y|}\prod_{i=1}^{N}\left(\sum_{j=1}^{N}\|\varphi_{ij}\|^2\right)^{1/2}, \tag{A21}$$

$$\text{(v)}\quad \|\circ\text{-det}\{\varphi_{ij}\}\| < (N+2)[D(\alpha,p)]^N \prod_{i=1}^{N}\left(\sum_{j=1}^{N}\|\varphi_{ij}\|^2\right)^{1/2}. \tag{A22}$$

The last inequality is a generalization of Hadamard's lemma in the case of the $\circ$-determinants.

For simplicity, we define the functions $\Phi_{ij}(x+iy)$ as follows:

$$\Phi_{ij}(x+iy) = (1/\pi)\int_0^\infty dt\,\varphi_{ij}(t)/(x+iy-t).$$

Hadamard's lemma applied in the second term of Eq. (A18) gives

$$|\Phi(x+i0)|^2 < \left(\sum_{j=1}^{N}\Phi_{1j}(x+i0)|^2\right) \times \prod_{i=2}^{N}\left(\sum_{j=1}^{N}\|\Phi_{ij}(\cdot + i0)\|_\infty{}^2\right).$$

If $p \geqslant 2$, the functions $|\Phi(x+i0)|^2$ and $|\Phi_{1j}(x+i0)|^2$ belong in the space $L_{p/2}(R)$. The Minkowski inequality[12] gives

$$\|\Phi(\cdot + i0)\|_p^2 < \left(\sum_{j=1}^{N}\|\Phi_{1j}(\cdot + i0)\|_p^2\right) \times \prod_{i=2}^{N}\left(\sum_{j=1}^{N}\|\Phi_{ij}(\cdot + i0)\|_\infty{}^2\right).$$

The inequality (A19) is an obvious result of the last relation.

From the definition of the equation $\Phi(x+i0)$ it is clear that

$$\begin{aligned}
&\Phi(x+i0) - \Phi(x'+i0)\\
&= \sum(-)^P(\Phi_{1k_1}(x+i0) - \Phi_{1k_1}(x'+i0))\\
&\quad\times\Phi_{2k_2}(x+i0)\cdots\Phi_{Nk_N}(x+i0)\\
&\quad+\sum(-)^P\Phi_{+k_1}(x'+i0)(\Phi_{2k_2}(x+i0) - \Phi_{2k_2}(x'+i0))\\
&\quad\times\Phi_{3k_3}(x+i0)\cdots\Phi_{Nk_n}(x+i0) + \cdots\\
&\quad+\sum(-)^P\Phi_{1k_1}(x'+i0)\Phi_{2k_2}(x'+i0)\cdots\\
&\quad\times(\Phi_{nk_N}(x+i0) - \Phi_{nk_N}(x'+i0)).
\end{aligned}$$

The difference $\Phi(x+i0) - \Phi(x'+i0)$ can be understood as a sum of determinants. If we divide the two parts of the last equality by $|x-x'|^\alpha$ and Hadamard's lemma is applied for each determinant, then we find

$$\|\Phi(\cdot + i0)\|_{d^\alpha} < \sum_{m=1}^{N}\left[\left(\sum_{j=1}^{N}\|\Phi_{mj}(\cdot + i0)\|_{d^\alpha}^2\right)\prod_{\substack{i=1\\ i\neq m}}^{N}\left(\sum_{j=1}^{N}\|\Phi_{ij}(\cdot i0)\|_\infty^2\right)\right]^{1/2}$$

The inequality (A20) is a trivial result of the last relation. Property 1 [relation (A7)] and the inequalities (A19) and (A20) imply the relation (A22). Relation (A21) is a result of the inequalities (A11) and (A19).

APPENDIX B

In this appendix the following integral equation will be studied:

$$\varphi(u;z) = f(u) + \lambda\sum_{n=1}^{\infty}f_n(u)\frac{1}{\pi}\int_0^\infty dt\,\frac{g_n(t)\varphi(t;z)}{z-t}, \tag{B1}$$

where z is a complex number.

The function f_n and g_n are assumed to obey the restrictions

$$||f_n|| < F \quad \text{for every } n \tag{B2}$$

and

$$G = \sum_{n=1}^{\infty} ||g_n||_{\text{Lip}} < +\infty. \tag{B3}$$

In this section we will prove that the sums of both Fredholm's series[9,10] converge in this case.

The kernel of the integral equation is

$$N(u,t;z) = \frac{1}{\pi} \sum_{n=1}^{\infty} \frac{f_n(u) g_n(t)}{z-t}. \tag{B4}$$

The first Fredholm's series is given by the formula

$$D(\lambda;z) = 1 + \sum_{m=1}^{\infty} \frac{(-\lambda)^m}{m!} c_m(z), \tag{B5}$$

where

$$c_m(z) = \left(\frac{1}{m}\right)^m \int_0^\infty ds_1 \int_0^\infty ds_2 \cdots \int_0^\infty ds_m$$

$$\times \sum_{n_1,n_2,\ldots,n_m} \begin{vmatrix} f_{n_1}(s_1) g_{n_1}(s_1)/(z-s_1) & f_{n_2}(s_1) g_{n_2}(s_2)/(z-s_2) \cdots f_{n_m}(s_1) g_{n_m}(s_m)/(z-s_m) \\ f_{n_1}(s_2) g_{n_1}(s_1)/(z-s_1) & f_{n_2}(s_2) g_{n_2}(s_2)/(z-s_2) \cdots f_{n_m}(s_2) g_{n_m}(s_m)/(z-s_m) \\ \vdots & \vdots \qquad\qquad \vdots \\ f_{n_1}(s_m) g_{n_1}(s_1)/(z-s_1) & f_{n_2}(s_m) g_{n_2}(s_2)/(z-s_2) \cdots f_{n_m}(s_m) g_{n_m}(s_m)/(z-s_m) \end{vmatrix}. \tag{B6}$$

For any determinant in Eq. (B6), all the elements of the jth column are multiplied by

$$g_n(s_j)/(z-s_j). \tag{B7}$$

Therefore all these factors could be placed outside of each determinant as multiplication factors. After this manipulation the jth row can be multiplied by the factor (B7) and finally the formula (B6) is transformed as follows:

$$c_m(z) = \sum_{n_1,n_2,\ldots,n_m} \det\left[\frac{1}{\pi}\int_0^\infty dt\, \frac{g_{n_i}(t) f_{n_j}(t)}{z-t}\right]. \tag{B8}$$

The formulas (B6) and (B8) are valid when the series are finite sums. In the following we will show that the restrictions (B2) and (B3) are sufficient for the convergence of the infinite series (B8).

The equation (A18) of Appendix A implies the following equation:

$$c_m(z) = \sum_{n_1,n_2,\ldots n_m} \frac{1}{\pi} \int_0^\infty dt \frac{(\circ\text{-det}\{g_n f_{n_j}\})(t)}{z-t}. \tag{B9}$$

The extension of Hadamard's lemma, Eq. (A22), gives the following results:

$$||c_m(\cdot + i0)|| < (m+2)[D(\alpha,p)]^m \times \sum_{n_1,n_2,\ldots n_m} \left[\prod_{i=1}^m \left(\sum_{j=1}^m ||g_n f_{n_j}||^2\right)^{1/2}\right], \tag{B10}$$

and

$$||c_m(\cdot + i0)||_p < \frac{[D(\alpha,p)]^m}{|y|} \times \sum_{n_1,n_2,\ldots n_m} \left[\prod_{i=1}^m \left(\sum_{j=1}^m ||g_{n_i} f_{n_j}||^2\right)^{1/2}\right]. \tag{B11}$$

Formula (A8), in Appendix A, implies

$$||g_{n_i} f_{n_j}|| < ||g_{n_i}||_{\text{Lip}} ||f_{n_j}||. \tag{B12}$$

Consequently, the first Fredholm's series, (B5), are bounded by

$$||D(\lambda,\cdot + i0)|| \leqslant H_1(|\lambda| FGD(\alpha,p)), \tag{B13}$$

and

$$||D(\lambda,\cdot + i0)|| \leqslant (1/|y|) H_2(|\lambda| FGD(\alpha,p)). \tag{B14}$$

The functions $H_1(z)$ and $H_2(z)$ are entire functions in the complex z plane, defined as follows:

$$H_1(z) = \sum_{m=1}^{\infty} \frac{(m+2) m^{m/2}}{m!} z^m, \tag{B15}$$

and

$$H_2(z) = \sum_{m=1}^{\infty} \frac{m^{m/2}}{m!} z^m. \tag{B16}$$

The space $X(R_+,\alpha,p)$ is a Banach (complete) space. This fact and the inequality (B13) imply the existence of some function $\rho(\lambda;s)$ such that

$$D(\lambda;z) = 1 + \frac{1}{\pi}\int_0^\infty dt \frac{\rho(\lambda;s)}{z-t}, \tag{B17}$$

where

$$\rho(\lambda;s) = \sum_{m=1}^{\infty} \frac{(-\lambda)^m}{m!} \sum_{n_1,n_2,\ldots n_m} (\circ\text{-det}\{g_{n_i} f_{n_j}\})(s). \tag{B18}$$

The second Fredholm's series is defined by the formula

$$D(u,t;\lambda;z) = N(u,t;z) + \sum_{m=1}^{\infty} \frac{(-\lambda)^m}{m!} c_m(u,t;z), \tag{B19}$$

where $N(u,t;z)$ is the kernel (B4) and the functions $c_m(u,t;z)$ take the following form:

$$c_m(u,t;z) = \frac{1}{\pi}\sum_n \frac{f_n(u) g_n(t) c_m(z)}{z-t} - m\frac{1}{\pi}\sum_{n,n'} f_n(u)\left[\frac{1}{\pi}\int_0^\infty ds\, \frac{t_{nn'}^{(m)}(s)}{z-s}\right]\frac{g_{n'}(t)}{z-t}. \tag{B20}$$

The functions $t_{nn'}^{(m)}$ in the second part of Eq. (B20) are defined as follows:

$$t_{nn'}^{(m)}(s) = \sum_{n_2,n_3,\ldots n_m} \circ\text{-} \left\{ \begin{matrix} g_n f_{n'} & g_n f_{n_2} & \ldots g_n f_{n_m} \\ g_{n_2} f_{n'} & g_{n_2} f_{n_2} & \ldots g_{n_2} f_{n_m} \\ g_{n_3} f_{n'} & g_{n_3} f_{n_2} & \cdots g_{n_3} f_{n_m} \\ & & \\ & & \\ g_{n_m} f_{n'} & g_{n_m} f_{n_2} & \cdots g_{n_m} f_{n_m} \end{matrix} \right\}(s). \tag{B21}$$

The extended Hadamard's lemma [see (A22)] when it is applied to the second part of (B21) gives the following inequality:

$$\| t_{nn'}^{(m)} \| < (m+2) m^{m/2} (F \cdot D(\alpha,p))^m G^{m-1} \| g_n \|,$$

where F and G are defined by the restrictions (B2) and (B3). The above inequality and Eqs. (B20) and (B21) imply the following relation:

$$D(u,t;\lambda;z) = N(u,t;z) D(\lambda;z) + \frac{1}{\pi} \sum_{n,n'} f_n(u) \left[\frac{1}{\pi} \int_0^\infty ds \frac{T_{n'}(s;\lambda)}{z-s} \right] \frac{g_{n'}(t)}{z-t}, \tag{B22}$$

where $D(\lambda;z)$ is the first Fredholm's series given by Eq. (B17) and

$$T_{nn'}(s;\lambda) = \sum_{m=1}^{\infty} \frac{(-\lambda)^m}{(m-1)!} t_{nn'}^{(m)}(s). \tag{B23}$$

The bounds of this function are given by

$$\| T_{nn'}(\cdot,\lambda) \| < |\lambda| F \cdot D(\alpha,p) H_3(|\lambda| F \cdot G \cdot D(\alpha,p)) \cdot \| g_n \|_{\mathrm{Lip}}, \tag{B24}$$

where the function $H_3(z)$ is an entire function in the z plane such that

$$H_3(z) = H_1'(z). \tag{B25}$$

The convergence of the second Fredholm's series implies that Fredholm's alternative is applicable to the case of the integral equation (B1), when the restrictions (B2) and (B3) are satisfied.

[1] M. K. Srivastava and D. W. Sprung, Adv. Nucl. Phys. **8**, 121 (1975).
[2] D. Gutowski and A. Scalia, J. Math. Phys. **9**, 588 (1968) and **10**, 2306 (1969).
[3] B. Talukdar, U. Das, and S. Mukhopadhyaya, J. Math. Phys. **20**, 372 (1979).
[4] K. Chadan, Nuovo Cimento **10**, 892 (1958) and Nuovo Cimento A **47**, 510 (1967).
[5] M. G. Fuda, Phys. Rev. C **1**, 1910 (1970).
[6] K. de Leeuw, Studia Mathematica **21**, 55 (1961/62).
[7] D. R. Shebert, Trans. Amer. Math. Soc. **111**, 240 (1964).
[8] J. A. Johnson, Trans. Amer. Math. Soc. **148**, 147 (1970).
[9] W. Pogorzelski, *Integral Equations and their Applications* (Pergamon Press-PWN-Polish Scientific Publishers, Warsawa, 1966).
[10] F. G. Tricomi, *Integral Equations* (Interscience, New York, 1970).
[11] R. G. Newton, *Scattering Theory of Waves and Particles* (McGraw-Hill, New York, 1966).
[12] K. Yoshida, *Functional Analysis* (Springer, New York, 1971).
[13] E. C. Titchmarch, *Introduction to the Theory of Fourier Integrals* (Oxford University, Oxford, 1937).
[14] T. A. Osborn, J. Math. Phys. **14**, 373, 1485 (1973).
[15] J. S. Levinger, J. Math. Phys. **14**, 1314 (1973).

On the stability of periodic orbits of two-dimensional mappings

Tassos Bountis
Department of Mathematics and Computer Science, Clarkson College of Technology, Potsdam, New York 13676

Robert H. G. Helleman
Twente University of Technology, Theoretical Physics, P. O. Box 217, 7500 AE Enschede, The Netherlands and The La Jolla Institute, P. O. Box 1434, La Jolla, California 92038

(Received 15 October 1980; accepted for publication 27 February 1981)

We present a closed form stability criterion for the periodic orbits of two-dimensional conservative as well as "dissipative" mappings which are analogous to the Poincaré maps of dynamical systems. Our stability criterion has a particularly simple form involving a finite, symmetric, nearly tridiagonal determinant. Its derivation is based on an extension of the stability analysis of Hill's differential equation to difference equations. We apply our criterion and derive a sufficient stability condition for a large class of periodic orbits of the widely studied "standard mapping" describing a periodically "kicked" free rotator. As another example we also obtain explicitly and in closed form the intervals of bounded (and unbounded) solutions of a discrete "Schrödinger equation" for the Kronig and Penney crystal model.

PACS numbers: 02.30.Sa

I. INTRODUCTION

In the study of dynamical systems mappings of the plane onto itself play a significant role[1-9]: several properties of two-degree-of-freedom Hamiltonian systems have been determined by considering only the intersection points of the orbits with a two-dimensional "section" of the (three-dimensional) energy surface, often called a "surface of section" or a Poincaré map[1-6] (cf. Fig. 1).

In this paper we present analytical results on the stability of periodic orbits of two dimensional mapppings T:

$$T:\begin{Bmatrix} x_{t+1} = f(x_t,y_t) \\ y_{t+1} = g(x_t,y_t) \end{Bmatrix}, \quad t = 0,1,2,\cdots. \tag{1.1}$$

Such mappings can model conservative or dissipative dynamical systems depending on whether T is "area-preserving" or "area-contracting" respectively.[7,10] As usual, in dynamical systems the most innocent-looking nonlinear functions f,g (e.g., x_t^2, y_t^2, etc.) may render (1.1) "nonintegrable", i.e., its general solution cannot be obtained in closed form or as convergent series.[1]

Our main result is a closed form stability criterion for periodic orbits of mappings of the type (1.1). Periodic orbits of arbitrary (integer) period m are repeating sequences of points (x_t,y_t) in the x,y plane, i.e.,

$$(x_{t+m},y_{t+m}) = T^m(x_t,y_t) = (x_t,y_t), \tag{1.2}$$

$t = 0,1,2,\cdots$. We consider the variational equations about a given m-periodic solution of (1.1) and ask whether the solutions of these variational equations are bounded or not (for all t). This identifies the given periodic solution as *stable* or unstable. Since the variational equations are linear, the (un)-boundedness of their solutions is decided by the value of the determinant of an $m \times m$ matrix $\mathbf{H}(0)$, which depends on the coordinates of the given m-periodic orbit, cf. Sec. IIA. For every such orbit we derive

$$|2 + \det \underset{=}{\mathbf{H}}(0)| \begin{Bmatrix} <2: & \text{stability} \\ >2: & \text{instability} \end{Bmatrix}, \tag{1.3}$$

in the area-preserving case, cf. Sec. IIB. A similar result is derived in Sec. IV for area-contracting mappings, which model dissipative systems.[10]

The main advantage of the closed expression (1.3) over the usual procedures[7-9] is that it can yield more globar stability results of which there is great need. Examples of such results are given in Sec. III and Appendix A.

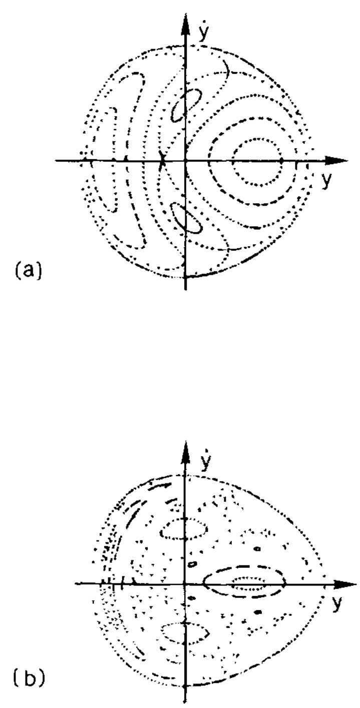


FIG. 1. Surfaces of section for the 2-D-system of Ref. 2. (a) $E = 1/24$. (b) $E = 1/8$. Note the "dotted" chaotic regions in (b) where solutions depend very sensitively on initial conditions.

 0022-2488/81/091867-11$01.00

The study of "nonintegrable" systems has revealed that there are large classes of orbits which depend *singularly* on their initial conditions and exhibit a highly "chaotic" and "irregular" behavior.[1–9,11] In fact, this dependence on initial conditions is so sensitive that some properties of certain solutions can be proven to be truly random.[6,12] The situation is not unlike a throw of dice where the outcome depends critically on how the sharp ("singular") edges of the dice hit the table, etc.—a deterministic process resulting in a "random" sequence of outcomes.

Numerical investigations and a number of rigorous results indicate that there exist two distinct types of motion[1–9]: *Regular* or ordered motion, represented in Fig. 1(a) and 2(a) by the concentric "curves" or "islands" and *irregular* or *chaotic* motion, represented by regions where the intersections of many orbits "scatter" about in a seemingly random and area-filling fashion as in Figs. 1(b) and 2(b).

These regions of chaotic behavior are of interest to a number of problems in physics: They may be useful in establishing the ergodic hypothesis[12–14] at least for some nonintegrable systems, i.e., that almost all orbits cover the energy surface densely and uniformly. The presence (or absence) of large chaotic regions is also crucial to long-term stability of planetary orbits in celestial mechanics,[1–3] the confinement of charged particles in fusion devices[15–17] and high energy accelerators,[18–19] the dynamics of molecular dissociation[20,21] and other areas of current research.

The onset of large scale chaotic behavior is associated with more and more *stable* periodic orbits turning unstable as one varies the values of the parameters of the problem.[1,6–9,11] Thus, it is important to determine analytically all

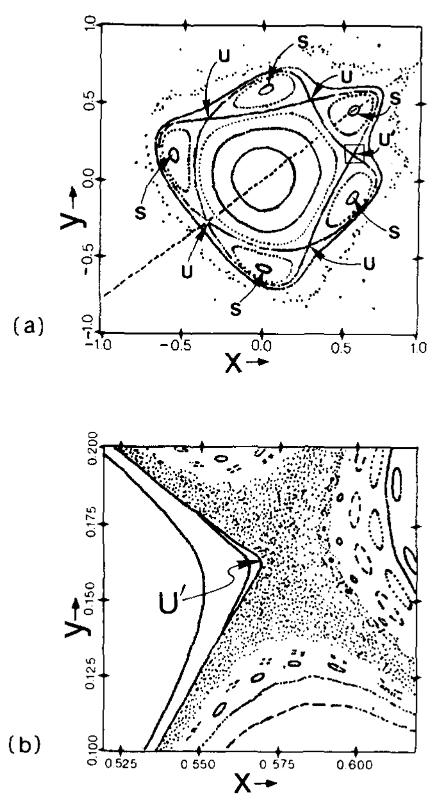


FIG. 2. Iterates of the mapping (2.1) for $\cos\alpha = 0.24$. (a) Observe the two 5-periodic orbits, a stable one (S) and an unstable one (U). (b) A magnification of the chaotic region about the point U' of (a), cf. Ref. 7.

ranges of parameter values corresponding to stable vs. unstable behavior. As we demonstrate in Sec. III on the so-called standard mapping,[9,22] we can now begin to do this with the aid of criterion (1.3).

We have extended these results to dissipative, or area-contracting mappings, cf. Ref. 1 (1981). Large scale chaotic behavior in such mappings is related to the presence of so-called "strange attractors" and the onset of "turbulent" motion.[16,10,23] In Sec. IV we derive a stability criterion analogous to (1.3) for dissipative mappings and discuss its dependence on a "damping" parameter b ($|b| < 1$). Finally, following the approach of Sec. III, we apply this criterion to a dissipative form of the standard mapping and obtain a sufficient stability criterion for its periodic orbits.

II. STABILITY CRITERIA FOR AREA-PRESERVING MAPPINGS

A. Earlier methods

We consider here the quadratic mapping

$$T:\begin{cases} x_{t+1} = x_t\cos\alpha - (y_t - x_t^2)\sin\alpha, \\ y_{t+1} = x_t\sin\alpha + (y_t - x_t^2)\cos\alpha \end{cases}, \tag{2.1}$$

$t = 0,1,2,\cdots$, due to Hénon.[7] Using this example we first review the usual stability analysis[7–9] of periodic orbits and point out its limitations. A new stability criterion is derived in Sec. IIB.

The quadratic mapping (2.1) is the simplest nontrivial mapping[7] which exhibits many of the interesting features of a nonintegrable Hamiltonian system, cf. Fig. 1 and 2. Eliminating y_t between Eqs. (2.1), we obtain the single second-difference equation

$$x_{t+1} + x_{t-1} - 2x_t\cos\alpha - x_t^2\sin\alpha = 0. \tag{2.2}$$

Orbits are obtained in the x_t, x_{t-1} plane by substituting x_t, x_{t-1} in (2.2) starting with some x_1, x_0 and solving for x_{t+1}, etc.

We consider small variations Δx_t about a given m-periodic orbit $\{\hat{x}_t = \hat{x}_{t+m}\}$ setting in (2.2) $x_t = \hat{x}_t + \Delta x_t$ and keep only first order terms in Δx_t to find

$$\Delta x_{t+1} + \Delta x_{t-1} - 2(\cos\alpha + \hat{x}_t\sin\alpha)\Delta x_t = 0. \tag{2.3}$$

In vector form this variational equation becomes

$$\begin{pmatrix}\Delta x_{t+1} \\ \Delta x_t\end{pmatrix} = \begin{pmatrix} 2(\cos\alpha + \hat{x}_t\sin\alpha) & -1 \\ 1 & 0\end{pmatrix}\begin{pmatrix}\Delta x_t \\ \Delta x_{t-1}\end{pmatrix}, \tag{2.4a}$$

or

$$\mathbf{\Delta x}_{t+1} = \underset{\approx}{\mathbf{M}}_t \mathbf{\Delta x}_t. \tag{2.4b}$$

Note that $\det\underset{\approx}{\mathbf{M}}_t = 1$, i.e., (2.1) is area-preserving indeed.[7] Choosing some $\mathbf{\Delta x}_1 = (\Delta x_1, \Delta x_0)$ and iterating (2.1) m times one calculates upon return near $(\hat{x}_1, \hat{x}_0)$ the resulting variation

$$\mathbf{\Delta x}_{m+1} = \underset{\approx}{\mathbf{M}}\mathbf{\Delta x}_1, \tag{2.5}$$

where $\underset{\approx}{\mathbf{M}}$ is the product of m 2×2 matrices:

$$\underset{\approx}{\mathbf{M}} \equiv \prod_{t=1}^{m} \underset{\approx}{\mathbf{M}}_t. \tag{2.6}$$

In this analysis one computes the eigenvalues λ_1, λ_2 of the matrix $\underset{\approx}{\mathbf{M}}$ and distinguishes two cases:

(a) $|\mathrm{Tr}\,\underset{\approx}{\mathbf{M}}| < 2$, whence $\lambda_1 = \lambda_2^*$ complex with

$|\lambda_1| = |\lambda_2| = 1$; therefore, $\Delta\mathbf{x}_{t+km}$ rotates with every mth return (remaining bounded) and $\hat{x}_t$ is stable or "elliptic", e.g., the orbit marked S in Fig. 2(a).

(b) $|\mathrm{Tr}\mathbf{M}| > 2$ whence λ_1, λ_2 real with, say, $|\lambda_1| < 1$ and $|\lambda_2| = |\lambda_1|^{-1} > 1$. For almost all $\Delta\mathbf{x}_1$, $\Delta\mathbf{x}_{t+km}$ becomes unbounded and $\hat{x}_t$ is unstable or "hyperbolic", e.g., the orbit marked μ in Fig. 2(a). [The highly exceptional case $|\mathrm{Tr}\,\mathbf{M}| = 2$ implies $\lambda_1 = \lambda_2 = \pm 1$ and $\hat{x}_t$ is referred to as marginally (un)stable or "parabolic".]

The familiar method outlined above requires cumbersome algebraic or numerical computations. Furthermore, its results are often not transparent and cannot be easily generalized to *families* of orbits. It amounts, in fact, to an expansion of the determinant in our criterion [see (2.19) and the comments below it] and does not offer as easily new analytic insight.

B. A closed form stability criterion

Our result has an analog in differential equations with which the reader may be more familiar: Consider a π-periodic solution of the differential equation

$$\frac{d^2x}{dt^2} + F(x) = 0. \tag{2.7}$$

i.e., an $\hat{x}(t)$ with $\hat{x}(t) = \hat{x}(t+\pi)$, whose stability type we wish to determine. Letting $x = \hat{x} + \xi$ in (2.7), we find to first order in ξ

$$\frac{d^2\xi}{dt^2} + Q(t)\xi(t) = 0, \tag{2.8}$$

where

$$Q(t) = \frac{dF}{dx}\Big|_{x=\hat{x}} = Q(t+\pi). \tag{2.9}$$

Clearly the boundedness properties of the solutions $\xi(t)$ of (2.8) will determine whether nor not $\hat{x}(t)$ is (linearly) stable under small perturbations.[24,25]

Similarly, for our mapping (2.2), we obtained the second order linear *difference* equation (2.5),

$$-\Delta x_{t+1} - \Delta x_{t-1} + d_t \Delta x_t = 0, \tag{2.10}$$

with

$$d_t \equiv 2(\cos\alpha + \hat{x}_t \sin\alpha) = d_{t+m} \tag{2.11}$$

in analogy with (2.8) and (2.9).

Floquet's theorem[24,25] establishes that Hill's equation (2.10–2.11) possesses two linearly independent solutions of the form

$$\xi_+(t) = e^{i\beta t}P(t), \quad \xi_-(t) = e^{-i\beta t}P^*(t), \tag{2.12}$$

in general, where $P(t)$ is also π-periodic (and hence bounded) as is $Q(t)$, i.e., $P(t) = P(t+\pi)$, cf. (2.9). The so-called Floquet characteristic exponent β is the important quantity here since, as we see from Eq. (2.12), it is the value of β real versus complex which determines whether the solutions of (2.10) are bounded or not.

The extension of the Floquet theory to difference equations[26] establishes that Eq. (2.10) similarly has two linearly independent solutions of the form

$$\Delta x_t^{(+)} = e^{i\beta t}P_t, \quad \Delta x_t^{(-)} = e^{-i\beta t}P_t^*, \tag{2.13}$$

with $P_t = P_{t+m}$. Again the stability type of the periodic orbit $\hat{x}_t$ depends on the value of the Floquet exponent β.

Inserting $\Delta x_t^{(+)}$ from (2.13) in (2.10), we find

$$-e^{+i\beta}P_{t+1} - e^{-i\beta}P_{t-1} + d_t P_t = 0, \tag{2.14}$$

or in vector form,

$$\underset{\approx}{\mathbf{H}}(\beta)\mathbf{P} = \mathbf{0}, \tag{2.15}$$

where

$$\underset{\approx}{\mathbf{H}}(\beta) \equiv \begin{pmatrix} d_1 & -e^{i\beta} & 0 & 0 & \cdots & e^{-i\beta} \\ -e^{-i\beta} & d_2 & -e^{i\beta} & & & \\ & & & & & \cdot \\ 0 & -e^{-i\beta} & d_3 & \cdot & & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & & \cdot & & & 0 \\ \cdot & & \cdot & \cdot & \cdot & \\ & & & & \cdot & -e^{i\beta} \\ e^{i\beta} & & \cdots & 0 & -e^{-i\beta} & d_m \end{pmatrix} \tag{2.16}$$

and $\mathbf{P} \equiv \mathrm{col}(P_1, P_2, \cdots, P_m)$. In order for (2.15) to have a nontrivial solution,

$$\det\underset{\approx}{\mathbf{H}}(\beta) = 0. \tag{2.17a}$$

From this we can explicitly solve for β: Multiplying the first row of $\mathbf{H}(\beta)$ by $\exp(i\beta)$ and the first column by $\exp(-i\beta)$ leaves $\det\underset{\approx}{\mathbf{H}}(\beta)$ invariant. Similarly, multiplying the second row by $\exp(2i\beta)$ and the second column by $\exp(-2i\beta)$, etc., finally gives

$$\det\begin{pmatrix} d_1 & -1 & 0 & \cdots & & & -e^{-im\beta} \\ -1 & d_2 & & \cdots & & 0 & 0 \\ 0 & & & \cdot & \cdot & & \cdot \\ \cdot & \cdot & \cdot & & \cdot & \cdot & \cdot \\ \cdot & & & \cdot & \cdot & \cdot & \cdot \\ \cdot & 0 & & \cdot & \cdot & & -1 \\ & & \cdots & \cdot & & & \\ -e^{+im\beta} & & & & 0 & -1 & d_m \end{pmatrix} = 0. \tag{2.17b}$$

Expanding this determinant with respect to its first and last columns yields the Floquet exponent explicitly:

$$2\cos(m\beta) = 2 + \det\underset{\approx}{\mathbf{H}}(0), \tag{2.18a}$$

where $\underset{\approx}{\mathbf{H}}(0)$ is particularly simple, cf. (2.16).

Equation (2.18a) allows us to determine whether the β in (2.13) is real or imaginary. Hence, from (2.13) and the discussion below (2.6) we conclude that the periodic orbit $\hat{x}_t$ is

$$|2+\det \underset{\approx}{\mathbf{H}}(0)| \begin{cases} <2: & \text{stable or "elliptic,"} \\ >2 & \text{unstable or "hyperbolic,"} \\ =2: & \text{marginally (un)stable, "parabolic."} \end{cases} \qquad (2.18\text{b})$$

Our stability cirtierion (2.18) is in closed form and thus lends itself more easily to analytic manipulations: It was recently used by Greene[9] to obtain results for the so-called standard mapping.[22] Applying (2.18) to this same mapping, we derive additional results in the next section. In Appendix A we use (2.18) to obtain the stability intervals for a Hill's difference equation driven by a pulse-shaped periodic function.

The connection between our criterion and that of Sec. IIA is given by the simple equation

$$2+\det \underset{\approx}{\mathbf{H}}(0)=\mathrm{Tr}(\underset{\approx}{\mathbf{M}}), \qquad (2.19)$$

where $\underset{\approx}{\mathbf{M}}$ was defined in (2.6). Equation (2.19) was derived in Appendix B of Ref. 9 cf. also Ref. 7 (1978). Since the $\det \underset{\approx}{\mathbf{H}}(0)$ can be written as the difference of two tridiagonal determinants [see Appendix A, Eq. (A1)] its expansion in terms of the elements of $\underset{\approx}{\mathbf{H}}(0)$ can be written down, cf. Eq. (4.13). This provides us with an expansion for the trace of a product of 2×2 matrices which is of interest to various important problems of physics, notably the 2-D Ising model in the presence of magnetic field. Recently,[27] in that connection, the trace of such a product was given as the integral

$$\mathrm{Tr}(\underset{\approx}{\mathbf{M}})=\lim_{S\to\infty}\frac{1}{S}\int_0^S ds\prod_{t=1}^{m}(-ie^{-ig_t s}+d_{t+1}-ie^{ig_{t+1}s}), \qquad (2.20)$$

where $g_t=\log P_t$, P_t being the t th prime number.

III. SUFFICIENT STABILITY CONDITION

With the aid of (2.18) we obtain below a sufficient stability condition for periodic orbits of the so-called standard mapping[22,18,9]

$$T_K:\begin{Bmatrix} r_{t+1}=r_t-(K/2\pi)\sin 2\pi\theta_t \\ \theta_{t+1}=\theta_t+r_{t+1} \end{Bmatrix}, \quad t=0,1,2,\cdots, \qquad (3.1)$$

which models the motion of charged particles in toroidal magnetic fields[18] and is also used in the study of nonlinear resonances.[22] Our result is: If an m-periodic orbit of (3.1) exists over a K interval including $K=0$, we find a *range* of K values

$$0<K\leqslant K_m$$

over which this periodic orbit is *stable*.

For $0<K\ll 1$ "most" orbits of (3.1) exhibit regular behavior and the system appears integrable; see Fig. 3(b). Near $K\gtrsim 1$, however, large regions of irregular or chaotic behavior exist, where "most" orbits are *unstable* (see the discussion in Sec. I and Fig. 3 below).

The mapping (3.1) is invariant under translations of r or θ by an integer. We therefore, restrict ourselves to the torus $0\leqslant r\leqslant 1$, $0\leqslant\theta\leqslant 1$. In addition, (3.1) also preserves area in the r, θ plane (its Jacobian is equal to 1). Note that, for $K=0$, T_K reduces to the "twist mapping"[6,1]

$$T_0:\begin{Bmatrix} r^{(0)}_{t+1}=r^{(0)}_t \ (=r^{(0)}_0), \\ \theta^{(0)}_{t+1}=\theta^{(0)}_t+r^{(0)}_{t+1} \ \ (=\theta^{(0)}_0+(t+1))r^{(0)}_0 \end{Bmatrix}, \qquad (3.2)$$

$t=0,1,2,\cdots$, which generates straight lines $r^{(0)}=$ const. everywhere in the r, θ plane, cf. Fig. 3(a). All lines corresponding to *rational* $r^{(0)}_0$

$$r^{(0)}_0=n/m \qquad (3.3)$$

consist *entirely* of m-periodic orbits of (3.2). According to a theorem by Poincaré and Birkhoff,[15,6,1] for any given period m there exists a $K=K(m)$ such that for $0<K<K(m)$ the continuous lines (3.3) break up into an even number of m-periodic orbits, half of which are stable and half unstable. In Appendix B we show in greater detail how the Poincaré–Birkhoff theorem applies here, and compute explicitly the first periodic orbits near $K=0$.

As K increases the stable periodic orbits turn unstable and larger regions appear where nearby orbits separate exponentially and successive mappings "scatter" about in an apparently erratic and chaotic manner,[1,22,9] cf. Fig. 3(c).

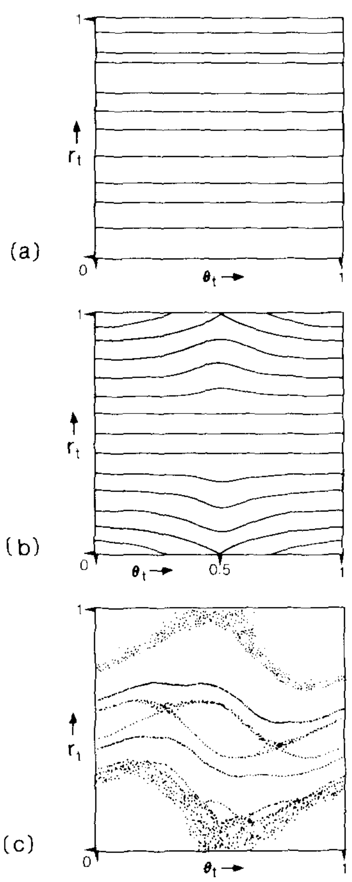


FIG. 3. Phase plane behavior of the standard mapping (3.1). (a) $K=0$, (b) $0<K\ll 1$, and (c) $K=0.97$ (taken from Ref. 9). Note the chaotic regions in (c).

Here we derive an explicit, closed form expression for K-intervals $(0,K_m)$ over which a *stable* m-periodic orbit will not yet have turned unstable. This provides us with an estimate of the range of K values over which one would *not* expect widespread irregular or chaotic behavior.

We first write (3.1) in the form of a second-difference equation, eliminating r_t:

$$\theta_{t+1}+\theta_{t-1}-2\theta_t+(K/2\pi)\sin 2\pi\theta_t=0, \tag{3.4}$$

cf. Eq. (2.2). Its variational equation about an m-periodic orbit $\hat{\theta}_t$ is

$$-\Delta\theta_{t+1}+(2-K\cos 2\pi\hat{\theta}_t)\Delta\theta_t-\Delta\theta_{t-1}=0, \tag{3.5}$$

cf. (2.9). According to criterion (2.18) the orbit $\hat{\theta}_t$ is stable (resp. unstable) iff $|2+\det\underset{\approx}{\mathbf{H}}|<2$ (resp. >2), where

$$\underset{\approx}{\mathbf{H}}\equiv\begin{pmatrix} 2-K\cos 2\pi\hat{\theta}_1 & -1 & 0 & \cdot & 0 & -1\\ -1 & 2-K\cos 2\pi\hat{\theta}_2 & -1 & & & 0\\ 0 & -1 & & \cdot & & \cdot\\ & & & & & \cdot\\ \cdot & & & & \cdot & \\ \cdot & & \cdot & \cdot\ \cdot & & \\ \cdot & & \cdot & \cdot & \cdot & 0\\ 0 & & & \cdot & \cdot & -1\\ -1 & 0\cdots & & 0 & -1 & 2-K\cos 2\pi\hat{\theta}_m \end{pmatrix}. \tag{3.6}$$

In the remainder of this section we derive upper and lower bounds for the eigenvalues of $\underset{\approx}{\mathbf{H}}$ and use them to obtain information about its determinant.

We define the $m\times m$ matrices $\underset{\approx}{\mathbf{A}}$, $\underset{\approx}{\mathbf{B}}$ by

$$\underset{\approx}{\mathbf{H}}=\underset{\approx}{\mathbf{A}}+K\underset{\approx}{\mathbf{B}}, \tag{3.7}$$

where

$$\underset{\approx}{\mathbf{A}}\equiv\begin{pmatrix} 2 & -1 & 0 & \cdots & 0 & -1\\ & & & & & 0\\ & & & & & \cdot\\ -1 & 2 & -1 & & & \cdot\\ 0 & & & \cdot & & \cdot\\ & \cdot & \cdot & & \cdot & \\ \cdot & & \cdot & \cdot & \cdot & 0\\ \cdot & & \cdot & & \cdot & -1\\ \cdot & & & & & \\ 0 & & & & & \\ -1 & 0 & \cdots & 0 & -1 & 2 \end{pmatrix},\quad \underset{\approx}{\mathbf{B}}\equiv(\cos 2\pi\theta_1,\cdots,\cos 2\pi\theta_m,)\underset{\approx}{\mathbf{I}} \tag{3.8}$$

$\underset{\approx}{\mathbf{I}}$ being the $m\times m$ identity matrix and denote the eigenvalues of $\underset{\approx}{\mathbf{A}}$ by $\lambda_j(0)$ and those of $\underset{\approx}{\mathbf{H}}$ by $\lambda_j(K)$, $j=1,2,\ldots,m$. Since $\underset{\approx}{\mathbf{A}}$ is symmetric, the following inequalities hold[28]

$$|\lambda_j(K)-\lambda_j(0)|<K\,\|\underset{\approx}{\mathbf{B}}\|,\quad j=1,2,\ldots,m, \tag{3.9}$$

where

$$\|\underset{\approx}{\mathbf{B}}\|\equiv\sup_{\|\mathbf{X}\|=1}\|\underset{\approx}{\mathbf{B}}\mathbf{X}\|,\quad \|\mathbf{X}\|\equiv\Big(\sum_{i=1}^{m}|x_i|^2\Big)^{1/2}. \tag{3.10}$$

Hence (3.9) immediately gives

$$|\lambda_j(K)-\lambda_j(0)|<K\max_k|\cos 2\pi\theta_k|\leqslant K$$

or

$$-K<\lambda_j(K)-\lambda_j(0)<K,\quad j=1,2\ldots,m. \tag{3.11}$$

Since $\underset{\approx}{\mathbf{A}}$ is a "circulant" matrix,[29] its eigenvalues are well known,

$$\lambda_j(0)=2-2\cos(2\pi j/m), \tag{3.12}$$

whence (3.11) becomes

$$2-2\cos(2\pi j/m)-K<\lambda_j(K)<2-2\cos(2\pi j/m)+K. \tag{3.13}$$

For $1\leqslant j\leqslant m-1$, we make all three parts of (3.13) positive by restricting the values of K:

$$0<K\leqslant 2-2\cos(2\pi/m)\leqslant 2-2\cos(2\pi j/m), \tag{3.14}$$

$j=1,2,\ldots,m-1$. With this restriction (3.13) are preserved upon taking the products

$$0<\prod_{j=1}^{m-1}\Big(2-2\cos\frac{2\pi j}{m}-K\Big)<\prod_{j=1}^{m-1}\lambda_j(K)<\prod_{j=1}^{m-1}\Big(2-2\cos\frac{2\pi j}{m}+K\Big). \tag{3.15}$$

With the aid of the formulas[30]

$$\prod_{j=1}^{m}\Big(2-2\cos\frac{2\pi j}{m}+K\Big)=4\sinh^2\Big(\frac{mz}{2}\Big),$$
$$\prod_{j=1}^{m}\Big(2-2\cos\frac{2\pi j}{m}-K\Big)=-4\sin^2\Big(\frac{m\delta}{2}\Big), \tag{3.16}$$

where z,δ are defined by

$$2\cosh z \equiv 2 + K, \quad 2\cos\delta \equiv 2 - K, \tag{3.17}$$

we find that (3.15) gives

$$\frac{4}{K}\sin^2\left(\frac{m\delta}{2}\right) < \prod_{j=1}^{m-1}\lambda_j(K) < \frac{4}{K}\sinh^2\left(\frac{mz}{2}\right). \tag{3.18}$$

It remains to use inequality (3.13) for the lowest eigenvalue $\lambda_m(K)$, i.e.,

$$-K < \lambda_m(K) < K. \tag{3.19}$$

For the *stable* m-periodic orbits which exist down to $K = 0$ we must have

$$-K < \lambda_m(K) < 0 \tag{3.20}$$

since, otherwise, all $\lambda_j(K)$ would be positive, $\det\underset{\approx}{\mathbf{H}}(0) > 0$ and the orbit would be unstable. Combining (3.20) with (3.18), we conclude that these m-periodic orbits *remain* stable with increasing K *at least* as long as

$$0 > \det\underset{\approx}{\mathbf{H}}(0) = \prod_{j=1}^{m}\lambda_j(K) > -4\sinh^2\left(\frac{mz}{2}\right) \geqslant -4, \tag{3.21}$$

cf. (2.18b), or as long as $\sinh^2(mz/2) \leqslant 1$, which is equivalent to

$$1 \leqslant \cosh(mz) \leqslant 3, \tag{3.22}$$

with z as defined in (3.17). This inequality is satisfied by the range of K values

$$0 \leqslant K \leqslant K_m \equiv 2\cosh[(1/m)\cosh^{-1}3] - 2. \tag{3.23}$$

Thus (3.23) constitutes a sufficient condition for the stability of m-periodic orbits [one should take the intersection between (3.23) and (3.14); (3.23) however, is contained in (3.14)].

It easily follows from (3.23) that $K_m > K_{m+1}$, which indicates that the longer the period of the orbit the sooner it may turn unstable, as K increases. This is indeed observed to be true not only for the mapping (3.1) but for many other Hamiltonian systems. However, the estimates of K_m obtained from (3.23) are significantly lower than the actual K values at which the corresponding orbit turns unstable. For instance, for the period 3 orbit, (3.23) yields $K_3 \cong 0.355$ while the actual orbit turns unstable at $K \cong 1.52$.

IV. STABILITY CRITERION FOR DISSIPATIVE MAPPINGS

Here the stability results of the previous sections are extended to area-contracting or dissipative mappings. Consider for example the quadratic mapping

$$x_{t+1} = y_t + 1 - ax_t^2,$$
$$y_{t+1} = -bx_t, \tag{4.1}$$

studied by Hénon[10] (with $b \to -b$), which "destroys" area if $|b| < 1$. Eliminating y_t in (4.1), we obtain

$$x_{t+1} + bx_{t-1} - 1 + ax_t^2 = 0. \tag{4.2}$$

The variational equation of an m-periodic orbit of (4.2) is

$$-\Delta x_{t+1} - b\Delta x_{t-1} + d_t\Delta x_t = 0, \tag{4.3}$$

where

$$d_t \equiv 2a\hat{x}_t = d_{t+m} \tag{4.4}$$

$t = 0,1,2,\cdots$. When $b > 0$, we define a new real variable z_t:

$$\Delta x_t \equiv b^{t/2}z_t. \tag{4.5}$$

[The case $b < 0$ is more complicated: Instead of (4.5) one writes $\Delta x_t = |b|^{t/2}\exp(i\pi t/2)z_t$ + c.c., with z_t complex, and solves two equations of the type (4.6) for $\mathrm{Re}(z_t)$ and $\mathrm{Im}(z_t)$.] Since $b < 1$, (4.5) is a contracting t-dependent transformation . Inserting (4.5) in (4.3), we find that z_t satisfies the Hill's difference equation

$$-z_{t+1} - z_{t-1} + b^{-1/2}d_t z_t = 0. \tag{4.6}$$

As before, in (2.13) the two linearly independent Floquet solutions of (4.6) are

$$z_t^+ = e^{i\beta t}P_t, z_t^- = e^{-i\beta t}P_t^* \quad (P_t = P_{t+m}),$$

for two linearly independent solutions of (4.3) are obtained

$$\Delta x_t^{(+)} = e^{(B+i\beta)t}P_t, \quad \Delta x_t^{(-)} = e^{(B-i\beta)t}P_t^*, \tag{4.7}$$

where

$$B \equiv \tfrac{1}{2}\ln b < 0 \quad (1 \geqslant b > 0). \tag{4.8}$$

The Floquet exponent β may be solved from the stability relation (2.18)

$$2\cos m\beta = 2 + \det\underset{\approx}{\mathbf{H}}(0), \tag{4.9}$$

where

$$\underset{\approx}{\mathbf{H}}(0) \equiv \begin{pmatrix} b^{-1/2}d_1 & -1 & 0 & 0 & -1 \\ -1 & b^{-1/2}d_2 & -1 & & 0 \\ & & \cdot & & \cdot \\ & & & & \cdot \\ 0 & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & & 0 \\ \cdot & \cdot & & & \\ 0 & & & & -1 \\ -1 & 0\cdots & 0 & -1 & b^{-1/2}d_m \end{pmatrix}, \tag{4.10}$$

cf. (2.16), (2.11). Clearly β can be either real or imaginary. As before, β real implies stability and $\hat{x}_t$ is a so-called "periodic attractor," cf. Ref. 1 (1981). However, β imaginary does not *necessarily* imply instability since this requires, in addition, $|\beta| > |B|$, cf. (4.7), (4.8). Thus our stability criterion (2.18) is generalized to

$$|2 + \det\underset{\approx}{\mathbf{H}}(0)| < 2\cosh(mB): \quad \hat{x}_t \text{ stable or "attracting,"}$$
$$|2 + \det\underset{\approx}{\mathbf{H}}(0)| > 2\cosh(mB): \quad \hat{x}_t \text{ unstable or "repelling."} \tag{4.11}$$

Near a stable periodic orbit $\hat{x}_t$ of the dissipative equation (4.3) solutions are attracted to $\hat{x}_t$, whereas in the conservative case [e.g., $|b| = 1$ in (4.3)] nearby solutions forever "circle" around $\hat{x}_t$.

In the remainder of this section we examine this result and point out its usefulness in a number of cases. Factoring $b^{-1/2}$ out of all the elements of $\underset{\approx}{\mathbf{H}}(0)$ and multiplying by $b^{m/2}$, we rewrite (4.11) in the more convenient form:

$$|2b^{m/2} + \det \mathbf{H}_b| \begin{pmatrix} < 1 + b^m: & \text{stability (attractor)} \\ > 1 + b^m: & \text{instability (repellor)} \end{pmatrix}, \quad (4.12)$$

where

$$\mathbf{H}_b \equiv \begin{pmatrix} d_1 & -\sqrt{b} & 0 & \cdots & 0 & -\sqrt{b} \\ -\sqrt{b} & d_2 & \cdot & & 0 & \\ & & & \cdot & & \cdot \\ & & & & & \cdot \\ 0 & & \cdot & \cdot & & \cdot \\ \cdot & & & & \cdot & 0 \\ \cdot & & \cdot & & \cdot & \\ \cdot & & & & \cdot & -\sqrt{b} \\ 0 & & & \cdot & & \\ -\sqrt{b} & 0 & \cdots & 0 & -\sqrt{b} & d_m \end{pmatrix}, \quad (4.12')$$

In the large dissipation limit $b \to 0$, $\det \mathbf{H}_b \to \prod_{t=1}^{m} d_t$, Eq. (4.2) becomes a first difference equation and (4.12) reduces to the well-known stability condition for such equations[31]:

$\left|\prod_t d_t\right| < 1$ (resp. > 1) implies stability (resp. instability).

For $b > 0$, but *small*, our criterion (4.12) offers a number of computational advantages: First, the matrix $\mathbf{H}_b$, for $|b| \ll 1$, becomes *diagonally dominant*, which is a highly attractive feature from the point of view calculating its eigenvalues (and hence its determinant) using a variety of iterative algorithms.[32] We may also approximate $\det \mathbf{H}_b$ by expanding it in powers of b

$$\det \mathbf{H}_b = -4 \sin^2(m\pi/4) + S_0 + bS_1 + b^2 S_2 + \cdots, \quad (4.13)$$

where

$$S_0 \equiv \prod_{t=1}^{m} d_t,$$

and $S_r \equiv (-1)^r \Sigma$ [all terms obtained by deleting any pair of consecutive d_t's—including the pair of first and last d—from all the terms of S_{r-1},] cf. (2.19) and Ref. 33. Of course, the number of terms required in (4.13) to achieve a certain accuracy depends on the magnitude of b and the "length" of orbit, i.e., the value of m.

Another example where (4.12) can be applied is Chirikov's dissipative mapping[34]

$$\left. \begin{aligned} r_{t+1} &= br_t - (K/2\pi)\sin 2\pi\theta_t \\ \theta_{t+1} &= \theta_t + r_{t+1} \end{aligned} \right\}, \quad t = 0,1,2,\cdots, \quad (4.14)$$

cf. (3.1), where, as before, b (with $|b| < 1$) is the rate at which (4.14) "destroys" area in the r_t, θ_t plane. Combining equations (4.14) into a single second difference equation and linearizing about an m-periodic orbit, we obtain the variational equation

$$-\Delta\theta_{t+1} - b\Delta\theta_{t-1} + (1 + b - K\cos 2\pi\hat{\theta}_t)\Delta\theta_t = 0, \quad (4.15)$$

cf. (2.10) and (3.5). Following the same approach as in Sec. III and using criterion (4.12), we derive a sufficient stability condition for $\hat{\theta}_t$ analogous to (3.21):

$$\begin{aligned} &[K - (1 - \sqrt{b})^2]\sinh^2(mz/2) \\ &\quad \leqslant [K + (1 - \sqrt{b})^2]\cosh^2(mB/2), \end{aligned} \quad (4.16)$$

where z is now defined by

$$2\cosh z \equiv (1 + b + K)/\sqrt{b} \quad (4.17)$$

cf. (3.17) and B is given in (4.8).

The above result guarantees that an m-periodic orbit of (4.14) which is stable near $K = 0$, remains stable *at least* over the range

$$(1 - \sqrt{b})^2 \leqslant K \leqslant K_m, \quad (4.18)$$

where K_m is the K value at which (4.16) becomes an equality. Such orbits exist in the area preserving case $b = 1$ (see Sec. III and Appendix B) and are expected to exist for $b < 1$ also. The condition (4.16)–(4.18) reduces, of course, to (3.21)–(3.23) at $b = 1$ and yields best estimates K_m in the small dissipation limit $b \lesssim 1$.

ACKNOWLEDGMENTS

We express our appreciation to John Greene for his enthusiastic comments and his encouragement. Many useful conversations with Joe Ford, Mel Month, Boris Chirikov, Giulio Casati, and Elliott Montroll are gratefully acknowledged. One of us (Tassos Bountis) acknowledges the help and support of Mark Ablowitz, Eytan Barouch, and Alan Newell of the Mathematics Department of Clarkson College. This study was partially supported by the D.O.E. under Grant EG-79-C-03-1538.

APPENDIX A: AN APPLICATION OF THE STABILITY CRITERION (2.18)

As we have seen in Sec. II, the criterion (2.18) can be used to determine the (un)boundedness of the solutions of second difference equations of the form

$$x_{t+1} + x_{t-1} + (\lambda + Q_t)x_t = 0, \quad (A1)$$

where Q_t is periodic with period m, i.e., $Q_t = Q_{t+m}$, $t = 0,1,2,\cdots$, and λ is a parameter independent of t [clearly (2.10) with $-d_t = \lambda + Q_t$ is of this type].

We obtain in this appendix analytic stability results for (A1) with Q_t a periodic pulse

$$Q_t \equiv \left\{ \begin{matrix} -V < 0, & t = 1,2,\ldots,\tau, \\ 0, & t = \tau + 1,\ldots,m \end{matrix} \right\}. \quad (A2)$$

Thus, (A1)–(A2) may be viewed as the analog of Hill's equation

$$\frac{d^2x}{dt^2} + [\lambda + Q(t)]x = 0 \quad (A3)$$

with

$$Q(t) \equiv \left\{ \begin{matrix} -V, & 0 < t < \tau, \\ 0, & \tau < t < T \end{matrix} \right\}, \quad (A4)$$

$Q(t + T) = Q(t)$, which is Schrödinger's equation for a periodic potential in the Kronig–Penney crystal model. Here we use our stability criterion to derive analytically the λ-intervals for which (A1)–(A2) has bounded solutions. The result is similar to the corresponding one[24] for (A3)–(A4). More stability results on mappings (A1) as well as on thier applicability to problems of accelerator physics can be found in Ref. 35.

In Sec. IIB it was derived that the solutions of (A1) are bounded iff

$$|2+\det\mathbf{H}(0)|<2, \tag{A5}$$

where, in the present case,

$$\det\mathbf{H}(0)=\det\begin{pmatrix}
-\lambda+V & -1 & 0 & \cdots & & 0 & & -1\\
-1 & -\lambda+V & -1 & & & & & 0\\
0 & -1 & \ddots & \ddots & & 0 & & \vdots\\
\vdots & & \ddots & \ddots & -1 & & & \vdots\\
 & & & -1 & -\lambda+V & -1 & & \\
 & 0 & & & -1 & \lambda & \ddots & \\
 & & & & & \ddots & \ddots & 0\\
 & & & & & & \ddots & \ddots\\
0 & & & & & & \ddots & -1\\
-1 & 0 & & \cdots & & 0 & -1 & -\lambda
\end{pmatrix}. \tag{A6}$$

[τ columns m-τ columns.]

Expanding this determinant about its upper right and lower left corner elements gives

$$\det\mathbf{H}(0)=(-1)^m(K_{\tau,m}-K_{\tau-1,m-2})-2, \tag{A7}$$

where the $K_{i,j}$ determinants are defined by

$$K_{i,j}\equiv\det\begin{pmatrix}
\lambda-V & -1 & & & & & \\
1 & \lambda-V & \ddots & & 0 & & \\
 & \ddots & \ddots & 1 & & & \\
 & & 1 & \lambda-V & 1 & & \\
 & 0 & & 1 & \lambda & \ddots & \\
 & & & & \ddots & \ddots & 1\\
 & & & & & 1 & \lambda
\end{pmatrix} \tag{A8}$$

[i columns (j-i) columns]

Equation (A7) combined with (A5) gives

$$|K_{\tau,m}-K_{\tau-1,m-2}|\leqslant 2 \tag{A9}$$

as the boundedness criterion for the solutions of (A1).

To evaluate the determinant $K_{\tau,m}$, we proceed by induction. Let $\tau=1$; expanding $K_{1,m}$ with respect to its first row, we find

$$K_{1,m}=(\lambda-V)K_{m-1}-K_{m-2}, \tag{A10}$$

where K_j is defined by

$$K_j\equiv\det\begin{pmatrix}
\lambda & 1 & & 0 & \\
1 & \lambda & 1 & & \\
 & \ddots & \ddots & \ddots & \\
 & & \ddots & \ddots & 1\\
 & 0 & & 1 & \lambda
\end{pmatrix}. \tag{A11}$$

j columns

Similarly, for $\tau = 2$ we get

$$K_{2,m} = (\lambda - V)K_{1,m-1} - K_{m-2}$$
$$= [(\lambda - V)^2 - 1]K_{m-2} - K_{m-3}.$$

Repeating this process for $\tau = 3,4$ finally yields

$$K_{\tau,m} = U_\tau K_{m-\tau} - U_{\tau-1}K_{m-\tau-1}, \tag{A12}$$

where U_τ satisfies

$$U_\tau = (\lambda - V)U_{\tau-1} - U_{\tau-2} \tag{A13}$$

with $U_0 = 1$, $U_2 = \lambda - V$. Defining θ by

$$2\cos\theta \equiv \lambda - V, \tag{A14}$$

we can write the general solution of (A13) as

$$U_\tau = \sin(\tau + 1)\theta / \sin\theta. \tag{A15}$$

These are trigonometric sines if $|\lambda - V| \leqslant 2$ and hyperbolic sines if $|\lambda - V| < 2$.

It is also easy to show that $K_{m-\tau}$ in (A11) satisfies

$$K_{m-\tau} = \lambda K_{m-\tau-1} - K_{m-\tau-2}, \tag{A16}$$

with $K_1 = \lambda$ and $K_0 = 1$. Defining again an angle ϕ by

$$\lambda \equiv 2\cos\phi, \tag{A17}$$

we write the solution of (A16) in the form

$$K_{m-\tau} = \sin(m - \tau + 1)\phi / \sin\phi, \tag{A18}$$

where cos, sin→cosh, sinh if $|\lambda| > 2$. Putting (A18) and (A15) in (A12) and then substituting back in (A9) finally gives the stability condition

$$\left| \sin\tau\theta \sin(m-\tau)\phi \frac{\cos\theta\cos\phi - 1}{\sin\theta\sin\phi} + \cos\tau\theta\cos(m-\tau)\phi \right| \leqslant 1. \tag{A19}$$

The final result, therefore, is that the solutions of (A1) and (A2) are *bounded* if λ is such that (A19) is satisfied and are unbounded otherwise. These intervals can be determined numerically or graphically, while analytic estimates are also easily obtained from (A19), cf. Ref. 35.

APPENDIX B: PERIODIC ORBITS OF THE MAPPING (3.1)

The existence of *stable* periodic orbits of the standard mapping (3.1) down to $K = 0$ is guaranteed by a theorem due to Poincaré and Birkhoff, cf. Ref. 15, p. 39. As stated there, the theorem applies directly to "sufficiently smooth" perturbations of the "twist" mapping

$$\left.\begin{aligned} r^{(0)}_{t+1} &= r^{(0)}_t \\ \theta^{(0)}_{t+1} &= \theta^{(0)}_t + \sigma(r^{(0)}_t) \end{aligned}\right\}, \quad t = 0,1,2,\cdots, \tag{B1}$$

where $\sigma'(r^{(0)}_t) \neq 0$, cf. Ref. 1 (1981). Such is the case with the mapping (3.1):

$$T_K \left\{\begin{aligned} r_{t+1} &= r_t - (K/2\pi)\sin 2\pi\theta_t \\ \theta_{t+1} &= \theta_t + r_{t+1} \end{aligned}\right\}, \quad t = 0,1,2,\cdots, \tag{B2}$$

where $\sigma(r^{(0)}_t) = r^{(0)}_t$, cf. (B1). The Poincaré–Birkhoff theorem states that given any *invariant circle* of the corresponding "twist" mapping (B1) with *rational* radius $r^{(0)}_t = n/m$, there is a *range of* K values, $K \in [0,K(m)]$, over which this circle breaks up into an equal number of stable and unstable m-periodic orbits.

Thus the existence of stable periodic orbits of the mapping (B2) in a neighborhood of $K = 0$ is established. In particular, for any $m = m^*$, we have the K interval for which stable m-periodic orbits with $m \leqslant m^*$ exist as $\cup_{m=1}^{m^*} [0,K(m)]$, i.e., as the intersection of all intervals of the Poincaré–Birkhoff theorem with $1 \leqslant m \leqslant m^*$. Finally, we note that in the proof of these statements one needs to restrict r_t to some annulus, e.g., $1 \leqslant r_t \leqslant 2$, whereas in this paper we take $0 \leqslant r_t \leqslant 1$. This is not an important difference since the mapping (B2) is invariant under $r_t \to r_t + 1$.

We now demonstrate below how one may explicitly compute the periodic orbits predicted by the Poincaré–Birkhoff theorem, near $K = 0$. We present results for $m = 1,2$, and 3, for which, in fact, the theorem need not apply.[1,6] Clearly, as m increases the calculations, although straightfoward, become quite cumbersome.

According to the symmetry arguments of Greene,[9] the initial conditions for periodic orbits of T_K, cf. (B2), fall in either one of two classes:

$$\begin{aligned} &\text{Class A:} \quad r_0 \text{ arbitrary}, \quad \theta_0 = 0 \text{ or } \tfrac{1}{2}, \\ &\text{Class B:} \quad r_0 \text{ arbitrary}, \quad \theta_0 = \tfrac{1}{2}r_0 \text{ or } \tfrac{1}{2}(r_0 + 1) \end{aligned} \tag{B3}$$

There are two $m = 1$ periodic "orbits" satisfying

$$r_0 = r_0 - (K/2\pi)\sin 2\pi\theta_0, \quad \theta_0 = \theta_0 + r_0,$$

$(r_0,\theta_0) = (0,0)$ and $(0,\frac{1}{2})$, see Fig. 3(b). Their stability is immediately obtained from the variational equations

$$\begin{pmatrix} \Delta\theta_1 \\ \Delta\theta_0 \end{pmatrix} = \begin{pmatrix} 2 \pm K & -1 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} \Delta\theta_0 \\ \Delta\theta_{-1} \end{pmatrix}, \tag{B4}$$

cf. (3.5), (2.4) the (+) sign corresponding to $(0,\frac{1}{2})$ and the (−) to (0,0). From (B4)

$$\operatorname{tr}(\mathbf{M}) = 2 \pm K, \tag{B5}$$

and hence, according to the discussion at th end of Sec. IIA, (0,0) is a stable $m = 1$ periodic "orbit" for $0 < K < 4$, while $(0,\frac{1}{2})$ is unstable for all $K > 0$.

Consider now the case $m = 2$. Applying T_K once to (r_0,θ_0) yields

$$\begin{aligned} r_1 &= r_0 - (K/2\pi)\sin 2\pi\theta_0 = r_0, \\ \theta_1 &= \theta_0 + r_1 = r_0 \end{aligned} \tag{B6}$$

for the class A, $\theta_0 = 0$ solutions, cf. (B3). Closing the orbit upon itself after one more application of T_K, we write

$$\begin{aligned} r_2 &= r_1 - (K/2\pi)\sin 2\pi\theta_1 = r_0 - (K/2\pi)\sin 2\pi r_0 = r_0, \\ \theta_2 &= \theta_1 + r_0 = 2r_0 = \theta_0 = 1, \end{aligned} \tag{B7}$$

since on the unit torus $\theta_0 = 1$ is equivalent to $\theta_0 = 0$. From (B7), we find $r_0 = \theta_1 = \frac{1}{2}$ and we thus obtain the 2-periodic orbit

$$(\tfrac{1}{2},0) \to (\tfrac{1}{2},\tfrac{1}{2}) \to (\tfrac{1}{2},1). \tag{B8}$$

This periodic orbit is also stable for $0 < K < 4$ since a calculation of the tr (M) gives tr (M) $= 2 - K$ as in (B5). Furthermore, note that the above $m = 1$ and 2 periodic orbits obviously exist down to $K = 0$ since they are independent of K. The corresponding unstable 2-periodic solution predicted by the Poincaré–Birkhoff theorem is obtained starting with class B initial conditions [the class A, $\theta_0 = \frac{1}{2}$ initial conditions lead to the same orbit (B8)].

These results from the $m = 2$ case are, in fact, common to all periodic orbits with even period $m = 2k$: I.e., class A initial conditions yield one and the same stable orbit, while class B conditions yield the corresponding unstable one.

The situation is different for m odd. It turns out that $m = 2k + 1$ periodic orbits are all given either by class A or class B initial conditions. In the case of the class A solutions $\theta_0 = 0$ yields the stable orbits while $\theta_0 = \frac{1}{2}$ yields the unstable ones.

Consider, for example, the case $m = 3$. With $\theta_0 = 0$ we find, as in (B6) and (B7):

$$(r_0,\theta_0) = (r_0,0),(r_1,\theta_1) = (r_0,r_0),$$

$$(r_2,\theta_2) = (r_0 - (K/2\pi)\sin 2\pi r_0, 2r_0 - (K/2\pi)\sin 2\pi r_0).$$

To get a periodic orbit of period 3, we require

$$r_3 = r_2 - (K/2\pi)\sin 2\pi\theta_2 = r_0, \tag{B9}$$

$$\theta_3 = \theta_2 + r_3 = 3r_0 - (K/2\pi)\sin 2\pi r_0 = n,$$

where n is a positive integer, prime relative to m and $1 \leqslant n < m$, cf. (3.3).

In the case of the ($m = 3$)-periodic orbit we fine that (B9) is satisfied if

$$3r_0 - (K/2\pi)\sin 2\pi r_0 = n \tag{B10}$$

with $n = 1,2$. Solving (B10) numerically we list below several values of r_0 corresponding to different K-values for the case $n/m = \frac{1}{3}$, $\theta_0 = 0$:

$K =$	0.0	0.01	0.2	0.5	1.0
$r_0 \cong$	0.3333	0.33379	0.34222	0.35436	0.37164

The stability of the resulting 3-periodic orbit is determined using our stability criterion of Sec. II. Thus we evaluate

$$\det\mathbf{H} = \det\begin{bmatrix} 2 - K\cos 2\pi r_0 & -1 & -1 \\ -1 & 2 - K\cos 2\pi r_0 & -1 \\ -1 & -1 & 2 - K \end{bmatrix},$$

whence

$$\det\mathbf{H} = K(3 - K\cos 2\pi r_0)$$
$$(K\cos 2\pi r_0 - 2\cos 2\pi r_0 - 1). \tag{B11}$$

As $K \to 0$ $r_0 \to \frac{1}{3}$ [cf. (B10) and the numerical solution below it] and to lowest order in K, (B11) yields

$$\det\mathbf{H} \cong -K^3 < 0. \tag{B12}$$

Thus the $\theta_0 = 0$, 3-periodic orbits are *stable*, cf. (2.18), for sufficiently small K.

For $\theta_0 = \frac{1}{2}$, all the above expressions are preserved, with one change only: $K \to -K$ [since letting $\theta_t \to \theta_t + \frac{1}{2}$ simply switches the sign of the sine term in (B12)]. Hence (B11) becomes

$$\det\mathbf{H} = K(3 + K\cos 2\pi r_0)$$
$$(K\cos 2\pi r_0 + 2\cos 2\pi r_0 + 1), \tag{B13}$$

which, as $K \to 0$, and $r_0 \to \frac{1}{3}$, gives to lowest order in K

$$\det\mathbf{H} \cong +K^3 > 0. \tag{B14}$$

Hence, these $\theta_0 = \frac{1}{2}$ orbits are *unstable* near $K = 0$, cf. (3.6b).

Similar results are obtained for $m = 4,5$, etc. As m increases, we find that we have to solve r_0 from more and more complicated transcendental equations of the type (B10). For example, for $m = 5$ the stable periodic orbits ($\theta_0 = 0$) are obtained solving r_0 from

$$5r_0 + (3K/2\pi)\sin 2\pi r_0 - (K/2\pi)$$
$$\times \sin 2\pi[2r_0 - (K/2\pi)\sin 2\pi r_0] = n,$$

$n = 1,2,3,4$. However complicated this equation may appear, it is easy to solve it numerically and obtain r_0 (for sufficiently small K) with $r_0 \to n/5$ as $K \to 0$.

We have thus demonstrated how to construct periodic orbits of (B2) existing down to $K = 0$. Our results indicate that pairs of stable and unstable periodic orbits with *any* period p exist for $K \geqslant 0$ as predicted by the Poincaré–Birkhoff theorem.

[1]M. V. Berry, "Regular and Irregular Motion," in *Topics in Nonlinear Dynamics*, edited by S. Jorna, A.I.P. Conf. Proc. **46** (Am. Inst. Phys., New York, 1978); for a review see also R. H. G. Helleman, "Self-Generated Chaotic Behavior in Nonlinear Mechanics," in *Fundamental Problems in Statistical Mechanics*, Vol. 5, edited by E. G. D. Cohen (North-Holland, Amsterdam, 1981).

[2]M. Hénon and C. Heiles, Astron. J. **69**, 73 (1964).

[3]G. Contopoulos, Astrophys. J. **138**, 1297 (1963); Astron. J. **68**, 1 (1963). See also, e.g., M. Hénon, Ann. Astrophys. **28**, 499, 992 (1965); Bull. Astron., Ser. 3, **1**, Fasc. 1, 57, Fasc. 2, 49 (1966).

[4]J. Ford with G. H. Walker, Phys. Rev. **188**, 416 (1969);with G. H. Lunsford, J. Math. Phys. **13**, 700 (1972); with S. D. Stoddard, Phys. Rev. A **8**, 1504 (1973); with G. Casati, Phys. Rev. A **12**, 1702 (1975).

[5]T. Bountis and R. H. G. Helleman, "Periodic Solutions, Variational Methods," in *Lecture Notes in Physics* **93**, edited G. Casati and J. Ford (Springer, New York, 1978); also T. Bountis, Ph.D. thesis (Physics Department, University of Rochester, New York, 1978).

[6]J. Moser, *Stable and Random Motions in Dynamical Systems* (Princeton U. P., Princeton, N. J., 1973); also V. I. Arnol'd, *Mathematical Methods of Classical Mechanics* (Springer, New York, 1978).

[7]M. Hénon, Quant. Appl. Math. **27**, 291 (1969); see also R. H. G. Helleman, "Iterative Solution of a Stochastic Mapping," in *Statistical Mechanics and Statistical Methods*, edited by U. Landman (Plenum, New York, 1978). The expansion (2.10) there is meant to be the standard expansion of a determinant.

[8]T. Bountis, "Period Doubling Bifurcations and Universality in Conservative Systems," Physica D (June 1981).

[9]J. M. Greene, J. Math. Phys. **20**, 1183 (1979); see also J. Math. Phys. **9**, 760 (1968).

[10]M. Hénon, Commun. Math. Phys. **50**, 69 (1976).

[11]M. Tabor, "The Onset of Chaotic Motion in Dynamical Systems," Adv. Chem. Phys. (to appear, Preprint: La Jolla Institute, P. O. Box 1434, La Jolla, Calif. 92038).

[12]V. I. Arnol'd and A. Avez, *Ergodic Problems of Classical Mechanics* (Benjamin, New York, 1968).

[13]J. Ford's review article in *Fundamental Problems in Statistical Mechanics* Vol. 3, edited by E. G. D. Cohen (North-Holland, Amsterdam, 1975); also Adv. Chem. Phys. **24**, 155 (1973).

[14] J. Lebowitz and O. Penrose, "Modern Ergodic Theory," Physics Today **26** (2), 23 (1973).

[15]J. Moser, *Lectures on Hamiltonian Systems*, Mem. Am. Math. Soc. **81**, 1–60 (1968).

[16]Y. Treve, "Theory of Chaotic Motion with Application of Controlled Fusion Research," in Ref. 1, p. 147.

[17]G. Smith and A. Kaufman, Phys. Rev. Lett. **34**, 1613 (1975); also Phys. Rev. Lett. **38**, 970 (1977).

[18]L. J. Laslett, "Some Illustrations of Stochasticity," in Ref. 1, p. 221.

[19]Several articles in *Nonlinear Dynamics and the Beam-Beam Interaction*, edited by M. Month and J. Herrera, Am. Inst. of Phys. Conf. Proc. **57** (Am. Inst. Phys., New York, 1980).
[20]I. Percival, Adv. Chem. Phys. **36**, 1 (1977).
[21]D. Noid and R. Marcus, J. Chem. Phys. **62**, 2119 (1975).
[22]B. V. Chirikov, "Universal Instability in Many Dimensional Oscillator Systems," Phys. Rep. **52**, 265 (1979).
[23]P. Holmes, Phil. Trans. Royal. Soc. Lond. **292**, 419 (1979).
[24]W. Magnus and S. Winkler, *Hill's Equation* (Interscience, New York, 1966).
[25]L. A. Pars, *Analytical Dynamics* (Wiley, New York, 1965), esp. Secs. 23.5, 23.6, and 26.3.
[26]T. Fort, *Finite Differences and Difference Equations in the Real Domain* (Oxford U. P., Oxford, 1948); also H. Levy and F. Lessamn, *Finite Difference Equations* (MacMillan, New York, 1961).
[27]E. Barouch, "On the Ising Model in the Presence of Magnetic Field," Physica D **1**, 333 (1980).
[28]I. Bendixson, Acta Math. **25**, 359 (1902).
[29]R. Bellman, *Introduction to Matrix Analysis* (McGraw-Hill, New York, 1970), Chap. 12.
[30]I. Gradshteyn and I. Ryzhik, *Tables of Integrals, Series and Products* (Academic, New York, 1965), Formula 1.394.
[31]R. May, Nature **261**, 459 (1976).
[32]J. Wilkinson, *The Algebraic Eigenvalue Problem* (Clarendon, Oxford, 1965).
[33]Our expansion (4.13) — with the comments below it — is derived using the fact that det**H** can be written as the diference of two tridiagonal determinants, cf. (A7); see e.g., T. Muir and W. Metzler, *A Treatise on the Theory of Determinants* (Dover, New York, 1960), Sec. 10.542–546.
[34]B. V. Chirikov, Coll. Intern., in CNRS No. 229, "Transformations ponctuelles et leurs Applications," (Toulouse, CNRS, Paris, 1976).
[35]R. H. G. Helleman, *Exact Results for Some Linear and Nonlinear Beam-Beam Effects*," in Ref. 19, p. 236.

Geodesic first integrals with explicit path-parameter dependence in Riemannian space–times

Gerald H. Katzin
Department of Physics, North Carolina State University, Raleigh, North Carolina 27650
Jack Levine
Department of Mathematics, North Carolina State University, Raleigh, North Carolina 27650

(Received 31 October 1980; accepted for publication 30 January 1981)

In a Riemannian space V_n general formulas are obtained for geodesic first integrals which are mth order polynomials in the tangent vector and which are assumed to depend *explicitly* on the path parameter s. It is found that such first integrals must also be polynomials in s. Necessary and sufficient conditions are found for the existence of these first integrals. The existence of many well-known symmetries such as homothetic motions (scale change), affine collineations, conformal motions, projective collineations, conformal collineations, or special curvature collineations are shown to be sufficient for the existence of such first integrals with explicit path-parameter dependence. To illustrate the theory, geodesic first integrals of this type have been calculated for four Riemannian space–times of general relativity.

PACS numbers: 02.40.Ky, 04.20.Me

I. INTRODUCTION

The second order differential equations[1]

$$Du^i/ds \equiv du^i/ds + \Gamma^i_{jk} u^j u^k = 0, \quad u^i \equiv dx^i/ds \tag{1.1}$$

which determine the geodesics of a Riemannian space V_n, admit $2n$ functionally independent first integrals of which at least one must depend explicitly on the path-parameter s.[2] In this paper we consider for such differential equations the problem of determining mth-order (in the tangent vector u^i) first integrals $I^{(m)}$, $m = 1, 2, \cdots$, of the form

$$\begin{aligned} I^{(m)} &\equiv M^{(m:0)} + M^{(m:1)}_{i_1} u^{i_1} + \cdots + M^{(m:m)}_{i_1 \cdots i_m} u^{i_1} \cdots u^{i_m} \\ &\equiv \sum_{l=0}^{m} M^{(m:l)}_{i_1 \cdots i_l} u^{i_1} \cdots u^{i_l}, \end{aligned} \tag{1.2}$$

where the functions $M^{(m:l)}_{i_1 \cdots i_l} = M^{(m:l)}_{i_1 \cdots i_l}(x,s)$ are completely symmetric on all indices $i_1 \cdots i_l$. *Such first integrals will in general be inhomogeneous in the tangent vector u^i and have explicit dependence on the path parameter s.*

For indefinite space–times of general relativity the geodesics may be separated into two types—those of the null type and those of the nonnull type, as characterized by the value of ϵ in the relation[3]

$$g_{ij} u^i u^j = \epsilon. \tag{1.3}$$

Accordingly, the formulation of the necessary and sufficient conditions for the existence of an mth-order geodesic first integral with explicit path-parameter dependence will depend upon the assumed type of geodesic. The procedure for formulating the conditions which determine such first integrals for geodesics of a specific type involves the use of constraints and is generally more complicated than the procedure for formulating first integrals for arbitrary (both types of) goedesics. Therefore, in this paper we shall primarily consider the simpler case of arbitrary geodesics and postpone to a later paper a systematic investigation of restricted type geodesic mth-order first integrals with explicit path-parameter dependence. Thus, throughout this paper unless otherwise stated the term geodesic shall mean *arbitrary* geodesic.

For generality in our analysis, unless indicated otherwise, we shall assume an n-dimensional indefinite Riemannian space V_n. However, for illustrative purposes we shall draw our examples from the V_4 space–times of general relativity.

For geodesics certain *homogeneous* first integrals with *no* explicit path-parameter dependence are known to be concomitant with the existence in the V_n of specific infinitesimal point mappings of the type[4,5]

$$\bar{x}^i = x^i + \xi^i(x)\delta a. \tag{1.4}$$

The vectors ξ^i which define these mappings are determined by "symmetry equations" which involve conditions on the Lie deformation of the metric tensor of the V_n. Since certain of the new first integrals and their associated conditions derived in this paper will be related to some of these known cases, we summarize for later reference the pertinent above-mentioned known "symmetry equations" and associated homogeneous first integrals in Table I.

In Sec. II it is shown that every geodesic first integral which is an mth-order (in general inhomogeneous) polynomial in the tangent vector u^i and has *explicit* dependence on the path-parameter s must be a polynomial in s. Necessary and sufficient conditions for the existence of such integrals are given. Detailed formulas for these explicit path-parameter dependent first integrals through third order are listed in tables in Sec. II.

In Sec. III it is shown that if there exists one mth-order first integral with explicit path-parameter dependence, then in general there will exist several such integrals of the same order. Tables are given which list such related integrals through the third order.

In Sec. IV it is shown that the existence of a parallel vector field is necessary and sufficient for the geodesics in a V_n to admit a linear first integral with explicit path-parameter dependence.

In Sec. V quadratic first integrals with explicit depen-

TABLE I. Well-known symmetry conditions and concomitant geodesic homogeneous first integrals with no explicit path-parameter dependence.[a]

	Symmetry	Notation	Defining equation	Concomitant first integral(s)
(a)	Motion	M	$h_{ij}=0$	$\xi_i u^i$
(b)	Homothetic motion	HM	$h_{ij}=2\sigma_0 g_{ij}$ $\sigma_0=\text{const}$	$\sigma_0 g_{ij}u^iu^j$ $\xi_i u^i$ (null geodesic only)
(c)	Affine collineation	AC	$h_{ij;k}=0$	$h_{ij}u^iu^j$
(d)	Projective collineation	PC	$h_{ij;k}=2g_{ij}\phi_{,k}+g_{jk}\phi_{,i}+g_{ik}\phi_{,j}$	$(h_{ij}-4\phi g_{ij})u^iu^j$
(e)	Special projective collineation	SPC	$h_{ij;k}=2g_{ij}\phi_{,k}+g_{jk}\phi_{,i}+g_{ik}\phi_{,j}$, $\phi_{;ij}=0$	$(h_{ij}-4\phi g_{ij})u^iu^j$ $\phi_{,i}u^i$
(f)	Conformal motion	CM	$h_{ij}=2\sigma g_{ij}$	$\xi_i u^i$ (null geodesics only)
(g)	Special conformal motion	SCM	$h_{ij}=2\sigma g_{ij}$ $\sigma_{;ij}=0$	$\xi_i u^i$ (null geodesics only) $\sigma_{,i}u^i$
(h)	Conformal collineation	CONFC	$h_{ij;k}=2\tau_{,k}g_{ij}$	$(h_{ij}-2\tau g_{ij})u^iu^j$
(i)	Special conformal collineation	SCONFC	$h_{ij;k}=2\tau_{,k}g_{ij}$ $\tau_{;ij}=0$	$(h_{ij}-2\tau g_{ij})u^iu^j$ $\tau_{,i}u^i$
(j)	Special curvature collineation	SCC	$h_{ij;kl}=0$	$h_{ij;k}u^iu^ju^k$ $h_{,k}u^k$, $h\equiv g^{ih}h_{ij}$

[a] $h_{ij}=\mathscr{L}_\xi g_{ij}$. The symbol $\mathscr{L}_\xi$ indicates Lie differentiation with respect to vector ξ^i.

dence on the path-parameter are shown to be of two basic types—those based on scalar fields and those based on vectors fields. For each type certain conditions must be satisfied by the field. It is shown that the vector condition is satisfied by an affine collineation or homothetic motion (scale change). *Hence the existence of an affine collineation implies the existence of an inhomogeneous quadratic first integral with explicit dependence on the path-parameter*, in addition to the well-known homogeneous quadratic first integral with no explicit path-parameter dependence. For the case of a homothetic motion (which is a subcase of an affine collineation) the above-mentioned quadratic first integral with explicit dependence on the path-parameter reduces by means of the metrical quadratic integral to an inhomogeneous linear first integral with explicit path-parameter dependence. For the case of a null geodesic this linear integral further reduces to the well-known homogeneous linear integral with no explicit path-parameter dependence.

If vectors which define projective collineations or conformal collineations are assumed to be solutions of the condition which is necessary and sufficient for the existence of the above-mentioned vector-based quadratic first integral with explicit path-parameter dependence, we find that both types of collineations necessarily reduce to affine collineations. By a similar analysis conformal motion vectors are shown to be limited to homothetic motions.

For a nonempty space–time of general relativity it is shown that if there exists a quadratic first integral with explicit path-parameter dependence of the vector-based type mentioned above, then there will also exist a conserved 4-current which is dependent upon the energy–momentum tensor and the vector.

In Sec. VI cubic first integrals with explicit path-parameter dependence are shown to be of three basic types—those based on scalar fields, vector fields, or second rank tensor fields. It is found that a vector which defines a special curvature collineation will satisfy the condition for the vector-based type cubic first integral. Such cubic integrals reduce to quadratic first integrals with explicit path-parameter dependence for the cases in which the vector field is either a special projective collineation or a special conformal collineation (both collineations being subcases of special curvature collineations). In a like manner the cubic integrals associated with a special conformal motion reduce to linear first integrals with explicit dependence on the path-parameter.

In Sec. VII it is shown that a null geodesic inhomogeneous quadratic first integral with explicit path-parameter dependence (which was derived in Sec. VI as a degenerate cubic first integral concomitant with the existence of either *special* projective or *special* conformal collineations) will exist whenever the space–time admits *any* (i.e., not necessarily special) projective collineation, conformal collineation, or seminull geodesic collineation. It is of particular interest to note that this new type of quadratic integral will exist in addition to the well-known homogeneous quadratic first integrals associated with these symmetries.

To illustrate several of the theorems, we obtain geodesic first integrals with explicit path-parameter dependence for the following space–times: gravitational plane wave, Einstein static, a perfect fluid, and a Friedmann–Lemaitre.

II. TYPES OF *M*TH-ORDER FIRST INTEGRALS WITH EXPLICIT PATH-PARAMETER DEPENDENCE

In this section we determine the types of mth-order first integrals with explicit dependence on the path-parameter s which could possibly be admitted by the geodesics in a Riemannian space V_n and then obtain necessary and sufficient conditions for the existence of such integrals.

We thus assume the geodesics admit an mth-order first integral of the form $I^{(m)}$ given by (1.2) and proceed to determine what restrictions must be placed on the coefficients $M^{(m:l)}_{i_1\cdots i_l}$, $l = 0,1,\cdots,m$, by requiring that along the geodesics[6]

$$\frac{D}{ds} I^{(m)} \overset{\circ}{=} 0. \tag{2.1}$$

By substituting (1.2) into (2.1) and carrying out the indicated differentiation [with use of (1.1)] we obtain

$$\sum_{l=0}^{m} \left[M^{(m:l)}_{i_1\cdots i_l;i_{l+1}} u^{i_1}\cdots u^{i_{l+1}} + M^{(m:l)}_{i_1\cdots i_l,s} u^{i_1}\cdots u^{i_l} \right] \overset{\circ}{=} 0. \tag{2.2}$$

In (2.2) the terms may be regrouped to give

$$M^{(m:0)}_{,s} + \sum_{l=1}^{m} \left[M^{(m:l-1)}_{i_1\cdots i_{l-1};i_l} + M^{(m:l)}_{i_1\cdots i_l,s} \right] u^{i_1}\cdots u^{i_l} + M^{(m:m)}_{i_1\cdots i_m;i_{m+1}} u^{i_1}\cdots u^{i_{m+1}} \overset{\circ}{=} 0. \tag{2.3}$$

Since (2.3) must hold for arbitrary geodesics, we require that (2.3) be identically zero in the tangent vectors u^i and obtain (after symmetrization[7]) the following necessary conditions on the functions $M^{(m:l)}_{i_1\cdots i_l}$ for the existence of an mth-order integral (1.2):

$$M^{(m:0)}_{,s} = 0, \tag{2.4}$$

$$M^{(m:l-1)}_{(i_1\cdots i_{l-1};i_l)} + M^{(m:l)}_{i_1\cdots i_l,s} = 0, \quad l = 1,\ldots,m, \tag{2.5}$$

$$M^{(m:m)}_{(i_1\cdots i_m;i_{m+1})} = 0. \tag{2.6}$$

Equations (2.4)–(2.6) are to be regarded as differential equations in the $n+1$ independent variables x^i, s.

The set of $m+1$ equations given by (2.4) and (2.5) may be integrated in sequence with respect to s to yield

$$M^{(m:l)}_{i_1\cdots i_l} = \sum_{\alpha=0}^{l} \frac{(-s)^{l-\alpha}}{(l-\alpha)!} C^{(m:\alpha)}_{(i_1\cdots i_\alpha;i_{\alpha+1}\cdots i_l)}, \tag{2.7}$$

where the α-rank tensors ($\alpha = 0,1,\cdots,m$) $C^{(m:\alpha)}_{i_1\cdots i_\alpha} = C^{(m:\alpha)}_{i_1\cdots i_\alpha}(x)$, which arise from the process of integration, are completely symmetric on all indices $i_1\cdots i_\alpha$.

By use of (2.7) in (2.6) we obtain

$$\sum_{\alpha=0}^{m} \frac{(-s)^{m-\alpha}}{(m-\alpha)!} C^{(m:\alpha)}_{(i_1\cdots i_\alpha;i_{\alpha+1}\cdots i_{m+1})} = 0. \tag{2.8}$$

From (2.8) it follows that

$$C^{(m:\alpha)}_{(i_1\cdots i_\alpha;i_{\alpha+1}\cdots i_{m+1})} = 0, \quad \alpha = 0,1,\ldots,m. \tag{2.9}$$

Hence a necessary condition that $I^{(m)}$ of the form (1.2) be a first integral of the geodesics is that the coefficients $M^{(m:l)}_{i_1\cdots i_l}(x,s)$ of (1.2) be expressible in the form (2.7), where the coefficients $C^{(m:\alpha)}_{i_1\cdots i_\alpha}(x)$ occurring in (2.7) satisfy (2.9).

Substitution of (2.7) into (1.2) shows that $I^{(m)}$ must have the form[8]

$$I^{(m)} = \sum_{l=0}^{m} \sum_{\alpha=0}^{l} \frac{(-s)^{l-\alpha}}{(l-\alpha)!} C^{(m:\alpha)}_{(i_1\cdots i_\alpha;i_{\alpha+1}\cdots i_l)} u^{i_1}\cdots u^{i_l}, \tag{2.10}$$

[where the $C^{(m:\alpha)}_{i_1\cdots i_\alpha}$ satisfy (2.9)].

By regrouping the terms in (2.10) we may express the mth-order first integral $I^{(m)}$ in the form

$$I^{(m)} = \sum_{\alpha=0}^{m} I^{(m:\alpha)}, \tag{2.11}$$

where

$$I^{(m:\alpha)} \equiv \sum_{l=0}^{m-\alpha} \frac{(-s)^l}{l!} C^{(m:\alpha)}_{i_1\cdots i_\alpha;i_{\alpha+1}\cdots i_{\alpha+l}} u^{i_1}\cdots u^{i_{\alpha+l}},$$
$$\alpha = 0,1,\ldots,m. \tag{2.12}$$

Each set of Eqs. (2.9) determined by a given α (say $\alpha = \alpha'$) does not involve C's defined by other choices of α. It follows that (2.9) may be satisfied by choosing all the C-tensors equal to zero except the one defined by $\alpha = \alpha'$. In this case $I^{(m)}$ reduces to $I^{(m)} = I^{(m:\alpha')}$, which implies $I^{(m:\alpha')}$ is a first integral for $\alpha' = 0,1,2,\ldots,m$. It also follows that a necessary condition for $I^{(m:\alpha')}$ to be a mth-order first integral is that (2.9) be satisfied for $\alpha = \alpha'$.

To show that for a given m and α that (2.9) is sufficient for $I^{(m:\alpha)}$ of (2.12) to be a first integral, one may calculate $DI^{(m:\alpha)}/ds$ and observe that all terms cancel in pairs except a term associated with $l = m - \alpha$, which vanishes by use of (2.9). It thus follows that (2.9) is also sufficient for $I^{(m)}$ of (2.10) to be a first integral.

We summarize the above in the following theorems and corollary.

Theorem 2.1: In a Riemannian space V_n a necessary and sufficient condition for the geodesic equations

$$Du^i/ds = 0, \quad u^i \equiv dx^i/ds \tag{1.1'}$$

to admit an mth-order first integral of the form

$$I^{(m)} = \sum_{l=0}^{m} M^{(m:l)}_{i_1\cdots i_l}(x,s) u^{i_1}\cdots u^{i_l}, \tag{1.2'}$$

(which in general is inhomogeneous in the tangent vectors u^i and has explicit dependence on the path parameter s) is that the coefficients $M^{(m:l)}_{i_1\cdots i_l}$ must be polynomials in s which have the form

$$M^{(m:l)}_{i_1\cdots i_l} = \sum_{\alpha=0}^{l} \frac{(-s)^{l-\alpha}}{(l-\alpha)!} C^{(m:\alpha)}_{(i_1\cdots i_\alpha;i_{\alpha+1}\cdots i_l)}, \quad l = 0,1,\ldots,m, \tag{2.7'}$$

where the α-rank symmetric tensors $C^{(m:\alpha)}_{i_1\cdots i_\alpha}(x)$ are functions of x^i which satisfy

$$C^{(m:\alpha)}_{(i_1\cdots i_\alpha;i_{\alpha+1}\cdots i_{m+1})} = 0, \quad \alpha = 0,1,\ldots,m. \tag{2.9'}$$

Theorem 2.2: For a given r the expression

$$I^{(m:r)} = \sum_{l=0}^{m-r} \frac{(-s)^l}{l!} C^{(m:r)}_{i_1\cdots i_r;i_{r+1}\cdots i_{r+l}} u^{i_1}\cdots u^{i_{r+l}} \quad (0 \leqslant r \leqslant m), \tag{2.12'}$$

TABLE II. Case $m = 1$. Linear first integrals. $I^{(1)} = I^{(1:0)} + I^{(1:1)}$.

$I^{(1)}$	$M^{(1:0)}$	$M^{(1:1)}_{i_1} u^{i_1}$	Condition
$I^{(1:0)}$	$C^{(1:0)}$	$-sC^{(1:0)}_{;i_1} u^{i_1}$	$C^{(1:0)}_{;i_1 i_2} = 0$
$I^{(1:1)}$		$C^{(1:1)}_{i_1} u^{i_1}$	$C^{(1:1)}_{(i_1;i_2)} = 0$

TABLE III. Case $m = 2$. Quadratic first integrals. $I^{(2)} = I^{(2:0)} + I^{(2:1)} + I^{(2:2)}$.

$I^{(2)}$	$M^{(2:0)}$	$M^{(2:1)}_{i_1} u^{i_1}$	$M^{(2:2)}_{i_1 i_2} u^{i_1} u^{i_2}$	Condition
$I^{(2:0)}$	$C^{(2:0)}$	$-sC^{(2:0)}_{;i_1} u^{i_1}$	$(s^2/2!)C^{(2:0)}_{;i_1 i_2} u^{i_1} u^{i_2}$	$C^{(2:0)}_{;(i_1 i_2 i_3)} = 0$
$I^{(2:1)}$		$C^{(2:1)}_{i_1} u^{i_1}$	$-sC^{(2:1)}_{i_1;i_2} u^{i_1} u^{i_2}$	$C^{(2:1)}_{(i_1;i_2 i_3)} = 0$
$I^{(2:2)}$			$C^{(2:2)}_{i_1 i_2} u^{i_1} u^{i_2}$	$C^{(2:2)}_{(i_1 i_2;i_3)} = 0$

will be an mth-order first integral of the goedesics if and only if the r-rank symmetric tensor $C^{(m:r)}_{i_1\cdots i_r}(x)$ satisfies

$$C^{(m:r)}_{(i_1\cdots i_r;i_{r+1}\cdots i_{m+1})} = 0. \tag{2.13}$$

Corollary 2.A: The integral $I^{(m)}$ of Theorem 2.1 can be represented in the form

$$I^{(m)} = \sum_{r=0}^{m} I^{(m:r)}, \tag{2.11$'$}$$

where $I^{(m:r)}$ are mth-order first integrals as defined in Theorem 2.2.

For illustration purposes and later reference we give in Tables II, III, and IV for the cases $m = 1,2,3$ respectively the detailed expansions of the formulas occurring in Theorems 2.1, 2.2, and Corollary 2.A.

Remarks concerning Tables II, III, IV: In each table the sum of the terms in the row (excluding the entry in the "condition" column) marked $I^{(m:r)}$ corresponds to (2.12′) of Theorem 2.2. For example, from Table III we read (for $m = 2$, $r = 1$)

$$I^{(2:1)} = C^{(2:1)}_{i_1} u^{i_1} - sC^{(2:1)}_{i_1;i_2} u^{i_1} u^{i_2}. \tag{2.14}$$

Also in each table the sum of terms in the column headed $M^{(m:l)}_{i_1\cdots i_l} u^{i_1}\cdots u^{i_l}$ (for a given m and l) corresponds to a term in (1.2′) of Theorem 2.1. For example, from Table III we read

$$M^{(2:2)}_{i_1 i_2} u^{i_1} u^{i_2} = \left(\frac{s^2}{2!} C^{(2:0)}_{;i_1 i_2} - sC^{(2:1)}_{i_1;i_2} + C^{(2:2)}_{i_1 i_2}\right) u^{i_1} u^{i_2}. \tag{2.15}$$

An entry in the last column marked "condition" corresponds to (2.13) of Theorem 2.2. For example, in Table III the equation

$$C^{(2:1)}_{(i_1;i_2 i_3)} = 0 \tag{2.16}$$

represents the necessary and sufficient condition that $I^{(2:1)}$ be a quadratic first integral.

III. SPECIAL *M*TH-ORDER FIRST INTEGRALS

We now shall show how the existence of an mth-order first integral $I^{(m:r)}$ $(0\leqslant r\leqslant m-1)$ of the type described by (2.12′) of Theorem 2.2 implies the existence of additional mth-order first integrals which we shall call *special* mth-order first integrals.

Assume then that for some r $(0\leqslant r\leqslant m-1)$ there exists an r-rank symmetric tensor $C^{(m:r)}_{i_1\cdots i_r}$ such that (2.13) of Theorem 2.2 is satisfied and hence an mth-order first integral $I^{(m:r)}$ (2.12′) exists. Based upon this $C^{(m:r)}_{i_1\cdots i_r}$ we define a symmetric tensor ${}^*C^{(m:r+k)}_{i_1\cdots i_{r+k}}$ of rank $r+k$ $(1\leqslant k\leqslant m-r)$ by the relationship

$${}^*C^{(m:r+k)}_{i_1\cdots i_{r+k}} \equiv C^{(m:r)}_{(i_1\cdots i_r;i_{r+1}\cdots i_{r+k})}. \tag{3.1}$$

From (3.1) we observe that covariant differentiation of ${}^*C^{(m:r+k)}_{i_1\cdots i_{r+k}}$ followed by symmetrization yields

$${}^*C^{(m:r+k)}_{(i_1\cdots i_{r+k};i_{r+k+1}\cdots i_{m+1})} = C^{(m:r)}_{(i_1\cdots i_r;i_{r+1}\cdots i_{r+k}i_{r+k+1}\cdots i_{m+1})}. \tag{3.2}$$

By the above assumption that $C^{(m:r)}_{i_1\cdots i_r}$ satisfies (2.13) it follows the right-hand side of (3.2) is zero and hence so also is the left-hand side, i.e.,

$${}^*C^{(m:r+k)}_{(i_1\cdots i_{r+k};i_{r+k+1}\cdots i_{m+1})} = 0. \tag{3.3}$$

By comparison of (3.3) with (2.13) it is observed that the $(r+k)$-rank symmetric tensor ${}^*C^{(m:r+k)}_{i_1\cdots i_{r+k}}$ defined by (3.1) will satisfy (2.13) with r replaced by $r+k$. By Theorem 2.2 Eq. (3.3) is a necessary and sufficient condition for the expression ${}^*I^{(m:r+k)}$ defined by

$${}^*I^{(m:r+k)} \equiv \sum_{l=0}^{m-(r+k)} \frac{(-s)^l}{l!} {}^*C^{(m:r+k)}_{i_1\cdots i_{r+k};i_{r+k+1}\cdots i_{r+k+l}} u^{i_1}\cdots u^{i_{r+k+l}}$$
$$(1\leqslant k\leqslant m-r, \quad 0\leqslant r\leqslant m-1), \tag{3.4}$$

to be an mth-order first integral. By use of (3.1) this integral can be rewritten in the form

$${}^*I^{(m:r+k)} = \sum_{l=0}^{m-(r+k)} \frac{(-s)^l}{l!} C^{(m:r)}_{i_1\cdots i_r;i_{r+1}\cdots i_{r+k}\cdots i_{r+k+l}} u^{i_1}\cdots u^{i_{r+k+l}}$$
$$(l\leqslant k\leqslant m-r), \quad 0\leqslant r\leqslant m-1), \tag{3.5}$$

which we refer to as a *special mth-order first integral of type* $(r+k)$.

The above discussion leads to the following theorem and corollaries.

Theorem 3.1: In a Riemannian space V_n the condition

$$C^{(m:r)}_{(i_1\cdots i_r;i_{r+1}\cdots i_{m+1})} = 0, \quad 0\leqslant r\leqslant m-1,$$

on the symmetric r-rank tensor $C^{(m:r)}_{i_1\cdots i_r}(x)$ is necessary and sufficient for the existence of each of the $N = m - r + 1$ mth-order geodesic first integrals

$$I^{(m:r)} \equiv \sum_{l=0}^{m-r} \frac{(-s)^l}{l!} C^{(m:r)}_{i_1\cdots i_r;i_{r+1}\cdots i_{r+l}} u^{i_1}\cdots u^{i_{r+l}} \quad (0\leqslant r\leqslant m-1), \tag{3.6}$$

TABLE IV. Case $m = 3$. Cubic first integrals. $I^{(3)} = I^{(3:0)} + I^{(3:1)} + I^{(3:2)} + I^{(3:3)}$.

$I^{(3)}$	$M^{(3:0)}$	$M^{(3:1)}_{i_1} u^{i_1}$	$M^{(3:2)}_{i_1 i_2} u^{i_1} u^{i_2}$	$M^{(3:3)}_{i_1 i_2 i_3} u^{i_1} u^{i_2} u^{i_3}$	Condition
$I^{(3:0)}$	$C^{(3:0)}$	$-sC^{(3:0)}_{;i_1} u^{i_1}$	$(s^2/2!)C^{(3:0)}_{;i_1 i_2} u^{i_1} u^{i_2}$	$-(s^3/3!)C^{(3:0)}_{;i_1 i_2 i_3} u^{i_1} u^{i_2} u^{i_3}$	$C^{(3:0)}_{;(i_1 i_2 i_3 i_4)} = 0$
$I^{(3:1)}$		$C^{(3:1)}_{i_1} u^{i_1}$	$-sC^{(3:1)}_{i_1;i_2} u^{i_1} u^{i_2}$	$(s^2/2!)C^{(3:1)}_{i_1;i_2 i_3} u^{i_1} u^{i_2} u^{i_3}$	$C^{(3:1)}_{(i_1;i_2 i_3 i_4)} = 0$
$I^{(3:2)}$			$C^{(3:2)}_{i_1 i_2} u^{i_1} u^{i_2}$	$-sC^{(3:2)}_{i_1 i_2;i_3} u^{i_1} u^{i_2} u^{i_3}$	$C^{(3:2)}_{(i_1 i_2;i_3 i_4)} = 0$
$I^{(3:3)}$				$C^{(3:3)}_{i_1 i_2 i_3} u^{i_1} u^{i_2} u^{i_3}$	$C^{(3:3)}_{(i_1 i_2 i_3;i_4)} = 0$

TABLE V. Linear first integrals concomitant with $I^{(1:0)}$.

Type	First integrals with nec. and suff. cond. $C^{(1:0)}_{;i_1 i_2} = 0$		$\doteq$
$I^{(1:0)}$	$C^{(1:0)}$	$-sC^{(1:0)}_{;i_1}u^{i_1}$	$k^{(1:0)}$
$*I^{(1:0+1)}$		$C^{(1:0)}_{;i_1}u^{i_1}$	$k^{(1:0+1)}$

$$*I^{(m:r+k)} \equiv \sum_{l=0}^{m-(r+k)} \frac{(-s)^l}{l!} C^{(m:r)}_{i_1 \cdots i_r ; i_{r+1} \cdots i_{r+k+l}} u^{i_1} \cdots u^{i_{r+k+l}}$$
$$(0 \leqslant r \leqslant m-1,\ k = 1,\ldots,m-r). \tag{3.5'}$$

Corollary 3.*A*: If the geodesics in a V_n admit at least one of the set

$$S^{(m:r)} \equiv \{I^{(m:r)}, *I^{(m:r+1)}, *I^{(m:r+2)}, \ldots, *I^{(m:r+(m-r))}\}$$

of $N = m - r + 1$ mth-order first integrals (described in Theorem 3.1), then the geodesics will admit all N of the integrals in the set $S^{(m:r)}$.

We illustrate Theorem 3.1 and Corollary 3.A for the values $m = 1,2,3$ and all applicable values of k and r by means of Tables V–X.

Remarks concerning Tables V–X: In each of the Tables V–X the row $I^{(m:r)}$, which represents an integral $I^{(m:r)}$, $r = 0,\ldots m-1$, has been selected from either Table II, III, or IV, depending upon the vaue of m. The remaining rows in Tables V–X are special first integrals and are designated by $*I^{(m:r+k)}$, $k = 1,\ldots,m-r$, as described in Theorem 3.1. For each row the entry in the column headed "$\doteq$" represents the numerical value which the integral in that row assumes along a geodesic. Note that all the integrals in a given table have the same necessary and sufficient condition as described in Theorem 3.1. This implies that if any row of a given table is known to be a first integral then all rows in that table will be first integrals as set forth in Corollaries 3.A and 3.B.

With reference to Tables V–X we shall now show how each set of mth-order first integrals associated with a particular necessary and sufficient condition may be expressed in a simplified alternative form. We shall illustrate this procedure for the set of integrals in Table VIII.

From Table VIII by means of the integral $*I^{(3:0+3)}$ the integral $*I^{(3:0+2)}$ may be expressed in the form $*\tilde{I}^{(3:0+3)}$, where

$$*\tilde{I}^{(3:0+2)} \equiv C^{(3:0)}_{;i_1 i_2}u^{i_1}u^{i_2} - sk^{(3:0+3)} \doteq k^{(3:0+2)}. \tag{3.7}$$

By means of $*I^{(3:0+3)}$ and (3.7) the integral $*I^{(3:0+1)}$ may be expressed in the form $*\tilde{I}^{(3:0+1)}$, where

$$*\tilde{I}^{(3:0+1)} \equiv C^{(3:0)}_{;i_1}u^{i_1} - sk^{(3:0+2)} - (s^2/2!)k^{(3:0+3)} \doteq k^{(3:0+1)}. \tag{3.8}$$

TABLE VI. Quadratic first integrals concomitant with $I^{(2:0)}$.

Type	First integrals with nec. and suff. cond. $C^{(2:0)}_{;(i_1 i_2 i_3)} = 0$			$\doteq$
$I^{(2:0)}$	$C^{(2:0)}$	$-sC^{(2:0)}_{;i_1}u^{i_1}$	$(s^2/2!)C^{(2:0)}_{;i_1 i_2}u^{i_1}u^{i_2}$	$k^{(2:0)}$
$*I^{(2:0+1)}$		$C^{(2:0)}_{;i_1}u^{i_1}$	$-sC^{(2:0)}_{;i_1 i_2}u^{i_1}u^{i_2}$	$k^{(2:0+1)}$
$*I^{(2:0+2)}$			$C^{(2:0)}_{;i_1 i_2}u^{i_1}u^{i_2}$	$k^{(2:0+2)}$

TABLE VII. Quadratic first integrals concomitant with $I^{(2:1)}$.

Type	First integrals with nec. and suff. cond. $C^{(2:1)}_{(i_1;i_2 i_3)} = 0$		$\doteq$
$I^{(2:1)}$	$C^{(2:1)}_{i_1}u^{i_1}$	$-sC^{(2:1)}_{i_1;i_2}u^{i_1}u^{i_2}$	$k^{(2:1)}$
$*I^{(2:1+1)}$		$C^{(2:1)}_{i_1;i_2}u^{i_1}u^{i_2}$	$k^{(2:1+1)}$

Finally by means of $*I^{(3:0+3)}$, (3.7), and (3.8) the integral $I^{(3:0)}$ may be expressed in the form $*\tilde{I}^{(3:0)}$, where

$$*\tilde{I}^{(3:0)} \equiv C^{(3:0)} - sk^{(3:0+1)} - (s^2/2!)k^{(3:0+2)} - (s^3/3!)k^{(3:0+3)}$$
$$\doteq k^{(3:0)}. \tag{3.9}$$

By a similar procedure the set of integrals in each of the Tables V–X may be rewritten to obtain Tables VA–XA, respectively.

With reference to the integrals in Tables VA–XA we next point out an alternative interpretation, which is suggested as a consequence of transposing to the right-hand side of the equal marks all terms in which s appears explicitly. For example, if in (3.8) (refer to Table VIII A) the terms involving s are transposed to the right-hand side, the resulting equation has the form

$$C^{(3:0)}_{;i_1}u^{i_1} \doteq k^{(3:0+1)} + sk^{(3:0+2)} + (s^2/2!)k^{(3:0+3)}. \tag{3.10}$$

The left-hand side of (3.10) may be regarded as a "pseudolinear first integral" in the sense that, although it is not constant along the geodesic, its variation along the geodesic is expressible as a (quadratic) polynomial (with constant coefficients) in the path parameter s. Similarly the leading term of each $*\tilde{I}$ or $\tilde{I}$ type first integral in Tables VA–XA may be expressed as a pseudofirst integral. For example, from Table IXA we obtain from $*\tilde{I}^{(3:1+1)}$ the pseudoquadratic first integral $C^{(3:1)}_{i_1;i_2}u^{i_1}u^{i_2}$ in that

$$C^{(3:1)}_{i_1;i_2}u^{i_1}u^{i_2} \doteq k^{(3:1+1)} + sk^{(3:1+2)}. \tag{3.11}$$

(We note this pseudoquadratic first integral varies linearly with s along a geodesic.)

It is of interest to observe that the derivative D/ds (along a geodesic) of a pseudofirst integral associated with row r of an A Table will give a pseudofirst (or first) integral associated with row $r+1$ of the table. For example, from the pseudo-integral (3.10) which is associated with row two of Table VIIIA we obtain by differentiation the pseudo-integral

$$C^{(3:0)}_{;i_1 i_2}u^{i_1}u^{i_2} \doteq k^{(3:0+2)} + sk^{(3:0+3)} \tag{3.12}$$

associated with row three.

It is clear how the Tables in Secs. 2 and 3 based on $m = 1,2,3$ and the above remarks concerning pseudofirst integrals can be extended to general values of m.

IV. LINEAR FIRST INTEGRALS WITH EXPLICIT PATH-PARAMETER DEPENDENCE

In this section we shall elaborate on linear first integrals of the type $I^{(1:0)}$ $(m = 1, r = 0)$ which appear in Table V. Such integrals are based on a scalar $C^{(1:0)}(x)$. A necessary and sufficient condition for the existence of these integrals is

TABLE VIII. Cubic first integrals concomitant with $I^{(3:0)}$.

Type	First integrals with nec. and suff. cond. $C^{(3:0)}_{;(i_1i_2i_3i_4)} = 0$				$\doteq$
$I^{(3:0)}$	$C^{(3:0)}$	$-sC^{(3:0)}_{;i_1}u^{i_1}$	$(s^2/2!)C^{(3:0)}_{;i_1i_2}u^{i_1}u^{i_2}$	$-(s^3/3!)C^{(3:0)}_{;i_1i_2i_3}u^{i_1}u^{i_2}u^{i_3}$	$k^{(3:0)}$
$*I^{(3:0+1)}$		$C^{(3:0)}_{;i_1}u^{i_1}$	$-sC^{(3:0)}_{;i_1i_2}u^{i_1}u^{i_2}$	$(s^2/2!)C^{(3:0)}_{;i_1i_2i_3}u^{i_1}u^{i_2}u^{i_3}$	$k^{(3:0+1)}$
$*I^{(3:0+2)}$			$C^{(3:0)}_{;i_1i_2}u^{i_1}u^{i_2}$	$-sC^{(3:0)}_{;i_1i_2i_3}u^{i_1}u^{i_2}u^{i_3}$	$k^{(3:0+2)}$
$*I^{(3:0+3)}$				$C^{(3:0)}_{;i_1i_2i_3}u^{i_1}u^{i_2}u^{i_3}$	$k^{(3:0+3)}$

$$C^{(1:0)}(x)_{;ij} = 0. \tag{4.1}$$

Equation (4.1) is recognized as the condition that the vector $C^{(1:0)}_{,i}$ defines a parallel field. Hence we may state the theorem.

Theorem 4.1: A necessary and sufficient condition that the geodesic equation (1.1) of a V_n admit a linear first integral with explicit dependence on the path parameter s is that the V_n admit a field of parallel vectors $C^{(1:0)}_{,i}$. Such integrals have the form

$$I^{(1:0)} \equiv C^{(1:0)} - sC^{(1:0)}_{,i}u^i \doteq k^{(1:0)}. \tag{4.2}$$

As is well known, every parallel vector field defines a motion (Killing vector) and hence the term $C^{(1:0)}_{,i}u^i$ appearing in (4.2) is also a linear first integral [refer to (a), Table I]. (This integral is denoted in Tables V and VA by $*I^{(1:0+1)}$.) Hence $I^{(1:0)}$ of (4.2) can be expressed in the form $\tilde{I}^{(1:0)}$ of Table VA.

We illustrate the above by means of the following example from general relativity. Consider the gravitational plane-wave space–time V_4 defined by the line element[9]

$$d\phi^2 = -(dx^1)^2 - (dx^2)^2 - (dx^3)^2 + (dx^4)^2 + 2f(x^1,x^2,Z)(dx^3 - dx^4)^2, \tag{4.3}$$

where $Z \equiv x^3 - x^4$ and the function f satisfies $f_{,11} + f_{,22} = 0$. Based on the scalar $C^{(1:0)} \equiv x^4 - x^3$ a (null) parallel vector field with components $C^{(1:0)}_{,i} = (0,0,-1,1)$ exists in this space–time.

The geodesic equations (1.1) in this V_4 take the form

$$\frac{du^1}{ds} + 2\frac{\partial f}{\partial x^1}u^3u^4 - \frac{\partial f}{\partial x^1}u^4u^4 = 0, \tag{4.4}$$

$$\frac{du^2}{ds} - \frac{\partial f}{\partial x^2}u^3u^3 + 2\frac{\partial f}{\partial x^2}u^3u^4 - \frac{\partial f}{\partial x^2}u^4u^4 = 0, \tag{4.5}$$

$$\frac{du^3}{ds} + 2\frac{\partial f}{\partial x^1}u^1u^3 - 2\frac{\partial f}{\partial x^1}u^1u^4 + 2\frac{\partial f}{\partial x^2}u^2u^3 - 2\frac{\partial f}{\partial x^2}u^2u^4 + \frac{\partial f}{\partial Z}u^3u^3 - 2\frac{\partial f}{\partial Z}u^3u^4 + \frac{\partial f}{\partial Z}u^4u^4 = 0, \tag{4.6}$$

$$\frac{du^4}{ds} + 2\frac{\partial f}{\partial x^1}u^1u^3 - 2\frac{\partial f}{\partial x^1}u^1u^4 + 2\frac{\partial f}{\partial x^2}u^2u^3 - 2\frac{\partial f}{\partial x^2}u^2u^4 + \frac{\partial f}{\partial Z}u^3u^3 - 2\frac{\partial f}{\partial Z}u^3u^4 + \frac{\partial f}{\partial Z}u^4u^4 = 0. \tag{4.7}$$

For the plane-wave space–time being considered the linear first integral with explicit path-parameter dependence (4.2) takes the form

$$I^{(1:0)} = x^4 - x^3 - s(u^4 - u^3), \tag{4.8}$$

and the first integral $*I^{(1:0+1)}$ of Table V takes the form

$$*I^{(1:0+1)} = u^4 - u^3. \tag{4.9}$$

Direct verification that (4.8) and (4.9) are first integrals follows immediately by use of the relation

$$\frac{du^4}{ds} - \frac{du^3}{ds} = \frac{d}{ds}(u^4 - u^3) \doteq 0, \tag{4.10}$$

obtained from the geodesic equations (4.6) and (4.7).

TABLE IX. Cubic first integrals concomitant with $I^{(3:1)}$.

Type	First integrals with nec. and suff. cond. $C^{(3:1)}_{(i_1;i_2i_3i_4)} = 0$			$\doteq$
$I^{(3:1)}$	$C^{(3:1)}_{i_1}u^{i_1}$	$-sC^{(3:1)}_{i_1;i_2}u^{i_1}u^{i_2}$	$(s^2/2!)C^{(3:1)}_{i_1;i_2i_3}u^{i_1}u^{i_2}u^{i_3}$	$k^{(3:1)}$
$*I^{(3:1+1)}$		$C^{(3:1)}_{i_1;i_2}u^{i_1}u^{i_2}$	$-sC^{(3:1)}_{i_1;i_2i_3}u^{i_1}u^{i_2}u^{i_3}$	$k^{(3:1+1)}$
$*I^{(3:1+2)}$			$C^{(3:1)}_{i_1;i_2i_3}u^{i_1}u^{i_2}u^{i_3}$	$k^{(3:1+2)}$

TABLE X. Cubic first integrals concomitant with $I^{(3:2)}$.

Type	First integrals with nec. and suff. cond. $C^{(3:2)}_{(i_1 i_2; i_3 i_4)} = 0$		$\doteq$
$I^{(3:2)}$	$C^{(3:2)}_{i_1 i_2} u^{i_1} u^{i_2}$	$- sC^{(3:2)}_{i_1 i_2; i_3} u^{i_1} u^{i_2} u^{i_3}$	$k^{(3:2)}$
$*I^{(3:2+1)}$		$C^{(3:2)}_{i_1 i_2; i_3} u^{i_1} u^{i_2} u^{i_3}$	$k^{(3:2+1)}$

V. QUADRATIC FIRST INTEGRALS WITH EXPLICIT PATH-PARAMETER DEPENDENCE

In this section we shall first examine more closely the explicit path-parameter dependent quadratic first integrals of the geodesics of a V_n. We observe from Tables VI and VII that the existence of such integrals depends upon the existence of certain scalar or vector fields.

With reference to Table VII we first consider the necessary and sufficient condition

$$C^{(2:1)}_{(i;jk)} = 0 \tag{5.1}$$

which the vector field $C^{(2:1)}_i$ must satisfy in order that the geodesics admit the quadratic first integrals $I^{(2:1)}$ and $*I^{(2:1+1)}$ of the above-stated type.

When the indicated symmetrization in (5.1) is carried out, the six terms which result may be grouped into three pairs in such a way that (5.1) is expressible in the equivalent form[10]

$$h_{ij;k} + h_{jk;i} + h_{ki;j} = 0, h_{ij} \equiv \mathscr{L}_{C^{(2:1)}} g_{ij}. \tag{5.2}$$

Alternatively, by use of the identity[11]

$$C^{(2:1)}_{i;jk} + C^{(2:1)}_{j;ki} + C^{(2:1)}_{k;ij} \equiv C^{(2:1)}_{i;kj} + C^{(2:1)}_{j;ik} + C^{(2:1)}_{k;ji}, \tag{5.3}$$

Eq. (5.1) [and hence (5.2)] can be given the equivalent form

$$C^{(2:1)}_{i;jk} + C^{(2:1)}_{j;ki} + C^{(2:1)}_{k;ij} = 0. \tag{5.4}$$

We combine in the theorem to follow the above results related to the vector-based quadratic first integrals of Table VII with the information related to scalar-based quadratic first integrals of Table VI.

Theorem 5.1: If the geodesics (1.1) in a V_n admit quadratic first integrals with explicit dependence on the path parameter s, then such integrals may be divided into two classes:

(i) those based upon the existence of a scalar $C^{(2:0)}(x)$ which satisfies a necessary and sufficient condition

$$C^{(2:0)}_{;(ijk)} = 0, \tag{5.5}$$

in which case the integrals are of the form

$$I^{(2:0)} \equiv C^{(2:0)} - sC^{(2:0)}_{,i} u^i + (s^2/2) C^{(2:0)}_{;ij} u^i u^j, \tag{5.6}$$

$$*I^{(2:0+1)} \equiv C^{(2:0)}_{,i} u^i - sC^{(2:0)}_{;ij} u^i u^j; \tag{5.7}$$

TABLE VIA. Combined quadratic first integrals concomitant with $I^{(2:0)}$.

Type	Combined first integrals with nec. and suff. cond. $C^{(2:0)}_{;(i_1 i_2 i_3)} = 0$			$\doteq$
$\tilde{I}^{(2:0)}$	$C^{(2:0)}$	$- sk^{(2:0+1)}$	$- (s^2/2!) k^{(2:0+2)}$	$k^{(2:0)}$
$*\tilde{I}^{(2:0+1)}$		$C^{(2:0)}_{;i_1} u^{i_1}$	$- sk^{(2:0+2)}$	$k^{(2:0+1)}$
$*I^{(2:0+2)}$			$C^{(2:0)}_{;i_1 i_2} u^{i_1} u^{i_2}$	$k^{(2:0+2)}$

(ii) those based upon the existence of a vector $C^{(2:1)}_i(x)$ which satisfies either of the equivalent necessary and sufficient conditions

$$h_{ij;k} + h_{jk;i} + h_{ki;j} = 0, \quad h_{ij} \equiv \mathscr{L}_{C^{(2:1)}} g_{ij}, \tag{5.2'}$$

or

$$C^{(2:1)}_{i;jk} + C^{(2:1)}_{j;ki} + C^{(2:1)}_{k;ij} = 0, \tag{5.4'}$$

in which case the integrals are of the form

$$I^{(2:1)} \equiv C^{(2:1)}_i u^i - sC^{(2:1)}_{i;j} u^i u^j. \tag{5.8}$$

We now consider two important special solutions of (5.2′) for vectors $C^{(2:1)}_i$ which define well-known geometric symmetries mentioned in Table I. By inspection of (b) and (c) of Table I it is seen that a proper homothetic motion or a proper affine collineation vector will satisfy (5.2′) in that for such vectors $h_{ij;k} = 0$. Hence by Theorem 5.1 (ii) each of these vectors will define a quadratic first integral with explicit path-parameter dependence in addition to the well-known concomitant first integral with no explicit path-parameter dependence given in Table I. We may therefore state the theorem which follows:

Theorem 5.2: If a V_n admits either

(i) a proper affine collineation in that there exists a vector $\xi^i(\mathrm{AC})$ such that $h_{ij;k} = 0$, $h_{ij} \equiv \mathscr{L}_{\xi(\mathrm{AC})} g_{ij} \neq 0$, or

(ii) a proper homothetic motion in that there exists a vector $\xi^i(\mathrm{HM})$ such that $h_{ij} = 2\sigma_0 g_{ij}$, $\sigma_0 \equiv \text{const} \neq 0$, $h_{ij} \equiv \mathscr{L}_{\xi(\mathrm{HM})} g_{ij}$, then the geodesics (1.1) of the V_n will admit an inhomogeneous quadratic first integral with explicit dependence on the path parameter s. In either case (i) or (ii) this integral is expressible in the form

$$I^{(2:1)} = \xi_i u^i - s\xi_{i;j} u^i u^j, \tag{5.9}$$

where in case (i) $\xi^i \equiv \xi^i(\mathrm{AC})$, and in case (ii) $\xi^i \equiv \xi^i(\mathrm{HM})$.

By use of the relation $\mathscr{L}_\xi g_{ij} \equiv \xi_{i;j} + \xi_{j;i}$ Eq. (5.9) is expressible in the form

$$I^{(2:1)} = \xi_i u^i - \tfrac{1}{2} s(\mathscr{L}_\xi g_{ij}) u^i u^j. \tag{5.10}$$

It then follows for case (ii), in which ξ^i defines a proper homothetic motion, that (5.10) can be written in the form

$$I^{(2:1)}_{\mathrm{HM}} = \xi_i(HM) u^i - s\sigma_0 g_{ij} u^i u^j. \tag{5.11}$$

TABLE VA. Combined linear first integrals concomitant with $I^{(1:0)}$.

Type	Combined first integrals with nec. and suff. con. $C^{(1:0)}_{;(i_1 i_2)} = 0$		$\doteq$
$\tilde{I}^{(1:0)}$	$C^{(1:0)}$	$- sk^{(1:0+1)}$	$k^{(1:0)}$
$*I^{(1:0+1)}$		$C^{(1:0)}_{;i_1} u^{i_1}$	$k^{(1:0+1)}$

TABLE VIIA. Combined quadratic first integrals concomitant with $I^{(2:1)}$.

Type	Combined first integrals with nec. and suff. cond. $C^{(2:1)}_{(i_1; i_2 i_3)} = 0$		$\doteq$
$\tilde{I}^{(2:1)}$	$C^{(2:1)}_{i_1} u^{i_1}$	$- sk^{(2:1+1)}$	$k^{(2:1)}$
$*I^{(2:1+1)}$		$C^{(2:1)}_{i_1; i_2} u^{i_1} u^{i_2}$	$k^{(2:1+1)}$

TABLE VIIIA. Combined cubic first integrals concomitant with $I^{(3:0)}$.

Type	Combined first integrals with nec. and suff. cond. $C^{(3:0)}_{;(i_1 i_2 i_3 i_4)} = 0$				$\doteq$
$\tilde{I}^{(3:0)}$	$C^{(3:0)}$	$-sk^{(3:0+1)}$	$-(s^2/2!)k^{(3:0+2)}$	$-(s^3/3!)k^{(3:0+3)}$	$k^{(3:0)}$
$*\tilde{I}^{(3:0+1)}$		$C^{(3:0)}_{;i_1}u^{i_1}$	$-sk^{(3:0+2)}$	$-(s^2/2!)k^{(3:0+3)}$	$k^{(3:0+1)}$
$*\tilde{I}^{(3:0+2)}$			$C^{(3:0)}_{;i_1 i_2}u^{i_1}u^{i_2}$	$-sk^{(3:0+3)}$	$k^{(3:0+2)}$
$*I^{(3:0+3)}$				$C^{(3:0)}_{;i_1 i_2 i_3}u^{i_1}u^{i_2}u^{i_3}$	$k^{(3:0+3)}$

By use of (1.3) in (5.11) we are led to the following corollary to Theorem 5.2 (ii).

Corollary 5.A: If a V_n admits a homothetic motion [$\xi^i = \xi^i$(HM), defined by (b) Table I] the expression

$$L(\text{HM}) = \xi_i(\text{HM})u^i - \epsilon\sigma_0 s \tag{5.12}$$

will be a linear first integral of nonnull geodesics when $\epsilon = \pm 1$ (in which case the parameter s is the arclength, or a linear first integral of null geodesics when $\epsilon = 0$. The integral L (HM) is a degenerate form of the quadratic first integral $I^{(2:1)}$ of Theorem 5.2.[12]

To illustrate Theorem 5.2(ii) and Corollary 5.A we consider the space–time with line element[13]

$$d\phi^2 = (x^4)^{2a_0}[(dx^1)^2 + (dx^2)^2 + (dx^3)^2] - (dx^4)^2, \quad a_0 \equiv \text{const.} \tag{5.13}$$

This space–time admits a homothetic motion defined by the vector ξ^i(HM) with components

$$\xi^\nu(\text{HM}) = (1 - a_0)x^\nu, \quad \nu = 1,2,3, \quad \xi^4(\text{HM}) = x^4, \tag{5.14}$$

and scale factor $\sigma_0 = 1$. For this space–time the geodesic equations (1.1) take the form

$$du^\nu/ds + 2a_0(x^4)^{-1}u^\nu u^4 = 0, \quad \nu = 1,2,3, \tag{5.15}$$

$$du^4/ds + a_0(x^4)^{2a_0 - 1}u^\nu u^\nu = 0. \tag{5.16}$$

Based on the homothetic motion vector (5.14) and scale factor σ_0 given above, we have from (5.12) that the geodesic equations (5.15) and (5.16) admit the first integral

$$I^{(2:1)}_{\text{HM}} = (1 - a_0)(x^4)^{2a_0}x^\nu u^\nu - x^4u^4 - \epsilon s. \tag{5.17}$$

It is easily verified that $I^{(2:1)}_{\text{HM}}$ is a first integral by showing $dI^{(2:1)}_{\text{HM}}/ds \doteq 0$ with the aid of (5.15), (5.16) and the metrical first integral $(x^4)^{2a_0}u^\nu u^\nu - u^4u^4 \doteq \epsilon$.

Another illustration of parameter-dependent quadratic first integrals is contained in the example given at the end of Sec. VI.

We now continue the investigation which led to Theorem 5.2 to determine if vectors ξ^i which define certain symmetries (refer to Table I) other than the above discussed affine collineation and homothetic motions could also satisfy (5.2′) [or equivalently (5.4′)] of Theorem 5.1.

First we consider whether or not (5.2′) could be satisfied by a vector $C^{(2:1)}_i \equiv \xi_i$(PC) which defines a projective collineation [refer to (d), Table I]. For this case it would be necessary that (5.2′) and the projective collineation condition

$$h_{ij;k} = 2g_{ij}\phi_{,k} + g_{jk}\phi_{,i} + g_{ik}\phi_{,j}, \quad h_{ij} \equiv \mathscr{L}_{\xi(\text{PC})}g_{ij}, \tag{5.18}$$

be satisfied by the same vector ξ^i. When $h_{ij;k}$ as determined by (5.18) is used in (5.2′), we find that $\phi_{,i} = 0$, which implies that $h_{ij;k} = 0$, and hence the assumed projective collineation is reduced to an affine collineation.

In a similar fashion it can be shown that if a conformal collineation [refer to (h), Table I] vector ξ^i(CONFC) is to satisfy (5.2′), the conformal collineation is reduced to an affine collineation.

It is also easily shown that if (5.2′) is to be satisfied by a conformal motion vector ξ^i(CM) [refer to (f), Table I], then the conformal motion reduces to a homothetic motion (which is a subcase of an affine collineation).

As discussed above every affine collineation vector will satisfy condition (5.2′). Hence we may state the following theorems:

Theorem 5.3: In a V_n a vector ξ^i(PC) which defines a projective collineation (refer to Table I) will also be a solution to (5.2′) of Theorem 5.1 if and only if the projective collineation is an affine collineation.

Theorem 5.4: In a V_n a vector ξ^i(CONFC) which defines a conformal collineation (refer to Table I) will also be a solution to (5.2′) of Theorem 5.1 if and only if the conformal collineation is an affine collineation.

Theorem 5.5: In a V_n a vector ξ^i(CM) which defines a conformal motion (refer to Table I) will also be a solution to (5.2′) of Theorem 5.1 if and only if the conformal motion is a homothetic motion.

To conclude our discussion of the relationships of certain well-known symmetries to those vectors which satisfy (5.2′), we now determine a *necessary* condition which must be satisfied by a curvature collineation vector in order for it to satisfy (5.2′).

TABLE IXA. Combined cubic first integrals concomitant with $I^{(3:1)}$.

Type	Combined first integrals with nec. and suff. cond. $C^{(3:1)}_{(i_1;i_2 i_3 i_4)} = 0$			$\doteq$
$\tilde{I}^{(3:1)}$	$C^{(3:1)}_{i_1}u^{i_1}$	$-sk^{(3:1+1)}$	$-(s^2/2!)k^{(3:1+2)}$	$k^{(3:1)}$
$*\tilde{I}^{(3:1+1)}$		$C^{(3:1)}_{i_1;i_2}u^{i_1}u^{i_2}$	$-sk^{(3:1+2)}$	$k^{(3:1+1)}$
$*I^{(3:1+2)}$			$C^{(3:1)}_{i_1;i_2 i_3}u^{i_1}u^{i_2}u^{i_3}$	$k^{(3:1+2)}$

TABLE XA. Combined cubic first integrals concomitant with $I^{(3:2)}$.

Type	Combined first integrals with nec. and suff. cond. $C^{(3:2)}_{(i_1 i_2; i_3 i_4)} = 0$		$\doteq$
$\tilde{I}^{(3:2)}$	$C^{(3:2)}_{i_1 i_2} u^{i_1} u^{i_2}$	$-sk^{(3:2+1)}$	$k^{(3:2)}$
$*I^{(3:2+1)}$		$C^{(3:2)}_{i_1 i_2; i_3} u^{i_1} u^{i_2} u^{i_3}$	$k^{(3:2+1)}$

A curvature collineation is defined by a vector ξ^i(CC) which satisfies the condition[5]

$$\mathscr{L}_\xi R^i_{\ klj} \equiv (\mathscr{L}_\xi \Gamma^i_{jk})_{;l} - (\mathscr{L}_\xi \Gamma^i_{lk})_{;j} = 0. \tag{5.19}$$

It can be shown that a necessary condition for a curvature collineation is

$$h_{ij;kl} - h_{ij;lk} = 0, \quad h_{ij} \equiv \mathscr{L}_\xi g_{ij}. \tag{5.20}$$

We first use the identity[14]

$$\mathscr{L}_\xi \Gamma^i_{jk} \equiv \tfrac{1}{2} g^{il}(h_{jl;k} + h_{kl;j} - h_{jk;l}) \tag{5.21}$$

to express (5.2′) in the equivalent form

$$\mathscr{L}_\xi \Gamma^i_{jk} = -g^{il} h_{jk;l}. \tag{5.22}$$

By use of (5.22) in (5.19) we find that if a vector ξ^i is to define a curvature collineation and also satisfy (5.2′), then it must satisfy

$$h_{lk;ij} - h_{jk;il} = 0. \tag{5.23}$$

By covariant differentiation of (5.2′) we obtain

$$h_{ij;kl} + h_{jk;il} + h_{ki;jl} = 0. \tag{5.24}$$

From (5.24) by use of (5.20), (5.23), and the fact that $h_{ij} = h_{ji}$ we find the curvature collineation must satisfy

$$h_{ij;kl} = 0, \tag{5.25}$$

which [refer to (j) of Table I] is the condition for a special curvature collineation. We may thus state the following theorem:

Theorem 5.6: In a V_n a necessary condition for a curvature collineation vector ξ^i(CC) to be a solution of (5.2′) of Theorem 5.1 is that the vector ξ^i(CC) define a special curvature collineation (refer to Table I).

We continue our analysis of (5.2′) by showing how the existence in Riemannian space–time of a vector ξ^i which satisfies (5.2′) implies the existence of a concomitant conserved 4-current.

It follows by use of (5.22) [the above-derived equivalent form of (5.2′)] in the identity given in (5.19) that a necessary condition for a vector ξ^i to satisfy (5.2′) is

$$\mathscr{L}_\xi R^i_{\ klj} = g^{im}(h_{lk;mj} - h_{jk;ml}). \tag{5.26}$$

Contraction of i and j in (5.26) gives

$$\mathscr{L}_\xi R_{kl} = g^{im}(h_{lk;mi} - h_{ik;ml}). \tag{5.27}$$

By use of (5.24) with a suitable change of indices we may rewrite (5.27) in the form

$$\mathscr{L}_\xi R_{kl} = -g^{im}(h_{km;li} + h_{ml;ki} + h_{ik;ml}). \tag{5.28}$$

From (5.28) we obtain the equation

$$g^{kl} \mathscr{L}_\xi R_{kl} = -3h^{\ kl}_{;kl}. \tag{5.29}$$

By use of (5.29) in the known identity[15]

$$(R^l_{\ k} \xi^k)_{;l} - \tfrac{1}{2} g^{kl} \mathscr{L}_\xi R_{kl} \equiv 0, \tag{5.30}$$

we obtain as a necessary condition for a vector ξ^i to satisfy (5.2′)

$$\left[R^l_{\ k} \xi^k + \tfrac{3}{2} h^{\ kl}_{;k}\right]_{;l} = 0. \tag{5.31}$$

By means of the Einstein field equations in the form[16]

$$R^l_{\ k} = T^l_{\ k} - \tfrac{1}{2} T \delta^l_k \tag{5.32}$$

condition (5.31) is expressible in the form of conserved 4-current

$$J^l_{\ ;l} = 0, \tag{5.33}$$

where the current vector J^i is defined by

$$J^l \equiv (T^l_{\ k} - \tfrac{1}{2} T \delta^l_k) \xi^k + \tfrac{3}{2} h^{\ kl}_{;k}. \tag{5.34}$$

An alternative form for J^l will now be obtained. From the identity $h^{kl} \equiv g^{ki} g^{lm} h_{im}$ and the definition $h_{ij} \equiv \xi_{i;j} + \xi_{j;i}$ we obtain the identity

$$h^{\ il}_{;k} \equiv g^{lm}(\Box^2 \xi_m + \xi^{\ k}_{;mk}) \quad (\Box^2 \xi_m \equiv g^{ij} \xi_{m;ij}). \tag{5.35}$$

By of contraction of (5.4′) with g^{ik} [and the notational change $C^{(2:1)}_i = \xi_i$] we obtain

$$\Box^2 \xi_m + \xi^{\ k}_{;mk} + \xi^{\ k}_{;km} = 0. \tag{5.36}$$

Use of (5.36) in (5.35) gives

$$h^{\ kl}_{;k} = -g^{lm} \xi^{\ k}_{;km}. \tag{5.37}$$

By means of (5.37) and (5.34) we obtain the above-mentioned alternative form of J^l:

$$J^l = (T^l_{\ k} - \tfrac{1}{2} T \delta^l_k) \xi^k - \tfrac{3}{2} g^{lm} \xi^{\ k}_{;km}. \tag{5.38}$$

We summarize the above in the theorem to follow:

Theorem 5.7: If the geodesics (1.1) in a Riemannian space–time of general relativity admit a quadratic first integral (with explicit dependence on the path parameter s) of the form

$$I^{(2:1)} \equiv \xi_i u^i - s \xi_{i;j} u^i u^j,$$

as described in Theorem 5.1(ii), then the space–time will admit a conserved 4-current $J^l(x)$ in that $J^l_{\ ;l} = 0$, where J^l is defined in terms of the energy–momentum tensor T^i_j and the vector ξ^i by (5.38) [or equivalently (5.34)].

Remark: We note that for the case in which the vector ξ_i of Theorem 5.7 defines an affine collineation ($h_{ij;k} = 0$) then the current vector J^l reduces to a form discussed elsewhere.[15]

Remark: By use of the Ricci identity the necessary condition (5.36) for the existence of a solution to (5.4′) may be expressed in the form[17]

$$\Box^2 \xi_m + 2\xi^{\ k}_{;km} - \xi^k R_{km} = 0. \tag{5.39}$$

VI. CUBIC FIRST INTEGRALS WITH EXPLICIT PATH-PARAMETER DEPENDENCE

In this section we examine in detail the explicit path-parameter dependent cubic first integrals of the geodesics which are given in Tables VIII, IX, and X. From these tables it is observed that the existence of such cubic integrals depends on the existence of certain scalar, vector, or second order tensor fields.

With reference to Table IX we first consider the necessary and sufficient condition

$$C^{(3:1)}_{(i;jkl)} = 0 \tag{6.1}$$

which the vector field $C^{(3:1)}_i$ must satisfy in order that the geodesics admit the cubic first integrals $I^{(3:1)}$ and $*I^{(3:1+1)}$.

The 24 terms in the expansion of (6.1) can be grouped into 12 pairs which may be written in the form

$$h_{ij;kl} + h_{ij;lk} + h_{ik;jl} + h_{ik;lj} + h_{il;jk} + h_{il;kj} + h_{jk;li} + h_{jk;il} + h_{jl;ki} + h_{jl;ik} + h_{kl;ij} + h_{kl;ji} = 0, \quad h_{ij} \equiv \mathscr{L}_{C^{(3:1)}} g_{ij}. \tag{6.2}$$

By use of the identity[11]

$$C^{(3:1)}_{i;jkl} + C^{(3:1)}_{j;kil} + C^{(3:1)}_{k;ijl} \equiv C^{(3:1)}_{i;kjl} + C^{(3:1)}_{j;ikl} + C^{(3:1)}_{k;jil} \tag{6.3}$$

Eq. (6.1) [and hence (6.2)] can be expressed in the equivalent form

$$C^{(3:1)}_{j;kli} + C^{(3:1)}_{k;lji} + C^{(3:1)}_{l;jki} + C^{(3:1)}_{k;lij} + C^{(3:1)}_{l;ikj} + C^{(3:1)}_{i;klj} + C^{(3:1)}_{l;ijk} + C^{(3:1)}_{i;jlk} + C^{(3:1)}_{j;lik} + C^{(3:1)}_{i;jkl} + C^{(3:1)}_{j;kil} + C^{(3:1)}_{k;ijl} = 0. \tag{6.4}$$

The above results relating to the vector-based cubic first integrals of Table IX may be combined with information obtained from Tables VIII and X (relating respectively to scalar-based and tensor-based cubic first integrals) into the following theorem.

Theorem 6.1: If the geodesics (1.1) in a V_n admit cubic first integrals with explicit dependence on the path parameter s, then such integrals may be divided into three classes:

(i) those based upon the existence of a scalar $C^{(3:0)}(x)$ which satisfies a necessary and sufficient condition

$$C^{(3:0)}_{;(ijkl)} = 0, \tag{6.5}$$

in which case the integrals are of the forms

$$I^{(3:0)} \equiv C^{(3:0)} - sC^{(3:0)}_{;i}u^i + (s^2/2)C^{(3:0)}_{;ij}u^iu^j - (s^3/6)C^{(3:0)}_{;ijk}u^iu^ju^k, \tag{6.6}$$

$$*I^{(3:0+1)} \equiv C^{(3:0)}_{;i}u^i - sC^{(3:0)}_{;ij}u^iu^j + (s^2/2)C^{(3:0)}_{;ijk}u^iu^ju^k, \tag{6.7}$$

$$*I^{(3:0+2)} \equiv C^{(3:0)}_{;ij}u^iu^j - sC^{(3:0)}_{;ijk}u^iu^ju^k; \tag{6.8}$$

(ii) those based upon the existence of a vector $C^{(3:1)}_i(x)$ which satisfies either of the equivalent necessary and sufficient conditions

$$h_{ij;kl} + h_{ij;lk} + h_{ik;jl} + h_{ik;lj} + h_{il;jk} + h_{il;kj} + h_{jk;li} + h_{jk;il} + h_{jl;ki} + h_{jl;ik} + h_{kl;ij} + h_{kl;ji} = 0, \quad h_{ij} \equiv \mathscr{L}_{C^{(3:1)}} g_{ij}, \tag{6.2$'$}$$

or

$$C^{(3:1)}_{j;kli} + C^{(3:1)}_{k;lji} + C^{(3:1)}_{l;jki} + C^{(3:1)}_{k;lij} + C^{(3:1)}_{l;ikj} + C^{(3:1)}_{i;klj} + C^{(3:1)}_{l;ijk} + C^{(3:1)}_{i;jlk} + C^{(3:1)}_{j;lik} + C^{(3:1)}_{i;jkl} + C^{(3:1)}_{j;kil} + C^{(3:1)}_{k;ijl} = 0, \tag{6.4$'$}$$

in which case the integrals are of the forms

$$I^{(3:1)} \equiv C^{(3:1)}_i u^i - sC^{(3:1)}_{i;j}u^iu^j + (s^2/2)C^{(3:1)}_{i;jk}u^iu^ju^k, \tag{6.9}$$

$$*I^{(3:1+1)} \equiv C^{(3:1)}_{i;j}u^iu^j - sC^{(3:1)}_{i;jk}u^iu^ju^k; \tag{6.10}$$

(iii) those based upon the existence of a symmetric tensor $C^{(3:2)}_{ij}(x)$ which satisfies a necessary and sufficient condition

$$C^{(3:2)}_{(ij;kl)} = 0, \tag{6.11}$$

in which case the integral is of the form

$$I^{(3:2)} \equiv C^{(3:2)}_{ij}u^iu^j - sC^{(3:2)}_{ij;k}u^iu^ju^k. \tag{6.12}$$

By inspection of Table I we observe that any vector ξ^i(SCC) which defines a (proper) special curvature collineation will be a solution to (6.2′) in that ξ^i(SCC) vectors satisfy $h_{ij;kl} = 0$. Hence by Theorem 6.1(ii) such vectors will define cubic first integrals with explicit path-parameter dependence. These integrals are in addition to the known concomitant first integrals with no explicit path parameter dependence given in Table I. We may therefore state the theorem which follows.

Theorem 6.2: If a V_n admits a proper special curvature collineation in that there exists a vector ξ^i(SCC) such that $h_{ij;kl} = 0$, $h_{ij;k} \neq 0$, $h_{ij} \equiv \mathscr{L}_{\xi(\mathrm{SCC})} g_{ij}$, then the geodesics (1.1) of the V_n will admit the inhomogeneous cubic first integrals (with explicit dependence on the path parameter s) of the types

$$I^{(3:1)} \equiv \xi_i u^i - s\xi_{i;j}u^iu^j + (s^2/2)\xi_{i;jk}u^iu^ju^k, \tag{6.13}$$

$$*I^{(3:1+1)} \equiv \xi_{i;j}u^iu^j - s\xi_{i;jk}u^iu^ju^k, \tag{6.14}$$

where in each integral $\xi_i \equiv \xi_i$(SCC).

We now investigate whether vectors ξ^i(PC), ξ^i(CONFC), or ξ^i(CM) which define projective collineation, conformal collineations, or conformal motions, respectively, (refer to Table I) could be solutions to (6.2′), and thus determine cubic first integrals with explicit path-parameter dependence.

We start with the assumption that the V_n admits a projective collineation determined by a vector ξ^i(PC) which hence satisfies the defining relation (5.18). Use of (5.18) in (6.2′) leads to

$$g_{ij}\phi_{;kl} + g_{jk}\phi_{;il} + g_{ik}\phi_{;jl} + g_{jl}\phi_{;ik} + g_{il}\phi_{;kj} + g_{kl}\phi_{;ij} = 0. \tag{6.15}$$

By contraction of (6.15) it follows that $\phi_{;ij} = 0$; hence the projective collineation must be a *special* projective collineation[5] [(e), Table I]. With $\phi_{;ij} = 0$, it follows from (5.18) that $h_{ij;kl} = 0$. Thus a sufficient condition for (6.2′) to be satisfied is for the vector ξ^i to define a special projective collineation. We state this in the form of a theorem.

Theorem 6.3: In a V_n a necessary and sufficient condition that a (proper) projective collineation vector [(d), Table I] satisfy (6.2′) of Theorem 6.1 is that the vector be a special projective collineation vector [(e), Table I].

By referring to [(j), Table I] it is seen from the discussion preceeding Theorem 6.3 that every special projective collineation vector [(e), Table I] is a special curvature collineation vector. Theorem 6.2 is therefore applicable for associating cubic first integrals which have explicit path-parameter dependence with the existence of special projective collineations.

By means of (5.18) the cubic expression $\xi_{i;jk}u^iu^ju^k$ occurring in the last term of both cubic integrals (6.13) and (6.14) may be expressed in the form $2\phi_{,k}u^kg_{ij}u^iu^j$. With use of the metrical first integral[3] $g_{ij}u^iu^j \doteq \epsilon$ this cubic expression reduces to $2\epsilon\phi_{,k}u^k$. This leads us to the theorem:

Theorem 6.4: If a Riemannian space–time admits a special projective collineation [$\xi^i \equiv \xi^i$(SPC) defined by (e), Table I] the functions

$$Q\,(\mathrm{SPC}) = (\xi_i + \epsilon s^2\phi_{,i})u^i - s\xi_{i;j}u^iu^j, \tag{6.16}$$

$$*Q\,(\mathrm{SPC}) = -2\epsilon s\phi_{,i}u^i + \xi_{i;j}u^iu^j, \tag{6.17}$$

will be quadratic first integrals of nonnull geodesics when $\epsilon = \pm 1$ (in which case the parameter s is the arc length) or quadratic first integrals of null geodesics when $\epsilon = 0$. The integrals Q (SPC) and $*Q$ (SPC) are degenerate forms of the cubic first integrals $I^{(3:1)}$ and $*I^{(3:1)}$, respectively, of Theorem 6.2.

Remark: Note that $*Q$ (SPC) reduces to an integral with no explicit path-parameter dependence when $\epsilon = 0$ [refer to the quadratic integral obtained from (e), Table I for null geodesics].

In a like manner results similar to those obtained above for projective collineations which satisfy (6.2′) may be shown to hold for conformal collineations [(h), Table I] or conformal motions [(f), Table I] which satisfy (6.2′). These results are given in the theorems below.

Theorem 6.5: In a V_n a necessary and sufficient condition that a (proper) conformal collineation vector [(h), Table I] satisfy (6.2′) of Theorem 6.1 is that the vector be a special conformal collineation vector [(i), Table I].

Theorem 6.6: If a Riemannian space–time admits a special conformal collineation [$\xi^i \equiv \xi^i$(SCONFC) defined by (i) of Table I], the functions

$$Q\,(\mathrm{SCONFC}) = [\xi_i(\epsilon/2)s^2\tau_{,i}]u^i - s\xi_{i;j}u^iu^j, \tag{6.18}$$

$$*Q\,(\mathrm{SCONFC}) = -\epsilon s\tau_{,i}u^i + \xi_{i;j}u^iu^j, \tag{6.19}$$

will be quadratic first integrals of nonnull geodesics when $\epsilon = \pm 1$ (in which case the parameter s is the arc length) or quadratic first integrals of null geodesics when $\epsilon = 0$. The integrals Q (SCONFC) and $*Q$ (SCONFC) are degenerate forms of the cubic first integrals $I^{(3:1)}$ and $*I^{(3:1)}$, respectively, of Theorem 6.2.

Remark: It is noted that $*Q$ (SCONFC) reduces to an integral with no explicit path-parameter dependence when $\epsilon = 0$ [refer to the quadratic integral obtained from [(i), Table I] for null geodesics.

Since every conformal motion is a conformal collineation [in which case $\tau \equiv \sigma$; refer to (f) and (h) of Table I] we have the following theorem.

Theorem 6.7: In a V_n a necessary and sufficient condition that a (proper) conformal motion vector [(f), Table I] satisfy (6.2′) of Theorem 6.1 is that the vector be a special conformal motion vector [(g), Table I].

By use of the conformal motion condition [(f), Table I] the quadratic expression $\xi_{i;j}u^iu^j$ occurring in (6.18) and (6.19) may be expressed in the form $\sigma g_{ij}u^iu^j$. We may thus state the theorem to follow.

Theorem 6.8: If a Riemannian space–time admits a special conformal motion [$\xi^i \equiv \xi^i$(SMC) defined by (g) of Table I], the functions

$$L\,(\mathrm{SCM}) \equiv -\epsilon s\sigma + [\xi_i + \epsilon(s^2/2)\sigma_{,i}]u^i, \tag{6.20}$$

$$*L\,(\mathrm{SCM}) \equiv \epsilon(\sigma - s\sigma_{,i}u^i), \tag{6.21}$$

will be linear first integrals of nonnull geodesics when $\epsilon = \pm 1$ (in which case the parameter s is the arc length) or linear first integrals of null geodesics when $\epsilon = 0$. The integrals L (SCM) and $*L$ (SCM) are degenerate forms of the cubic first integrals $I^{(3:1)}$ and $*I^{(3:1)}$, respectively, of Theorem 6.2.

Remark: For the case in which $\sigma = \sigma_0 =$ const (i.e., the special conformal motion is taken to be a homothetic motion) the integral L (SCM) [(6.20)] reduces to L (HM) [(5.12)].

Remark: For the case $\epsilon = 0$ it is noted that $*L$ (SCM) is trivial and L (SCM) reduces to the well-known integral with no explicit path-parameter dependence [refer to (g), Table I].

Remark: With reference to Theorems 6.4, 6.6, and 6.8 the existence of a special projective collineation, a special conformal collineation, or a special conformal motion implies the existence of the respective parallel vector fields $\phi_{,i}$, $\tau_{,i}$, or $\sigma_{,i}$. Hence by Theorem 4.1 there will exist in each case a concomitant linear first integral of the form (4.2), which for the case of the special conformal motion is equivalent to the integral (6.21), $*L$ (SCM), of Theorem 6.8 (with $\epsilon = \pm 1$).

We conclude this section with an example that illustrates Theorems 6.1 and 6.2 of this section in addition to several of the theorems contained in preceeding sections. Consider then the Einstein cosomological space–time

$$d\phi^2 = -\psi^{-2}[(dx^1)^2 + (dx^2)^2 + (dx^3)^2] + (dx^4)^2, \tag{6.22}$$

where

$$\psi \equiv 1 + (K_0/4)[(x^1)^2 + (x^2)^2 + (x^3)^2]. \tag{6.23}$$

The geodesics in this space–time take the form

$$\frac{du^\mu}{ds} + \frac{K_0}{\psi}\left(\frac{x^\mu}{2}u^\nu u^\nu + \frac{x^\mu}{2}u^\mu u^\mu - 2x^\nu x^\nu u^\mu\right) = 0, \quad \mu,\nu = 1,2,3, \tag{6.24}$$

$$\frac{du^4}{ds} = 0. \tag{6.25}$$

In order to formulate the various first integrals with explicit dependence on the path parameter s that illustrates the above-mentioned theorems, we first summarize the required prerequisite symmetries which are admitted by the space–time (6.22).

In addition to the Killing vector

$$C^{(1:0)i} = b\delta^i_4, \quad b = \text{const}, \tag{6.26}$$

which is a parallel vector field based on the scalar $C^{(1:0)}$,

$$C^{(1:0)} \equiv bx^4, \quad C^{(1:0)}_{;ij} = 0, \tag{6.27}$$

this space–time admits a six-parameter group of motions defined by the vectors

$$\xi^i(\mathrm{M}) \equiv [\eta^1, \eta^2, \eta^3, 0], \tag{6.28}$$

where the three-dimensional vectors η^μ (parameters $\omega^\mu_\nu = -\omega^\nu_\mu;\ a_\mu;\ \mu,\nu = 1,2,3$)

$$\eta^\mu \equiv \omega^\mu_\nu x^\nu - a_\nu[(2x^\nu x^\mu - \delta^\mu_\nu)K_0/4 - \delta^\mu_\nu] \tag{6.29}$$

define a six-parameter group of motions in the constant curvature subspace K_3 with line element

$$dl^2 = \psi^{-2}[(dx^1)^2 + (dx^2)^2 + (dx^3)^2]. \tag{6.30}$$

It can also be shown[18] that the V_4 (6.22) admits a (proper) special curvature collineation based on the vector

$$\xi^i(\mathrm{SCC}) \equiv [\eta^\mu, \alpha_1(x^4)^2 + \beta_1 x^4 + \gamma_1],$$
$$\alpha_1 \neq 0, \quad \alpha_1, \beta_1, \gamma_1 = \text{const}, \tag{6.31}$$

in that

$$h_{ij;kl} = 0, h_{ij;k} = 4\alpha_1 \delta_i^4 \delta_j^4 \delta_k^4, h_{ij} \equiv \mathscr{L}_{\xi(\mathrm{SCC})} g_{ij}. \tag{6.32}$$

It is also easily verified that the V_4 (6.22) admits a (proper) affine collineation based on the vector

$$\xi^i(\mathrm{AC}) \equiv [\eta^\mu, \alpha_2 x^4 + \beta_2], \alpha_2 \neq 0, \quad \alpha_2, \beta_2 = \text{const}, \tag{6.33}$$

in that

$$h_{ij;k} = 0, \quad h_{ij} = 2\alpha_2 \delta_i^4 \delta_j^4, \quad h_{ij} \equiv \mathscr{L}_{\xi(\mathrm{AC})} g_{ij}. \tag{6.34}$$

In both (6.31) and (6.33) η^μ is the motion vector (6.29) of the subspace K_3.

The space–time (6.22) admits scalars $C^{(2:0)}$ and $C^{(3:0)}$, where

$$C^{(2:0)} \equiv \alpha_3(x^4)^2 + \beta_3 x^4 + \gamma_3, \quad \alpha_3 \neq 0,$$
$$C^{(2:0)}_{;ijk} = 0, \quad C^{(2:0)}_{;ij} = 2\alpha_3 \delta_i^4 \delta_j^4; \tag{6.35}$$
$$C^{(3:0)} \equiv \alpha_4(x^4)^3 + \beta_4(x^4)^2 + \gamma_4 x^4 + \delta_4, \quad \alpha_4 \neq 0,$$
$$C^{(3:0)}_{;ijkl} = 0, \quad C^{(3:0)}_{;ijk} = 6\alpha_4 \delta_i^4 \delta_j^4 \delta_k^4, \tag{6.36}$$

where $\alpha_3, \beta_3, \gamma_3$ and $\alpha_4, \beta_4, \gamma_4, \delta_4$ are constants.

With reference to the above symmetries we may now formulate certain integrals with explicit path-parameter dependence which exemplify the various theorems.

Based on the existence of the parallel field given by (6.26), (6.27) we obtain by means of Theorem 4.1 the linear first integral

$$I^{(1:0)} = b(x^4 - su^4). \tag{6.37}$$

As a consequence of the existence of the scalar $C^{(2:0)}$ given in (6.35) we obtain the quadratic integrals $I^{(2:0)}$ and $*I^{(2:0+1)}$ described by Theorem 5.1. These can be expressed in the forms

$$I^{(2:0)} = \alpha_3(x^4 - su^4)^2 - \beta_3(x^4 - su^4) + \gamma_3, \tag{6.38}$$

$$*I^{(2:0+1)} = 2\alpha_3(x^4 - su^4)u^4 + \beta_3 u^4. \tag{6.39}$$

Concomitant with the existence of the affine collineation vector (6.33) we obtain by Theorem 5.2(i) [by use of (6.28), (6.29), (6.34)] the quadratic integral[19]

$$I^{(2:1)} = \alpha_2(x^4 - su^4)u^4 + \beta_2 u^4 + \xi_\mu(\mathrm{M})u^\mu, \quad \alpha_2 \neq 0. \tag{6.40}$$

As a result of the existence of the scalar $C^{(3:0)}$ given by (6.36) we may obtain the cubic integrals $I^{(3:0)}$, $*I^{(3:0+1)}$, and $*I^{(3:0+2)}$ described in Theorem 6.1(i). These integrals can be expressed in the forms

$$I^{(3:0)} = -\alpha_4(x^4 - su^4)^3 + \beta_4(x^4 - su^4)^2 + \gamma_4(x^4 - su^4) + \delta_4, \tag{6.41}$$

$$*I^{(3:0+1)} = [3\alpha_4(x^4 - su^4)^2 + 2\beta_4(x^4 - su^4) + \gamma_4]u^4, \tag{6.42}$$

$$*I^{(3:0+2)} = [6\alpha_4(x^4 - su^4) + 2\beta_4]u^4u^4. \tag{6.43}$$

From the existence of the special curvature collineation vector (6.31) we evaluate [with use of (6.32)] the cubic integrals $I^{(3:1)}$ and $*I^{(3:1+1)}$ given by Theorem 6.1. (ii). These integrals may be expressed in the forms[19]

$$I^{(3:1)} = [\alpha_1(x^4 - su^4)^2 + \beta_1(x^4 - su^4) + \gamma_1]u^4 + \xi_\mu(M)u^\mu, \tag{6.44}$$

$$*I^{(3:1+1)} = [2\alpha_1(x^4 - su^4) + \beta_1]u^4u^4. \tag{6.45}$$

Associated with the seven Killing vectors (6.26) and (6.28) it is easily shown [refer to (a), Table I] that the well-known concomitant linear first integrals without explicit path-parameter dependence take the form

$$I(b) = u^4, \tag{6.46}$$

$$I(\omega^\mu_\nu, a_\nu) = \xi_\nu(\mathrm{M})u^\nu. \tag{6.47}$$

Inspection of the explicit path-parameter dependent integrals (6.38)–(6.45) obtained above shows that they are functions of the linear integrals (6.46), (6.47), and (6.37). In general such a functional dependence will not occur.

VII. NULL GEODESIC QUADRATIC FIRST INTEGRALS WITH EXPLICIT DEPENDENCE ON THE PATH-PARAMETER

In this section we shall examine in more detail the null-geodesic quadratic first integrals of the form

$$Q \equiv \xi_i u^i - s\xi_{i;j}u^i u^j, \quad \xi_i \equiv \xi_i(x), \tag{7.1}$$

which were shown to arise [refer to (6.16) of Theorem 6.4 and (6.18) of Theorem 6.6] as degenerate cubic first integrals (based on special curvature collineations) whenever the space–time admits special projective collineations [(e), Table I] or special conformal collineations [(i), Table I].

By forming the absolute derivative of Q along a null geodesic we immediately obtain

$$\frac{DQ}{ds} \doteq \tfrac{1}{2}h_{ij;k}u^i u^j u^k, \quad h_{ij} \equiv \mathscr{L}_\xi g_{ij} = \xi_{i;j} + \xi_{j;i}. \tag{7.2}$$

Since on null geodesics $g_{ij}u^i u^j = 0$, it follows that $h_{ij;k}u^i u^j u^k$ vanishes for those h_{ij} based upon either projective collineation vectors $\xi_i \equiv \xi_i(\mathrm{PC})$ [refer to (d), Table I] or conformal collineation vectors $\xi_i \equiv \xi_i(\mathrm{CONFC})$, [see (h), Table I]. It is noted that these collineations are more general than the *special* projective collineations or *special* conformal collineations which lead to Theorems 6.4 or 6.6, respectively.

In a similar manner (7.2) will vanish if the space–time admits a seminull geodesic collineation. Such collineation map null geodesics into nonnull geodesic and are defined by vectors which satisfy the condition[4]

$$h_{ij;k} = g_{jk}\psi_{,i} + g_{ik}\psi_{,j}, \quad h_{ij} \equiv \mathscr{L}_{\xi(\mathrm{SNGC})} g_{ij}. \tag{7.3}$$

We summarize the above in the theorem to follow.

Theorem 7.1: The null geodesics (1.1) in a Riemannian space–time will admit a quadratic first integral (with explicit dependence on the path parameter s) of the form

$$Q = \xi_i u^i - s\xi_{i;j}u^i u^j \tag{7.1'}$$

if the space–time admits

(i) a projective collineation [defined by the vector $\xi_i(\mathrm{PC})$, (d), Table I] in which case $\xi_i \equiv \xi_i(\mathrm{PC})$ in (7.1′), or

(ii) a conformal collineation [defined by the vector ξ_i(CONFC), (h), Table I] in which case $\xi_i \equiv \xi_i$(CONFC) in (7.1′), or

(iii) a seminull geodesic collineation [defined by the vector ξ_i(SNGC), (7.3)] in which case $\xi_i \equiv \xi_i$(SNGC) in (7.1′).

As an illustration of Theorem 7.1 we shall formulate the null geodesic quadratic first integral with explicit dependence on the path parameter (7.1′) for a Friedmann–Lemaitre cosmological space–time which admits a projective collineation.

Consider the space–time with fundamental form

$$d\phi^2 = \frac{-e^{-x^4/a_0}}{W^2}[(dx^1)^2 + (dx^2)^2 + (dx^3)^2] + (dx^4)^2, \tag{7.4}$$

where

$$W \equiv 1 + K_0 r^2, \quad r \equiv (x^1)^2 + (x^2)^2 + (x^3)^2, \tag{7.5}$$

and K_0 and a_0 are constant, $a_0 \neq 0$.

The null geodesics for this space–time are given by

$$\frac{du^\mu}{ds} + \frac{2K_0 x^\mu}{W}u^\nu u^\nu - \frac{4K_0}{W}x^\nu u^\nu u^\mu - \frac{1}{a_0}u^\mu u^4 = 0, \tag{7.6}$$

$$\frac{du^4}{ds} - \frac{e^{-x^4/a_0}}{2a_0 W^2}u^\nu u^\nu = 0, \tag{7.7}$$

where the tangent vector to the null geodesic $u^i = dx^i/ds$ satisfies the null vector condition $g_{ij}u^i u^j = 0$, which by (7.4) takes the form

$$\frac{-e^{-x^4/a_0}}{W^2}u^\nu u^\nu + u^4 u^4 = 0. \tag{7.8}$$

By use of (7.8) the geodesics equation (7.7) may be reduced to

$$\frac{du^4}{ds} - \frac{1}{2a_0}u^4 u^4 = 0. \tag{7.9}$$

It is a straightforward calculation to verify that this space–time admits a (proper) projective collineation defined by the vector[20,21]

$$\xi^i(\text{PC}) = (0,0,0,a_0 e^{-x^4/a_0}), \tag{7.10}$$

and that this projective collineation is *not* a special projective collineation, in that $\phi_{;ij} \neq 0$ [with reference to (5.18) it may be shown by contraction with g_{ij} that $\phi_{,k} = \frac{1}{5}\xi^i_{\ ;ik}(\text{PC})$].

Based on the vector (7.10), the quadratic integral (7.1′) for the null geodesics of the space–time (7.4) takes the form

$$Q(\text{PC}) = a_0 e^{-x^4/a_0}u^4 - \frac{s}{2}\frac{e^{-2x^4/a_0}}{W^2}u^\nu u^\nu + s e^{-x^4/a_0}u^4 u^4. \tag{7.11}$$

With use of the null-vector condition (7.8) this integral is reducible to the form

$$Q'(\text{PC}) = e^{-x^4/a_0}[a_0 + (s/2)u^4]u^4. \tag{7.12}$$

VIII. CONCLUSION

For arbitrary geodesics in Riemannian spaces V_n we have found the basic forms of *all* mth-order (in general inhomogeneous in the tangent vector) first integrals with *explicit* dependence on the path parameter and obtained necessary and sufficient conditions for the existence of such integrals. It was shown that these integrals must have polynomial structure in the path parameter. If one mth-order first integral with explicit dependence on the path parameter existed, then it was shown in general there would exist a whole family of these mth-order integrals, all associated with the same "symmetry condition."

It was observed that the existence of some well-known symmetries such as homothetic motions (scale change), affine collineations, conformal motions, projective collineations, special curvature collineations, etc. was sufficient for the existence of certain of the first integrals of the above-mentioned type. We find this result to be of particular interest in that previously it was only known that such geometric symmetries led to homogeneous first integrals without explicit path-parameter dependence.

Aside from the four given examples of space–times which were known to admit some of the above well-known symmetries, we have made no systematic attempt to solve (for a particular space–time) any of the necessary and sufficient conditions for the existence of mth-order first integrals with explicit dependence on the path parameter. There is thus the possibility that general solutions to some of the necessary and sufficient conditions exist and can be obtained for those space–times we considered, as well as for other space–times of physical interest—possibly even for those space–times which do not admit any of the well-known geometric symmetries.

As yet, we have made no attempt to interpret physically those integrals with explicit path-parameter dependence which we derived for illustration purposes. However, such interpretations appear to be possible particularly for the examples based upon the cosmological space–times.

In this paper we considered the problem of obtaining integrals with explicit path-parameter dependence primarily for *arbitrary* geodesics. With reference to Riemannian space–times this led to integrals for *both* null and nonnull geodesics. However, in certain portions of our analysis some immediate results were obtained concerning the existence of first integrals with explicit dependence on the path parameter for restricted type geodesics (i.e., either null *or* nonull). Of particular interest is the result that null geodesics admit a quadratic first integral with explicit path-parameter dependence whenever the space–time admits a projective collineation. At the present time we are making a detailed analysis of the restricted geodesic case.

An additional investigation of interest would be to apply the Noether theoretical approach to the problem of obtaining geodesic first integrals with explicit path-parameter dependence. By means of this approach we have analyzed (in a separate investigation) the related problem of determining explicit path-parameter (time-) dependent quadratic and linear first integrals of dynamical systems with simple velocity-dependent potentials in Riemannian configuration space, and have derived necessary and sufficient conditions for the existence of such integrals along with the form of the associated Noether symmetry mappings. It is clear that the geodesic case for linear and quadratic integrals is included in the above-mentioned Noether analysis.

[1]Unless indicated otherwise, lower case Latin indices range from 1 to n, where n is the dimensionality of the space. The lower case Greek indices μ and ν have the range 1,2,3 unless otherwise indicated. The Einstein summation convention is employed. Absolute differentiation along a curve with path parameter s is indicated by D/ds. Covariant differentiation with respect to the Christoffel symbol Γ^{i}_{jk} is indicated by a semicolon (;). Partial differentiation is indicated by a comma (,). A comma followed by s (,s) means partial differentiation with respect to the path parameter s for those cases (noted in the text) in which the n coordinates x^i and s are to be regarded as $n+1$ independent variables.

[2]For nonnull geodesics we assume s is the arc length, in which case the tangent vector u^i is a unit vector.

[3]For space–times with signature -2, timelike geodesics will have $\epsilon = +1$, and spacelike geodesics will have $\epsilon = -1$; for space–times with signature $+2$ the values of ϵ are interchanged. Null geodesics are characterized by $\epsilon = 0$.

[4]G. H. Katzin and J. Levine, Colloq. Math. (Wroclaw) **26**, 21 (1972).

[5]G. H. Katzin, J. Levine, and W. R. Davis, J. Math. Phys. **10**, 617 (1969).

[6]Equality on a geodesic is indicated by the symbol $\overset{\circ}{=}$.

[7]Covariant tensor indices enclosed in parentheses indicate complete symmetrization of the tensor with respect to those indices. For example $A_{(i;jk)} \equiv (1/3!)[A_{ij;k} + A_{ji;k} + A_{jk;i} + A_{kj;i} + A_{ki;j} + A_{ik;j}]$.

[8]Since $u^{i_1}\cdots u^{i_l}$ is completely symmetric on all indices, it is easily shown that

$$C^{(m)}_{(i_1\cdots i_a;i_{a+1}\cdots i_l)}u^{i_1}\cdots u^{i_l} \equiv C^{(m)}_{i_1\cdots i_a;i_{a+1}\cdots i_l}u^{i_1}\cdots u^{i_l};$$

hence the symmetrization parenthesis can be omitted in such expressions.

[9]H. Takeno, "The Mathematical Theory of Plane Gravitational Waves in General Relativity," Sci. Res. Inst. Teor. Phys., Hiroshima Univ. **1**, 1 (1961).

[10]Lie differentiation with respect to the vector $C^{(2)i}$ is denoted by $\mathscr{L}_{C^{(2)}}$.

[11]This identity is easily verified by use of the Ricci identity and an algebraic property of the curvature tensor.

[12]For the null geodesic case ($\epsilon = 0$) L (HM) reduces to the well-known linear first integral (with no explicit path-parameter dependence) associated with a homothetic motion and listed in Table I.

[13]C. B. G. McIntosh, Phys. Lett. A **50**, 429 (1974).

[14]K. Yano, *The Theory of Lie Derivatives and its Applications* (North-Holland, Amsterdam, 1957).

[15]G. H. Katzin and J. Levine, J. Math. Phys. **19**, 1918 (1978).

[16]The energy–momentum tensor and its trace are denoted by T^i_j and T, respectively.

[17]The form of this equation suggests that (5.2) may in some sense be considered a generalization of Maxwell's equation. With regard to (5.39) as a curved space–time form of Maxwell's equation, refer to C. W. Misner, K. S. Thorne, and J. A. Wheeler, *Gravitation* (Freeman, San Francisco, 1973), p. 569.

[18]G. H. Katzin, J. Levine, and W. R. Davis, Tensor N.S. **21**, 51 (1970).

[19]$\xi_\mu(M) = g_{\mu\mu}\xi^\mu(M) = -\psi^{-2}\eta^\mu$.

[20]C. F. Martin, "Spaces in General Relativity Admitting Geodesic Correspondence," PhD thesis (North Carolina State University, Raleigh, N.C., 1962).

[21]W. R. Davis and M. K. Moss, Nuovo Cimento, Ser. X **38**, 1558 (1965).

Global definition of nonlinear sigma model and some consequences

R. Percacci
Istituto di Fisica Teorica dell'Università di Trieste and Istituto Nazionale di Fisica Nucleare, Sezione di Trieste, Italy

(Received 27 February 1980; accepted for publication 20 June 1980)

The (nonlinear) sigma model is defined as a field theory whose configurations are sections of a nontrivial fiber bundle over space–time. The action functional is a generalization of the "energy" used in the theory of harmonic maps. This definition requires minimal coupling to a Yang–Mills field, and the solutions of the coupled equations exhibit spontaneous symmetry breaking. It is shown that in a Higgs phenomenon making use of a sigma model instead of the Higgs fields, no scalars would survive symmetry breaking.

PACS numbers: 02.40.Ma, 11.10.Lm

1. INTRODUCTION

The main motivation for this paper was the belief that at least at the classical level, nature should be described in geometrical terms. This philosophy[1] is substantiated by the observation that the two most successful physical theories, general relativity (describing the spin-2 sector) and gauge theories (describing the spin-1 sector), are both of essentially geometric character. Thus it would seem natural to try to construct a realistic theory in which the bosonic sector is totally geometrized; this is the case if the scalars in the theory are sigma fields, whose self-interaction is due to the Riemannian structure of the internal space.

In the first part of the paper, I give a global definition of sigma model, which allows the existence of so-called "twisted"[2] field configurations. The logical necessity of introducing twisted fields can be made clear by means of the following simple argument. Let us take a very general attitude and regard a field configuration simply as a mapping $\varphi: X \rightarrow Y$ where X is space–time and Y some "internal" space (the field space). To any such mapping there is associated another mapping (the "graph" of φ) $\sigma: X \rightarrow X \times Y$ defined by

$$\sigma(x) = (x, \varphi(x)) \quad \forall x \in X, \tag{1.1}$$

which is a cross section of the trivial bundle $X \times Y$. If we focus our attention on the graph σ instead of φ, it becomes natural to generalize the definition of field configuration in order to include sections of nontrivial bundles: these are the so-called twisted fields.

Another motivation comes from the fiber bundle formulation of gauge theories[3]: if we regard a Yang–Mills field as a connection in a principal fiber bundle over space–time, the minimally coupled matter fields are naturally interpreted as sections of associated fiber bundles.

Twisted scalar ($Y = \mathbb{R}$ or $\mathbb{C}$) and spinor ($Y = \mathbb{C}^2$) fields have been studied in some detail both from the classical and the quantum viewpoints[4]; besides their intrinsic simplicity, this choice was motivated by two facts: (1) vector bundles always admit sections and (2) in the case of the scalars, the structure group is $\mathbb{Z}_2$ and thus there is no need of introducing connections. While these special features allowed to perform explicit computations, it might be of some interest to begin studying the properties of more complicated or more general models which do not share them; it is precisely the purpose of this paper to examine the case when Y is a Riemannian homogeneous space. The possibility of doing this has already been noticed [Refs. (1) and (2)]; although the procedure to be followed proves fairly straightforward, there are some interesting facts which I believe are worth being spelled out in detail. As a by-product, I obtain the recipe for coupling minimally a G/H-valued sigma model to a G Yang–Mills field. This might be of some interest in itself as is shown by the sigma-model-induced Higgs phenomenon, mentioned in Sec. 4; the fact that there are no scalar fields in the theory when the symmetry has been broken might be of some use in bringing order in the jungle of grand-unified theories.

Throughout this paper, I will use the following terminology: by "sigma model" I mean a nonlinear sigma model, and by a "linear theory" I mean a theory for which Y is a linear vector space, irrespective of any interaction term in the Lagrangian.

2. PRELIMINARIES

I will collect here the terminology and some facts on fiber bundles; the reader is referred to the classical texts[5] for more information. A differentiable fiber bundle with total space B, base space X, projection $\beta: B \rightarrow X$, fiber Y, and structure group G will be denoted $(B, \beta, X; Y, G)$ or simply B; the associated principal bundle is $(P, \pi, X; G)$. The principal map is $\chi: P \times Y \rightarrow B$; fixing the first entery we have a map $\chi_p: Y \rightarrow \beta^{-1}(\pi(p))$ defined by $\chi_p(y) = \chi(p, y)$ and fixing the second entry we have a bundle homomorphism $\chi_y: P \rightarrow B$ defined by $\chi_y(p) = \chi(p, y)$. If $\{U_A\}$ is a family of coordinate neighborhoods for P and $\psi_A: U_A \times G \rightarrow \pi^{-1}(U_A)$ are local trivializations of P, the principal map can be used to induce local trivializations of B $\bar{\psi}_A: U_A \times Y \rightarrow \beta^{-1}(U_A)$ in the following way: if $p = \psi_A(x, g) \in P$ and $b = \chi(p, y) \in B$, then $b = \bar{\psi}_A(x, gy)$. A local trivialization of P (a local gauge) can be fixed giving a local section $s_A: U_A \rightarrow \pi^{-1}(U_A)$ and requiring that it has the form $s_A(x) = \psi_A(x, e)$, where $x \in U_A$ and e is the identity of G. A gauge transformation is a change of trivialization, i.e., a change of local section

$$s_A(x) \rightarrow s'_A(x) = R_{g(x)}(s_A(x)). \tag{2.1}$$

If ω is a connection form in P, the YM potential in the gauge defined by S_A is[6]

$$B_A = B_A{}^i e_i = s_A^* \omega. \tag{2.2}$$

Similarly, if $\Omega = \mathscr{D}\omega$ is the curvature form of the connection,

$$F_A = F_A{}^i e_i = s_A^* \Omega, \tag{2.3}$$

is the YM field strength. Under (2.1) they transform in the well-known way

$$\begin{aligned} B'_A &= g^{-1} B_A g + g^{-1} dg, \\ F'_A &= g^{-1} F_A g. \end{aligned} \tag{2.4}$$

The sections $\sigma: X \rightarrow B$ are in one-to-one correspondence with G-equivariant mappings $\tilde{\sigma}: P \rightarrow Y$ (Husemoller in Ref. 5, p. 46):

$$\tilde{\sigma} \circ R_g = L_g^{-1} \circ \tilde{\sigma}. \tag{2.5}$$

Given $\tilde{\sigma}$, σ is defined by

$$\sigma(\pi(p)) = \chi(p, \tilde{\sigma}(p)), \tag{2.6}$$

The local representative of $\tilde{\sigma}$ on U_A is $\varphi_A : U_A \rightarrow Y$ defined by

$$\varphi_A = \tilde{\sigma} \circ s_A. \tag{2.7}$$

Choosing $p = s_A(x)$ in (2.6) we find

$$\sigma(x) = \bar{\psi}_A(x, \varphi_A(x)), \tag{2.8}$$

and thus φ_A is also a local representative for σ. Under (2.1) φ_A transforms as

$$\varphi'_A = L_g^{-1} \circ \varphi_A. \tag{2.9}$$

The left action $L: G \times Y \rightarrow Y$ is represented in coordinates[6] by functions[7]

$$(L_g(y))^\alpha = L^\alpha(g^i, y^\beta). \tag{2.10}$$

The generators of G realized on Y have components

$$L_i^\alpha(y) = \frac{\partial L^\alpha(g,y)}{\partial g^i}\bigg|_{g=e}. \tag{2.11}$$

3. TWISTED SIGMA MODELS

Let $Y = G/H$, be a homogeneous space with a left-G invariant Riemannian structure h, and X be space–time, with a Riemannian structure g. A *field configuration* for a locally G/H-valued sigma model is a section $\sigma: X \rightarrow B$ of a fiber bundle $(B, \beta, X; G/H, G)$. When B is trivial, the *energy* of the configuration σ is given by[8]

$$E[\sigma] = \frac{1}{2} \int_X (d\varphi, d\varphi) \cdot \eta, \tag{3.1}$$

where $\varphi: X \rightarrow G/H$ is related to σ by (1.1), η is the volume element $\sqrt{\det g}\, dx^1 \wedge dx^2 \wedge dx^3 \wedge dx^4$ canonically defined by g and $(d\varphi, d\varphi) = g^{\mu\nu} \partial_\mu \varphi^\alpha \partial_\nu \varphi^\beta h_{\alpha\beta}$.[6] In order to generalize this to the case when B is nontrivial, we have to define in a sensible way the derivative of a cross section; this is well known in the case of vector bundles but not for general fiber bundles, so I will outline the procedure. If Γ is a connection in the principal bundle $(P, \pi, X; G)$ associated to B, the vertical and horizontal subspaces V_b and H_b of $T_b(B)$ are defined as the images of the vertical and horizontal subspaces V_p and H_p of $T_p(P)$ under the map $\chi_{0*}: T_p(P) \rightarrow T_b(B)$, where $b = \chi_0(p)$ and 0 is the "origin" of G/H, i.e., the distinct point whose isotropy group is H. Thus Γ defines a parallelism in B: if $c(t)$, $0 \leqslant t \leqslant 1$ is a curve in X with $c(0) = \pi(p) = \beta(b) = x$ and $c(1) = x'$ we can define the horizontal lift of $c(t)$ in B, $\tilde{c}(t)$, $0 \leqslant t \leqslant 1$ with $\tilde{c}(0) = b$ and the point $\tilde{c}(1) \in \beta^{-1}(x')$ will be called the parallel translate of b along $c(t)$. Varying b we obtain thus a diffeomorphism between fibers: $\tau_c(x, x'): \beta^{-1}(x) \rightarrow \beta^{-1}(x')$. Let $\sigma: X \rightarrow B$ be a cross section and v be the vector $v = (d/dt)c(t)|_{t=0}$ tangent to $c(t)$ at x. Varying t, $\sigma(c(t))$ is a curve in B and $\tau_c(c(t), c(0)) \cdot \sigma(c(t))$ is a curve in the fiber $\beta^{-1}(x)$; we define the covariant derivative of σ at x along v to be the vertical vector

$$\mathscr{D}_v \sigma = \frac{d}{dt} \tau_c(c(t), x) \sigma(c(t))\bigg|_{t=0}. \tag{3.2}$$

It is clear how this definition at a point x has to be generalized to all of X: if v is a vector field on X, $\mathscr{D}_v \sigma|_x = \mathscr{D}_{v(x)} \sigma$ and $\mathscr{D}_v \sigma$ is a section of the vertical bundle $V = \cup_{b \in B} V_b \subseteq T(B)$. If B were a vector bundle, V would be canonically isomorphic to B and thus $\mathscr{D}_v \sigma$ would be another section of B.

Formula (3.2) is not very practical in order to obtain an explicit form, so we turn to the equivariant mapping $\tilde{\sigma}: P \rightarrow G/H$ associated to σ. The covariant differential of $\tilde{\sigma}$ is defined by

$$\mathscr{D} \tilde{\sigma} = \mathrm{hor} d\tilde{\sigma}: H_p \rightarrow T_{\tilde{\sigma}(p)}(G/H), \tag{3.3}$$

and the covariant derivative of $\tilde{\sigma}$ along $\tilde{v} \in T_p(p)$

$$\mathscr{D}_{\tilde{v}} \tilde{\sigma} = \mathscr{D} \tilde{\sigma}(\tilde{v}) = d\tilde{\sigma}(\mathrm{hor} \tilde{v}) \in T_{\tilde{\sigma}(p)}(G/H). \tag{3.4}$$

The definitions (3.4) and (3.2) are related by

$$\mathscr{D}_v \sigma = \chi_{p*} \mathscr{D}_{\tilde{v}} \tilde{\sigma}, \tag{3.5}$$

where $\chi_{p*}: T_{\tilde{\sigma}(p)}(G/H) \rightarrow V_{\chi_p(\tilde{\sigma}(p))} = V_{\sigma(\pi(p))}$ and $v = \pi_* \tilde{v}$. The proof of this is the direct generalization of the proof of the lemma on p. 116, Vol. I of Kobayashi and Nomizu.[5] From (3.4) one obtains an explicit formula; defining $\mathscr{D}_v \varphi_A = \mathscr{D}_{s_{A*} v} \tilde{\sigma}$, $\mathscr{D}_\mu = \mathscr{D}_{e_\mu}$, and $B_A{}^i{}_\mu = B_A{}^i(e_\mu)$ we have

$$\mathscr{D}_\mu \varphi_A^\alpha = \partial_\mu \varphi_A^\alpha + B_A{}^i{}_\mu L_i^\alpha(\varphi_A), \tag{3.6}$$

where φ_A^α are the coordinates of the image of φ_A. The explicit computation is given elsewhere.[9] Since φ_A is also a representative for σ, we may regard (3.6) as the explicit form of $\mathscr{D}\sigma$.

In the familiar case when Y is a vector space supporting a representation $\rho: G \rightarrow GL(Y)$,

$$L_i^\alpha(\varphi_A) = (T_i)^\alpha_\beta \varphi_A^\beta, \tag{3.7}$$

where $T_i = (d/dt)\rho(\exp t e_i)|_{t=0}$ are the images of the generators e_i under ρ.

The transformation law of $\mathscr{D}\sigma$ under (2.1) is easily found from Eq. (3.3); from the differential of (2.5) and $H_{Rg(p)} = R_{g*} H_p$ we have $\mathscr{D}\tilde{\sigma} \circ R_{g*} = L_{g*}^{-1} \circ \mathscr{D}\tilde{\sigma}$, and thus

$$\mathscr{D}_v \varphi'_A = L_{g*}^{-1} \mathscr{D}_v \varphi_A, \tag{3.8}$$

as expected. This could also be checked directly using (2.4) and the properties of the auxiliary functions L_i^α.

Having done this, it is now simple to find the generalization of (3.1). Let φ_A and φ_B be local representatives of σ on U_A and U_B respectively: $\sigma(x)|_{U_A} = \bar{\psi}_A(x, \varphi_A(x))$, $\sigma(x)|_{U_B} = \bar{\psi}_B(x, \varphi_B(x))$ where $\varphi_B(x) = g_{BA}(x) \varphi_A(x)$ $\forall x \in U_A \cap U_B$. Then the transformation law (3.8) and the fact that G is an isometry group for Y imply

$$(\mathscr{D}\varphi_A, \mathscr{D}\varphi_A) = (\mathscr{D}\varphi_B, \mathscr{D}\varphi_B).$$

Therefore, if $\{ f_A \}$ is a partition of unity subordinate to the covering $\{ U_A \}$, we define

$$E[\sigma] = \frac{1}{2} \sum_A \int_{U_A} f_A (D\varphi_A, D\varphi_A) \cdot \eta. \tag{3.9}$$

Symbolically, we will also write

$$E[\sigma] = \frac{1}{2} \int_X (D\sigma, D\sigma) \cdot \eta. \tag{3.10}$$

The minima of $E[\sigma]$ in the space of sections for a given connection Γ in P will be called *harmonic sections* of B. Physically, it is natural to suppose that the Yang–Mills field is also dynamically active, so we can add to (3.10) the term

$$E[\omega] = -\frac{1}{4} \int_X g^{\mu\nu} g^{\rho\sigma} F^i_{\mu\rho} F^j_{\nu\sigma} \gamma_{ij} \cdot \eta. \tag{3.11}$$

Following de Witt,[6] the YM coupling constant has been absorbed into the scalar product γ in the Lie algebra $\mathscr{G}$.

4. SYMMETRY BREAKING

In the previous sections we assumed implicitly that the fiber bundle B admits cross sections; indeed, it is obvious from physical considerations that B must admit continuous, global cross sections if it has to be physically interesting. But not all fiber bundles do admit global cross sections; the key theorem is the following (in Ref. 5 see Husemoller, p. 71; Kobayashi and Nomizu Vol. 1, p. 57):

Theorem: The bundle $(B,\beta,X;G/H,G)$ admits a cross section if and only if the associated principal bundle $(P,\pi,X;G)$ admits a reduction to a principal bundle $(P',\pi',X;H)$; in this case, B is associated to P'. Furthermore, there is a one-to-one correspondence between cross sections of B and reduced bundles.

Let us see how P' is defined. First of all we have the commutative diagram

$$\begin{tikzcd} P \arrow[r, "\tau"] \arrow[dr, "\pi"'] & B \arrow[d, "\beta"] \\ & X \end{tikzcd} \tag{4.1}$$

where $\tau: P \to B = P \bmod H$ is the natural projection (Husemoller p. 70; Kobayashi and Nomizu p. 57).[5] P' is the inverse image of the section σ under τ

$$P' = \tau^{-1}(\sigma(X)) = \{ p \in P \mid \tau(p) = \sigma(\pi(p)) \}, \tag{4.2}$$

and π' is the restriction of π to P'. $(P,\tau,B;H)$ is a principal H-bundle and P' can be regarded as the pull-back of this bundle to X induced by σ. In fact

$$\sigma^*(P \to B) = \{ (x,p) \in X \times P \mid \sigma(x) = \tau(p) \}, \tag{4.3}$$

and the projection of $\sigma^*(P \to B)$ maps $(x,p) \mapsto x$. But $\sigma(x) = \tau(p)$ implies $x = \pi(p)$ and therefore (4.3) coincides with (4.2).

Since homotopic maps $X \to B$ induce X-isomorphic bundles, the topology of the reduced bundle P' depends only on the topology of P and on the homotopy class of the section σ. The classification of principal H-bundles over four dimensional manifolds is dealt with in Ref. 10. In physical terms, the theorem above asserts that every solution (indeed, every configuration) of the sigma model leads to a spontaneous breaking of the symmetry (gauge) group from G to H.

When solving the equations of motion deriving from (3.10) and (3.11) one can think from the outset that the Yang–Mills potential B is a connection in the reduced bundle P'.

Let us now consider two special cases in which a connection in the bundle P plays a role.

Let ω be a connection form in P; by a well-known theorem (Kobayashi and Nomizu Vol. 1, p. 88)[5] the restriction ω' of ω to P' defines a connection in P' if and only if the section $\sigma: X \to B$ related to the reduction from P to P' is covariantly constant with respect to the connection ω, i.e.,

$$\mathscr{D}\sigma = 0.$$

In this case σ is a minimum for (3.10); if ω is a solution of the YM field equations, so is ω' and thus (ω',σ) is a solution of the coupled YM-sigma field equations. An explicit example of these solutions is given elsewhere.[11]

The second example is a sigma-model version of the Higgs phenomenon. Start with a G-YM field and suppose we want to break down the gauge symmetry to H. To this end, couple minimally the YM field to a G/H-valued sigma model as in Eqs. (3.10) and (3.11).

We require G/H to be a (weakly) reductive homogeneous space;[12] that is to say, there exists an $Ad_{\mathscr{G}}(H)$ invariant linear subspace $\mathscr{P} \subseteq \mathscr{G}$ such that $\mathscr{G} = \mathscr{H} \oplus \mathscr{P}$. Then, we can identify $\mathscr{P}$ with $T_0(G/H)$ (0 being the coset H) and if γ is an inner product in $\mathscr{G}$, we can take h to be the unique left-invariant metric in G/H that coincides with the restriction of γ to $\mathscr{P}$ at 0. L^α_i is then the αth component of the ith Killing vector of G/H in the metric h. Let $\{ e_i \}$ be a basis for $\mathscr{G}$ such that $e_i, i = 1,\ldots,d$ form a basis for $\mathscr{P}$ and $e_i, i = d+1,\ldots,n$ form a basis for $\mathscr{H}$. The first d Killing vectors form a field of bases on G/H:

$$h_{\alpha\beta} L^\alpha_i L^\beta_j = \gamma_{ij} \quad i, j = 1,2,\ldots,d.$$

We are now ready to break the symmetry. Choose the "constant" global section $\varphi_A = 0 \ \forall A$; then

$$E[\sigma] = \frac{1}{2} \sum_A \int_{U_A} f_A g^{\mu\nu} B^i_{A\mu} B^j_{A\nu} L^\alpha_i(0) L^\beta_j(0) h_{\alpha\beta}(0) \eta.$$

It is always possible to perform local gauge transformations by elements of H, such that B_A has vanishing components on the subspace $\mathscr{H}$ (this is the "unitary" gauge); then, by the discussion above

$$E[\sigma] = \frac{1}{2} \sum_A \int_{U_A} f_A \, g^{\mu\nu} \sum_{i,j=1}^{d} B^i_{A\mu} B^j_{A\nu} \gamma_{ij} \eta$$

has become a pure mass term for the $\mathscr{P}$-components of the YM field. The remarkable fact here is that unlike in the usual case there are no scalar fields surviving the symmetry breaking. The condition of (weak) reductivity holds in any one of the following cases: H is discrete, H is compact, H is semisimple and connected.

[1]C. W. Misner, Phys. Rev. D **18**, 4510 (1978). Of course, the idea is much older and goes back at least to Riemann.

[2]C. J. Isham, Proc. R. Soc. London Ser. A **362**, 383 (1978).

[3]W. Drechsler and M. E. Mayer, "Fiber bundle techniques in Gauge Theories," *Lecture Notes in Physics* (Springer-Verlag, Berlin 1977), Vol. 67; A.

Trautman, "Elementary introduction to fibre bundles and gauge fields," Warsaw Univ. preprint; M. Daniel and C. M. Viallet, "The geometrical setting of gauge theories of the Yang–Mills type," Rev. Mod. Phys. **52,** 175 (1980).

[4]Besides Ref. 1 see S. J. Avis and C. J. Isham, Proc. R. Soc. London Ser. A **363**, 581 (1978) for classical solutions and the following for quantum properties: B. S. De Witt, C. F. Hart, and C. J. Isham, "Topology and Quantum Field Theory," to appear in Schwinger Festschrift; L. H. Ford, "Vacuum Polarization in a Non simply Connected Space Time," King's College preprint (1979); N. D. Birrell and L. H. Ford, "Renormalization of Self-Interacting Scalar Field Theories in a Non simply Connected Space Time," Kings College preprint (1979); G. Denardo and E. Spallucci, "Dynamical Mass Generation in $S^1 \times \mathbb{R}^3$," Univ. Trieste preprint (1979).

[5]N. Steenrod, *The Topology of Fibre Bundles* (Princeton U. P., Princeton, NJ, 1951); D. Husemoller, *Fibre bundles*, 2nd ed. (Springer-Verlag, New York, 1975); partly also, S. Kobayashi and K. Nomizu, *Foundations of Differential Geometry* (Interscience, New York, 1963), where the theory of connections is also treated.

[6]The following conventions are used for the indices: X: coord. x^μ, $\mu = 1,2,3,4$; Y: coord. y^α, $\alpha = 1,2,...,d$; G: coord. g^i, $i = 1,2,...,n$. The bases for vectors and 1-forms are $\{e_\cdot\}$ and $\{e^\cdot\}$; in the case of G they are taken left invariant, so $\{e_i\}$ is a basis for the Lie Algebra $\mathscr{G}$ of G.

[7]B. de Witt "Dynamical theory of groups and fields" in *Relativity Groups and Topology*, edited by B. de Witt and C. de Witt. The functions L^α and L_i^{α} are de Witt's F^j and R^{j}_{α}.

[8]I will use the term "energy" rather than "action" since X is supposed to be Riemannian; also, this is the term currently used in the mathematical literature to define harmonic maps. See J. Eells and J. H. Sampson, Am. J. Math. **86**, 109 (1964); and J. J. Eells and L. Lemaire, Bull. London Math. Soc. **10**, 1 (1978).

[9]R. Percacci, "Geometrical aspects of nonlinear sigma models," report of Istituto Nationale di Fisica Nucleare INFN/AE-80/1.

[10]S. J. Avis and C. J. Isham, "Quantum field theory and fibre bundles in a general spacetime," Cargese Lectures (1978).

[11]K. G. Akdeniz, M. Goodman, and R. Percacci, "Monopoles and twisted sigma models," IC/80/89 (Trieste).

[12]See Kobayashi and Nomizu,Vol. II in Ref. 5; or A. Lichnerowicz, *Geometry of groups of transformations* (Noordhoff, Leyden, 1977).

On the inverse problem of the calculus of variations

Sergio Hojman and Luis F. Urrutia
Centro de Estudios Nucleares, Universidad Nacional Autónoma de México, Apdo. Postal 70-543, México 20, D. F., México

(Received 9 October 1980; accepted for publication 10 April 1981)

We consider the inverse problem of the calculus of variations for any system by writing its differential equations of motion in first-order form. We provide a way of constructing infinitely many Lagrangians for such a system in terms of its constants of motion using a covariant geometrical approach. We present examples of first-order Lagrangians for systems for which no second-order Lagrangians exist. The Hamiltonian theory for first-order (degenerate) Lagrangians is constructed using Dirac's method for singular Lagrangians.

PACS numbers: 03.20. + i

1. INTRODUCTION

Even though in classical mechanics the dynamical evolution of a system is completely characterized by Newton's equations, the idea of formulating the theory in terms of a variational principle (Lagrangian or Hamiltonian approach) has proved to be very useful in suggesting very effective similar constructions in other areas of physics like quantum mechanics or field theory. It is also well known that the standard prescription $L = T - V$ for constructing the Lagrangian only works either for conservative systems (where the potential energy V is a function of position only) or for some very special velocity-dependent forces like the Lorentz force in electromagnetism. There still remains a great many classical systems which do not correspond to the above-mentioned cases and which are consequently lacking a variational formulation.[1]

Having in mind possible extensions to systems possessing some kind of gauge freedom like the classical relativistic particle (or the electromagnetic field), for example, where it is impossible to solve for the acceleration in the equations of motion and recover a Newton-like equation of the form $\ddot{x} = F(x,\dot{x},t)$, one realizes that it is also convenient to enlarge the class of dynamical differential equations under discussion. To this end, we will pursue the idea that a mechanical system is characterized by a complete set of trajectories in configuration space and we will consider as perfectly admissible any dynamical set of differential equations which reproduces the given orbits as the complete set of their solutions. Let us remark that this point of view defines a class of equivalent Lagrange functions which is broader than the usual one where its elements differ only by the total time derivative of an arbitrary function. These are the so-called s-equivalent Lagrangians and they have been studied recently by several authors.[2–6]

It is worth reminding the reader that two Lagrangians which are related by a total time derivative yield the *same* set of differential equations. On the other hand, s-equivalent Lagrangians[2,4,5] give rise to families of differential equations, which in general *are not the same*, but their complete set of solutions coincide.

A closely related (although wider) subject is that of the inverse problem of the calculus of variations, which consists essentially in trying to find all Lagrangians that under variation will give rise to a system of differential equations with a given complete set of solutions. Much work along this line has been published lately.[3,6–11]

In this note we consider the inverse problem of the calculus of variations for any system of differential equations in which the highest derivatives can be algebraically solved for. In other words, we restrict ourselves to regular (i.e., nongauge, nonconstrained) systems only,[12] and we look for Lagrangians which reproduce such systems of equations when written in first-order form.

The inverse problem of the calculus of variations was solved for the one-dimensional case by Darboux in 1894[13] and the extension to two dimensions was carried on by Douglas in 1941.[1] They both considered a second-order formulation of the differential equations and Douglas exhibited some examples of them for which the Lagrangian simply did not exist.

Recently Havas,[14] Santilli,[3] Sarlet,[6] and others started looking at this problem using a first-order formalism for the differential equations which, of course, leads to first-order (degenerate) Lagrangians.

We emphasize that the use of first-order Lagrangians is widespread in physics, even though they are degenerate. As a matter of fact, in some cases they are more convenient than the corresponding second-order versions. One of these instances is found in the success and simplicity of the first-order version of supergravity.[15] Some other well-known examples of their use are the description of fermionic degrees of freedom, the so called Palatini variational principle, and Schwinger's action principle in quantum mechanics. It is also worth mentioning the fact that even for such degenerate Lagrangians it is possible to construct a Hamiltonian theory by using Dirac's method for singular Lagrangians.[16,17]

As is well known, any system of differential equations can always be equivalently written in first-order form by introducing an adequate number of new variables, which are functions of the first derivatives of the initial variables.[6] A familiar example of this procedure is the Hamiltonian formulation of classical mechanics where the momenta are the new variables needed. Another possible choice of new variables is the velocities themselves.

Havas[14] has shown that when a given system of differential equations is written in first-order form it is possible to find more than one Lagrangian which reproduces the corre-

 0022-2488/81/091896-08$01.00

sponding set of solutions. Our work can be regarded as a continuation and further elaboration of Havas's paper.

The work contained in Refs. 3 and 6 is restricted to the so-called self-adjoint systems and for such cases an algorithm to construct the Lagrangian is provided. Only for some particular cases is it shown how to write a system in self-adjoint form and a general prescription to perform this transformation is not given.

In Sec. 2 we discuss different variational principles used in physics and we justify the adoption of the so-called Weiss's principle for first-order Lagrangians.

Section 3 of this paper is devoted to the construction of a first-order formulation for second-order dynamical systems. There we establish our notation and also make contact with some previous work along this line done by other authors.

In Sec. 4 we present a different approach to the inverse problem of the calculus of variations in its first-order form. The central objects of our formulation are the constants of motion associated with the differential equations together with a (*covariant*) geometrical interpretation related to the fact that the dynamics for a given system is uniquely determined by a given vector which corresponds to the direction of the tangent to the solution curve in some specified space. In terms of the constants of motion, whose existence is guaranteed by some very general assumptions (but whose explicit construction might usually prove difficult in practice), we provide an explicit local method for constructing infinitely many first-order Lagrangians for a given system of curves.

Section 5 contains two examples of first-order Lagrangians for systems for which no second-order Lagrangian exists.

Section 6 is a summary of the work together with some comments related to the further use of these ideas and methods.

Finally, the Appendix deals with the construction of the Hamiltonian theory for first-order Lagrangians using Dirac's method.[16,17]

2. VARIATIONAL PRINCIPLES

To describe a physical system one may define different variational principles. Two of them are the so-called Hamilton's and Weiss's principles, the latter being also known as Schwinger's action principle in its quantum-mechanical version.

Hamilton's (or fixed end points) principle establishes that the desired equations of motion are obtained from a given action S,

$$S = \int_{t_1}^{t_2} L(q^i, \dot{q}^i, t)\, dt, \tag{2.1}$$

by requiring

$$\delta S = 0 \tag{2.2}$$

for arbitrary

$$\delta q^i(t), \quad t_1 < t < t_2 \tag{2.3}$$

and

$$\delta q^i(t) = 0, \tag{2.4}$$

for $t = t_1$ and $t = t_2$. The equations of motion have the well-known form

$$\frac{d}{dt}\left(\frac{\partial L}{\partial \dot{q}^i}\right) - \frac{\partial L}{\partial q^i} = 0, \tag{2.5}$$

and conditions (2.4) agree with the boundary conditions needed for the integration of system (2.5) for a nondegenerate Lagrangian $\det(\partial^2 L/\partial \dot{q}^i \partial \dot{q}^j) \neq 0$.

On the other hand, Weiss's principle requires δS to be a function of the end points only and identifies the coefficients of δq^i and δt, at the end points, with the momenta (p_i) conjugated to q^i and (minus) the Hamiltonian $(-H)$, respectively. The equations of motion obtained are the same as before, i.e., system (2.5). The definitions of p_i and H coincide with the ones usually adopted for the transition from the Lagrangian to the Hamiltonian theory.[17]

Consider now the first-order action principle which is used for defining the usual Hamiltonian theory, i.e.,

$$p_i = \partial L/\partial \dot{q}^i, \tag{2.6}$$

$$H = H(q^i, p_j, t) = p_i \dot{q}^i - L(q^i, \dot{q}^i, t), \tag{2.7}$$

$$\bar{L} = \bar{L}(q^i, \dot{q}^i, p_j, t) = p_i \dot{q}^i - H(q^i, p_j, t). \tag{2.8}$$

The new action principle defined by $\bar{L}$ is now based on $2n$ independent variables q^i and p_j and the equations of motion are obtained by requiring

$$\delta \bar{S} = 0 \tag{2.9}$$

for arbitrary

$$\delta q^i(t), \delta p_j(t), \quad t_1 < t < t_2 \tag{2.10}$$

and

$$\delta q^i(t) = 0, \tag{2.11}$$

for $t = t_1$ and $t = t_2$. This variational principle yields the well-known Hamilton's equations

$$\dot{p}_i = -\partial H/\partial q^i, \quad \dot{q}^j = \partial H/\partial p_j. \tag{2.12}$$

Nevertheless it is worth noting that, strictly speaking, Hamilton's principle is not wide enough to allow canonical transformations in the Hamiltonian theory. In fact, a canonical transformation to variables Q^i, P_j,

$$Q^i = Q^i(q,p,t), \quad P_j = P_j(q,p,t), \tag{2.13}$$

implies the definition of a new Hamiltonian H' and a new Lagrangian L',

$$\begin{aligned} L' &= \dot{Q}^i P_i - H'(Q,P,t) \\ &= \dot{q}^i p_i - H(q,p,t) + dF(q,p,t)/dt. \end{aligned} \tag{2.14}$$

It is clear that in order to get the equations of motion for Q^i and P_j we must require

$$\delta S' = 0 \tag{2.15}$$

for arbitrary

$$\delta Q^i(t), \quad \delta P_j(t), \quad t_1 < t < t_2, \tag{2.16}$$

and

$$\delta Q^i(t) = 0, \tag{2.17}$$

for $t = t_1$ and $t = t_2$. If we consider a nontrivial transformation, the conditions (2.11) and (2.17) are not equivalent, according to Eq. (2.13). Therefore, using Hamilton's principle in a strict sense does not allow for nontrivial canonical transformations (i.e., those for which $\partial Q^i/\partial p_j \neq 0$).

In order to be able to perform canonical transformations it is more convenient to adopt Weiss's principle. Due to the fact that Lagrangians (2.8) and (2.14) are first order (the same kind we are going to be dealing with in this work) we prefer to choose Weiss's principle to get the equations of motion.

In the appendix we prove that the definitions of the momenta and Hamiltonian obtained from this first-order principle can be used to construct a sound Hamiltonian theory based on Dirac's method for degenerate Lagrangians. In this way we will prove that Weiss's variational principle provides a starting point for constructing infinitely many Lagrangians (and Hamiltonian theories) for any system of differential equations.

3. FIRST-ORDER FORMULATION FOR DYNAMICAL SYSTEMS

Any system of differential equations in which the highest derivatives can be solved for algebraically may be written in a first-order form.[18,19] For purposes of exposition we consider a general system of n coupled second-order differential equations

$$F_i(\ddot{q}^k,\dot{q}^k,q^k,t) = 0, \quad i,k = 1,...,n, \tag{3.1}$$

with q^k being generalized coordinates and the dots denoting total derivatives with respect to the time t. We assume that Eq. (3.1) can be solved for $\ddot{q}^k$. It is always possible to define n new variables q^{n+k} in such a way that the system (3.1) reduces to the first-order form

$$\dot{q}^i = a^i(q^k,q^{n+k},t), \tag{3.2}$$

$$F_i(\dot{q}^{n+k},q^k,q^{n+k},t) = 0. \tag{3.3}$$

We require now that Eq. (3.2) is such that we can solve for the generalized coordinates q^{n+k} and therefore Eqs. (3.2) and (3.3) can be written in compact form

$$\dot{q}^a = f^a(q^b,t), \quad a,b = 1,...,2n \tag{3.4}$$

which is equivalent to Eq. (3.1). We note in passing that a Newtonian system with arbitrary velocity-dependent forces can be brought to form (3.4). One can always choose the function a^i in such a way that $q^{n+k} = \dot{q}^k$. We adopt this choice from now on.

It is worth noticing that any curve $q^i = q^i(t)$ in configuration space can be naturally mapped in a curve of the form $q^a = h^a(t)$, where $h^i(t) = g^i(t)$ for $a = i = 1,...n$ and $h^{n+i}(t) = \dot{g}^i(t)$ for $a = n + i = n + 1,...,2n$. With this prescription the state of a given physical system at time t is characterized by a point $q^a(t)$ and its time evolution will generate a one-parameter family of trajectories in this $2n$-dimensional space. In other words, the dynamics is defined by the tangent vector $f^a(q^b,t)$, which in some sense plays the role of the generator of time displacements. Any two trajectories having the same dynamics but different initial conditions will never intersect in this space and thus the set of all possible solutions to a given problem can be imagined as an infinite collection of curves filling the whole space.

From now on, we restrict ourselves to equations of motion which can be written in the form (3.4). In order to have a Lagrangian system equivalent to that represented by Eq. (3.4), it is necessary that

$$\partial^2 L/\partial\dot{q}^a\partial\dot{q}^b \equiv 0, \quad \forall a,b,q^a,\dot{q}^a,t. \tag{3.5}$$

In fact, the left-hand side of the Euler–Lagrangian equations for a Lagrangian $L = L(q^a,\dot{q}^a,t)$ is

$$L_a \equiv \frac{d}{dt}\left(\frac{\partial L}{\partial\dot{q}^a}\right) - \frac{\partial L}{\partial q^a} \tag{3.6}$$

$$= \frac{\partial^2 L}{\partial\dot{q}^a\partial\dot{q}^b}\ddot{q}^b + \frac{\partial^2 L}{\partial\dot{q}^a\partial q^b}\dot{q}^b + \frac{\partial^2 L}{\partial\dot{q}^a\partial t} - \frac{\partial L}{\partial q^a}, \tag{3.7}$$

which in general contains accelerations. Therefore, Eq. (3.5) has to be fulfilled in order to reproduce Eq. (3.4). That is to say, L must be at most linear in the velocities having the form

$$L = l_a(q^b,t)\dot{q}^a + l_0(q^b,t). \tag{3.8}$$

The corresponding Euler–Lagrange equations are

$$(\partial l_a/\partial q^b - \partial l_b/\partial q^a)\dot{q}^a = \partial l_0/\partial q^b - \partial l_b/\partial t \tag{3.9}$$

and if we require systems (3.9) and (3.4) to be equivalent it is necessary to assume that $(\partial l_a/\partial q^b - \partial l_b/\partial q^a)$ is invertible and to find the solutions l_a and l_0 to the following system of partial differential equations:

$$(\partial l_a/\partial q^b - \partial l_b/\partial q^a)f^a(q,t) = \partial l_0/\partial q^b - \partial l_b/\partial t. \tag{3.10}$$

Using Koenig's theorem[14,20] it can proved that, for any given l_0, a solution for l_a exists. As a matter of fact, Havas considered a system closely related to Eqs. (3.10) [see Eqs. (B12) of this paper] and he proved the existence of solutions. In other words, the inverse problem of the calculus of variations always has a solution when formulated using first-order differential equations for the time evolution. Nevertheless, there is no general prescription on how to solve system (3.10) and one does not know of any possible relationship among the different Lagrangians that will arise. We remind the reader that when using a second-order formulation for the dynamics. the existence of a Lagrange function is not guaranteed at all. In particular, there are explicit examples in which the Lagrangian simply does not exist.[1,21]

In the next section we present a different approach which solves the inverse problem of the calculus of variations in its first-order form for an arbitrary set of complete trajectories characterized by a given dynamics. We give a definite prescription on how to construct all possible Lagrangians that arise and we also show that they are infinite in number.

4. THE INVERSE PROBLEM OF THE CALCULUS OF VARIATION IN FIRST-ORDER FORM

It will be convenient for our purposes to consider the time as another coordinate by enlarging our space to $2n + 1$ dimensions with the notation $x^0 = t(\tau), x^a = q^a(\tau)$. The complete set of trajectories corresponding to a given dynamics is characterized now by the tangent

$$dx^\mu/d\tau = f^\mu(x), \quad \mu = 0,1,...,2n, \tag{4.1}$$

where the ratios f^a/f^0 are given functions and $f^0 = dt/d\tau$ is arbitrary. The action is written in general as

$$S = \int l_\mu(x)\frac{dx^\mu}{d\tau}\,d\tau, \tag{4.2}$$

where the freedom in the τ-parametrization is self-evident.

We can always choose our parameter τ as the time t and this is equivalent to the normalization $f^0(x) = 1$. The equations of motion obtained from (4.2) are then written as

$$(\partial l_\mu/\partial x^\nu - \partial l_\nu/\partial x^\mu)dx^\nu/d\tau \equiv M_{\mu\nu}dx^\nu/d\tau = 0. \quad (4.3)$$

Due to the antisymmetry of $M_{\mu\nu}$ ($= -M_{\nu\mu}$) together with the odd dimensionality of the space considered, the determinant of $M_{\mu\nu}$ vanishes. We assumed in Sec. 3 that

$$\det M_{ab} \neq 0; \quad (4.4)$$

therefore the rank of $M_{\mu\nu}$ is $2n$. Thus, Eq. (4.3) implies that $dx^\nu/d\tau$ is the *only* eigenvector of $M_{\mu\nu}$ with zero eigenvalue. The eigenvector with zero eigenvalue of an antisymmetric matrix $M_{\mu\nu}$ in a space with an odd number of dimensions is proportional to the vector

$$v^\mu = \epsilon^{\mu\mu_1\mu_2\cdots\mu_{2n-1}\mu_{2n}} M_{\mu_1\mu_2}\cdots M_{\mu_{2n-1}\mu_{2n}}. \quad (4.5)$$

If two Lagrangians are s-equivalent (i.e., their complete set of solutions coincides), the eigenvectors with zero eigenvalues associated with them are then parallel (not necessarily equal) to each other in order to define the same solution. Furthermore, the freedom of τ-parametrization does not allow us to compute $dx^\mu/d\tau$ from the Lagrangian only; therefore, this vector is parallel to v^μ,

$$dx^\mu/d\tau = \lambda(\tau) l(x) v^\mu, \quad (4.6)$$

where $\lambda(\tau)$ is related to the τ-parametrization and $l(x)$ depends on which s-equivalent Lagrangian is chosen. Nevertheless, the physical velocity is well defined,

$$\frac{dx^\mu}{dt} = \frac{dx^\mu}{dx^0} = \frac{dx^\mu/d\tau}{dx^0/d\tau} = \frac{\epsilon^{\mu\cdots}M\cdots M}{\epsilon^{0\cdots}M\cdots M}. \quad (4.7)$$

Using this notation, the inverse problem of the calculus of variations reduces to finding all the possible functions $l_\mu(x)$ such that the extremal requirement $\delta S = 0$ reproduces Eqs. (4.3) with the given direction defined by the vector $f^\mu(x)$. In order to construct the most general Lagrangian we consider l_μ to be a covariant vector in this $(2n+1)$-dimensional space, where the only explicitly defined direction is that of the tangent vector $f^\mu(x)$. The problem thus arises of constructing a suitable basis in this space. In particular, it is necessary to define the subspace orthogonal to the given tangent to the solution curve. A very natural basis for this subspace is the one generated by the $2n$ constants of motion associated with Eq.(4.1). Let us remark that system (4.1) possesses $2n$ independent functions $C^{(a)}$, which depend upon the coordinates x^μ and that correspond to the initial values $x^a(\tau_0)$ which completely specify the curves for a given dynamics.[22] These constants of motion can be obtained by inverting the solutions $x^a = x^a(C^{(b)},\tau)$ of the system (4.1) and they satisfy the conservation equations.

$$\frac{\partial C^{(a)}}{\partial x^\mu} f^\mu(x) = \frac{\partial C^{(a)}}{\partial x^\mu}\frac{dx^\mu}{d\tau} = \frac{dC^{(a)}}{d\tau} = 0. \quad (4.8)$$

In other words, we have generated $2n$ vectors $C^{(a)}{}_{,\mu} = \partial C^{(a)}/\partial x^\mu$, which are orthogonal to $f^\mu(x)$.

We choose $C^{(a)}(x^\mu)$ to be $2n$ independent functions so that the determinant of the matrix $C^{(a)}{}_{,b}$ is different from zero, implying that the $2n$ vectors $C^{(a)}{}_{,\mu}$ are linearly independent. Thus, the set $\{C^{(a)}{}_{,\mu}\}$ constitutes a basis for the subspace orthogonal to $f^\mu(x)$. The vector $C^{(0)}{}_{,\mu}$ where $C^{(0)} = t$, is such that

$$\det(C^{(\alpha)}{}_{,\mu}) = \det(C^{(a)}{}_{,b}) \neq 0. \quad (4.9)$$

This means that the $2n+1$ vectors $C^{(\alpha)}{}_{,\mu}$ are linearly independent and constitute a basis for the $2n+1$ dimensional space considered.

The most general vector $\bar{l}_\mu(x)$ can be written as a linear combination of the basis vectors $C^{(\alpha)}{}_{,\mu}$. For our purposes, it is more convenient to consider the coefficients of this linear combination to be functions of the $2n+1$ functionally independent "new coordinates" $C^{(\alpha)}(x)$. Therefore, we write

$$\bar{l}_\mu = \bar{l}_{(a)}(C^{(b)},C^{(0)})C^{(a)}{}_{,\mu} + \bar{l}_0(C^{(b)},C^{(0)})C^{(0)}{}_{,\mu}. \quad (4.10)$$

Among all s-equivalent Lagrangians there are some which are trivially related by a total time derivative which, in terms of $\bar{l}_\mu$, means the gradient of an arbitrary function Λ. Without losing generality, we can then consider

$$l_\mu = \bar{l}_\mu - \partial_\mu \Lambda = \bar{l}_{(a)}C^{(a)}{}_{,\mu} + \bar{l}_0 C^{(0)}{}_{,\mu} - \frac{\partial \Lambda}{\partial C^{(a)}}C^{(a)}{}_{,\mu} - \frac{\partial \Lambda}{\partial C^{(0)}}C^{(0)}{}_{,\mu} \quad (4.11)$$

and choose

$$\frac{\partial \Lambda}{\partial C^{(0)}}(C^{(b)},C^{(0)}) = \bar{l}_0(C^{(b)},C^{(0)}), \quad (4.12)$$

which implies that

$$l_\mu = \left(\bar{l}_{(a)}(C^{(b)},C^{(0)}) - \frac{\partial \Lambda}{\partial C^{(a)}}(C^{(b)},C^{(0)})\right)C^{(a)}{}_{,\mu} \quad (4.13)$$

$$\equiv l_{(a)}(C^{(b)},C^{(0)})C^{(a)}{}_{,\mu}, \quad (4.14)$$

where $l_{(a)}$ still depends on the $2n+1$ variables $C^{(\alpha)}$.

The next step in the construction is to require that the Euler–Lagrange equations arising from (4.2) are satisfied by virtue of the equations of motion (4.1). That is to say, we require

$$L_\mu \equiv \frac{d}{d\tau}(l_{(a)}C^{(a)}{}_{,\mu}) - \frac{\partial}{\partial x^\mu}(l_{(a)}C^{(a)}{}_{,\nu})\frac{dx^\nu}{d\tau} \quad (4.15)$$

$$= \frac{\partial l_{(a)}}{\partial C^{(b)}}C^{(a)}{}_{,\mu}\frac{dC^{(b)}}{d\tau} + \frac{\partial l_{(a)}}{\partial C^{(0)}}C^{(0)}{}_{,\mu}\frac{dC^{(0)}}{d\tau} + l_{(a)}C^{(a)}{}_{,\mu\nu}\frac{dx^\nu}{d\tau} - \frac{\partial l_{(a)}}{\partial C^{(b)}}C^{(b)}{}_{,\mu}\frac{dC^{(a)}}{d\tau} - l_{(a)}C^{(a)}{}_{,\nu\mu}\frac{dx^\nu}{d\tau} \quad (4.16)$$

to be zero. In Eq. (4.16) the third and fifth terms cancel each other and the first and fourth terms vanish due to Eqs. (4.8). Then we are left with

$$\frac{\partial l_{(a)}}{\partial C^{(0)}}(C^{(b)},C^{(0)})C^{(a)}{}_{,\mu}\frac{dC^{(0)}}{d\tau} = 0. \quad (4.17)$$

The factor $dC^{(0)}/d\tau$ can never be zero for any proper parametrization ($dC^{(0)}/d\tau = dt/d\tau$). The matrix $C^{(a)}{}_{,\mu}$ has rank $2n$; therefore, we have $2n$ equations.

$$\frac{\partial l_{(a)}}{\partial C^{(0)}}(C^{(b)}, C^{(0)}) = 0, \quad (4.18)$$

which imply that $l_{(a)}$ are arbitrary functions of the constants of motion $C^{(b)}$ only. Thus the most general Lagrangian can be written as

$$L = l_{(a)}(C^{(b)})C^{(a)}{}_{,\mu}\frac{dx^\mu}{d\tau}. \tag{4.19}$$

It is worth mentioning that L vanishes when the equations of motion hold.

There is still one more condition on the components $l_{(a)}$ which is derived by examining the converse problem of obtaining the equations of motion (4.1) from the stationary condition on the action (4.2). In other words, we demand that

$$\left(\frac{\partial l_{(a)}}{\partial C^{(b)}} - \frac{\partial l_{(b)}}{\partial C^{(a)}}\right) C^{(a)}{}_{,\mu} C^{(b)}{}_{,\nu}\frac{dx^\nu}{d\tau} = 0 \tag{4.20}$$

implies $dC^{(a)}/d\tau = 0$ [which is equivalent to Eqs. (4.1)]. Because the vectors $C^{(a)}{}_{,\mu}$ are linearly independent, Eq. (4.20) tells us that

$$\left(\frac{\partial l_{(a)}}{\partial C^{(b)}} - \frac{\partial l_{(b)}}{\partial C^{(a)}}\right)\frac{dC^{(b)}}{d\tau} = 0. \tag{4.21}$$

In order to recover the equations of motion (4.1) we must construct the functions $l_{(a)}$ in such a way that the determinant of the $2n\times 2n$ antisymmetrical matrix

$$\eta_{(a)(b)} = \partial l_{(a)}/\partial C^{(b)} - \partial l_{(b)}/\partial C^{(a)} \tag{4.22}$$

is different from zero. Then we can deduce

$$\frac{dC^{(b)}}{d\tau} = C^{(b)}{}_{,\mu}\frac{dx^\mu}{d\tau} = 0 \tag{4.23}$$

from Eq. (4.21). Recalling that the vectors $C^{(b)}{}_{,\mu}$ satisfy

$$C^{(b)}{}_{,\mu} f^\mu(x) = 0 \tag{4.24}$$

and that the rank of $C^{(\nu)}{}_{,\mu}$ is $2n$, we finally conclude from (4.23) and (4.24) that

$$dx^\mu/dx^0 = f^\mu/f^0, \tag{4.25}$$

which is equivalent to Eq. (4.1).

The condition upon $\det \eta_{(a)(b)}$ can be implemented in an infinite number of ways. We can see this by considering a subclass of all possible ways of constructing the functions $l_{(a)}$. Let us first remind the reader that the determinant of an antisymmetric matrix η in a space of even dimensionality is proportional to the square of its Pfaffian,

$$\mathrm{Pf}\eta = \epsilon^{a_1 a_2 \cdots a_{2n-1} a_{2n}}\eta_{a_1 a_2}\cdots\eta_{a_{2n-1} a_{2n}}. \tag{4.26}$$

Therefore, it is enough to require that $\mathrm{Pf}\eta$ is different from zero to ensure $\det \eta\equiv 0$. Equation (4.26) suggests the following construction for the functions $l_{(a)}$:

$$l_{(a)} = \frac{1}{2p+1}(C^{(a+1)})^{2p+1} + \rho_{(a)}C^{(a+1)}; \qquad a = 1,3,\ldots,2n-1,$$

$$l_a = 0, \quad a = 2,4,\ldots,2n, \tag{4.27}$$

where p is an arbitrary integer and $\rho_{(a)}$ are positive numbers. The only matrix elements different from zero are then

$$\eta_{(2k+1)(2k+2)} = (C^{(2k+2)})^{2p} + \rho_{(2k+1)} > 0, \qquad k = 0,1,\ldots,n-1, \tag{4.28}$$

which ensure that expression (4.26) is strictly positive. In this way we have provided a prescription to construct an infinitely denumerable set of Lagrangians which reproduce the equations of motion (4.1) and which obviously do not cover all the possible ways of defining adequate functions $l_{(a)}$.

As can be seen from Eq. (4.19), our method is useful whenever a system of differential equations can be equivalently written as $\dot{C}^{(a)} = 0$, where $C^{(a)}$ are $2n$ functionally independent constants of motion of the original system. There may be cases (when f^a is singular) where the constants of motion are not well defined everywhere and in such cases our method will not be directly applicable. Although the method is based on a local (i.e., nonglobal) approach its content is nontrivial, as shown by the fact that we are able to present in the next section examples of first-order Lagrangians for two systems which do not possess second-order Lagrangians.

Finally, a comment may be added regarding the matrix $\Lambda_a{}^b$ given by

$$\bar{L}_a = \Lambda_a{}^b L_b,$$

which relates the equations of motion of two s-equivalent Lagrangians $\bar{L}$ and L.[5] The expressions for $\bar{L}_a$ and L_b are defined by Eq. (3.6). Using Eqs. (4.3), (4.19), and (4.20) it can be proved directly that $\mathrm{tr}\,(\Lambda^k)$ are constants of motion for any integer k, with $\Lambda_a{}^b = \bar{M}_{ac}(M^{-1})^{cb}$. This theorem is the first-order counterpart of the one for second-order Lagrangians presented in Ref. 5.

5. EXAMPLES

In this section we are going to construct first-order Lagrangians for two systems of second-order differential equations for which no second-order Lagrangian exists. The construction is based on the method developed in Sec. 4.

Example 1: Consider the system of two differential equations for the variables $x = x(t)$ and $y = y(t)$,

$$\ddot{x} + \dot{y} = 0,$$
$$\ddot{y} + y = 0. \tag{5.1}$$

It may be easily proved that system (5.1) corresponds to the case III-b of Douglas's classification and therefore has no second-order Lagrangian.[21] Nevertheless, a first-order Lagrangian for such a set of equations can be found by writing it in first-order form with the definitions

$$x_1 = x, \quad x_2 = y, \quad x_3 = \dot{x}, \quad x_4 = \dot{y}. \tag{5.2}$$

In this notation system (5.1) becomes

$$\dot{x}_1 = x_3, \quad \dot{x}_2 = x_4,$$
$$\dot{x}_3 = -x_4, \quad \dot{x}_4 = -x_2, \tag{5.3}$$

and its general solution is

$$\begin{aligned} x_1 &= -A\sin t + B\cos t + Ct + D,\\ x_2 &= A\cos t + B\sin t,\\ x_3 &= -A\cos t - B\sin t + C,\\ x_4 &= -A\sin t + B\cos t. \end{aligned} \tag{5.4}$$

The constants of motion A, B, C, and D can be explicitly written in terms of x_1, x_2, x_3, x_4 and t:

$$
\begin{aligned}
A &= x_2\cos t - x_4\sin t,\\
B &= x_2\sin t + x_4\cos t,\\
C &= x_2 + x_3,\\
D &= x_1 - x_4 - (x_2 + x_3)t.
\end{aligned} \tag{5.5}
$$

One possible Lagrangian for the system is

$$L = \tfrac{1}{2}(A\dot{B} - B\dot{A} + C\dot{D} - D\dot{C}), \tag{5.6}$$

where we have used Eq. (4.27) with $\rho_{(a)} = 0$ and $p = 0$. The Lagrangian (5.6) expressed in terms of x_1, x_2, x_3, x_4 and their time derivatives becomes (up to a total time derivative)

$$L = (x_2 + x_3)\dot{x}_1 + x_4\dot{x}_3 + \tfrac{1}{2}(x_4^2 - 2x_2x_3 - x_3^2). \tag{5.7}$$

In fact, varying x_1, x_2, x_3, and x_4 we obtain the equations of motion

$$
\begin{aligned}
&\dot{x}_2 + \dot{x}_3 = 0,\\
&-\dot{x}_1 + x_3 = 0,\\
&-\dot{x}_1 + \dot{x}_4 + x_2 + x_3 = 0,\\
&-\dot{x}_3 - x_4 = 0,
\end{aligned} \tag{5.8}
$$

which are equivalent to the system (5.3).

We have, therefore, found a first-order Lagrangian for the set of equations (5.1) or (5.3) which cannot be obtained from a usual (second-order) Lagrangian.

It is worth noting that in the Lagrangian (5.7) only half of the time derivatives of the variables (i.e., $\dot{x}_1$ and $\dot{x}_3$) occur. This is exactly the same situation encountered in Lagrangian (2.8), where only the time derivatives of the coordinates (and not the ones of the momenta) appear.

Example 2: Now let us consider the system of differential equations

$$
\begin{aligned}
&\ddot{x}_1 + 2\gamma_1\dot{x}_1 + \omega_1^2 x_1 - \xi x_2 = 0,\\
&\ddot{x}_2 + 2\gamma_2\dot{x}_2 + \omega_2^2 x_2 - \eta x_1 = 0,
\end{aligned} \tag{5.9}
$$

which describes two coupled damped oscillators with different frequencies and friction coefficients. As mentioned in Ref. 21 such a system is not derivable from a second-order Lagrangian when the condition

$$\xi\eta(\gamma_1 - \gamma_2)(\omega_1^2 - \gamma_1^2 + \omega_2^2 - \gamma_2^2) \neq 0 \tag{5.10}$$

is met. This last equation says that the system under consideration is type IV in the Douglas classification.

In order to exhibit a relatively simple example of a first-order Lagrangian that reproduces a set of equations like (5.9) we have considered the case

$$
\begin{aligned}
&\ddot{x}_1 - \tfrac{1}{4}x_2 = 0,\\
&\ddot{x}_2 + \tfrac{4}{3}\dot{x}_2 - \tfrac{1}{3}x_1 = 0,
\end{aligned} \tag{5.11}
$$

which corresponds to the choice of parameters $\gamma_1 = \omega_1 = \omega_2 = 0, \gamma_2 = \frac{2}{3}, \eta = \frac{1}{3}$, and $\xi = \frac{1}{4}$ that obviously satisfy the condition (5.10).

Writing the system (5.11) in first-order form,

$$
\begin{aligned}
&\dot{x}_1 = x_3, \quad \dot{x}_2 = x_4,\\
&\dot{x}_3 = \tfrac{1}{4}x_2,\\
&\dot{x}_4 = -\tfrac{4}{3}x_4 + \tfrac{1}{3}x_1,
\end{aligned} \tag{5.12}
$$

and using a procedure similar to that of Example 1, we obtain the following first-order Lagrangian (up to a total time derivative):

$$
\begin{aligned}
L = e^t\,[&(6x_4 - x_2)\dot{x}_1 + (2x_3 + 12x_4 - 3x_1)\dot{x}_2\\
&+ (18x_4 - 4x_1)\dot{x}_3 + (6x_3 + 3x_2)\dot{x}_4 + x_1^2 + \tfrac{1}{4}x_2^2\\
&- 2x_3^2 - \tfrac{9}{2}x_4^2\,]\\
&+ e^{t/3}[(3x_2 + 2x_4)\dot{x}_1 + (x_1 + 6x_3 - 4x_4)\dot{x}_2\\
&+ (4x_1 + 2x_4)\dot{x}_3 + (6x_3 - 3x_2)\dot{x}_4\\
&+ \tfrac{1}{3}x_1^2 + \tfrac{3}{4}x_2^2 + 2x_3^2 + \tfrac{1}{2}x_4^2\,].
\end{aligned} \tag{5.13}
$$

It is straightforward to show that upon variations we obtain the equations of motion

$$
\begin{aligned}
&e^t\,[6x_4 - x_2 - 2x_1 + 2\dot{x}_2 + 6\dot{x}_4 + 4\dot{x}_3]\\
&\quad + e^{t/3}[\,x_2 + \tfrac{2}{3}x_4 - \tfrac{2}{3}x_1 + 2\dot{x}_2 - 4\dot{x}_3 + 2\dot{x}_4] = 0,\\
&e^t\,[2x_3 + 12x_4 - 3x_1 - \tfrac{1}{2}x_2 - 2\dot{x}_1 + 9\dot{x}_4 + 2\dot{x}_3]\\
&\quad + e^{t/3}[\,2x_3 - \tfrac{4}{3}x_4 + \tfrac{1}{3}x_1 - \tfrac{3}{2}x_2 - 2\dot{x}_1 - \dot{x}_4 + 6\dot{x}_3] = 0,\\
&e^t\,[4x_3 + 18x_4 - 4x_1 - 4\dot{x}_1 - 2\dot{x}_2 + 12\dot{x}_4]\\
&\quad + e^{t/3}[\,-4x_3 + \tfrac{2}{3}x_4 + \tfrac{4}{3}x_1 + 4\dot{x}_1 - 6\dot{x}_2 - 4\dot{x}_4] = 0,\\
&e^t\,[6x_3 + 9x_4 + 3x_2 - 6\dot{x}_1 - 9\dot{x}_2 - 12\dot{x}_3]\\
&\quad + e^{t/3}[2x_3 - x_4 - x_2 - 2\dot{x}_1 + \dot{x}_2 + 4\dot{x}_3] = 0.
\end{aligned} \tag{5.14}
$$

Now, in order to prove the equivalence of (5.14) with (5.12), we have to solve for the time derivatives $\dot{x}_a$ $(a = 1, 2, 3, 4)$. This can be done because the matrix of the coefficients has a nonvanishing determinant and the result is indeed system (5.12).

With these examples we have shown that there are instances where, in spite of the local nature of the method, first-order Lagrangians can be found for systems which are not derivable from usual second-order Lagrangians. We interpret this fact as a clear indication of the advantages of using a first-order formulation for the variational principle.

6. SUMMARY AND CONCLUSIONS

The main results of this paper are contained in Sec. 4, where we show how to construct infinitely many first-order Lagrangians for any given system of differential equations such that the highest derivatives can be algebraically solved for. The construction is achieved by rewriting such a system, introducing a suitable definition of new variables, in an equivalent form which consists of first-order differential equations only. The Lagrangian is then found using very simple geometrical arguments.

The first step is to map the general solution to the original system in the space defined by the original coordinates, their velocities, and the time (in the case the original system was second-order). In this space the solution is defined by the direction of its tangent vector at each point.

When the constants of motion related to the original system are globally well defined, this method guarantees that the construction of the Lagrangian [Eq. (4.19)] may always be achieved in infinitely many different ways. However, the explicit construction can only be performed when all the (functionally independent) constants of motion are known. Therefore, the method presented here should not be viewed only as a practical constructive procedure but rather as an explicit existence theorem which clearly shows the relationship among the infinitely many different first-order Lagrangians of a system. This situation is in contrast with the usual prescription ($L = T - V$) for second-order Lagrangians where the knowledge of the total energy ($T + V$) is enough to construct L. Nevertheless, for the second-order case, this prescription does not always work[1,21] and, in any case, provides only one of the many possible equivalent Lagrangians.[1–3,5,13]

Section 5 illustrates one way of explicitly constructing first-order Lagrangians for system for which no second-order Lagrangian exists. This fact seems to indicate that a first-order formulation of the variational principle has some advantages over the usual second-order one.

We should perhaps note in passing that Eq. (4.19) implies that any set of differential equations may be regarded as a variational problem. Furthermore, Eq. (4.2) constitutes a geometrization of the problem in the sense that Eqs. (4.3) are the geodesics equations of a degenerate Finsler space defined by the metric vector $l_\mu(x)$.

It is also worth noting that the geometrical approach developed in this paper is fully covariant under arbitrary coordinate transformations of the form $x'^\mu = x'^\mu(x^\nu)$.

We further remark that the whole family of Lagrangians represented by Eq. (4.19) does not depend on the choice of the set of constants of motion. In fact, if one chooses a different set $D^{(a)} = D^{(a)}(C^{(b)})$, say, then

$$l_\mu = l_{(a)}(D^{(b)})D^{(a)}{}_{,\mu} = l_{(a)} \frac{\partial D^{(a)}}{\partial C^{(b)}} C^{(b)}{}_{,\mu} = \tilde{l}_{(b)} C^{(b)}{}_{,\mu},$$

where

$$\tilde{l}_{(b)} = l_{(a)}(D^{(c)}) \partial D^{(a)} / \partial C^{(b)},$$

are still constants of motion and can be written in terms of the original set $C^{(a)}$.

The Hamiltonian theory associated with (degenerate) first-order Lagrangians is constructed in the Appendix using Dirac's method.[16,17] There we prove that the Euler–Lagrange and Hamiltonian equations of motion for those Lagrangians agree. Furthermore, the Hamiltonian theory of one of the first-order Lagrangians agree with the usual one, when a second-order Lagrangian exists for the system under consideration.

Finally, a comment on the problem of quantization. In the second-order formalism, the existence of a Lagrangian is not guaranteed. On the other hand, in the first-order formulation there are infinitely many of them. It is known[10,23] that the usual quantization procedure gives rise to nonequivalent quantum theories when using different (classically) s-equivalent Lagrangians. Therefore, one needs either a criterion to single out one among infinitely many Lagrangians (in the first-order formalism) or a quantum theory which is not based on the Lagrangians (which are not physically-measurable entities) but on other objects that remain invariant under the change of Lagrangians. One possible choice of such objects is the set of all curves which satisfy Eq. (4.1) [and/or Eq. (4.1) itself]. Such a quantum theory would be, by definition, invariant under the change of Lagrangians and would, in principle, allow one to quantize systems whose equations of motion cannot be derived from a (second-order) Lagrangian.

ACKNOWLEDGMENTS

We wish to express our deepest gratitude to Larry Shepley for thoughtful comments and for bringing Ref. 21 to our attention. One of us (S. H.) would like to thank Luis de la Peña, Santiago López de Medrano, Rafael Pérez-Pascual, and Sevín Recillas for useful suggestions.

APPENDIX

We construct here the Hamiltonian theory for the (degenerate) Lagrangian (3.8),

$$L = l_a(q^b)\dot{q}^{(a)} + l_0(q^b), \tag{A.1}$$

using Dirac's method. For simplicity, we have dropped the explicit time dependence of l_a and l_0.

From Weiss's principle (or the usual definitions), it is straightforward to obtain the canonical momenta π_{q^a} and the Hamiltonian h,

$$\pi_{q^a} \equiv \pi_a = l_a(q^b), \tag{A.2}$$

$$h = -l_0(q^b). \tag{A.3}$$

Due to the fact that Eqs. (A.2) do not contain the velocities $\dot{q}^a$ at all, there are $2n$ primary constraints ϕ_a,

$$\phi_a = \pi_a - l_a(q^b) \approx 0. \tag{A.4}$$

The Hamiltonian h has to be modified in order to get the right equations of motion with the usual Poisson brackets. The modified Hamiltonian h_T is given by

$$h_T = h + \lambda^a \phi_a, \tag{A.5}$$

where λ^a are Lagrange multipliers. The Poisson brackets are

$$[q^a, \pi_b] = \delta^a_b. \tag{A.6}$$

The consistency requirement for the constraints is that their time derivatives vanish,

$$\dot{\phi}_b = [\phi_b, h_T] \approx 0. \tag{A.7}$$

Condition (A.7) implies

$$(\partial l_a / \partial q^b - \partial l_b / \partial q^a)\lambda^a + \partial l_0 / \partial q^b = 0, \tag{A.8}$$

i.e.,

$$\lambda^a = (M^{-1})^{ab}(-\partial l_0 / \partial q^b), \tag{A.9}$$

where M is the nonsingular antisymmetric matrix defined in Eq. (4.3). There are no secondary constraints in the theory. The Lagrange multipliers λ^a are determined by Eq. (A.9), which means that no arbitrary (gauge) functions appear in the theory. The equations of motion for q^b can be obtained in the usual way

$$\begin{aligned} \dot{q}^b &= [q^b, h_T] \\ &= [q^b, -l_0(q^a) + \lambda^a(\pi_a - l_a)(q^c)] \\ &= \lambda^b, \end{aligned} \tag{A.10}$$

and they agree with Eqs. (3.9), obtained directly from the

Lagrangian (A.1), when Eqs. (A.9) are used. We now prove that all the ϕ_a constraints are second class. In fact,

$$[\phi_a,\phi_b] = [\pi_a - l_a(q),\pi_b - l_b(q)] = \partial l_b/\partial q^a - \partial l_a/\partial q^b = M_{ba} = -M_{ab}. \tag{A.11}$$

Therefore, the matrix $[\phi_a,\phi_b]$ is invertible and no linear combination of the ϕ_a's is first class.

The Dirac bracket for any two functions $f = f(g,t)$ and $q = q(g,t)$ is defined by

$$[f,g]^* = [f,g] - [f,\phi_a](-M^{-1})^{ab}[\phi_b,q], \tag{A.12}$$

$$[f,g]^* = (M^{-1})^{ab}\frac{\partial f}{\partial q^a}\frac{\partial q}{\partial q^b}, \tag{A.13}$$

which agrees with the results found (in a different way) in Ref. 17. Let us mention that it is enough to consider only functions of q^a and t because the momenta π_a can be completely written in terms of the coordinates q^b due to the fact that the $2n$ Eqs. (A.4) now become strong equations with the Dirac brackets. The equations of motion (A.10) remain the same because h_T is first class, i.e.,

$$\dot{q}^a = [q^a,h_T]^* = [q^a,h_T]. \tag{A.14}$$

Of course the numerical values of h_T and h coincide.

In this way we have proved that Weiss's principle gives rise to the desired equations of motion together with a reasonable Hamiltonian theory.

The results obtained above can be used to prove that when a second-order, nondegenerate, Lagrangian for a system of n variables Q^i exists, then its usual Hamiltonian theory defined by

$$P_i = \partial L/\partial \dot{Q}^i,$$
$$H = H(Q^i,P_j,t) = (\partial L/\partial \dot{Q}^i)\dot{Q}^i - L(Q^i,\dot{Q}^i,t), \tag{A.15}$$

agrees with the one obtained (in the way we have just done) from the first-order Lagrangian with $2n$ variables Q^i,P_j given by

$$\bar{L} = \bar{L}(Q^i,\dot{Q}^i,P_j,t) = \dot{Q}^iP_i - H(Q^i,P_j,t). \tag{A.16}$$

One has, for this case,

$$\begin{aligned} &q^a = Q^a, \quad l_a = P_a, \quad \text{for } a = 1, 2, \ldots, n\\ &q^a = P_a, \quad l_a = 0, \quad \text{for } a = n+1,\ldots,2n\\ &l_0 = -H. \end{aligned} \tag{A.17}$$

The canonical momenta conjugate to Q^i and P_j are

$$\pi_{q^a} \equiv \pi_a = P_a \quad \text{for} \quad a = 1, 2, \ldots, n,$$
$$\pi_{q^a} \equiv \pi_a = 0 \quad \text{for} \quad a = n+1, \ldots, 2n. \tag{A.18}$$

The constraints are

$$\phi_a = \pi_a - P_a \approx 0 \quad \text{for } a = 1, 2, \ldots, n,$$
$$\phi_a = \pi_a \simeq 0 \quad \text{for } a = n+1,\ldots,2n, \tag{A.19}$$

and the nonsingular matrix $[\phi_a,\phi_b]$ is

$$[\phi_a,\phi_b] = -M_{ab} = \begin{pmatrix} 0 & -1\\ 1 & 0\end{pmatrix}. \tag{A.20}$$

Therefore, the Dirac brackets for $f = f(Q,P,t)$ and $g = g(Q,P,t)$ are

$$[f,g]^* = \frac{\partial f}{\partial Q^i}\frac{\partial g}{\partial P_i} - \frac{\partial f}{\partial P_i}\frac{\partial g}{\partial Q^i}, \tag{A.21}$$

the equations of motion are Hamilton's, and h_T and H coincide numerically. In this way, we have proved not only that the Hamiltonian and Euler–Lagrange theories constructed from first-order Lagrangians agree, but also that one of the Hamiltonian theories obtained from first-order Lagrangians coincide with the usual Hamiltonian theory when a second order Lagrangian for the system exists.

[1] J. Douglas, Trans. Am. Math. Soc. **50**, 71 (1941).
[2] D. F. Currie and E. J. Saletan, J. Math. Phys. **7**, 967 (1966).
[3] R. M. Santilli, *Foundations of Theoretical Mechanics* 1 (Springer, Heidelberg, 1978).
[4] S. Okubo, Phys. Rev. D (to appear) and "Canonical Quantization for Some Dissipative Systems and Non-uniqueness of Lagrangians," University of Rochester preprint, 1980.
[5] S. Hojman and H. Harleston, "Equivalent Lagrangians: Multidimensional Case," J. Math. Phys. **22**, 1414 (1981).
[6] W. Sarlet, Hadronic J. **2**, 407 (1979).
[7] R. M. Santilli, Ann. Phys. (NY) **103**, 354 (1977); **103**, 409 (1977); **105**, 227 (1977).
[8] R. M. Santilli, Hadronic J. **1**, 223 (1978); **1**, 574 (1978).
[9] Y. Gelman and E. J. Saletan, Nuovo Cimento B **18**, 62 (1973).
[10] G. Marmo and E. J. Saletan, Hadronic J. **1**, 955 (1978).
[11] W. Sarlet and F. Cantrijn, Hadronic J. **1**, 101 (1978).
[12] A. Hanson, T. Regge, and C. Teitelboim, *Constrained Hamiltonian Systems* (Accademia Nazionale Dei Lincei, Rome, 1976).
[13] G. Darboux, *Leçons sur la Théorie Générale des Surfaces* (Gauthier-Villars, Paris, 1894).
[14] P. Havas, Acta. Phys. Aust. **38**, 145 (1973).
[15] S. Deser and B. Zumino, Phys. Lett. B **62**, 335 (1976).
[16] P. A. M. Dirac, "Lectures on Quantum Mechanics," Belfer Graduate School of Science, Yeshiva University, New York (1964).
[17] E. C. G. Sudarshan and N. Mukunda, *Classical Dynamics: A Modern Perspective* (Wiley, New York, 1974).
[18] V. I. Arnold, *Mathematical Methods of Classical Mechanics* (Springer, Heidelberg, 1978).
[19] Y. Choquet-Bruhat, C. De Witt-Morette, and M. Dillard-Bleick, *Analysis, Manifolds, and Physics* (North-Holland, Amsterdam, 1977).
[20] A. R. Forsyth, *Theory of Differential Equations* (Cambridge U. P., Cambridge, 1906), Vol. V, Art. 160-161.
[21] V. V. Dodonov, V. I. Man'ko, and V. D. Skarzhinsky, "Arbitrariness in the Choice of Action and Quantization of the Given Classical Equations of Motion," P. N. Lebedev Physical Institute Preprint No. 216 (1978).
[22] L. S. Pontryagin, *Ordinary Differential Equations* (Addison-Wesley, Reading, Mass., 1962).
[23] S. Hojman and R. Montemayor, Hadronic J. **3**, 1644 (1980).

Dynamical Noether invariants for time-dependent nonlinear systems

R. S. Kaushal [a] and H. J. Korsch
Fachbereich Physik, Universität Kaiserslautern, D-6750 Kaiserslautern, Federal Republic of Germany

(Received 20 March 1980; accepted for publication 18 March 1981)

Dynamical invariants are derived for time-dependent systems with nonlinear equations of motion including nonharmonic damped systems. The concept of a dynamical algebra is discussed and its utility for the construction of dynamical invariants for nonharmonic systems is demonstrated. Finally we show the existence of dynamical invariants for some nonlinear quantum systems.

PACS numbers: 03.20. + i, 03.65. − w

1. INTRODUCTION

The study of time-dependent oscillator systems has attracted considerable interest in the literature, both in classical [1–13] and quantum [11–15] mechanics. The origin of this development was no doubt the discovery of an exact invariant by Lewis,[2,11] which was previously known as an approximate adiabatic invariant.[1] The existence of a conserved quantity, i.e., an invariant, is of importance in many physical problems. For example, its utility for the motion of charged particles in time varying electromagnetic fields has already been known for a long time, and very recently the invariant has been applied to some models for cosmological particle production.[16]

Since the basic work by Lewis[2,11] various derivations of the dynamic invariant have been presented in the literature. Günther and Leach[13] and Leach[5,6] used time-dependent canonical transformations, Lutzky[7] applied Noether's theorem, and Ray and Reid[8,9] obtained the invariant by Ermakov's technique. Very recently one of the present authors constructed the invariant by means of the dynamical algebra.[10] This algebraic technique provides a direct and unsophisticated derivation of the dynamical invariant. Furthermore, it allows a straightforward transition from classical to quantum systems, because in the algebraic treatment the formulation of classical and quantum dynamics is almost identical.[17]

The existence of invariants for nonharmonic systems[3,6] was recently demonstrated by Ray and Reid,[8,9] who derived a family of invariants for a special class of systems with nonlinear equations of motion. It is obvious, however, that the study of invariants for nonharmonic time-dependent systems is only beginning and is far from being understood. In Sec. 2 we give a short discussion of Ray and Reid's[8,9] results for nonharmonic systems and extend them to the case of time-dependent damped oscillators. In Sec. 3 an alternative treatment of nonharmonic systems is presented, which is based on the dynamical algebra.[10] This approach allows — contrary to Noether's theorem — a direct extension to more general systems of a type recently investigated by Ray and Reid.[8,9] In Sec. 4 we show that dynamical invariants for nonharmonic systems can also be constructed in quantum mechanics. Section 5 concludes with a short summary.

[a] A. v. Humboldt foundation fellow on leave from Ramjas College, University of Delhi, Delhi-110007, India.

2. NOETHER'S THEOREM FOR DAMPED, TIME-DEPENDENT SYSTEMS

A formulation of Noether's theorem in terms of one-parameter Lie groups has been recently considered by Lutzky.[7] This method, when applied to the time-dependent harmonic oscillator, was not only found[8,9] to be simpler for obtaining the dynamical invariant,

$$I = \tfrac{1}{2}[k\,(q/\rho)^2 + (\dot\rho q - \rho\dot q)^2] \tag{2.1}$$

for the system described by

$$\ddot q + \omega^2(t)q = 0, \tag{2.2}$$

but has also offered a clue to solve the nonlinear differential equations of the type

$$\ddot\rho + \omega^2(t)\rho = k/\rho^3. \tag{2.3}$$

In fact, the invariant I plays the role of providing a link between the solutions to (2.2) and (2.3).

The construction of the invariant for time-dependent harmonic systems has a long history [1–15] and almost all the approaches used in the past deal with rather involved methods. More recently, Ray and Reid[9] have applied Lutzky's method to the Lagrangian (note the change of notations from that of Ray and Reid)

$$L = \tfrac{1}{2}[\dot q^2 - \omega^2(t)q^2 + 2g(t)G(q)], \tag{2.4}$$

and have shown that the invariant obtained in this way is a special case of their earlier[8] results obtained by generalizing Ermakov's method. As a matter of fact, an account of the damping term in the equation of motion may as well provide the solution to a more general type of nonlinear equations in terms of solutions to a linear system. In this section, we account for such a term and list more general results. In Sec. 3, we shall return to a further, simpler method (dynamical algebra approach) for the construction of the invariants, which will also enable us to look into other general cases when (i) q- and t-dependence in the third term of (2.4) is nonseparable or (ii) this term contains p-dependence (p is the conjugate of q) instead of q-dependence, however, again in a separate form.

In order to account for the damping we start with the Lagrangian

$$L = \tfrac{1}{2}e^{B(t)}[\dot q^2 - \omega^2(t)q^2 + 2g(t)G(q)], \tag{2.5}$$

which yields the equation of motion as

$$\ddot q + b(t)\dot q + \omega^2(t)q = g(t)G'(q), \quad [b(t) = dB/dt]. \tag{2.6}$$

Note that the factor $e^{B(t)}$ in (2.5) leads to the damping term in (2.6). Now, following the same steps as those of Ray and

 0022-2488/81/091904-05$01.00

Reid,[9] the auxiliary equation in this case turns out to be

$$\ddot{\rho} + b(t)\dot{\rho} + \omega^2(t)\rho = (k/\rho^3)e^{-2B(t)}. \tag{2.7}$$

Note that this nonlinear equation is independent of the term $g(t)G(q)$ in (2.5), which must be of the form

$$G(q) = G_0 q^{-2m},\ G_0 = \text{const}, \tag{2.8}$$

$$g(t) = g_0\rho^{2m-2}e^{-2B(t)}. \tag{2.9}$$

The invariant now becomes

$$I = \tfrac{1}{2}[(c/m)(\rho/q)^{2m} + k(q/\rho)^2 + (\dot{\rho}q - \rho\dot{q})^2 e^{2B(t)}], \tag{2.10}$$

where $c = -2mG_0g_0$ is an arbitrary constant. This result reduces for $B(t) = 0$ to that of Ray and Reid[9] and for $c = 0$ to that of Eliezer and Gray.[4]

Alternatively, linear friction can be introduced into the equations of motion by changing the independent variable $dt \to e^{-B(t)}dt$ and the invariant (2.10) and the auxiliary equation (2.7) can be derived from the nondamped case dealt with by Ray and Reid[9] by means of the transformation

$$\begin{aligned} &\omega^2 \to e^{2B}\omega^2, \\ &dt \to e^{-B}dt, \\ &\frac{d}{dt} \to e^{B}\frac{d}{dt}, \\ &g \to e^{2B}g. \end{aligned} \tag{2.11}$$

3. DYNAMICAL ALGEBRA

Recently one of the authors[10] presented a simple and straightforward derivation of the dynamical invariant for the time-dependent forced and damped harmonic oscillator. This approach makes use of an algebraic treatment, which is generally more common in quantum mechanics.[14,15] For the Hamiltonian

$$H = \sum_N h_n(t)\Gamma_n(p,q) \tag{3.1}$$

a dynamical Lie algebra of phase-space functions Γ_n is constructed which is closed with respect to the Poisson bracket,

$$\{\Gamma_n,\Gamma_m\} = \sum_r C^r_{nm}\Gamma_r \tag{3.2}$$

[this may, of course, introduce new Γ_k, which are originally absent in the Hamiltonian; these new Γ_k can be formally included in (3.1) by setting $h_k(t) = 0$]. Now the invariant $dI/dt = 0$ is written as a member of the dynamical algebra

$$I = \sum_r \lambda_k(t)\Gamma_r, \tag{3.3}$$

and by means of

$$0 = dI/dt = \{I,H\} + \partial I/\partial t \tag{3.4}$$

and comparison of coefficients a system of first-order linear differential equations for the unknown λ_r in (3.3) is obtained

$$\dot{\lambda}_r + \sum_n\Big[\sum_m C^r_{nm}h_m(t)\Big]\lambda_n = 0, \tag{3.5}$$

where the structure constants C^r_{nm} of the Lie algebra are defined in (3.2).

For the damped harmonic oscillator

$$H = \tfrac{1}{2}(e^{-B(t)}p^2 + e^{B(t)}\omega^2(t)q^2) \tag{3.6}$$

this approach leads to a finite algebra[10] containing only

$$\Gamma_1 = \tfrac{1}{2}p^2,\ \Gamma_2 = pq,\ \Gamma_3 = \tfrac{1}{2}q^2, \tag{3.7}$$

with the Poisson brackets

$$\{\Gamma_1,\Gamma_2\} = -2\Gamma_2,\ \{\Gamma_2,\Gamma_3\} = -2\Gamma_3,\quad \{\Gamma_3,\Gamma_1\} = \Gamma_2. \tag{3.8}$$

The differential equations (3.5) read in this case

$$\begin{aligned} \dot{\lambda}_1 &= -2e^{-B(t)}\lambda_2, \\ \dot{\lambda}_2 &= \omega^2(t)e^{B(t)}\lambda_1 - e^{-B(t)}\lambda_3, \\ \dot{\lambda}_3 &= 2\omega^2(t)e^{B(t)}\lambda_2, \end{aligned} \tag{3.9}$$

which can be simplified by setting

$$\lambda_1 = \rho^2 \tag{3.10}$$

to give $[b(t) = dB/dt]$

$$\ddot{\rho} + b(t)\dot{\rho} + \omega^2\rho = (k/\rho^3)e^{-2B} \tag{3.11}$$

and

$$\begin{aligned} \lambda_2 &= -e^{B}\rho\dot{\rho}, \\ \lambda_3 &= \dot{\rho}^2e^{2B}. \end{aligned} \tag{3.12}$$

The invariant can be written in the form ($\dot{q} = pe^{-B}$)

$$I = \tfrac{1}{2}[k(q/\rho)^2 + (\rho\dot{q} - \dot{\rho}q)^2e^{2B}]. \tag{3.13}$$

A. Application to nonharmonic systems

Here we apply the dynamical algebra approach described above to the case discussed in Sec. 2.

Let us consider the nonharmonic system (2.9),

$$\begin{aligned} H &= \tfrac{1}{2}[e^{-B(t)}p^2 + \omega^2(t)e^{B(t)}q^2 - 2g(t)e^{B(t)}G(q)] \\ &= e^{-B(t)}\Gamma_1 + \omega^2(t)e^{B(t)}\Gamma_2 - g(t)e^{B(t)}\Gamma_4, \end{aligned} \tag{3.14}$$

with

$$\Gamma_4 = G(q). \tag{3.15}$$

The Poisson brackets are in the first round,

$$\begin{aligned} \{\Gamma_1,\Gamma_4\} &= -pG'(q), \\ \{\Gamma_2,\Gamma_4\} &= -qG'(q), \\ \{\Gamma_3,\Gamma_4\} &= 0, \end{aligned} \tag{3.16}$$

so that $pG'(q)$ and $qG'(q)$ must be included in the dynamical algebra. In the next step one obtains p^2G'', pqG'', q^2G'', $G'^2 + qG'G''$ as additional elements of the algebra, and so on. With the exception of some rare cases the dynamical algebra becomes infinite. As an example the infinite dynamical algebra generated by $G = q^4$ (harmonic oscillator with quartic anharmonicity) is discussed in some detail in Appendix A. An exceptional example which leads to a finite dynamical algebra is presented in Appendix B. In the general case the system of linear differential equations (3.5), which determines the invariant I, is infinite and there are questions of convergence and existence of solutions, which are related to the existence or nonexistence of dynamical invariants. These problems may be solved in the future, but for the moment we confine ourselves to a more modest question: Are there special choices of $g(t)$ and $G(q)$, which yield a closed finite set of coupled differential equations (3.5) for the λ_ν ? Assuming that an invariant can be constructed in the subset

$(\Gamma_1,\Gamma_2,\Gamma_3,\Gamma_4)$ of the dynamical algebra, i.e.,

$$I = \sum_{r=1}^{4} \lambda_r(t)\Gamma_r, \tag{3.17}$$

we obtain

$$\begin{aligned}\sum_{r=0}^{4} \dot{\lambda}_r \Gamma_r &= \{H,I\} \\ &= -2e^{-B}\lambda_2\Gamma_1 + (\omega^2 e^B \lambda_1 - e^{-B}\lambda_3)\Gamma_2 \\ &\quad + 2\omega^2 e^B \lambda_2 \Gamma_3 + (e^{-B}\lambda_4 + g e^B \lambda_1)\{\Gamma_1,\Gamma_4\} \\ &\quad + g e^B \lambda_2 \{\Gamma_2,\Gamma_4\}.\end{aligned} \tag{3.18}$$

This equation can only be satisfied if

$$e^{-B}\lambda_4 + g e^B \lambda_1 = 0 \tag{3.19}$$

and if the Poisson bracket $\{\Gamma_2,\Gamma_4\}$ is a linear combination of $\Gamma_1,\ldots,\Gamma_4$. From (3.16) we see that $\{\Gamma_2,\Gamma_4\} = -qG'(q)$ is a function of q only, so that we get $\{\Gamma_2,\Gamma_4\} \sim \Gamma_3$ or $\{\Gamma_2,\Gamma_4\} \sim \Gamma_4$. The first possibility leads to $G' \sim q$, i.e., $G \sim \frac{1}{2}q^2$, which is nothing new as it is already in the algebra. The second possibility gives $qG'(q) \sim G$, a relation analogous toEq. (2.8), which provides

$$G(q) = G_0 q^{-2m},\ G_0 = \text{const} \tag{3.20}$$

for an arbitrary constant m. It may be emphasized that the form (3.20) was obtained after several manipulations in the approach of Ray and Reid[9] (cf.Sec.2), whereas here it appears in quite a natural way as a restricted closure property of a dynamical algebra subset. Comparison of coefficients in (3.18) gives

$$\dot{\lambda}_1 = -2e^{-B(t)}\lambda_2, \tag{3.21a}$$

$$\dot{\lambda}_2 = \omega^2 e^{B(t)}\lambda_1 - e^{-B(t)}\lambda_3, \tag{3.21b}$$

$$\dot{\lambda}_3 = 2\omega^2 e^{B(t)}\lambda_2, \tag{3.21c}$$

$$\dot{\lambda}_4 = 2mg(t)e^{B(t)}\lambda_2. \tag{3.21d}$$

The last equation is decoupled from the other three equations, which are identical to (3.9) and lead again to the auxiliary differential equation (3.11) for ρ. Using (3.19) to eliminate g from (3.21d) and expressing λ_2,λ_1 in terms of ρ by (3.10) and (3.12) we obtain

$$\dot{\lambda}_4/\lambda_4 = 2M\,\dot{\rho}/\rho, \tag{3.22}$$

i.e,

$$\lambda_4 = (c/2m)\rho^{2m}, \tag{3.23}$$

with an arbitrary constant c and the invariant

$$I = \tfrac{1}{2}[(c/m)(\rho/q)^{2m} + k(q/\rho)^2 + (\rho\dot{q} - \dot{\rho}q)^2 e^{2B(t)}], \tag{3.24}$$

in agreement with (2.10). Finally, g can be obtained from (3.21d)

$$g(t) = -(c/2m)e^{-2B}\rho^{2m-2} \tag{3.25}$$

and the Hamiltonian (3.14) reads

$$\begin{aligned}H = \tfrac{1}{2}[&e^{-B(t)}p^2 + \omega^2(t)e^{B(t)}q^2 \\ &+ (c/m)e^{-B(t)}(1/\rho^2)(\rho/q)^{2m}],\end{aligned} \tag{3.26}$$

in complete agreement with the result of Sec.2. Thus we see that the dynamical invariant is obtained here using a simpler and more straightforward approach of dynamical algebras, which may perhaps provide more physical insight.

In the special case $m = 1$ the Hamiltonian becomes independent of ρ and the invariant (3.24) agrees with the invariant (B 6) derived for this case in a different manner in Appendix B.

The Hamiltonian (3.26) can, of course, be generalized to include a sum of terms with different values of m and c,

$$H = \tfrac{1}{2}\left[e^{-B}p^2 + \omega^2 e^B q^2 + \frac{1}{\rho}\sum_i \frac{c_i}{m_i} e^{-B}(\rho/q)^{2m_i}\right] \tag{3.27}$$

(see also Ray and Reid[9] and the discussion at the end of Sec.2).

B. More general cases

In Subsec. 3A we considered the case when the time and coordinate dependences in the nonharmonic term are separable. Here we extend the application of the dynamical algebra approach to more general cases. For simplicity we restrict ourselves to the undamped case. Let us consider the Hamiltonian

$$\begin{aligned}H &= \tfrac{1}{2}[p^2 + \omega^2(t)q^2 + \phi(q,t)] \\ &= \Gamma_1 + \omega^2(t)\Gamma_3 + \Gamma_4\end{aligned} \tag{3.28}$$

with $\Gamma_4 = \frac{1}{2}\phi$. The Poisson brackets now become

$$\begin{aligned}\{\Gamma_1,\Gamma_4\} &= -\tfrac{1}{2}p\,\partial\phi/\partial q, \\ \{\Gamma_2,\Gamma_4\} &= -\tfrac{1}{2}q\,\partial\phi/\partial q, \\ \{\Gamma_3,\Gamma_4\} &= 0.\end{aligned} \tag{3.29}$$

Assuming that an invariant can be constructed in the subset $(\Gamma_1,\Gamma_2,\Gamma_3,\Gamma_4)$of the dynamical algebra, Eqs. (3.3) and (3.4) [in analogy with (3.17) and (3.18)] imply

$$\begin{aligned}\frac{\partial I}{\partial t} &= \sum_{\nu=1}^{4} \dot{\lambda}_\nu \Gamma_\nu + \tfrac{1}{2}\lambda_4 \frac{\partial\phi}{\partial t} = \{H,I\} \\ &= -2\lambda_2\Gamma_1 + (\omega^2\lambda_1 - \lambda_3)\Gamma_2 + 2\omega^2\lambda_2\Gamma_3 \\ &\quad + \lambda_2\{\Gamma_4,\Gamma_2\} + (\lambda_1 - \lambda_4)\{\Gamma_4,\Gamma_1\}.\end{aligned} \tag{3.30}$$

This Equation can only be satisfied if $\lambda_1 = \lambda_4$. Now, when we equate the coefficients of Γ_i on either side of this equation the results for $i = 1,2,3$ will lead to the auxiliary equation $\ddot{\rho} + \omega^2(t)\rho = k/\rho^3$, whereas for $i = 4$ we obtain

$$\dot{\lambda}_4\Gamma_4 + \tfrac{1}{2}\lambda_4\frac{\partial\phi}{\partial t} = \lambda_2\{\Gamma_4,\Gamma_2\}$$

or

$$\dot{\lambda}_1\phi + \lambda_1\partial\phi/\partial t = \lambda_2 q\partial\phi/\partial q. \tag{3.31}$$

Now setting $\lambda_1 = \rho^2$ (as before) and using $\dot{\lambda}_1 = 2\rho\dot{\rho}$ and $\lambda_2 = -\rho\dot{\rho}$, we are left with a partial differential equation

$$\dot{\rho}(2\phi + q\partial\phi/\partial q) + \rho\partial\phi/\partial t = 0, \tag{3.32}$$

whose solution would provide ϕ as $\phi(q,\rho,\dot{\rho},t)$. For the choice $\phi(q,t) = \phi(q,\rho(t))$, which implies that $\partial\phi/\partial t = \dot{\rho}\,\partial\phi/\partial\rho$, Eq. (3.32) becomes

$$2\phi + q\partial\phi/\partial q + \rho\partial\phi/\partial\rho = 0. \tag{3.33}$$

For an *ansatz* $\phi(q,\rho) = (1/\rho^2)\psi(q,\rho)$, this equation reduces to the form

$$q\partial\psi/\partial q = -\rho\partial\psi/\partial\rho, \tag{3.34}$$

which is satisfied by the functional form $\psi(\rho/q)$. Thus for the Hamiltonian

$$H = \tfrac{1}{2}[p^2 + \omega^2(t)q^2 + (1/\rho^2)\psi(\rho/q)], \tag{3.35}$$

the corresponding invariant becomes

$$I = \tfrac{1}{2}[k(q/\rho)^2 + \psi(\rho/q) + (\rho p - \dot\rho q)^2] \tag{3.36}$$

and the equation of motion is given by

$$\ddot q + \omega^2(t)q = (1/\rho q^2)\psi'(\rho/q). \tag{3.37}$$

Another general case which can be outlined briefly here is that in which the nonharmonic terms in (3.28) are momentum dependent instead of q dependent. We assume, however, that they are separable in p and t, i.e., of the type

$$H = \tfrac{1}{2}[p^2 + \omega^2(t)q^2 + n(t)G(p)].$$

In this case, in order that an invariant can be constructed in the subset of the dynamical algebra as before it turns out that $G(p) \sim p^{-2m}$, where m is an arbitrary constant. Finally, for the Hamiltonian

$$H = \tfrac{1}{2}[p^2 + \omega^2(t)q^2 + C\omega^2(t)(k/\rho^2 + \dot\rho^2)^{m-1}p^{-2m}], \tag{3.38}$$

where C is an integration constant, the invariant turns out to be

$$I = \tfrac{1}{2}[k(q/\rho)^2 + C(k/\rho^2 + \dot\rho^2)^m p^{-2m} + k(q/\rho)^2 + (\rho p - \dot\rho q)^2]. \tag{3.39}$$

For a specific choice of ψ (i.e., power form), the results obtained above for the first general case are similar to those of Ray and Reid[8] obtained by using a different method. However, it may be remarked that the present method provides some justification for this specific functional dependence of ψ on ρ/q, instead of choosing it in an *ad hoc* manner as was done by Ray and Reid in their generalization of Eqs. (2.2) and (2.3). Further, for the second general case discussed above, their method is not very transparent.

4. QUANTUM INVARIANTS

It is well known[11–15] that for the harmonic oscillator dynamical invariants do also exist in quantum mechanics. In fact they are identical to the classical invariants (2.1), where $p = \dot q$ and q are simply replaced by the corresponding quantum operators $\hat p$ and $\hat q$ (we denote quantum operators by a caret)

$$\hat I = \tfrac{1}{2}[(k/\rho^2)\hat q^2 + (\rho\hat p - \dot\rho\hat q)^2]. \tag{4.1}$$

(To simplify the discussion we consider only the undamped case in this section.) The origin of this correspondence is the identity between the classical dynamical algebra (3.7) of phase-space functions and the quantum mechanical operator algebra

$$\hat\Gamma_1 = \tfrac{1}{2}\hat p^2, \quad \hat\Gamma_2 = \tfrac{1}{2}(\hat p\hat q + \hat q\hat p), \quad \hat\Gamma_3 = \tfrac{1}{2}\hat q^2, \tag{4.2}$$

where the Lie bracket is the commutator [,]

$$\begin{aligned}[][\hat\Gamma_1,\hat\Gamma_2] &= -2i\hbar\hat\Gamma_1,\\ [\hat\Gamma_2,\hat\Gamma_3] &= -2i\hbar\hat\Gamma_3,\\ [\hat\Gamma_3,\hat\Gamma_1] &= i\hbar\hat\Gamma_2.\end{aligned} \tag{4.3}$$

For the more general Hamiltonian corresponding to the classical Ray and Reid *ansatz* (3.14),

$$\hat H = \tfrac{1}{2}[\hat p^2 + \omega^2(t)\hat q^2 - 2g(t)\hat G(\hat q)], \tag{4.4}$$

the one-to-one correspondence breaks down.

In this section the treatment of Sec. 3 will be carried over to the quantum case. Introducing

$$\hat\Gamma_4 = \hat G(\hat q), \tag{4.5}$$

we obtain the Hamiltonian

$$\hat H = \hat\Gamma_1 + \omega^2(t)\hat\Gamma_3 - g(t)\hat\Gamma_4 \tag{4.6}$$

and the commutators

$$\begin{aligned}[][\hat\Gamma_1,\hat\Gamma_4] &= -(i\hbar/2)(\hat p\hat G'(\hat q) + \hat G'(\hat q)\hat p),\\ [\hat\Gamma_2,\hat\Gamma_4[&= -i\hbar\hat q\hat G'(\hat q),\\ [\hat\Gamma_3,\hat\Gamma_4] &= 0.\end{aligned} \tag{4.7}$$

With

$$\hat I = \sum_{r=1}^{4}\lambda_r\hat\Gamma_r$$

and

$$0 = d\hat I/dt = (1/i\hbar)[\hat I,\hat H] + \partial\hat I/\partial t, \tag{4.8}$$

one finds

$$\begin{aligned}\sum_{r=0}^{4}\lambda_r\hat\Gamma_r = & -2\lambda_2\hat\Gamma_1 + (\omega^2\lambda_1 - \lambda_3)\hat\Gamma_2\\ & + 2\omega^2\lambda_2\hat\Gamma_3 + (\lambda_4 + g\lambda_1)(1/i\hbar)[\Gamma_1,\Gamma_4]\\ & + g\lambda_2(1/i\hbar)[\hat\Gamma_2,\hat\Gamma_4].\end{aligned} \tag{4.9}$$

Again identity (4.9) requires

$$\lambda_4 + g\lambda_1 = 0 \tag{4.10}$$

and

$$(1/i\hbar)[\hat\Gamma_2,\hat\Gamma_4] \sim \hat\Gamma_3 \text{ (which again means } \hat G_4 = \hat q^2)$$

or

$$(1/i\hbar)[\hat\Gamma_2,\hat\Gamma_4] \sim \hat\Gamma_4, \text{ which gives } -\hat q\hat G'(\hat q) \sim \hat G(\hat q),$$

and therefore

$$\hat G(\hat q) = G_0\hat q^{-2m}, \tag{4.11}$$

exactly as in the classical case.

The system of differential equations for the λ_r obtained in the quantum case is identical to the classical equations (3.21) [with $B = 0$ in (3.21)] and the invariant of the Hamiltonian

$$\hat H = \tfrac{1}{2}[\hat p^2 + \omega^2(t)\hat q^2 + (c/m)\rho^{2m-2}\hat q^{-2m}] \tag{4.12}$$

is given by

$$\hat I = \tfrac{1}{2}[(c/m)\rho^{2m}\hat q^{-2m} + (k/\rho^2)\hat q^2 + (\rho\hat q - \dot\rho\hat q)^2], \tag{4.13}$$

where ρ is a solution of

$$\ddot\rho + \omega^2(t)\rho = k/\rho^3. \tag{4.14}$$

It has thus been shown that the Ray and Reid invariants[9] do also exist in quantum mechanics and furthermore it has been demonstrated that the dynamical algebra formalism provides a natural way for a transition from classical to quantum mechanics.

5. CONCLUDING REMARKS

In this paper we have demonstrated the utility of the dynamical algebra approach in constructing invariants for harmonic and special cases of nonharmonic time-dependent systems. This method is based on the Lie algebra of phase-space functions and can be carried over naturally to quantum mechanics (cf.Sec.4). We hope that the algebra approach will provide more insight into the role of the dynamical invariants both classically and quantum mechanically. Futhermore, we would anticipate that the formalism developed here could also be successfully applied to higher-dimensional problems. Such work is in progress.

APPENDIX A

As an example of a dynamical algebra generated by p^2, q^2, and $G(q)$ (see Sec.3) it is instructive to look at the case $G(q) = q^4$ in more detail [a complementary discussion is given in Appendix B for $G(g) = q^{-2}$]. Trying to close the dynamical algebra, one finds in the first round the new member pq^3, in the second round p^2q^2 and q^6, in the third round p^3q, pq^5, p^2q^4, and so on. In the following we will show that the dynamical algebra for this system is the set of *all* $p^\mu q^\nu$ (μ,ν non-negative integers) with even degree $d = \mu + \nu$. To prove this one observes that the Poisson bracket of two $p^\mu q^\nu$ with even degree,

$$\{p^{\mu_1}q^{\nu_1},p^{\mu_2}q^{\nu_2}\} = (\mu_2 + \nu_1 - \mu_1\nu_2)p^{\mu_1+\mu_2-1}q^{\nu_1+\nu_2-1}, \tag{A1}$$

yields again a $p^{\bar\mu}q^{\bar\nu}$ with even degree,

$$\bar d = \mu_1 + \mu_2 + \nu_1 + \nu_2 - 2 = d_1 + d_2 - 2.$$

If, on the other hand, the algebra does not contain all $p^\mu q^\nu$ of even degree, there must be a $p^\mu q^\nu$ with minimum even degree, $\bar d = \bar\mu + \bar\nu$ and $\bar\mu \neq \bar\nu$, which is not in the algebra. (The case $\bar\mu = \bar\nu$ is trivial because $\{pq^{\bar\mu},p^{\bar\mu}q\} = \bar\mu^2 p^{\bar\mu}q^{\bar\mu}$). Observing that $p^{\bar\mu-1}q^{\bar\nu-1}$ (even degree $< \bar d$) and p^2q^2($\{p^2,\{p^2,q^4\}\} = 48p^2q^2$) are in the algebra and calculating the Poisson bracket

$$\{p^{\bar\mu-1}q^{\bar\nu-1},p^2q^2\} = 2(\bar\nu - \bar\mu)p^{\bar\mu}q^{\bar\nu}, \tag{A2}$$

one finds the $p^{\bar\mu}q^{\bar\nu}$ is in the algebra, in contradiction to the assumption. The harmonic quartic oscillator algebra is therefore infinite, which is expected to be typically the case. An example which yields a finite algebra is discussed in Appendix B.

APPENDIX B

In Sec. 3 we discussed in some detail a Hamiltonian which generates an infinite dynamical algebra. This is expected to be typical. There are, however, Hamiltonians which have a finite dynamical algebra. One example is, of course, the harmonic oscillator.

Another example is provided by

$$H = \tfrac{1}{2}(p^2 + \omega^2(t)q^2 + k/q^2) \tag{B1}$$

with constant k. The Hamiltonians (B1) can be interpreted as the radial part of a three-dimensional harmonic oscillator, where k is related to the angular momentum. Redefining now $\Gamma_1 = p^2/2$ of the harmonic oscillator [Eq.(3.7)] as

$$\Gamma_1^{(k)} = \tfrac{1}{2}(p^2 + k/q^2), \tag{B2}$$

we obtain the Poisson brackets

$$\{\Gamma_1^{(k)},\Gamma_2\} = -2\Gamma_1^{(k)}, \qquad \{\Gamma_3,\Gamma_1^{(k)}\} = \Gamma_2, \tag{B3}$$

i.e., the algebra $(\Gamma_1^{(k)},\Gamma_2,\Gamma_3)$ is identical to the oscillator algebra $(\Gamma_1,\Gamma_2,\Gamma_3)$.

The Hamiltonian (B1) reads

$$H = \Gamma_1^{(k)} + \omega^2(t)\Gamma_3 \tag{B4}$$

and the invariant

$$I = \lambda_1\Gamma_1^{(k)} + \lambda_2\Gamma_2 + \lambda_3\Gamma_3 \tag{B5}$$

can be evaluated in the same manner as for the simple harmonic oscillator, which yields

$$I = \tfrac{1}{2}[\bar k(q/\bar q)^2 + k(\bar q/q)^2 + (\bar q\dot q - \dot{\bar q}q)^2], \tag{B6}$$

where $\bar q$ is a solution of the "auxiliary" equation

$$\ddot{\bar q} + \omega^2(t)\bar q = \bar k/\bar q^3 \tag{B7}$$

[compare Eq.(2.7)], which can be derived from the Hamiltonian

$$\bar H = \tfrac{1}{2}(\bar p^2 + \omega^2(t)\bar q^2 + \bar k/\bar q^2). \tag{B8}$$

There is a complete symmetry between the Hamiltonians H and $\bar H$. For $k = 0$ (or $\bar k = 0$) we recover the well known invariant (2.1) for the time-dependent harmonic oscillator. The invariant (B6) is a constant of motion with respect to H and $\bar H$, i.e., we have

$$\{I,H\}_{q,p} = \{I,\bar H\}_{\bar q,\bar p}. \tag{B9}$$

I generates a mapping between the Hamiltonians H and $\bar H$, which is, of course, canonical, because the equations of motion are conserved.

[1]M. Kruskal, J. Math. Phys. **3**, 806 (1962).
[2]H. R. Lewis, Jr., J. Math. Phys. **9**, 1976 (1968).
[3]K. R. Symon, J. Math. Phys. **11**, 1320 (1970).
[4]C. J. Eliezer and A. Gray, SIAM J. Appl. Math. **30**, 463 (1976).
[5]P. G. L. Leach, J. Math. Phys. **18**, 1608 (1977); SIAM J. Appl. Math. **34**, 496 (1978).
[6]P. G. L. Leach, J. Math. Phys. **20**, 96 (1979).
[7]M. Lutzky, J. Phys. A **11**, 249 (1978); Phys. Lett A **68**, 3 (1978); J. Phys. A **12**, 973 (1979).
[8]J. R. Ray and J. L. Reid, Phys. Lett. A **71**, 317 (1979).
[9]J. R. Ray and J. L. Reid, J. Math. Phys. **20**, 2054 (1979).
[10]H. J. Korsch, Phys. Lett. A **74**, 294 (1979).
[11]H. R. Lewis, Jr., Phys. Rev. Lett. **18**, 510 (1967).
[12]P. G. L. Leach, J. Math. Phys. **18**, 1902 (1977); Am. J. Phys. **46**, 1247 (1978).
[13]N. J. Günther and P. G. L. Leach, J. Math. Phys. **18**, 572 (1977).
[14]H. R. Lewis, Jr. and W. B. Riesenfeld, J. Math. Phys. **10**, 1458 (1969).
[15]D. C. Khandekar and S. V. Lawande, J. Math. Phys. **20**, 1870 (1979); R. R. Puri and S. V. Lawande, Phys. Lett. A **70**, 69 (1979).
[16]J. R. Ray, Phys. Rev. D **20**, 2632 (1979).
[17]See, for example, E. C. G. Sudarshan and N. Mukunda, *Classical Dynamics: A Modern Perspective* (Wiley, New York, 1974).

Dispersion relations for linear wave propagation in homogeneous and inhomogeneous media

Richard L. Weaver and Yih-Hsing Pao
Department of Theoretical and Applied Mechanics, Cornell University, Ithaca, New York 14853

(Received 29 April 1980; accepted for publication 19 September 1980)

For the dispersion of waves in a homogeneous medium there exist the Kramers–Kronig relations for the wave number $K(\omega) = \omega/c(\omega)$. The usual mathematical proof of such relations depends on assumptions for the asymptotic behavior of $c(\omega)$ at high frequency, which for electromagnetic waves in dielectrics can be evaluated from the microphysical properties of the medium. In this paper such assumptions are removed and the necessary asymptotic behavior is shown to follow the representation of $K(\omega)$ as a Herglotz function. From the linear, causal, and passive properties of the media we thus establish the Kramers–Kronig relations for all linear wave disturbances including acoustic, elastic, and electromagnetic waves in inhomogeneous as well as homogeneous media without any reference to the microphysical structure of the medium.

PACS numbers: 03.40.Kf

I. INTRODUCTION

The Kramers–Kronig relations have been established for the dispersion of electromagnetic waves in dielectrics since 1927.[1,2] If the wave in the medium is represented by $\exp i[zK(\omega) - \omega t]$ for a real circular frequency ω and a complex wave number $K(\omega)$, the imaginary part of $K(\omega)$, $\mathrm{Im}K(\omega)$, defines the attenuation coefficient of the wave along the spatial axis z, and the real part of $K(\omega)$, $\mathrm{Re}K(\omega)$, when divided by ω, equals the reciprocal of the phase velocity. The Kramers–Kronig relations state that the real and imaginary parts of $K(\omega)$ are related by a pair of Hilbert transforms.

Similar relations have been applied for sound-wave propagation. Recently Horton[3] and O'Donnell *et al.*[4] have employed the equations derived by Ginzberg[5] and Goldberger[6] to investigate the applicability of similar relations in acoustics. In the literature, Ginzberg is credited with establishing the Kramers-Kronig relations for sound waves in homogeneous fluid media. Upon a close reexamination, we found that the proof given in Ref. 5, and in several other sources, depends on assumed asymptotic behavior of the phase velocity c_∞ as ω approaches infinity.

While the asymptotic behavior of the phase velocity can be determined for electromagnetic waves from the dynamics of electrons in dielectrics, this cannot be done rigorously for a sound wave. Furthermore, it is difficult to ascertain the behavior of c_∞ for stress waves in solids, especially in inhomogeneous media. The purpose of this paper is to derive the dispersion equations (Kramers–Kronig relations) valid for a general class of linear homogeneous or inhomogeneous media. The proof proceeds without *a priori* knowledge of c_∞ of the medium that supports the wave.

Our investigation is motivated by the search for an alternative method to determine the attenuation coefficients of stress waves in solids. Measurements of the attenuation coefficients in solids are known to be very difficult whereas the determination of the dispersion ($\mathrm{Re}K$ as a function of ω) is comparatively easy even for waves in inhomogeneous media such as fiber-reinforced composite materials.[7] If a Kramers–Kronig type equation can be rigorously established for the inhomogeneous solids, one could then calculate the attenuation coefficient from a measurement of the dispersion.

The inhomogeneous media considered in this paper must support a plane wave of the form $\exp i[zK(\omega) - \omega t]$. Thus it includes a random medium which is statistically homogeneous, $K(\omega)$ being the wave number of the averaged field, and a periodic medium for which $K(\omega)$ is the Floquet wave number. Other than that, the derivation is general for a medium which is linear and causal, homogeneous or inhomogeneous.

In the next section, Sec. II, we review the existing proofs of a dispersion relation, which proceed along the lines used by Titchmarsh[8] appropriate to the real and imaginary parts of a generalized system function. A somewhat stronger result has been obtained by Toll.[9] In Sec. III we discuss the usual derivation of Kramers–Kronig relations for the wave number $K(\omega)$ for wave propagation in three types of linear homogeneous media, electromagnetic waves in dielectrics, acoustic waves in fluids, and stress waves in solids. These proofs will be seen to have certain unsatisfying aspects, amongst them being difficulties in making generalizations to inhomogeneous media. Section IV discusses the phase problem as it has appeared in the literature, and relates its contribution to our problem. Section V will present the newly-constructed proof for the dispersion equations in a general class of linear inhomogeneous media.

II. CAUSALITY AND DISPERSION

The literature on dispersion relations and causality is extensive. An excellent introduction and review of the subject and its applications is provided by Nussenzveig.[10] Some of his results which are related to the subsequent discussions are repeated in this section. A more thorough treatment of the logical foundations of causality and dispersion relations is given by Toll.[9]

Consider a general linear system with an output $w(t)$, a function of time, which is a linear causal functional of the

 0022-2488/81/091909-10$01.00

input $f(t)$,

$$w(t) = \int_{-\infty}^{t} g(t-\tau) f(\tau)\, d\tau. \tag{1}$$

The system is assumed invariant under translations in time. Hence the system function $g(t-\tau)$ depends only on the difference, $t-\tau$, and not on t and τ separately. Equation (1) states that the output at time t depends only on the input at times $\tau \leqslant t$. Hence the system is termed *causal.*

Equation (1) may be rewritten in terms of the Fourier transforms of the input and output functions if they are square-integrable functions of time,

$$W(\omega) \equiv \int_{-\infty}^{\infty} w(t) e^{i\omega t}\, dt,$$

$$F(\omega) \equiv \int_{-\infty}^{\infty} f(t) e^{i\omega t}\, dt. \tag{2}$$

Thus,

$$W(\omega) = G(\omega) F(\omega), \tag{3}$$

where

$$G(\omega) \equiv \int_{0}^{\infty} g(\tau) e^{i\omega\tau}\, d\tau. \tag{4}$$

In the previous equations and the sequel, we shall use capital letters to denote the Fourier transforms of the corresponding time function in lower case letters.

Note that because of the causality condition, $g(\tau) = 0$ for $\tau < 0$, the lower limit of integration in Eq. (4) is zero, instead of $-\infty$. $G(\omega)$ can be considered a system function (transfer function) of a complex, but linear, system or a generalized scattering amplitude which converts an incident field to a scattered field. The change of lower limit of integration for $G(\omega)$ from $-\infty$ to 0 has far reaching consequence. It implies that $G(\omega)$ has a regular analytic continuation in the upper half ω plane. This connection between causality and analyticity is at the root of all dispersion relations.

If in addition $G(\omega)$ is square-integrable along the real axis of the ω plane, $\omega = \omega_r + i\omega_i$,

$$\int_{-\infty}^{\infty} |G(\omega)|^2\, d\omega < C, \tag{5}$$

where C is a constant, then $G(\omega_r + i\omega_i)$ $(\omega_i \geqslant 0)$ is also a square-integrable function of ω_0. By integrating the function $G(\omega')/(\omega' - \omega)$ over a complex contour as shown in Fig. 1 and letting Ω approach infinity, one finds that the real part of $G(\omega)$, $\mathrm{Re}G(\omega)$, and the imaginary part, $\mathrm{Im}G(\omega)$, form a pair of Hilbert transforms. This result is summarized by Titchmarsch's theorem as stated in Ref. 10:

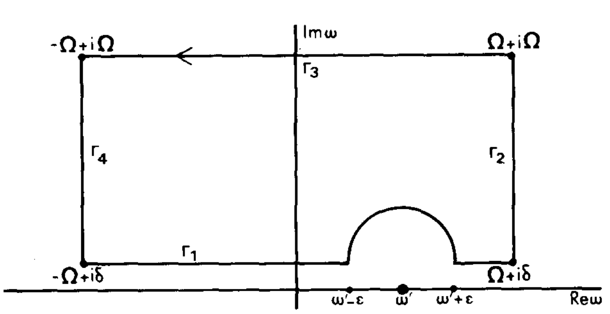


FIG. 1. The contour Γ. We take the limits $\Omega \to \infty$, $\epsilon \to 0$, and $\delta \to 0$.

If a square integrable-function $G(\omega)$ fulfills one of the four conditions below it fulfills all of them:

(i) The inverse Fourier transform, $g(t)$, equals 0 for $t < 0$

(ii) $G(\omega)$ is, for almost all real ω, the limit as $\mathrm{Im}\omega \to 0^+$ of a function which is analytic throughout the upper half plane and is square-integrable over any line parallel to and above the real axis

$$\int_{-\infty}^{\infty} |G(\omega_r + i\omega_i)|^2 d\omega_r < C \text{ for } \omega_i > 0. \tag{6}$$

$$\text{(iii) } \mathrm{Re}G(\omega) = \frac{1}{\pi} \fint_{-\infty}^{\infty} \frac{\mathrm{Im}G(\omega')}{\omega' - \omega}\, d\omega'. \tag{7}$$

$$\text{(iv) } \mathrm{Im}G(\omega) = -\frac{1}{\pi} \fint_{-\infty}^{\infty} \frac{\mathrm{Re}G(\omega')}{\omega' - \omega}\, d\omega'. \tag{8}$$

Equations (7) and (8) are a Hilbert-transform pair. The slash through the integral sign indicates that the Cauchy principal values of the integrals along the real axis are to be taken.

Equations (7) and (8) are dispersion relations for the system function $G(\omega)$. They are a consequence of linearity, causality, and square-integrability. Since the condition of the square-integrability (6) is to assure that the top and sides of the contour integral Γ vanish as $\Omega \to \infty$, Eqs. (7) and (8) may also be derived under

$$\lim_{|\omega| \to \infty} G(\omega) \to 0 \text{ uniformly, } \pi \geqslant \arg\omega \geqslant 0.$$

As often happens, the square-integrability condition on $G(\omega)$ cannot be satisfied, but rather the weaker condition that $|G(\omega)|$ is bounded, i.e.,

$$|G(\omega)|^2 \leqslant C.$$

For such cases, we may construct a new function $H(\omega)$,

$$H(\omega) = (G(\omega) - G(\omega_0))/(\omega - \omega_0),\ \mathrm{Im}\omega_o \geqslant 0.$$

$H(\omega)$ is square-integrable and has no poles in the upper half plane, and hence satisfies a pair of equations like Eqs. (7) and (8). Substituting $H(\omega)$ as defined above for $G(\omega)$ in Eqs. (7) and (8) and taking ω_0 to be real and then rearranging terms, we obtain

$$\mathrm{Re}G(\omega) = \mathrm{Re}G(\omega_0) + \frac{(\omega - \omega_0)}{\pi} \times \fint_{-\infty}^{\infty} \mathrm{Im}\left[\frac{G(\omega') - G(\omega_0)}{\omega - \omega_0}\right] \frac{d\omega'}{\omega' - \omega}, \tag{9}$$

$$\mathrm{Im}G(\omega) = \mathrm{Im}G(\omega_0) - \frac{(\omega - \omega_o)}{\pi} \times \fint_{-\infty}^{+\infty} \mathrm{Re}\left[\frac{G(\omega') - G(\omega_0)}{\omega' - \omega_0}\right] \frac{d\omega'}{\omega' - \omega}. \tag{10}$$

This is known as a *dispersion relation for* $G(\omega)$ *with one subtraction*. Further subtractions may be taken if $G(\omega)$ is bounded by a polynomial function of ω. Details are given in Ref. 10.

III. THE KRAMERS–KRONIG RELATION FOR HOMOGENEOUS MEDIA

A. The canonical proofs

A dispersion relation, similar to Eqs. (9) and (10), was first established by Kramers[1] and Kronig[2] for the complex index of refraction of light waves in homogeneous dielectric media. Later on, a similar relation was deduced for acoustic waves in homogeneous fluid media.[5] In both cases, such a relation can be expressed in terms of the complex propagation constant $K(\omega)$ of a time-harmonic wave.

Consider a plane harmonic wave with angular frequency ω propagating in the direction of the z space axis with a complex amplitude function $A(\omega)$,

$$u(t,z) = A(\omega)\, e^{i[-\omega t + K(\omega)z]} = A(\omega)\, e^{-i\omega[t - z/c(\omega)]}. \quad (11)$$

The complex wave number $K(\omega)$ is related to the complex phase velocity $c(\omega)$ by $K = \omega/c$. The Kramers–Kronig relations can be taken in the form

$$\mathrm{Re}K(\omega) = \frac{\omega}{c_\infty} + \frac{\omega}{\pi} \fint_{-\infty}^{\infty} \frac{\mathrm{Im}K(\omega')}{\omega'} \frac{d\omega'}{\omega' - \omega} + \mathrm{Re}K(0), \quad (12)$$

$$\mathrm{Im}K(\omega) = -\frac{\omega}{\pi} \fint_{-\infty}^{\infty} \left[\frac{\mathrm{Re}K(\omega')}{\omega'} - \frac{1}{c_\infty}\right] \frac{d\omega'}{\omega' - \omega} + \mathrm{Im}K(0), \quad (13)$$

where $c_\infty = \lim (\omega/K)$ as $\omega \to \infty$. Equations (12) and (13) are of the form of a dispersion relation for $K(\omega)$ with two subtractions, the point at $\omega = 0$ and the point at $\omega = \infty$.

To show that Eqs. (12) and (13) are valid, we must establish that $K(\omega)$ is analytic in the upper ω plane and that the real limit c_∞ exists. The proof is not as straightforward as that for $G(\omega)$ in Eqs. (7) and (8) because $k(t)$, the inverse Fourier transfrom of $K(\omega)$, is neither a causal function nor a physically meaningful function in the time domain. The proof usually proceeds as follows[10]:

Consider a material slab of thickness z_0 (Fig. 2), and let the input function at $z = 0$ be $f(t,0)$ and output at $z = z_0$ be $w(t,z_0)$. From Eq. (1) we write

$$w(t,z_0) = \fint_{-\infty}^{t} g(t - \tau, z_0) f(\tau, 0)\, d\tau, \quad (14)$$

where $g(t,z_0)$ is a causal function in time. If the input is $f(t,0) = f_0 \exp(-i\omega t)$, it generates a plane harmonic wave as represented by (11), propagating through the thickness z_0. The output should be $u(t,z_0) = A(\omega) \exp[-i\omega t + iK(\omega)z_0]$, where $A(\omega)$ is a thickness independent amplitude. The Fourier transform of Eq. (14) is

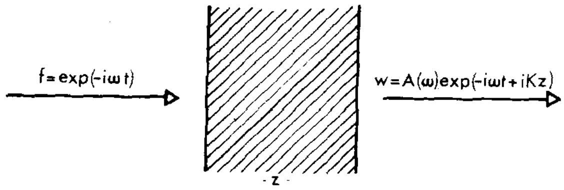


FIG. 2. A slab of thickness z with plane wave $f(t)$ and plane wave output $w(t)$.

$$W(\omega,z_0) = G(\omega,z_0) F(\omega,0),$$

where

$$G(\omega,z_0) = A(\omega) e^{iK(\omega)z_0}. \quad (15)$$

Since $G(\omega,z_0)$ is the Fourier transform of a causal function, it is analytic in the upper half ω plane. Furthermore, $|W(\omega,z_0)| \leqslant |F(\omega,0)|$ because energy is not generated within the medium, i.e., the system is assumed passive. This implies that $G(\omega,z_0)$ is bounded,

$$|G(\omega,z_0)| \leqslant 1.$$

These two conditions assure a dispersion relation, Eqs. (9) and (10), for $G(\omega,z_0)$ with one subtraction. Note that though $G(\omega,z_0)$ satisfies a dispersion formula, there is no *a priori* reason to expect its logarithm to obey the same type of formula.

To establish the analyticity of $K(\omega)$ in the upper ω plane, we consider two slabs of thickness z_0 and $z_1 = z_0 + d$, and apply Eq. (15) twice for two system-functions $G(\omega,z_0)$ and $G(\omega,z_1)$ with the same amplitude $A(\omega)$ and wave number $K(\omega)$. From the ratio of the two system-functions, we obtain

$$K(\omega) = \frac{-i}{d} \ln\left[\frac{G(\omega,z_1)}{G(\omega,z_0)}\right]. \quad (16)$$

Since the logarithm of an analytic function is analytic in the same region except at the zeros of the argument, and since the quotient of two analytic functions is also analytic except at the zeros of the denominator, we conclude that in the absence of zeros for $G(\omega,z_0)$ or $G(\omega,z_1)$, $K(\omega)$ is analytic in the upper half ω plane. That $G(\omega,z)$ for all finite z has no zeros can be seen from observing that no slab is a perfect reflector, i.e., there is always some penetration although often exponentially little. Equation (14) leaves $K(\omega)$ ambiguous by an integral multiple of $2\pi/d$ and one might therefore suspect the presence of branch cuts. But since K is independent of d and d is arbitrary, we see that there can be no such cuts.

The analyticity of $K(\omega)$ alone is not sufficient to establish Eqs. (12) and (13). We need to show further that either $K(\omega)$ or a new function

$$H(\omega) = K(\omega)/\omega - 1/c$$

is square-integrable,

$$\int_{-\infty}^{\infty} |H(\omega_r + i\omega_i)|^2\, d\omega_r < C, \quad \omega_i \geqslant 0.$$

This part of the proof is usually established by appealing to the field equations and constitutive relations of the media in which the plane harmonic wave is propagating. We proceed first with light waves in dielectric media, then discuss acoustic waves in a fluid, and conclude this section with a discussion of stress waves in solids.

B. Light waves in dielectric media

For an isotropic dielectric medium, the polarization $\mathbf{P}$ at a point is related to the history of the applied electric field $\mathbf{E}$ at that point by a convolution integral[10]

$$\mathbf{P}(t) = \int_{-\infty}^{t} \chi(t - \tau) \mathbf{E}(\tau)\, d\tau.$$

The susceptibility function $\chi(t)$ is causal in time. The Fourier transform of $\chi(t)$ is related to the dielectric constant $\epsilon(\omega)$

by

$$\epsilon(\omega) - 1 = 4\pi X(\omega), \tag{17}$$

where

$$X(\omega) = \int_0^\infty \chi(t) e^{i\omega t} dt. \tag{18}$$

Hence, the complex dielectric constant $\epsilon(\omega)$, or $\epsilon(\omega) - 1$, is analytic in the upper half ω plane. From Maxwell's equation for light waves in a medium with dielectric constant $\epsilon(\omega)$ and unit magnetic permeability ($\mu = 1$), we have

$$\epsilon(\omega) = c_0^2 \, [K(\omega)/\omega]^2, \tag{19}$$

where c_0 is the light speed in vacuum. Thus we can confirm that $K(\omega)/\omega$ is analytic in the upper half plane, except perhaps at its zeros.

Much is gained when use is made of the known physical properties for $\epsilon(\omega)$, or $X(\omega)$, at high frequencies. As shown by Nussenzveig (p. 44 of Ref. 10), based on the consideration of microscopic motion of electrons in an electric field, the susceptibility $X(\omega)$ is of the order of n_e/ω^2 as ω approaches infinity where, n_e is the electron density. Thus $\epsilon(\omega) - 1$ and, consequently $(K/\omega)^2 - 1/c_0^2$, are square-integrable in the form of Eq. (6). It is then simple to show that $K/\omega - 1/c_0$ is also square-integrable in the form of Eq. (6). From Titchmarsch's theorem we arrive at the Kramers–Kronig relations as given by Eqs. (12) and (13).

C. Acoustic waves in fluid

Ginzberg[5] has been quoted extensively as having provided a proof of Eqs. (12) and (13) for pressure waves in a fluid. Let $v(t)$ be the particle velocity of the fluid medium with ambient pressure p_0 and density ρ_0 and let $p(t)$ be the perturbation pressure. Again a heredity integral relation is assumed,[5]

$$\rho_0 v(t) = \int_{-\infty}^{t} s(t-\tau) p(\tau)\, d\tau. \tag{20}$$

The function $s(t)$ is casual and its transform $S(\omega)$ is analytic over the upper half ω plane. The Fourier transform of the above equation gives rise to

$$p_0 V(\omega) = S(\omega) P(\omega). \tag{21}$$

From the balance equation of linear momentum one finds

$$p_0 V(\omega) = -P(\omega)/c(\omega), \tag{22}$$

where $c(\omega)$ is the complex phase velocity of the plane harmonic wave given by $\omega/K(\omega)$. Hence $1/c(\omega) = -S(\omega)$ is also analytic in the upper half plane. This establishes the analytic property of $K(\omega)/\omega$.

The behavior of $K(\omega)$ or $1/c(\omega)$ as $\omega \to \infty$ is difficult to estimate as there is lacking a microscopic relation for $S(\omega)$ like that for $X(\omega)$ in the electromagnetic case. A discussion of the behavior of $1/c(\omega)$ for ω larger than a critical high frequency, say $\omega_0 (> 10^{12}\text{Hz})$, is given by Ginzberg. We shall not repeat his discussion here except to point out that he essentially assumed that $1/c(\omega)$ exists as $\omega \to \infty$, and that it approaches a limiting value uniformly, independent of $\arg\omega$.

Equation (20) may be criticized on the grounds that it is not a constitutive relation. The fluid velocity of a point is given by Eq. (20) as a linear functional only of the pressure at that point. At the very least there exists also an implied dependence on wave-propagation direction. The function $s(t)$ has no immediate physical significance.

D. Dispersion relations for homogeneous viscoelastic media

If we attempt a similar proof for stress waves in material media we find for the same reasons as in the previous section that $K(\omega)$ is analytic throughout the upper half ω plane; and again we find that the high frequency behavior of $K(\omega)$ can only be ascertained by appeal to physical argument. We can, however, improve the derivation by replacing Eq. (20) with a genuine constitutive equation.

For an anisotropic viscoelastic linear continuum the stress tensor σ_{ij} at a point is related to the local strain tensor ϵ_{kl} at the same point by a convolution integral[11]

$$\sigma_{ij}(t) = \int_{-\infty}^{t} c_{ijkl}(t-\tau)\epsilon_{kl}(\tau)\, d\tau. \tag{23}$$

The stiffness function $c(t)$ is causal. If ϵ and σ have a harmonic time dependence $\exp(-i\omega t)$ we may write

$$\Sigma_{ij}(\omega) = C_{ijkl}(\omega) E_{kl}(\omega), \tag{24}$$

with

$$C_{ijkl}(\omega) = \int_0^\infty c_{ijkl}(t) e^{i\omega t}\, dt. \tag{25}$$

From the balance equation of linear momentum

$$\rho \ddot{u}_i = \sigma_{ij,j},$$

where u_i is the particle displacement vector field, we conclude

$$-\rho\omega^2 U_i(\omega) = C_{ijkl}(\omega)\, \Sigma_{kl,j}(\omega). \tag{26}$$

Since $2\epsilon_{kl} = u_{k,l} + u_{l,k}$ and since $C_{ijkl} = C_{ijlk}$, we find

$$-\rho\omega^2 U_i(\omega) = C_{ijkl}(\omega)\, U_{k,lj}(\omega). \tag{27}$$

If the particle displacements are assumed to have a plane-wave dependence on position $\exp(i\mathbf{K}\cdot\mathbf{x})$, then

$$\rho\omega^2 U_i(\omega) = C_{ijkl}(\omega) K_l K_j U_k(\omega). \tag{28}$$

For propagation in direction $\mathbf{n}$ with components n_l and $\mathbf{K} = K\mathbf{n}$ the solutions to Eq. (28) are the eigenvectors of the matrix $C_{ijkl} n_l\, n_j = M_{ik}$ and $\rho\omega^2/K^2(\omega)$ are the associated eigenvalues.

One concludes, in analogy with Eq. (19), that $K(\omega)$ is given by the eigenvalues of the matrix M_{ik}.

$$M_{ik}(\omega)\, U_k(\omega) = (\rho\omega^2/K^2)\, U_i(\omega). \tag{29}$$

Thus knowledge of the high frequency behavior of $C(\omega)$ would determine the high frequency behavior of $K(\omega)$. Unfortunately this behavior appears to be very difficult to estimate. In the electromagnetic case there is a nondispersive continuum (the vacuum) which underlies the dielectric. At high frequency the dielectric has a negligible response and the propagation becomes that characteristic of the vacuum. In the stress-wave case there is no such underlying continuum, and the dispersive medium cannot be divorced from the wave quantities. Strain and displacement fields, unlike electric fields cannot be defined at high frequencies. In short, the continuum approximation inherent in Eq. (23) is invalid in the very region of high frequency where we wished to employ that equation. Locality and linearity may likewise be failing in this regime.

One could avoid these questions and, following Ginzberg,[5] assume the existence of the real uniform limit

$$c_\infty = \lim_{\omega\to\infty} \omega/K(\omega), \quad \omega_i \geqslant 0 \tag{30}$$

independent of $\arg\omega$ in the upper half plane. Using the analyticity of $K(\omega)$, and its assumed high frequency behavior, and assuming that $K'(0)$ exists, the integration of the following quantity along the contour Γ shown in Fig. 1 is taken

$$\frac{H(\omega)}{\omega-\omega'} = \left[\frac{K(\omega)-K(0)}{\omega-0} - \frac{1}{c_\infty}\right]\frac{1}{\omega-\omega'},$$

yielding Eq. (12) and (13).

We do not find this proof entirely satisfactory. It is perhaps possible to rigorously derive Eq. (30) from considerations on the high frequency behavior of the stiffness $C(\omega)$. We have been unable to do so convincingly. In this context it is perhaps appropriate to point out that the assumption of homogeneity made in subsection B also breaks down at high frequency, where the electron density cannot be considered a constant.

We admit further dissatisfaction with any proof based on appeal to the high-frequency behavior of a generalized susceptability such as $X(\omega)$ or $C(\omega)$. It is only in a homogeneous medium that one may make identifications such as (19) or (29). Thus such proofs cannot be generalized to inhomogeneous media. In a periodic medium $K(\omega)$ is a Floquet wave number and not directly related to a generalized local susceptability. In a homogeneously random medium where $K(\omega)$ is the wave number of the ensemble averaged field, again there is no direct connection with a generalized susceptability.

In a certain sense the high-frequency behavior of $K(\omega)$ is irrelevant to actual experimental work. The plausible assumption that $|K(\omega)|$ is bounded by some power of $|\omega|$ as $|\omega|\to\infty$ with $\text{Im}\omega\geqslant 0$, together with the analyticity of $K(\omega)$, already established at least for homogeneous media, suffices to give a dispersion relation for $K(\omega)$ with an unspecified number of subtractions. In practical work though, the number of subtractions employed is determined by convenience and may greatly exceed the required number.[3] Thus the theoretical determination of the required number may be moot. Nevertheless it is of considerable significance theoretically to determine the required number of subtraction, or indeed if there is any number which is sufficient.

We conclude this section with the statement of a certain property of $K(\omega)$. This property will be useful from place-to-place in the following and we state it here for reference.

Given that $g(t,z_0)$ is a real function, a property which follows from Eq. (14), we conclude, in view of Eq. (15), that $G(\omega,z_0)$ has the property

$$G(\omega^*,z_0) = G^*(-\omega,z_0)$$

and that $K(\omega)$ has the property

$$K(-\omega^*) = -K^*(\omega), \tag{31}$$

where an asterisk denotes the complex conjugate.

IV. THE PHASE PROBLEM

In many cases a physical experiment will measure only the absolute value $|G(\omega)|$ of a complex system function. It is the concern of the phase problem to reconstruct the complex number $G(\omega)$ from knowledge of $|G(\omega)|$. Two references in this regard are the papers by Toll,[9] and Burge *et al.*[12] Here we will quote Toll's result and show its relevance for our problem.

For a quantity $G(\omega)$ which is bounded and causal in the form of Eq. (4), we consider the complex quantity $\eta(\omega)$ defined by

$$G(\omega) = e^{i\eta(\omega)} \tag{32}$$

or

$$\text{Re}\eta(\omega)\equiv\eta_r = \arg G(\omega),$$
$$\text{Im}\eta(\omega)\equiv\eta_i = -\ln|G(\omega)|. \tag{33}$$

We ask if knowledge of $\eta_i(\omega)$, which is an even function of ω for all real ω, suffices to determine $\eta_r(\omega)$ for all real ω. Toll's answer is that it suffices to determine η_r to within two real constants, if (i) $G(\omega)$ has no zeros in the upper half plane, if (ii) the integral

$$\int_0^\infty \frac{\eta_i(\omega)}{1+\omega^2}\,d\omega < \infty \tag{34}$$

exists, and if (iii) $\eta_r(\omega)$ is continuous. He obtains the result,

$$\eta_r(\omega') = \frac{2\omega'}{\pi}\,⨍_0^\infty \frac{\eta_i(\omega)}{\omega^2-\omega'^2}\,d\omega + B\omega' + J, \tag{35}$$

where B is positive and real and J is real. The condition on the zeros of $G(\omega)$ and the condition on the continuity of $\eta_r(\omega)$ may be relaxed at the expense of generating additional degrees of freedom for $\eta_r(\omega)$.

As $G(\omega,z)$ for electromagnetic or stress-wave propagation in homogeneous media satisfies most of the three conditions assumed in Toll's derivation we may conclude that the quantities $\text{Re}K(\omega)$ and $\text{Im}K(\omega)$ as defined in Eq. (15) satisfy Eq. (35) if $\text{Re}K(\omega)$ is continuous and if

$$\int_0^\infty \frac{\text{Im}K(\omega)}{1+\omega^2}\,d\omega < \infty. \tag{34'}$$

This result is very interesting but it leaves one wondering if the conditions on the continuity of $\text{Re}K(\omega)$ and the convergence of the integral (34') may be relaxed. One also wonders if an inverse relation giving $\text{Im}K(\omega)$ in terms of $\text{Re}K(\omega)$ can be obtained. In the next section we present a new proof, applicable to any type of linear wave propagation. The new proof establishes an equation like (35) for $\text{Re}K(\omega)$ and a reciprocal equation for $\text{Im}K(\omega)$. Furthermore Eq. (34') is shown to be a consequence of first principles rather than an initial constraint.

V. DISPERSION RELATIONS FOR A GENERAL LINEAR MEDIUM

In this section we will establish dispersion relations for $K(\omega)$ in general linear media independent of any appeal to detailed physical structure. We will show that there must exist a real quantity, which plays the role of c_∞, such that formulas equivalent to Eqs. (12) and (13) are valid. These relations will be established for homogeneous media and for that class of inhomogeneous media which admit plane waves with wave number $K(\omega)$.

We begin by considering a slab of thickness D such as in Fig. 2 supporting some unspecified type of linear disturbance at frequency ω. As discussed in Sec. III the transfer amplitude–or transmission coefficient–through the slab should be of the form

$$G(\omega,D) = A(\omega)e^{iK(\omega)D}. \tag{36}$$

This is a form appropriate to any homogeneous medium, where $K(\omega)$ is the wave number in that medium. For a random inhomogeneous medium which is statistically homogeneous, $K(\omega)$ is the wave number of the average field and $G(\omega,D)$ is the ensemble-averaged transmission coefficient. For a periodic inhomogeneous medium, $K(\omega)$ is a Floquet wave number and D is restricted to a set of slab thicknesses differing by an integral multiple of the periodic spacing. Note that here we make no reference to the type of linear disturbance.

A. The analyticity of $K(\omega)$ for inhomogeneous media

We will need to show that $K(\omega)$, as defined above, is analytic in the upper half ω plane. For a homogeneous medium the discussion of Sec. III suffices.

In a random inhomogeneous medium which is sufficiently statistically homogeneous to allow the wave form Eq. (36), the analyticity of $K(\omega)$ may be established by the following proof. The same proof is also applicable to homogeneous media and is found in the book by Nussenzveig.[10] He points out that $K(\omega)$ will have the same domain of analyticity as that of $G(\omega,D)$ except at the zeros of $\exp[iK(\omega)D]$. For any given realization in the ensemble of random media, the field transmitted through the slab, though not a plane wave, is causal. The average of these fields is a plane wave, and of course still causal. Hence $G(\omega,D)$ has the required analyticity in random as well as in deterministic media and $K(\omega)$ will have the same domain of analyticity except at the zeros of $\exp[iK(\omega)D]$. But any zeros in the exponential factor at, say $\omega = \zeta$ would imply a branch cut in $G(\omega,D')$ at $\omega = \zeta$ for some $D' \neq D$. The possibility of a branch cut is excluded by the analyticity of $G(\omega,D)$ for all D. Thus there are no such zeros and $K(\omega)$ is analytic throughout the upper half ω plane.

In the case of a periodic medium, we may conclude that there are no zeros in the factor $\exp[iK(\omega)D]$ at any finite value of ω in the upper half plane because a zero in $G(\omega,D)$ would imply that there is some frequency (perhaps complex) for which a vanishing field to the right of the slab is consistent with a nonvanishing field to the left. This can only happen at a frequency at which one of the constituents of the medium has singular properties. But these properties are causal transforms and analytic in the upper half plane. Thus $G(\omega,D)$ has no zeros in the upper half ω plane and $K(\omega)$ may be assumed analytic there except for possible branch cuts where $K(\omega)$ jumps by an integral multiple of $2\pi/h$, h being the periodicity of the medium. While such branch cuts may exist, they may be swung into the lower half plane without loss of generality by noting that there can be no branch points of $K(\omega)$ in the upper half plane. Any such points for $K(\omega)$ would imply branch points for $G(\omega,D)$ and these are disallowed by its analyticity. Van Kampen[13] has also discussed the analyticity of the Floquet wave number as a function of ω.

Thus we conclude for this class of media that $K(\omega)$ is analytic in the upper half plane. We have not yet discussed its behavior at large ω and cannot yet conclude with a dispersion relation for $K(\omega)$. The discussion of the high frequency limit of $K(\omega)$ will occupy the remainder of this communication.

B. $K(\omega)$ as a Herglotz function

$K(\omega)$ has a further property which has so far not been exploited, that is, its imaginary part may be assumed nonnegative. Since the system under consideration is assumed passive and no energy can be added to the wave, we conclude that

$$|G(\omega,D)| \leqslant 1 \tag{37}$$

for all real ω and all positive D. It is appropriate here to point out that logical connections between passivity and causality have been widely noted [10,14,15], often in the context of deriving dispersion relations for system functions.[16,17] By referring to Eq. (4) we may extend Eq. (37) to all values of ω in the upper half plane and on the real axis. By noting that Eq. (37) must hold for arbitrarily large thicknesses, D, we conclude

$$|e^{iK(\omega)D}| \leqslant 1, \quad \mathrm{Im}\omega \geqslant 0 \tag{38}$$

and thus

$$\mathrm{Im}K(\omega) \geqslant 0, \quad \mathrm{Im}\omega \geqslant 0. \tag{39}$$

Equation (39) together with the analyticity of $K(\omega)$ in the upper half plane form the definition of a Hergoltz function.[10,18]

Any Herglotz function admits a representation in the upper half ω plane in terms of a bounded (from above and below) nondecreasing real function $\alpha(t)$ of a real variable t, the real numbers B, J, with B positive

$$K(\omega) = B\omega + J + \int_{-\infty}^{\infty} \frac{1+t\omega}{t-\omega}\, d\alpha(t) \quad (\mathrm{Im}\omega > 0). \tag{40}$$

The integral is a Stieltjes integral.

There is a slight resemblance between Eq. (40) and Eq. (7). The resemblance can be explored by taking the limit $\mathrm{Im}\omega \to 0$ and assuming that $\alpha(t)$ is sufficiently differentiable everywhere. One obtains, for real ω,

$$K(\omega) = B\omega + J - i\pi\alpha'(\omega)(1+\omega^2) + \fint_{-\infty}^{+\infty} \frac{1+t\omega}{t-\omega}\,\alpha'(t)\, dt \quad (\mathrm{Im}\omega = 0). \tag{41}$$

Thus α' may be identified

$$\alpha'(\omega) = -\frac{1}{\pi}\frac{\mathrm{Im}K(\omega)}{1+\omega^2} \tag{42}$$

and Eq. (41) becomes

$$\mathrm{Re}K(\omega) = B\omega + J - \frac{1}{\pi}\fint_{-\infty}^{\infty} \frac{1+t\omega}{1+t^2}\frac{\mathrm{Im}K(t)}{t-\omega}\, dt. \tag{43}$$

With a little more manipulation which exploits the property that $\mathrm{Im}K(\omega)$ is an even function of real ω, Eq. (43) may be made into the form (12) with $c_\infty = 1/B$ and $J = \mathrm{Re}K(0)$. This approach, similar to that of Wu,[17] whose concern is with despersion relations for impedances, is not adequate for the present purposes. We have assumed here that $\alpha(t)$ is differentiable. The meaning of such an assumption is not clear. Furthermore we have derived only one of the Hilbert transform-pair, and not the reciprocal relation which gives $\mathrm{Im}K(\omega)$ in terms of $\mathrm{Re}K(\omega)$.

From $K(\omega)$ as described in Eq. (40), we construct another analytic function

$$H(\omega) = \frac{K(\omega) - K(\omega_0)}{\omega - \omega_0} - B = \int_{-\infty}^{\infty} d\alpha(t)\frac{1+t^2}{(t-\omega)(t-\omega_0)}, \tag{44}$$

where ω_0 will be specified later. Clearly $H(\omega)$ is analytic in the upper half plane. We will show in the following that as $\omega \to \infty$, $H(\omega)$ vanishes *uniformly* in a sector of the upper half plane, $3\pi/4 \geqslant \arg\omega \geqslant \pi/4$ and approaches zero in a possibly *nonuniform* manner on and in a neighborhood of the real axis. It is in this limited sense only that $c_\infty = 1/B$ can be said to exist for general linear media. Fortunately, the nonuniform properties of $H(\omega)$ will be seen to be sufficiently unpathological that equations effectively equivalent to Eqs. (12) and (13) will be valid.

An illustration of the possible nonuniform properties of $K(\omega)$ is provided by the consideration of a medium which embeds sharply-resonant scatterers. We take their resonances to occur at frequencies $\mathrm{Re}\omega_n$, $n = 1,2,3...,\infty$ and each resonance to have width $\mathrm{Im}\omega_n > 0$. Each oscillator will be assumed to have strength a_n. Thus $K(\omega)$ will include a pathological term of the form

$$K_p(\omega) = \sum_n a_n/(\omega - \omega_n).$$

If, for example, $\mathrm{Im}\omega_n \to 0$ as $n \to \infty$ such that

$$\lim_{n\to\infty} a_n(\mathrm{Im}\omega_n)^{-1}\omega_n^{-1}$$

does not exist, i.e., if the resonances are becoming increasingly sharp as $n \to \infty$, then $H(\omega)$ may converge to zero on most sequences $\omega \to \infty$, but will diverge on sequences close to the sequence $\{\omega_n\}$. It may be that the behavior described here is unphysical. It is not clear, though, that all pathological behavior can be ruled out.

C. The contour integral

We consider the integral along the closed contour Γ of Fig. 1 where $\Gamma = \Gamma_1 + \Gamma_2 + \Gamma_3 + \Gamma_4$,

$$\int_\Gamma \frac{H(\omega)}{\omega - \omega'} d\omega = 0. \tag{45}$$

By taking the limit of this integral as Ω goes to infinity, and as ϵ goes to zero, we wish to establish that the contribution to the contour integral from the top, Γ_3, left- and right-hand sides, Γ_2 and Γ_4, vanish in the limit $\Omega \to \infty$. Clearly if $|H(\omega)|$ vanishes uniformly as $\Omega \to \infty$ on all points of a given section of the contour, the contribution from that section will vanish in this limit.

1. The top of the contour Γ_3

The proof that $|H(\omega)|$ vanishes as $\Omega \to \infty$ on Γ_3 is also found in Ref. 14. We write $H(\omega)$ in the following form:

$$H(\omega) = \int_{|t| \leqslant T} \frac{1+t^2}{(t-\omega)(t-\omega_0)} d\alpha(t) + \int_{|t|>T} \frac{1+t^2}{(t-\omega)(t-\omega_0)} d\alpha(t), \tag{46}$$

where T will be taken as arbitrarily large but fixed as $\Omega \to \infty$ and ω in understood to be on Γ_3.

In the limit $\Omega \to \infty$ with $|\omega| \geqslant \Omega$, the first integral of Eq. (46) vanishes because it is of the order $1/\omega$ times a fixed quantity. In the domain of the second integral where $|t| \gg |\omega_0|$ the integrand is of the order of unity,

$$|(1+t^2)/(t-\omega)(t-\omega_0)| \cong 1. \tag{47}$$

The second term of Eq. (46) then becomes

$$\left| \int_{|t|>T} \frac{1+t^2}{(t-\omega)(t-\omega_0)} d\alpha(t) \right| \cong \int_{|t|>T} d\alpha(t). \tag{48}$$

Since T is arbitrarily large and as $\alpha(t)$ is monotonic and bounded from above and below this integral is arbitrarily small. We conclude that $|H(\omega)|$ vanishes on Γ_3 as $\Omega \to \infty$ and that the top section of Γ contributes zero in the limit $\Omega \to \infty$.

2. The sides of the contour Γ_2 and Γ_4

We now let ω lie on Γ_2 and break the integral representation for $H(\omega)$ into four terms

$$H(\omega) = \int_{-\infty}^{T} + \int_{T}^{\Omega - \beta\Omega} + \int_{\Omega-\beta\Omega}^{\Omega+\beta\Omega} + \int_{\Omega+\beta\Omega}^{\infty} \frac{1+t^2}{(t-\omega)(t-\omega_0)} d\alpha(t), \tag{49}$$

where $0 < \beta < 1$ and T is arbitrarily large but fixed as $\Omega \to \infty$.

The first integral vanishes because it is of the order $1/\omega$ times a fixed quantity; the second and fourth integrals vanish as $\Omega \to \infty$ in manners similar to that of Eq. (48). The third integral is readily seen to be bounded for sufficiently large Ω,

$$|H(\omega)| \leqslant [\alpha(\Omega + \beta\Omega) - \alpha(\Omega - \beta\Omega)](1+\beta)\Omega/\mathrm{Im}\omega. \tag{50}$$

It is not apparent that the expression above vanishes as $\Omega \to \infty$. The first factor vanishes, but the factor Ω renders the convergence questionable.

We can, however, through the bound on $|H(\omega)|$ along Γ_2, construct the consequent bound on the contribution to the contour integral from this segment. For large Ω and for $\omega = \Omega + i\omega_i$

$$\left| \int_{\Omega+i\delta}^{\Omega+i\Omega} \frac{H(\omega)}{\omega-\omega'} d\omega \right| \leqslant [\alpha(\Omega+\beta\Omega) - \alpha(\Omega-\beta\Omega)](1+\beta)[\ln\Omega - \ln\delta]. \tag{51}$$

This expression will now be shown to vanish as $\Omega\rightarrow\infty$ at fixed δ but in general, not in a uniform manner.

We state that there exists a sequence of increasing numbers Ω_n, $n = 1,2,3...$ with $\Omega_n\rightarrow\infty$ as $n\rightarrow\infty$ such that the right-hand side of inequality (51) vanishes as $n\rightarrow\infty$. This will be proved by asserting the contrary, i.e., that there is no such sequence, and discovering a contradiction. It is clear that if there is no such sequence then there exists some (large) Ω such that for all $X>\Omega/1-\beta$

$$\alpha(X+\beta X)-\alpha(X-\beta X)\geqslant\frac{1}{\ln X\ln\ln X}. \tag{52}$$

This implies

$$\alpha(\infty)-\alpha(\Omega)=\int_\Omega^\infty d\alpha(t)=\sum_{m=0}^\infty\int_{X_m}^{X_{m+1}}d\alpha(t)$$

$$=\sum_{m=0}^\infty[\alpha(X_{m+1})-\alpha(X_m)]=\sum_{m=0}^\infty\left[\left\{\alpha(1+\beta)\frac{X_M}{1-\beta}\right\}-\alpha\left\{(1-\beta)\frac{X_m}{1-\beta}\right\}\right]. \tag{53}$$

where $X_0=\Omega$ and $X_{m+1}=[(1+\beta)/(1-\beta)]X_m$. Combining Eqs. (52) and (53) we conclude, with $\gamma=(1+\beta)/(1-\beta)$,

$$\alpha(\infty)-\alpha(\Omega)\geqslant\sum_{m=0}^\infty\left[\left(m\ln\gamma+\ln\frac{\Omega}{1-\beta}\right)\ln\left(m\ln\gamma+\ln\frac{\Omega}{(1-\beta)}\right)\right]^{-1}. \tag{54}$$

The sum does not converge. But $\alpha(\infty)-\alpha(\Omega)$ must exist. Thus we have a contradiction and there does exist a sequence of Ω_n with the desired properties.

We conclude that

$$\lim_{n\rightarrow\infty}\int_{\Omega_n+i\delta}^{\Omega_n+i\Omega_n}\frac{H(\omega)}{\omega-\omega'}d\omega=0. \tag{55}$$

Similarly we can conclude for the left-hand side

$$\lim_{n\rightarrow\infty}\int_{iZ_n-Z_n}^{-Z_n+i\delta}\frac{H(\omega)}{\omega-\omega'}d\omega=0, \tag{56}$$

where Z_n is some sequence, not necessarily the same as Ω_n, which goes to infinity as n goes to infinity. It will be shown, however, that Z_n can be taken equal to Ω_n. This will follow from the parity of $K(\omega)$ under $\omega\rightarrow\omega^*$.

3. The bottom side of the contour

We now take the limit $\delta\rightarrow0$. The limit can be taken only if $K(\Omega_n)$ exists in a suitably well-defined way. The validity of the limit is related to the validity of taking the bottom section of the integration along the real axis in the first place. $K(\omega)$ has not been guaranteed analytic, or even everywhere defined, for real ω. But since $K(\omega_r+i\delta)$ is analytic for all $\delta>0$, it is reasonable to assume that its integral transforms with respect to ω_r are continuous functions of δ as $\delta\rightarrow0$. Hence $K(\omega_r)$, though perhaps not a good function, is in this sense locally integrable and thus a distribution. See Beltrami and Wohlers[19] for a discussion of distributions as boundary values of analytic functions.

Thus we conclude that

$$\lim_{n\rightarrow\infty}\int_{-Z_n}^{\Omega_n}\frac{H(\omega)}{\omega-\omega'}d\omega=0 \tag{57}$$

and, for $\mathrm{Im}\omega'\rightarrow0$,

$$\lim_{n\rightarrow\infty}⨍_{-Z_n}^{\Omega_n}\frac{H(\omega)}{\omega-\omega'}d\omega=i\pi H(\omega'); \tag{58}$$

the real and imaginary parts may be taken now.

$$\mathrm{Im}H(\omega')=\lim_{n\rightarrow\infty}\frac{-1}{\pi}⨍_{-Z_n}^{\Omega_n}\frac{\mathrm{Re}H(\omega)}{\omega-\omega'}d\omega, \tag{59}$$

$$\mathrm{Re}H(\omega')=\lim_{n\rightarrow\infty}\frac{1}{\pi}⨍_{-Z_n}^{\Omega_n}\frac{\mathrm{Im}H(\omega)}{\omega-\omega'}d\omega. \tag{60}$$

Since $\mathrm{Im}K(\omega)\geqslant0$ for all real ω, the limit in Eq. (60) is clearly independent of the details of the sequences Ω_n and Z_n. We therefore may take, for this integral, $\Omega_n=Z_n\rightarrow\infty$

$$\mathrm{Re}H(\omega')=\frac{1}{\pi}⨍_{-\infty}^{+\infty}\frac{\mathrm{Im}H(\omega)}{\omega-\omega'}d\omega, \tag{61}$$

where there can be no ambiguity in regard to the limits of integration. Furthermore, taking $\omega_0=0$, $\mathrm{Im}H(\omega)$, except for a part which gives zero upon being Hilbert-transformed as in Eq. (61), becomes an odd function of real ω [(see Eq. (31)] and the above integral can be rewritten as

$$\mathrm{Re}H(\omega')=\frac{2}{\pi}⨍_0^\infty\frac{\omega\mathrm{Im}H(\omega)}{\omega^2-\omega'^2}d\omega \tag{62}$$

or

$$\mathrm{Re}K(\omega')=B\omega'+\frac{2\omega'}{\pi}⨍_0^\infty\frac{\mathrm{Im}K(\omega)}{\omega^2-\omega'^2}d\omega+\mathrm{Re}K(0). \tag{63}$$

Note that Eq. (63) implies the inequality (34').

Equation (59) is not in a useful form. The purpose of the remaining parts of this section will be to replace Eq. (59) by Eq. (68). This will be done by considering the right- and left-hand sides of Γ simultaneously and showing that the real part of the possibly nonuniformly convergent part of the sum of the contributions from the right and left sides of Γ vanishes for all Ω.

4. The uniform limit

Again we assume $\omega_0=0$. We first show that the $\omega'\neq0$ integrals on the sides differ from the $\omega'=0$ integrals only by a uniformly convergent part. By referring to Eq. (50) one may write, for ω on the left or right sides of Γ,

$$\left|\frac{H(\omega)}{\omega-\omega'}\right|\leqslant[\alpha(\Omega+\beta\Omega)-\alpha(\Omega-\beta\Omega)]\times\frac{(1+\beta)}{\mathrm{Im}\omega}\left[1+O\left(\frac{\omega'}{\Omega}\right)\right]. \tag{64}$$

If this quantity is now integrated along the right or left sides of Γ, it is only the leading term, which is independent of ω', that is of questionable convergence. The integral of the remaining terms clearly vanishes uniformly as $\Omega\to\infty$. Thus for an investigation of the nonuniform covergence of the side integrals, we may take $\omega'=0$ without loss of generality.

Now setting $\omega'=0$, we consider the integral up the right-hand side of the contour Γ_2 and down the left-hand side Γ_4. The integral along Γ_2 is

$$\int_{\Gamma_2}\frac{H(\omega)}{\omega}\,d\omega=\int_{\delta}^{\Omega}\frac{H(\Omega+iy)}{\Omega+iy}\,i\,dy. \tag{65}$$

The integral along Γ_4 is

$$\int_{\Gamma_4}\frac{H(\omega)}{\omega}\,d\omega=\int_{\Omega}^{\delta}\frac{H(-\Omega+iy)}{-\Omega+iy}\,i\,dy. \tag{66}$$

Equation (66) becomes, on recognizing that $H(\omega)=H^*(-\omega^*)$ except for the unimportant term $K(0)/\omega$,

$$\int_{\Gamma_4}\frac{H(\omega)}{\omega}\,d\omega=\int_{\delta}^{\Omega}\frac{H^*(\Omega+iy)}{(\Omega+iy)^*}\,i\,dy$$

and the sum of the integrals on Γ_2 and Γ_4 becomes

$$\int_{\Gamma_2}\frac{H(\omega)}{\omega}\,d\omega+\int_{\Gamma_4}\frac{H(\omega)}{\omega}\,d\omega =i\int_{\delta}^{\Omega}dy\left\{\frac{H(\Omega+iy)}{\Omega+iy}+\frac{H^*(\Omega+iy)}{(\Omega+iy)^*}\right\}, \tag{67}$$

which has zero real part.

Therefore the integral

$$\lim_{\Omega\to\infty}\mathrm{Re}\left\{\int_{\Gamma_2}\frac{H(\omega)}{\omega-\omega'}\,d\omega+\int_{\Gamma_4}\frac{H(\omega)}{\omega-\omega'}\,d\omega\right\}$$

is uniformly convergent to zero as $\Omega\to\infty$.

We rewrite Eq. (59) as

$$\mathrm{Im}H(\omega')=\lim_{\Omega\to\infty}\frac{-1}{\pi}\fint_{-\Omega}^{+\Omega}\frac{\mathrm{Re}H(\omega)}{\omega-\omega'}\,d\omega. \tag{68}$$

Since $\mathrm{Re}H(\omega)=\mathrm{Re}H(-\omega)$ for ω real except for a part which gives zero upon being Hilbert transformed, we may recast Eq. (68) into

$$\mathrm{Im}H(\omega')=\frac{-2\omega'}{\pi}\fint_{0}^{\infty}\frac{\mathrm{Re}H(\omega)}{\omega^2-\omega'^2}\,d\omega \tag{69}$$

or

$$\mathrm{Im}K(\omega')=\frac{-2\omega'^2}{\pi}\int_{0}^{\infty}\frac{\mathrm{Re}K(\omega)/\omega-B}{\omega^2-\omega'^2}\,d\omega +\mathrm{Im}K(0). \tag{70}$$

By subtracting from Eqs. (63) and (70) the appropriate multiples of

$$\fint_{0}^{\infty}\frac{1}{\omega^2-\omega'^2}\,d\omega=0$$

one obtains the sometimes more convenient formulas

$$\mathrm{Re}K(\omega')=B\omega'+\frac{2\omega'}{\pi}\int_{0}^{\infty}\frac{\mathrm{Im}K(\omega)-\mathrm{Im}K(\omega')}{\omega^2-\omega'^2}\,d\omega +\mathrm{Re}K(0), \tag{71}$$

$$\mathrm{Im}K(\omega')=\frac{-2\omega'^2}{\pi}\int_{0}^{\infty}\left[\frac{\mathrm{Re}K(\omega')}{\omega}-\frac{\mathrm{Re}K(\omega')}{\omega'}\right]\times\frac{d\omega}{\omega^2-\omega'^2}+\mathrm{Im}K(0), \tag{72}$$

where now the integrands contain no explicit poles. B is an unknown positive real number, equal to, when the limit exists unambiguously,

$$\lim_{\omega\to\infty}K(\omega)/\omega=B.$$

In lieu of the definition (44) of $H(\omega)$ one could consider the quantity

$$H(\omega)=\left[\frac{K(\omega)-K(0)}{\omega}-K'(0)\right]\frac{1}{\omega}, \tag{73}$$

which, by analysis similar to that of Sec. VC1 and VC2 can be seen to be uniformly convergent to zero on Γ_2, Γ_3, and Γ_4 as $\Omega\to\infty$. It follows immediately that one may write a dispersion relation for $H(\omega)$. After taking the imaginary part of this dispersion relation and recognizing that $\mathrm{Im}K'(0)=0$, one obtains Eq. (72). After taking the real part one obtains, in place of Eq. (71), the possibly more useful relation

$$\mathrm{Re}K(\omega')=\mathrm{Re}K(0)+\omega'\mathrm{Re}K'(0) +\frac{2\omega'^3}{\pi}\int_{0}^{\infty}\left[\frac{\mathrm{Im}K(\omega)}{\omega^2}-\frac{\mathrm{Im}K'(\omega')}{\omega'^2}\right]\frac{d\omega}{\omega^2-\omega'^2}.$$

If $\mathrm{Re}K(0)=0$, as it must if $\mathrm{Re}K(\omega)$ is an odd function, Eq. (74) becomes Horton's[3] Eq. (5).

VI. CONCLUSIONS

To summarize, the Kramers–Kronig relations have been established without reference to the exact physical nature of the medium, and independent of any assumptions regarding the high frequency behavior of the medium. This high frequency behavior, until now a prerequisite to Kramers–Kronig relations, is seen to follow logically from the first principles of causality, linearity, and passivity. The Kramers–Kronig relations for the wave number $K(\omega)$ have thus been found to hold for a wider class of homogeneous and inhomogeneous media, for which *a priori* high frequency behavior is difficult to judge, than has heretofore been thought to be the case.

The surprising result is perhaps not the validity of some form of a Kramers–Kronig relation for general linear media. They have been employed in a wide range of circumstances for many years. It would have been remarkable had it been shown that for some media there is no dispersion relation for the wave number $K(\omega)$. The more surprising result is that two subtractions Eqs. (71) and (72), where the second subtraction is a point at infintiy, are always sufficient, or, speaking more loosely, that the attenuation, $\mathrm{Im}K(\omega)$, cannot rise as $\omega\to\infty$ as fast as a linear function of ω, nor can the real part, $\mathrm{Re}K(\omega)$, rise faster than a linear function of ω. It is remarkable that these properties follow from only causality, passivity, and linearity with no appeal to specific physical argument.

ACKNOWLEDGMENTS

The authors wish to thank Professor W. Fuchs of Cornell University and Professor H. M. Nussenzveig of Universidade De São Paulo for helpful discussions. This work was conducted under a contract from the Air Force Office of Scientific Research.

[1]H. A. Kramers, *Estratto diagli Atti del Congresso Internazionale de Fisica Coma* (Nicolo Zonichelli, Bologna, 1927).
[2]R. Kronig, Ned. Tijdschr. Natuurkd. **9**, 402 (1942).
[3]C. W. Horton Sr., J. Acoust. Soc. Am. **55**, 547 (1974).
[4]M. O'Donnell, E. T. Jaynes, and J. G. Miller, J. Acoust. Soc. Am. **63**, 1935 (1978).
[5]V. L. Ginzberg, Sov. Phys. Acoust. **1**, 32 (1955).
[6]M. L. Goldberger, in *Relations de Dispersions et Particules Élemaentaires*, edited by Cecile DeWitte (Hermann, Paris, 1960), pp. 15–157.
[7]W. Sachse and Y. H. Pao, J. Appl. Phys. **49**, 4320 (1978).
[8]E. C. Titchmarsch, *Theory of Fourier Integrals* (Clarendon, Oxford, 1948).
[9]J. S. Toll, Phys. Rev. **104**, 1760 (1956).
[10]H. M. Nussenzveig, *Causality and Dispersion Relations* (Academic, New York, 1972).
[11]B. Gross, *Theories of Viscoelasticity* (Hermann, Paris, 1953).
[12]R. E. Burge, M. A. Fiddy, A. H. Greenaway, and G. Ross, Proc. Roy. Soc. A **350**, 191 (1976).
[13]N. G. van Kampen, Ned. Tijdschr. Natuurkd. **24** 1, (1958).
[14]D. Youla, L. Castriota and H. Carlin, IEEE Trans. Circuit Theory **CT-6**, 102 (March 1959).
[15]A. Zemanian, *Distribution Theory and Transform Analysis* (McGraw–Hill, New York, 1965).
[16]W. Güttinger, SIAM J. Appl. Math. **15**, 964 (1967).
[17]T. T. Wu, J. Math Phys. **3**, 262 (1962).
[18]J. A. Shohat and J. D. Tamarkin, *The Problem of Moments* (Am. Math. Soc., Providence, R. I., 1943), p. 23.
[19]E. J. Beltrami and M. R. Wohlers, *Distributions and the Boundary Values of Analytic Functions*, (Academic, New York, 1966).

Electromagnetism and the holomorphic properties of spacetime

Nikos Salingaros
Physics Department, University of Crete, Iraklion, Crete, Greece, and Physics Department, University of Massachusetts, Boston, Massachusetts 02125

(Received 29 October 1980; accepted for publication 1 May 1981)

The Cauchy–Riemann equations of holomorphy are extended to fields in higher-dimensional spaces in a framework of Clifford algebras. The equations of holomorphy in Minkowski spacetime turn out to be the Maxwell equations in vacuum. The Lorentz gauge condition is a result of the holomorphy. Sources can be included in an extension of the residue theorem, where charges correspond to the residues.

PACS numbers: 03.50.De, 02.40. + m,

I. INTRODUCTION

Ever since the creation of the beautiful branch of mathematics dealing with holomorphic functions and residue integration in the complex plane,[1,2] there have been efforts to extend those results to higher dimensions. The direction which concerns us is the formulation of holomorphic fields, using, as a basis, an algebra with anticommuting elements and not several commuting complex bases. This latter approach has been adopted to create the theory of "several complex variables."[3,4] The theorems of Frobenius[5,6] and Hurwitz and Albert[7–10] give all normed division algebras without singular inverses other than the zero element as ℝ, ℂ, and H. This result has motivated a group of researchers to look into "quaternion holomorphy"[11,12] as the natural extension of complex holomorphic fields. (Ref. 12 contains an extensive bibliography.)

Recently, we have demonstrated that all Clifford algebras up to order 8 are either division algebras, or "singular" division algebras.[13,14] Furthermore, any larger Clifford algebra has division defined for each rank antisymmetric tensor field.[15,16] This property, as well as the manipulatory ease of Clifford algebras, leads us to formulate a theory of holomorphic fields using Clifford algebras. Related but distinct efforts can be found in Refs. 17–20. Also of related interest is the work of Penrose and his school on holomorphic twistor fields.[21–23]

In this paper, we show how to construct the generalized Dirac operator D [24–26] in any flat Riemannian space as a vector operator in the Clifford algebra (Sec. III). D is defined using the realization of Clifford algebras in terms of the differential forms of each space, introduced in Refs. 15 and 16.

In Sec. IV, we show that the Cauchy–Riemann equations of holomorphy correspond to the expression $Df=0$ in two dimensions. By using the Clifford algebra in four-dimensional Minkowski spacetime,[27–29] the corresponding expression $Df=0$ in four dimensions gives rise to two distinct sets of equations. When f is a vector field, it can be identified with the electromagnetic potential, and the condition $Df=0$ is equivalent to the Lorentz gauge condition plus the differential equations expressing a zero electromagnetic field in terms of the potential. When f is an antisymmetric rank-2 tensor field, f can be identified with the electomagnetic field, and the condition $Df=0$ is equivalent to the Maxwell equations in vacuum.

It is possible to include singular points, in which case the fields are nonholomorphic. An integral formalism derived from the differential treatment is introduced. Using the analogy with electromagnetic fields, we show how a residue in Minkowski space can be evaluated via an extension of the usual Residue theorem. We apply this method to (i) the vector field $\mathbf{r}|\mathbf{r}|^{n}$ and (ii) the electric and magnetic multipole fields. The result obtained is that the only nonholomorphic field is in fact that corresponding to the electric field of a point charge. These examples illustrate the close connection between electromagnetism and the holomorphic properties of Minkowski spacetime.

II. PROPERTIES OF THE CLIFFORD ALGEBRA IN MINKOWSKI SPACETIME

We review her the "vee" representation of Clifford algebras in terms of differential forms, which was introduced in Refs. 15 and 16. The reader is referred to there for details; here we give a summary of those results that are necessary in the following discussion. In particular, we study the Clifford algebra in Minkowski spacetime, denoted by $A^{1,3}=N_4$.[27–29]

Consider the differential oneforms $\sigma^{\mu}=dx^{\mu}$; $\mu=1,2,3,4$, of the Minkowski space M.[1,3] We construct a set of $2^4=16$ basis p-forms using the Cartan exterior product[30]:

$$\{1,\sigma^{\mu},\sigma^{\mu}\wedge\sigma^{\nu},\sigma^{\mu}\wedge\sigma^{\nu}\wedge\sigma^{\lambda},\quad \omega^4=\sigma^1\wedge\cdots\wedge\sigma^4\},$$
$$\mu,\nu,\lambda=1,\ldots,4,\quad \mu\neq\nu\neq\lambda. \tag{1}$$

The volume element in four dimensions is labelled ω^4. In the space $M^{1,3}$, define a metric scalar form

$$g^{\mu\nu}=(\sigma^{\mu},\sigma^{\nu})=\mathrm{diag}(-1,-1,-1,+1). \tag{2}$$

In general, the metric can have p plus signs, q minus signs, and $p+q=n$. In that case, the construction is extended to the Clifford algebra $A^{p,q}$, which is of dimension 2^{p+q}.[15,16]

We define an associative multiplication $\vee$, "vee," between all the basis forms in (1) in terms of the Cartan exterior product and the contractions (2). The "vee" multiplication between a basis r-form and a basis (s)-form is defined as a sum of permutations of basis forms in (1), as follows[15,16]:

Definition 1:

$$(\sigma^{\lambda_1}\wedge\cdots\wedge\sigma^{\lambda_r})\vee(\sigma^{\mu_1}\wedge\cdots\wedge\sigma^{\mu_s})$$
$$=\sum_{k=0}\frac{1}{k!(r-k)!(s-k)!}\sum_{\Pi_r}\sum_{\Pi_s}(-1)^{\Pi_r}(-1)^{\Pi_s}$$
$$\times g^{\nu_1\rho_1}\cdots g^{\nu_k\rho_k}\sigma^{\nu_{k+1}}\wedge\cdots\wedge\sigma^{\nu_r}\wedge\sigma^{\rho_{k+1}}\wedge\cdots\wedge\sigma^{\rho_s}. \tag{3}$$

 0022-2488/81/091919-07$01.00

Here $(-1)^{\Pi_r}$ and $(-1)^{\Pi_s}$ are the signs of the permutations

$$\Pi_r = \begin{pmatrix} \nu_1 & \cdots & \nu_r \\ \lambda_1 & \cdots & \lambda_r \end{pmatrix}, \quad \Pi_s = \begin{pmatrix} \rho_1 & \cdots & \rho_s \\ \mu_1 & \cdots & \mu_s \end{pmatrix}. \tag{4}$$

In actual practice, the rules of manipulation of the vee product are very simple. For example, it follows from definition (3) that

$$\sigma^\mu \vee \sigma^\nu = g^{\mu\nu} + \sigma^\mu \wedge \sigma^\nu, \tag{5a}$$

$$\sigma^\mu \vee (\sigma^\nu \wedge \sigma^\lambda) = g^{\mu\nu}\sigma^\lambda - g^{\mu\lambda}\sigma^\nu + \sigma^\mu \wedge \sigma^\nu \wedge \sigma^\lambda. \tag{5b}$$

Define the "tensor types" f_k, which are antisymmetric tensor fields of rank k expanded onto a basis of differential forms in (1). The crucial difference between the "tensor types" and the usual differential forms is that the bases are here endowed with the vee product; hence the tensor types possess intrinsic algebraic properties in addition to those expected from the theory of differential forms. The most novel property is the existence of a unique two-sided inverse of each tensor type.[15,16] In four dimensions, the ranks of the tensor fields can be only 0,1,2,3, and 4. The most general element of the algebra A [1,3] is a combination of all distinct tensor types f_k:

$$\begin{aligned} \alpha &= f_0 + f_1 + f_2 + f_3 + f_4 \\ &= f_0 + \sum_\mu f_1{}^\mu \sigma^\mu + \tfrac{1}{2}\sum_{\mu,\nu} f_2{}^{\mu\nu}\sigma^\mu \wedge \sigma^\nu \\ &\quad + \frac{1}{3!}\sum_{\mu,\nu,\lambda} f_3{}^{\mu\nu\lambda}\sigma^\mu \wedge \sigma^\nu \wedge \sigma^\lambda + f_4{}^0\omega^4, \\ &\qquad \mu, \nu, \lambda = 1,2,3,4, \quad \mu \neq \nu \neq \lambda. \end{aligned} \tag{6}$$

The coefficients of the tensor types are all *real*; the total number of scalar components is 16, which is the dimension of the algebra.

In the discussion of forms, the notation of the dual[30] plays an important role. A result of particular practical significance is the ability to express the dual notation algebraically using the vee product as follows[15,16]:

Theorem 1:

$$\underset{k}{*} f_p = (-1)^t \omega^k \vee f_p = f_{(k-p)}. \tag{7}$$

The index t is different for each space and for each rank. The duality theorem [Eq. (7)] can be used in the three-dimensional subspace of Minkowski space to reduce a second rank tensor F in four dimensions into space and time components. This is known as the "canonical decomposition"[29]:

$$F = \mathbf{E} \wedge \sigma^4 - \underset{3}{*}\,\mathbf{B} = \mathbf{E} \wedge \sigma^4 + \omega^3 \vee \mathbf{B},$$

$$E^i = F^{i4}, \quad B^i = -\tfrac{1}{2}\sum_{i,j,k} \epsilon^{ijk} F^{jk}, i, j,k = 1,2,3. \tag{8}$$

In the case of the electromagnetic field, this is just the familiar reduction into the vector fields $\mathbf{E}$ and $\mathbf{B}$ in the three-dimensional space.

In our notation, a and b will represent vectors in four dimensions, with scalar product $(a,b) = \Sigma_{\mu=1}^4 a_\mu b^\mu$. We use the Minkowski metric (2) $g^{\mu\mu} = (-1, -1, -1, +1)$. The quantities $\mathbf{a}$ and $\mathbf{b}$ are the *spatial* part of the *same* vectors, and have a Euclidean scalar product denoted by $(\mathbf{a}\cdot\mathbf{b}) = \Sigma_{i=1}^3 a^i b^i$. The usual vector cross product in the three-dimensional subspace is defined as $\mathbf{a}\times\mathbf{b} = \Sigma_{i,j,k=1}^3 a^i b^j \epsilon^{ijk}\sigma^k$.

We list the vee products between vectors a, b and the tensor type-2 F, as follows; they are easily calculated using (5):

$$a \vee b = (a,b) + (b^4\mathbf{a} - a^4\mathbf{b}) \wedge \sigma^4 - \omega^3 \vee \mathbf{a}\times\mathbf{b}, \tag{9a}$$

$$\begin{aligned} a \vee F = &-a^4\mathbf{E} - \mathbf{a}\times\mathbf{B} - (\mathbf{a}\cdot\mathbf{E})\sigma^4 + \omega^4 \vee [-(\mathbf{a}\cdot\mathbf{B})\sigma^4 \\ &+ \mathbf{a}\times\mathbf{E} - a^4\mathbf{B}]. \end{aligned} \tag{9b}$$

The hypercomplex character of the vee product is explicit in (9a). (Compare this product with the well-known quaternion product. It has a similar form, but is quite distinct.) In (9b) we have used the canonical decomposition (8) to separate the tensor F into the space part $\mathbf{B}$ and the space-time part $\mathbf{E}$. Identities (9) demonstrate that fields in space-time can be described in our formalism by employing the traditional vector notation.

III. CONSTRUCTING THE GENERALIZED DIRAC OPERATOR

Consider a field f that is a function of the n variables $x^1,\ldots,x^n$. Then, any general derivation D can be written in terms of the chain rule as

Definition 2:

$$Df[x^1,\ldots,x^n] = \sum_{\alpha=1}^n dx^\alpha \partial_\alpha f[x^1,\ldots,x^n]. \tag{10}$$

Motivated by this rule (10), proceed by making the following identifications: (i) Interpret the coordinate differentials dx^α as the basis one-forms σ^α endowed with the vee product, and (ii) interpret f as any "tensor type" in (6). We can consequently use the Leibnitz chain rule to define a first-order differential operator. The interpretation given here has the feature that the differential operator D actually generates the algebra in the space of variables.

It is possible to identify the partial differential operators ∂_α with the covariant components of a vector in n dimensions expanded on the σ-basis, as follows:

$$Df = D \vee F, \quad D = \sum_{\alpha=1}^n \partial_\alpha \sigma^\alpha. \tag{11}$$

With this identification (11), the purely operational definition (10) becomes an algebraic definition of D as a vector differential operator. The obvious advantage to this construction is that the properties of D can be deduced very simply by using the algebraic vee structure. We list some of the properties here:

$$D{:}f_p \rightarrow f_{(p-1)} + f_{(p+1)}, \tag{12a}$$

$$D \vee (f_p + f_k) = D \vee f_p + D \vee f_k, \tag{12b}$$

$$D \vee (D \vee f) = (D \vee D) \vee f, \tag{12c}$$

$$D \vee D = \Box. \tag{12d}$$

These properties (12), along with the universality of the Clifford algebra, demonstrate that D is in fact isomorphic to the generalized Dirac operator.

In the Grassmann algebra, D may be identified with the operator $(d + \delta)$, which is the sum of the Cartan exterior derivative d with the Hodge coderivative δ, as follows[25,26]:

Theorem 2:

$$D\vee f \leftrightarrow (d+\delta)\wedge f = d\wedge \pm {}^*(d\wedge({}^*f)). \tag{13}$$

It is to be stressed that, although some results are expressible using d and δ separately, it is crucial in the following analysis to consider the operator D as a vector operator.

Practical calculations in Minkowski spacetime are done as follows. In keeping with the traditional usage of vector calculus, we denote the vector derivative in the three-dimensional subspace of Minkowski space as ∇, using *contravariant* partial derivatives of the space coordinates.

$$\nabla = \sum_{i=1}^{3} \partial^i \sigma^i, \quad D = -\nabla + \partial^4\sigma^4. \tag{14}$$

Using (14) and the vee product rules (9), we explicitly write down the D derivatives of the tensor types in Minkowski spacetime:

$$D\vee a = (\nabla\cdot\mathbf{a}) + \partial^4 a^4 + (-\partial^4\mathbf{a} - \nabla a^4)\wedge\sigma^4 + \omega^3\vee\nabla\times\mathbf{a}, \tag{15a}$$

$$D\vee F = \nabla\cdot\mathbf{B} - \partial^4\mathbf{E} + (\nabla\cdot\mathbf{E})\sigma^4 + \omega^4\vee[-\partial^4\mathbf{B} - \nabla\times\mathbf{E} + (\nabla\cdot\mathbf{B})\sigma^4]. \tag{15b}$$

We have utilized the canonical decomposition (8) in (15b). One sees from (15) that fields in spacetime can be described by using the language of classical vector analysis. The properties of D acting as a differential operator on vee and wedge products of fields $f\vee g$ and $f\wedge g$ are easily obtained by writing everything in coefficient form and using the algebraic rules (5), (11). Often, it is easiest to first reduce the vee products between fields using (8),(9), then apply D using (14), (15).

IV. HOLOMORPHY IN FOUR DIMENSIONS. MAXWELL'S EQUATIONS

In this section, we investigate the extension of the concept of holomorphic fields to any dimension. The specific interest of this paper is in four-dimensional spacetime with Minkowski metric, but our construction is quite general.

We introduce the notion of holomorphy in n dimensions, using the D operator as defined in the previous section.

Definition 3: A tensor type f is holomorphic iff

$$D\vee f = 0. \tag{16}$$

By utilizing the identification (13) in the Grassmann algebra, the condition of holomorphy (16) is equivalent to the two separate conditions (f is a tensor field of homogeneous rank).

$$d\wedge f = 0, \tag{17a}$$

$$d\wedge({}^*f) = 0. \tag{17b}$$

In two dimensions, condition (16) gives the Cauchy–Riemann equations. Of interest are the specific expressions of holomorphy for fields in four-dimensional spacetime. The tensors of type 1 and 2 in spacetime will give two distinct sets of differential equations of holomorphy. (The tensor type 3 is dual to the vector, and gives the same equations.)

The D derivative of the vector a is given by (15a). Setting the coefficients of each basis form in (1) equal to zero, we obtain the following set of seven equations as the holomorphy conditions of a vector field a:

$$(\nabla\cdot\mathbf{a}) + \partial^4 a^4 = 0, \tag{18a}$$

$$\partial^4 a^i + \partial^i a^4 = 0, \tag{18b}$$

$$(\nabla\times a)^i = 0, \tag{18c}$$

Via the electromagnetic analogy, (18a) is the Lorentz gauge condition for the vector potential, which is implied by the holomorphy. Equations (18b) and (18c) define the derivatives of the vector potential when the electromagnetic field is identically zero.[31]

Equations (18) are a generalization of the Cauchy–Riemann equations, as can be seen by considering a two-dimensional subspace of Minkowski spacetime. Take the $A^{0,2}$ subalgebra of $A^{1,3}$ generated by σ^1 and σ^2. The expressions of holomorphy for a two-dimensional vector $u = u^1\sigma^1 - u^2\sigma^2$ are the Cauchy-Riemann equations[1,2]

$$\partial^1 u^1 = \partial^2 u^2, \qquad \partial^2 u^1 = -\partial^1 u^2. \tag{19}$$

The sign results from the conjugation in the complex product which does not arise in the vee product.

What is of considerable physical interest is the fact that the expressions of holomorphy for a tensor of type 2 in Minkowski spacetime are precisely the Maxwell equations in vacuum. This is seen by using (15b) and (16), and setting the coefficients of each basis form equal to zero to obtain the set of eight equations

$$(\nabla\times B)^i - \partial^4 E^i = 0, \tag{20a}$$

$$(\nabla\cdot\mathbf{E}) = 0, \tag{20b}$$

$$(\nabla\times E)^i + \partial^4 B^i = 0, \tag{20c}$$

$$(\nabla\cdot\mathbf{B}) = 0, \qquad i = 1,2,3. \tag{20d}$$

These are the Maxwell equations in vacuum. In deriving this result, we have used nothing more than the holomorphic structure of Minkowski spacetime, formulated in terms of Clifford algebras and the generalized Dirac operator.

One can, of course, obtain Eq. (20) by requiring the two separate conditions (17); in this case, however, both the transition to integral holomorphy and the algebraic framework are lost.

Equations (20) reduce to the Cauchy–Riemann equations in two dimensions in exactly the same way as do Eqs. (18).

This can be seen as follows. In two dimensions, the electromagnetic field is either an electrostatic or magnetostatic field in the plane; $E = E^1\sigma^1 - E^2\sigma^2$ or $B = B^1\sigma^1 - B^2\sigma^2$. Equation (20) then becomes

$$\partial^1 B^2 + \partial^2 B^1 = 0, \qquad \partial^1 B^1 - \partial^2 B^2 = 0, \tag{21a}$$

or

$$\partial^1 E^1 - \partial^2 E^2 = 0, \qquad \partial^1 E^2 + \partial^2 E^1 = 0, \tag{21b}$$

which are the Cauchy–Riemann equations. (The sign again results from having to use the conjugate field).

We would like to clarify the fact that, whereas Maxwell's equations have been written in terms of Clifford algebras in the past,[32–34] the derivation given here as a result of holomorphy in spacetime is quite distinct. Also, while it is known that the static Maxwell equations for electrostatic fields in two dimensions can be cast in a form analogous to the Cauchy–Riemann equations,[35] we have demonstrated that one can generalize the Cauchy–Riemann equations to four-dimensional spacetime to obtain the full time-varying Maxwell equations.

Related but distinct formulations of the Maxwell equations in vacuum have been given in terms of analytic twistor fields by Penrose and his school,[21–23] in terms of spinor fields in Refs. 36 and 37, and in the context of the 3×3 matrix representation of the Lie algebra SO (3).[38,39] In the latter case, these matrices do not satisfy anticommutation relation required of any representation of a Clifford albegra; hence the algebraic basis is quite distinct.

This completes the discussion of fields which are strictly holomorphic. We proceed in the next section to include sources.

V. INTEGRAL EXPRESSIONS OF HOLOMORPHY: RESIDUES

It is possible to extend the discussion of holomorphy given in the preceding sections to show that certain results of potential theory in three and four dimensions are analogous to residue integration in the complex plane.

We seek to obtain integral expressions corresponding to the differential formulation of the holomorphy condition (16). In this, we are motivated by Theorem 2, Eq. (13), and the duality Theorem 1, Eq. (7). The novelty of expressing the dual as the vee product with the volume element ω^n, along with the fact that ω^n can be treated via its purely algebraic properties, leads to a key identification. We define integral forms corresponding to the Cartan exterior derivative and the Hodge coderivative as follows:

Definition 4: The integral forms corresponding to $d\wedge f$ and $\delta\wedge f$ are

$$\int d\wedge f=\oint f \tag{22a}$$

and

$$\omega^n\vee\int d\wedge\Big(\underset{n}{*}f\Big)=\omega^n\vee\oint\underset{n}{*}f. \tag{22b}$$

We have applied the Stokes theorem to express the integrals over closed hypersurfaces of the appropriate dimension.[30] As always in this discussion, f is a tensor type, i.e., a tensor field of homogeneous rank. The domain of integration is determined by the differential form basis of f, and will in general extend to an infinite domain. In actual practice, the integrals are evaluated in a finite domain, and then limits are taken, following the standard procedure of potential theory (see below).

It is now possible to give a definition of "integral holomorphy" as follows:

Definition 5: (a) A tensor type f is holomorphic iff the sum of integrals (22) goes to zero in the limit of an infinite domain of integration.

(b) f is nonholomorphic if the sum of integrals (22) is a constant volume, and

(c) f is divergent nonholomorphic if the sum of integrals (22) diverges in the limit of an infinite domain of integration.

If a tensor field f is holomorphic, then from (16), (17), and (22) the following conditions are true separately (the limit is implied):

$$\oint f=0\quad\text{and}\quad\oint\underset{n}{*}f=0. \tag{23}$$

The case of particular interest is when a tensor field f is nonholomorphic, i.e., possesses a singularity. This case can be examined by proceeding with the electromagnetic analogy.

Consider the electromagnetic field in the canonical decomposition, $F=\mathbf{E}\wedge\sigma^4-\underset{3}{*}\mathbf{B}$(8). In this case, the integrals (22a), (22b) are easily evaluated and are equal to the magnetic and electric flux, respectively:

$$\oint F=\oint(\mathbf{E}\wedge\sigma^4-\underset{3}{*}\mathbf{B})=-\oint B^i\,dS^i=0, \tag{24a}$$

$$\oint\underset{4}{*}F=\oint(\mathbf{B}\wedge\sigma^4+\underset{3}{*}\mathbf{E})=\oint E^i\,dS^i=\Phi_e. \tag{24b}$$

The electric flux equals the sum of the charges; therefore, integral (24b) can be written as

$$\oint\underset{4}{*}F=4\pi\sum_\alpha q_\alpha. \tag{25}$$

We now apply Theorem 5 to the electromagnetic field F to obtain the following expression from (22), (24), and (25):

$$\int D\vee F\leftrightarrow\oint F+\omega^4\vee\oint\underset{4}{*}F=4\pi\omega^4\sum_\alpha q_\alpha. \tag{26}$$

Hence the electromagnetic field is in general nonholomorphic.

We now show that expression (26) is analogous to the usual residue theorem in two dimensions. In that case, the field F is either a pure electric, or pure magnetic field in the plane, i.e.,

$$F=E^1\sigma^1+E^2\sigma^2\quad\text{or}\quad F=B^1\sigma^1+B^2\sigma^2. \tag{27}$$

Since our construction is valid in any space, the expression corresponding to (26) in two dimensions is just

$$\oint F+\omega^2\vee\oint\underset{2}{*}F=4\pi\omega^2\sum_\alpha q_\alpha. \tag{28}$$

One can naturally indentify ω^2, the unit pseudoscalar in two dimensions with the complex unit $i=\surd-1$. Note that $\omega^2\vee\omega^2=-1$; the commutation properties are here irrelevant. After explicitly performing the two-dimensional dual of (27), we have from (28)

$$\oint(E^1\,dx^1+E^2\,dx^2)+i\oint(E^2\,dx^1-E^1\,dx^2)=4\pi i\sum_\alpha q_\alpha. \tag{29}$$

This expression is identical to the residue theorem. By combining the two integrals, we can express (29) as a single integral using complex multiplication. The complex electric field is $E=E^1+iE^2$, $dz=dx^1+i\,dx^2$; the tilde denotes complex conjugation, and we have the residue theorem:

$$\oint\tilde{E}\,dz=4\pi i\sum_\alpha q_\alpha. \tag{30}$$

This analysis demonstrates that, by considering the Dirac derivative as a vector operator, one obtains a correspondence with holomorphic functions in the complex plane. The charges correspond to the residues. The factor of 2 arises from the special geometry involved, and is discussed below. In the analogous case of a pure magnetic field in two dimensions, the integral (28) vanishes, and gives Cauchy's theorem for the holomorphic magnetic field:

$$\oint \widetilde{B}\, dz = 0. \tag{31}$$

We note that, in the special geometry involved, the electric field is due to infinite long wires perpendicular to the plane, and is equal to $E = 2\lambda / r$. Here, λ is the linear charge density which corresponds to a two-dimensional charge. The residue of the electric field at the origin is equal to 2λ, since $\mathrm{Res}_{z=0}(1/z) = 1$. This is the factor of 2 that appears in Eq. (30), in addition to the 2π of the residue theorem.

This completes the definition of integral holomorphy in n dimensions, and its relation to the residue theory in the complex plane.

VI. HOLOMORPHIC FIELDS IN THREE DIMENSIONS:MULTIPOLES

The questions one now asks is, what are the holomorphic functions? We first examine functions in the three-dimensional subspace of Minkowski space. The basic functions are polynomials of the radius vector $\mathbf{r}$, which can then be used to build any other field through a three-dimensional Fourier expansion. The scalar fields are

$$\ln|\mathbf{r}|, \tag{32a}$$

$$|\mathbf{r}|^n, \quad n = 0, \pm 1, \pm 2, \cdots, \tag{32b}$$

$$(\mathbf{k}\cdot\mathbf{r})|\mathbf{r}|^n, \quad \mathbf{k} = \text{constant vector}. \tag{32c}$$

The vector fields are

$$\mathbf{r}|\mathbf{r}|^n, \tag{33}$$

and the tensor fields are

$$\underset{3}{*}\mathbf{r}|\mathbf{r}|^n. \tag{34}$$

From these expressions, we can construct any tensor field in three dimensions.

By substituting the vector field (33) in the integral forms (22), we obtain direct expressions. Integral (22a) is evaluated on the perimeter of a circle of radius R around the origin:

$$\oint \mathbf{r}|\mathbf{r}|^n = 0. \tag{35a}$$

Integral (22b) is evaluated on the surface of a sphere of radius R around the origin:

$$\oint \underset{3}{*}\mathbf{r}|\mathbf{r}|^n = 4\pi R^{\,n+3}. \tag{35b}$$

This expression (35b) diverges when the radius R is taken to infinity for $n > -3$. It is equal to 4π for $n = -3$, and goes to 0 as $R \to \infty$ for $n < -3$. We can apply Definition 5 to determine the cases when the field (33) is holomorphic. From (22) and (35) we have the formula

$$\oint \mathbf{r}|\mathbf{r}|^n + \omega^3 \vee \oint \underset{3}{*}\mathbf{r}|\mathbf{r}|^n = 4\pi\omega^3 R^{\,n+3}. \tag{36}$$

Applying the prescription of Theorem 5, we see that of the vector fields (33), the holomorphic ones are for $n < -3$; the divergent ones are for $n > -3$; and the nonholomorphic field corresponds to $n = -3$. In that case, we have the analog to the residue theorem in three dimensions as

$$\oint \frac{\mathbf{r}}{|\mathbf{r}|^3} + \omega^3 \vee \oint \underset{3}{*} \frac{\mathbf{r}}{|\mathbf{r}|^3} = 4\pi\omega^3. \tag{37}$$

Returning to the differential definition of holomorphy, we can obtain the analogous expressions using (14) on (33). These correspond to Definition 3, Eq. (16) in the limit $r \to \infty$:

$$D \vee \mathbf{r}|\mathbf{r}|^n = -\nabla \vee \mathbf{r}|\mathbf{r}|^n$$

$$= (3+n)|\mathbf{r}|^n, \quad \begin{cases} n = 0, \pm 1, \pm 2, \cdots, \\ n \neq -3, \end{cases} \tag{38a}$$

$$= 4\pi\delta(\mathbf{r}), \quad n = -3. \tag{38b}$$

We see therefore the existence of the singularity for $n = -3$, the coefficient 4π, which is the surface area of the unit sphere S^3, and the three dimensional residue at the point $\mathbf{r} = 0$, which is equal to 1. This case corresponds to the electric field of a point charge.

A deeper connection to electromagnetism is established by considering the potential φ of an arbitrary charge distribution, expanded as a series of 2^n-pole scalar potentials φ_n:

$$\varphi = \sum_{n=0}^{\infty} \varphi_n. \tag{39}$$

Each term in the expansion is given by

$$\varphi_n = \frac{1}{n!}(D_n, R_n) = \frac{1}{n!} \sum_{i_1,\ldots,i_n=1}^{3} D_n^{\,i_1,\ldots,i_n} R_n^{\,i_1,\ldots,i_n}, \tag{40}$$

where D_n is the 2^n-pole moment of the charge distribution defined by the symmetric tensor

$$D_n = \int \mathbf{r}' \otimes \cdots \otimes \mathbf{r}'\rho[\mathbf{r}'](\omega^3)' \quad (n \text{ times}), \tag{41}$$

and $R_n[\mathbf{r}]$ is the spatial multipole tensor which gives the space dependence of the 2^n pole:

$$R_n^{\,i_1,\ldots,i_n} = \partial_{i_n} \cdots \partial_{i_n} \frac{1}{|\mathbf{r}|}. \tag{42}$$

We note that $R_n[\mathbf{r}]$ is a function of $1/|\mathbf{r}|^{n+1}$; hence the 2^n-pole potential varies as $1/|\mathbf{r}|^{n+1}$.

The electric field of each 2^n pole is obtained as the gradient of expression (40):

$$\mathbf{E}_n = -\nabla\varphi_n. \tag{43}$$

Substituting the electric field (43) into the holomorphy integrals (22), we obtain via the Stokes theorem,

$$\oint \mathbf{E}_n = \int \nabla \wedge \mathbf{E}_n = \int \nabla \wedge \nabla\varphi_n = 0, \tag{44a}$$

$$\oint \underset{3}{*}\mathbf{E}_n = \int \underset{3}{*}(\nabla\cdot\mathbf{E}_n) = \int \underset{3}{*}\nabla^2\varphi_n = \int \nabla^2\varphi_n\omega^3. \tag{44b}$$

We proceed to calculate the Laplacian of the potential

φ_n, using (40):

$$\nabla^2\varphi_n = \frac{1}{n!}\sum_{i_1,\ldots,i_n=1}^{3} D_n^{\,i_1,\ldots,i_n}\nabla^2\partial_{i_1}\cdots\partial_{i_n}\frac{1}{|\mathbf{r}|}. \tag{45}$$

We can exchange the order of differentiation to obtain the expression

$$\nabla^2\partial_{i_1}\cdots\partial_{i_n}\frac{1}{|\mathbf{r}|} = \partial_{i_1}\cdots\partial_{i_n}\nabla^2\frac{1}{|\mathbf{r}|} = -4\pi\partial_{i_1}\cdots\partial_{i_n}\delta(\mathbf{r}). \tag{46}$$

$$\oint_3 {*}\mathbf{E}_n = -\frac{4\pi}{n!}\sum_{i_1,\ldots,i_n=1} D_n^{\,i_1,\ldots,i_n}\int\partial_{i_1}\cdots\partial_{i_n}\delta(\mathbf{r})\omega^3. \tag{47}$$

Using a formula from the theory of generalized functions,[40] for any polynomial $P[x^1,x^2,x^3]$ we have the identity

$$\int f(\mathbf{x})P[\partial_1,\partial_2,\partial_3]\delta(\mathbf{x}-\mathbf{a})\omega^3 = P[-\partial_1,-\partial_2,-\partial_3]f(\mathbf{x})]_{\mathbf{x}=\mathbf{a}}. \tag{48}$$

In the case of interest, we have $f(\mathbf{x}) = 1$; hence the integral (47) is identically zero, except for the case $n = 0$, i.e., the point charge. We therefore have

$$\oint_3 {*}\mathbf{E}_n = \begin{cases} 4\pi e, & n = 0, \\ 0, & n > 0. \end{cases} \tag{49}$$

Equivalently, we can write this as the residue theorem (26) in three dimensions:

$$\oint \mathbf{E}_n + \omega^3 \vee \oint_3 {*}\mathbf{E}_n = \begin{cases} 4\pi\omega^3 e, & n = 0, \\ 0, & n > 0. \end{cases} \tag{50}$$

In the same manner, it can be shown that all the magnetic multipole fields are holomorphic.

We have shown in this framework how all the higher-order multipole fields are holomorphic. The only nonholomorphic field is that due to the point charge. A point worth noting is that the holomorphy of the vector fields $\mathbf{r}|\mathbf{r}|^n$, (33) requires the discussion of the limit to an infinite domain. In contradistinction, the multipole fields $\mathbf{E}_n$ (43) are in exact agreement with the differential conditions of holomorphy (16), since this condition in fact implies Maxwell's equations.

This analysis illustrates the intrinsic connection between electromagnetism and holomorphic fields in three dimensions.

VII. CONCLUSION

In this paper, we have indicated how the theory of holomorphic functions in the complex plane can be extended to spaces of any dimension.

In conclusion, we recall the main points of this paper.

First, we have shown how one can realize the generalized Dirac operator in a calculationally useful algebraic setting. Using this operator, we introduced a definition of holomorphy for fields in any dimension, and showed that the differential expressions of holomorphy in Minkowski spacetime are precisely the Maxwell equations. Second, we have obtained an integral expression for holomorphy in any dimension, and shown how it corresponds to residue integration and the Cauchy theorem when the dimension of the space is equal to 2. Third, we have determined the holomorphic fields in three-dimensional space, and have found that the nonholomorphic field is in fact that corresponding to the electric field of a point charge.

These results underline the connection between electromagnetism and the structure of spacetime. We have recast, combined, and extended certain results from potential theory in an essential manner in order to provide this construction.

Our analysis has been based on the following key points:

(i) We have used the tensor fields of a given rank which correspond to physical fields. This is in contrast to other work in this area, where one usually considers the most general element of the algebra, i.e., a linear combination of all rank tensor fields. This geometrical distinction has a profound consequence on our results. This same distinction was utilized in the definition of inverses in a Clifford algebra. We recall that a linear combination of tensor fields may not always have a inverse, while a specific tensor field always has.[15,16]

(ii) The identification of the differential forms as the bases of the algebra is crucial in the construction of the generalized Dirac operator as a vector operator in the Clifford algebra. This feature also enables us to utilize the entire apparatus of the exterior calculus in our discussion. Our use of a geometrical basis has as a consequence that most of our results are independent of the analytic properties of the individual field components; they follow from the underlying geometric structure.

We believe that this paper has illustrated how electromagnetism is intrinsically related to the holomorphic properties of fields in Minkowski spacetime.

[1]L. Ahlfors, *Complex Analysis*, 2nd ed. (McGraw-Hill, New York, 1966).
[2]R. Nevanlinna and V. Paatero, *Introduction to Complex Analysis* (Addison-Wesley, Reading, Mass., 1969).
[3]V. S. Vladimirov, *Methods in the Theory of Functions of Several Complex Variables* (M.I.T. Press, Cambridge, Mass., 1966).
[4]L. Hormander, *An Introduction to Complex Analysis in Several Variables*, 2nd ed. (North-Holland, Amsterdam, 1973).
[5]F. G. Frobenius, J. Reine Angew. Math. **84**,59 (1877).
[6]A. G. Kurosh, *Lectures on General Algebra* (Chelsea, New York, 1963).
[7]A. Hurwitz, Nach. Ges. Wiss. Göttingen, 309 (1898).
[8]A. A. Albert, Ann. Math. **43**, 161 (1942).
[9]A. A. Albert, Bull. Amer. Math. Soc. **55**, 763 (1949).
[10]C. W. Curtis, in *Studies in Modern Algebra*, edited by A. A. Albert (Prentice-Hall, Englewood Cliffs, N.J., 1963).
[11]R. Fueter, Comment. Math. Helv. **8**, 371 (1935).
[12]F. Gürsey and H. Chia Tze, Ann. Phys. (N. Y.) **128**, 29 (1980).
[13]Y. Ilamed and N. Salingaros, "Algebras with Three Anticommuting Elements I," J. Math. Phys. **22**, 2091 (1981).
[14]N. Salingaros, "Algebras with Three Anticommuting Elements II," J. Math. Phys. **22**, 2096 (1981).
[15]N. Salingaros, in "Proceedings of the Second Workshop on Lie-Admissible Formulations," Harvard University, 1979, Hadronic J. **3**, 339 (1979).
[16]N. Salingaros, J. Math. Phys. **22**, 226 (1981).
[17]D. Hestenes, J. Math. Anal. Appl. **24**, 313 (1968).
[18]D. Hestenes, J. Math. Anal. Appl. **24**, 467 (1968).
[19]R. Delanghe, Math. Ann. **185**, 91 (1970).
[20]R. Delanghe, Math. Ann. **196**, 293 (1972).
[21]R. Penrose, in *Quantum Gravity: Oxford Symposium*, edited by C. J. Isham, R. Penrose, and D. W. Sciama (Clarendon, Oxford, 1975).
[22]L. P. Hughston and R. S. Ward (Eds.), *Advances in Twistor Theory*, Research Notes in Math No. 37 (Pitman, London, 1979).
[23]M. G. Eastwood, R. Penrose, and R. O. Wells, Jr., Commun. Math. Phys. **78**, 305 (1981).

[24]P. A. M. Dirac, Proc. R. Soc. London Ser. A **117**, 610 (1928).
[25]M. F. Atiyah and I. M. Singer, Ann. Math. **87**, 546 (1968).
[26]F. Hirzebruch, *Topological Methods in Algebraic Geometry* (Springer, New York, 1966).
[27]N. Salingaros and M. Dresden, Phys. Rev. Lett. **43**, 1 (1979).
[28]N. Salingaros, in "Proceedings of the Third Workshop on Lie-Admissable Formulations," University of Massachusetts at Boston, 1980, Hadronic J. **4**, 949 (1981).
[29]N. Salingaros and M. Dresden, "An Algebra of Fields in Minkowski Spacetime" (to appear).
[30]H. Flanders, *Differential Forms* (Aacademic, New York, 1963).
[31]L. D. Landau and E. M. Lifshitz, *The Classical Theory of Fields* (Pergamon, Oxford, 1971).
[32]D. Hestenes, *Space-time Algebra* (Gordon and Breach, New York, 1966).
[33]A. Mercier, "Expression des Equations de l'Electromagnetisme au Moyen des Nombres de Clifford," thesis, University of Geneva, 1935 (unpublished).
[34]M. Riesz, *Clifford Numbers and Spinors*, Institute for Fluid Dynamics and Appl. Math. No. 38 (University of Maryland Press, College Park, Md., 1958).
[35]P. Morse and H. Feshbach, *Methods of Theoretical Physics* (McGraw-Hill, New York, 1953), Vol. 1.
[36]O. Laporte and G. E. Uhlenbeck, Phys. Rev. **37**, 1380 (1931).
[37]J. R. Oppenheimer, Phys. Rev. **38**, 725 (1931).
[38]E. Majorana, unpublished work; see R. Mignani, E. Recami, M. Baldo, Lett. Nuovo Cimento **11**, 568 (1974).
[39]R. H. Good, Jr., Phys. Rev. **105**, 1914 (1957).
[40]E. M. deJager, in *Mathematics Applied to Physics*, edited by E. Roubine (Springer-Verlag, Berlin, 1970).

A Hamiltonian structure of the interacting gravitational and matter fields

Wiktor Szczyrba [a)]
Institute for Theoretical Physics, University of Cologne, Zülpicherstrasse 77, 5000 Cologne, West Germany

(Received 7 November 1979; accepted for publication 20 June 1980)

We present a Hamiltonian formulation for classical field theories. In a general case we write the Hamilton equation by means of the energy–momentum function E and the symplectic 2-form Ω. We investigate thoroughly an important example, the gravitational field coupled to a matter tensor field. It will be shown that the energy-momentum differential 3-form yields a generalization of the Komar energy formula. We prove that the energy–momentum function E, the symplectic 2-form Ω, the Hamilton equation, and four constraint equations for initial values of canonical variables give rise to the system which is equivalent to the Euler–Lagrange variational equations. We also discuss relations between the Hamilton equation of evolution, the degeneracy of the symplectic 2-form Ω, and the action of the diffeomorphism group of spacetime in the set of solutions.

PACS numbers: 03.50.Kk

I. INTRODUCTION

The classical approach to the field equations in physical theories is based on variational principles and on the Einstein principle of general covariance. This combination gives rise to the geometric formulation of the field equations for a comprehensive class of field theories (scalar field, electrodynamics, Proca field, gravitation, Yang–Mills fields). In the 4-covariant picture the space of solutions of the field equations is a subspace of the space of sections of some bundle over spacetime M. This approach is static; a state (a solution of the field equations) represents the entire history of the system under consideration (the Heisenberg picture). We shall show that the space of solutions—Sol—is endowed with a closed differential 2-form Ω (in general degenerate). The static 4-covariant approach is not convenient for the discussion of the initial value problem. Therefore it is interesting to formulate a given classical field theory in terms of the space (Id) of initial data and their evolutions [curves in (Id)]. This picture corresponds to the Hamiltonian form of mechanics (the Schrödinger picture in quantum mechanics). In this paper we present a Hamiltonian formulation for classical field theories and prove that the evolution of an initial data $f\in(\mathrm{Id})$ is generated by the action of the group of diffeomorphisms of spacetime M. Therefore our construction suits those theories which are invariant with respect to the action of Diff M (e.g., the gravitational field coupled to a matter field).

The general scheme of this paper is the following. Let τ: $\mathscr{P}\to M$ be a bundle over spacetime M and $\Theta_{\text{H–C}}$ be a differential 4-form on $\mathscr{P}$. $(\mathscr{P}, \Theta_{\text{H–C}})$ is the multisymplectic bundle of a classical field theory in the sense of Refs. 1–4. The space of solutions—Sol—consists of four-dimensional submanifolds of $\mathscr{P}$ (images of sections of $\mathscr{P}$) which satisfy the field equations [cf. (2.1)]. We assume, additionally, that Diff M acts in $\mathscr{P}$ and that $\Theta_{\text{H–C}}$ is invariant with respect to this action. For a fixed three-dimensional surface σ in M (Cauchy surface) we define the restriction $\mathscr{P}(\sigma)$ of $\mathscr{P}$ to σ. The space of initial data (Id) (σ) consists of these sections of $\mathscr{P}(\sigma)$ which can be extended to sections of $\mathscr{P}$ satisfying field equations. The space (Id) (σ) is equipped with a symplectic 2-form $\Omega(\sigma)$ and an energy–momentum function $E(\sigma)$. These quantities satisfy the Hamilton equation

$$dE(\sigma)\,'\hat{V} = -\Omega(\sigma)('\hat{Y}\wedge '\hat{V}), \tag{1.1}$$

where $'\hat{Y}\in T((\mathrm{Id})(\sigma))$ is the evolution vector (Hamiltonian vector) and $'\hat{V}$ is an arbitrary vector tangent to the space of sections of $\mathscr{P}(\sigma)$.

If $f\in(\mathrm{Id})(\sigma)$ and $'Y$ is the vector field on $f(\sigma)\subset\mathscr{P}(\sigma)$ representing the vector of evolution $'\hat{Y}$, then the evolution $t\to f_t$ of $f=f_0$ is given by equation

$$\frac{df_t}{dt}(\mathbf{x}) = 'Y(f_t(\mathbf{x})), \quad \mathbf{x}\in\sigma. \tag{1.2}$$

We see from (1.1)–(1.2) that the energy–momentum function E and the symplectic 2-form Ω generate the dynamics in the same way as the Hamiltonian $H = H(p_i, q^j)$ and the symplectic 2-form $\omega = dp_i\wedge dq^i$ generate the canonical equations of classical mechanics

$$\frac{dp_i}{dt} = -\frac{\partial H}{\partial q^i},\quad \frac{dq^i}{dt} = \frac{\partial H}{\partial p_i}\quad \text{(cf. Refs. 5, 6)}$$

$$\left(\text{In this case } 'Y = \frac{dq^i}{dt}\frac{\partial}{\partial q^i} + \frac{dp_i}{dt}\frac{\partial}{\partial p_i}.\right)$$

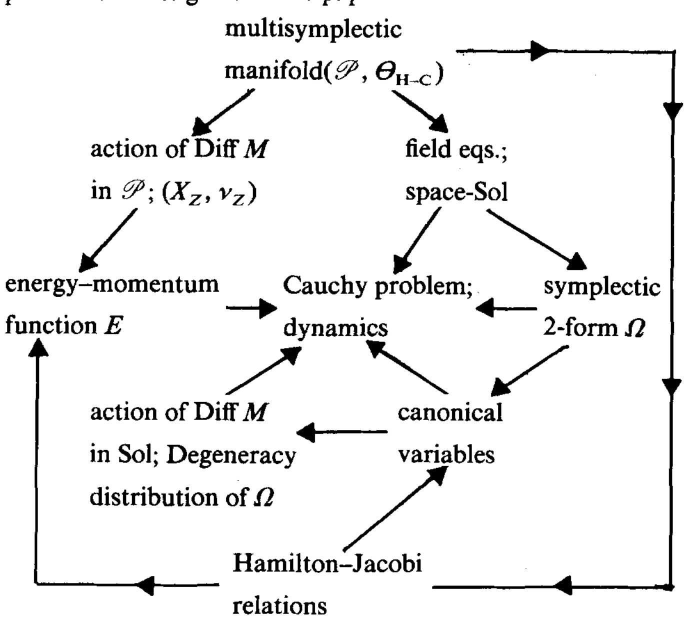
We give a geometric construction of the energy–momentum function E for an arbitrary classical field theory based on a multisymplectic structure $(\mathscr{P}, \Theta_{\text{H–C}})$. Our construction applied to the gravitational field gives rise to a generalization of the Komar energy formula[7] (cf. also Appendix D and Sec. 8). The part of the paper devoted to the Hamiltonian dynamics presents ideas which are close to those of Fischer and Marsden,[8,9] who, in the seventies, reformulated the classical results of Arnowitt, Deser, and Misner[10] concerning the dynamics of the Einstein theory of relativity. In our more general case the connection on spacetime is not Riemannian. Therefore we need two types of equations for the gravitational field, one system of which describes the

[a)]Permanent address: Institute of Mathematics, Polish Academy of Sciences, ul. Sniadeckich 8, 00-050 Warsaw, Poland. Alexander von Humboldt Fellow.

 0022-2488/81/091926-18$01.00

evolution of the metric and the second, the evolution of the connection. We get these equations by means of the variational principle of the Palatini type; that is, variations with respect to the metric and the connection are independent. Such a procedure is now widely accepted in theories of gravity. We are, however, able to get the inverse result: We can recover from (1.1) the variational gravitational equations and also the Euler–Lagrange matter field equations. We postulate the energy–momentum function E, the symplectic 2-form Ω and try to solve Eq. (1.1) with respect to an unknown vector of evolution $'\hat{Y}$. The essential difficulty is that the symplectic 2-form is degenerate and a solution $'\hat{Y}$, if it exists, is not unique. We solve this problem in the following way: let Sbm (4) be the space of all four-dimensional (sufficiently smooth) submanifolds of $\mathscr{P}$ (which are images of sections of $\mathscr{P}$) and let Sbm_σ (3) be the corresponding set of three-dimensional submanifolds in $\mathscr{P}(\sigma)$. We define E and Ω on Sbm (4) [and on Sbm_σ (3)], investigate Eq. (1.1) for arbitrary (sample) vectors $'\hat{V}\in T(\mathrm{Sbm}_\sigma\ (3))$ and show that it has a unique solution $'\hat{Y}\in T((\mathrm{Id})(\sigma))$. As one could expect, knowing E and Ω is not sufficient to determine $'\hat{Y}$. We have also to assume the constraint equations for initial values of the canonical variables. [We shall prove in Sec. 5 that the constraint equations (5.1) give the necessary and sufficient conditions for the solvability of (1.1).]

The following diagram shows relations among several problems investigated in the paper:



Two fundamental geometrical objects appear in the scheme: the energy–momentum function E and the symplectic 2-form Ω. The symplectic 2-form Ω for an arbitrary classical field theory based on a variational principle (multisymplectic bundle) has been constructed by Kijowski and the present author[3] (cf. also Ref 4). The diagonalization procedure for Ω and the geometrical definition of the canonical variables in theories of gravity were presented for the Einstein and Einstein–Maxwell theories[11–13] and for a generalized theory of gravity with the presence of a tensor matter field.[14] In the present paper we use a stronger version of the results given in Ref. 14 (cf. Proposition 4). The geometric definitions of the canonical variables, the Belinfante–Rosenfeld identities, the contracted Bianchi identities, and some other technical results of[14,15] are used throughout the paper. We also give the definition of the canonical variables by means of variations of the action integral (the Hamilton–Jacobi relations) (cf. Ref. 16 and Sec. 7).

All constructions in the paper are performed under the assumption that topologically M is the product of the real line $\mathbb{R}$ and a compact, three-dimensional manifold σ (without boundary). Therefore we can neglect all integrals of three divergencies on σ. If σ is a noncompact manifold, some surface integrals can give nontrivial contributions. We discuss this problem briefly in Sec. 8.

There are several papers in the literature devoted to the Hamiltonian (canonical) formulation of gravity (especially the Einstein theory). The most popular among physicists is the approach in the language of "constrained Hamiltonian systems" and the Dirac brackets; cf. papers by Dirac,[17] Bergmann,[18] Arnowitt–Deser–Misner,[10] Faddeev,[19] Kuchař,[20] Hojman–Kuchař–Teitelboim,[21] Hanson–Regge–Teitelboim,[22] Nelson–Teitelboim,[23,24] Pilati,[25] Nester–Isenberg,[26–28] and Murchadha–York[29,30] (see also the papers by Sniatycki,[31] Tulczyjew,[32] and Gotay–Nester–Hinds[33] devoted to the mathematics of the Dirac theory of constraints). As we have already mentioned above, there is the series of papers by Fischer–Marsden, Moncrief, and Arms,[8,9,34–36] who treat the Einstein equations, the coupled Einstein–Maxwell, and Einstein–Yang–Mills equations, respectively, as infinite-dimensional Hamiltonian systems (cf. also Ref. 5 and the review paper by Francaviglia[37]). These authors investigate such problems as the Cauchy problem, the linearization stability, the structure of the manifold of solutions, and singularities of this manifold. The third, recently developed approach is given in the interesting book by Kijowski and Tulczyjew[38] (see also Ref. 39).

We present here an alternative theory which has certainly many points in common with all the above-mentioned papers but tends rather towards the Fischer–Marsden approach. Some questions about relations between our theory and other papers are discussed in Sec. 8.

2. THE SYMPLECTIC 2-FORM Ω AND THE ENERGY-MOMENTUM FUNCTION E

For a given classical field theory the basic notions are the field potentials which are sections of some geometric bundle over spacetime M. The dynamics is determined by a Lagrange function L and the Euler–Lagrange equations of the variational problem for L. The Lagrangian L depends on values of the field potentials and their first partial derivatives (theories with derivatives of higher orders are not discussed here). In the geometric approach to the calculus of variations, solutions of a variational problem are sections of some bundle τ: $\mathscr{P}\to M$ over spacetime M. A point of the fibre of $\mathscr{P}$ at $x\in M$ determines values of the field potentials and their first partial derivatives. By means of the Lagrangian L we construct a differential 4-form $\Theta_{\text{H-C}}$ on $\mathscr{P}$ (the Hamilton–Cartan 4-form) and formulate the variational principle in terms of $\Theta_{\text{H-C}}$. Detailed descriptions of such a construction were given in papers by Dedecker,[40] Goldschmidt-Sternberg,[2] Kijowski–Szczyrba,[1,3,41,12] Garciá,[42] and Kijowski–Tulczyjew.[38] The Euler–Lagrange equations can be written in the following geometrical formulation:

A section $F:M\to\mathscr{P}$ is a solution of the E–L equations if and only if for every (τ-vertical) vector field X, tangent to $\mathscr{P}$ and defined at points of the submanifold $C_4 = F(M) \subset \mathscr{P}$, the condition

$$F^*(X \lrcorner d\Theta_{\text{H-C}}) = 0, \tag{2.1}$$

or equivalently

$$(X \lrcorner d\Theta_{\text{H-C}})\,|\,C_4 = 0, \tag{2.1'}$$

holds.

Remark: The symbols $\lrcorner$, d denote the interior product (contraction) between vectors and forms and the exterior derivative, respectively. F^* denote the pullback operation for differential forms generated by F and $(X \lrcorner d\,\Theta_{\text{H-C}})|C_4$ is the pullback of the 4-form $X \lrcorner d\Theta_{\text{H-C}}$ onto the submanifold $C_4 \subset \mathscr{P}$ (cf. Refs. 5, 6).

Four-dimensional submanifolds $C_4 = F(M) \subset \mathscr{P}$ which satisfy (2.1) form the space of solutions Sol. The space Sol is a subspace in the space Sbm(4) of all (sufficiently smooth) four-dimensional submanifolds of $\mathscr{P}$ which are the images of sections of τ. The space Sbm(4) carries the natural structure of an infinite-dimensional manifold (either Banach or Fréchet–Schwartz cf. Refs. 43, 44). The space of solutions Sol is not a manifold for generic classical field theories, however. It has singular points, that is, such points that no neighborhood of them can be parametrized by means of vectors of the tangent space. For the Einstein theory of gravity this problem was investigated by Fischer, Marsden, Moncrief, and Arms.[35,36,45,46] However, a weaker structure is sufficient for our considerations. We endow the spaces Sbm(4) and Sol with a kind of pseudodifferential structure.[3,4,12] The most important definition we need in the present paper is that of a tangent vector. A vector $\widehat{Y}$ tangent to Sbm(4) at $C_4 \in \text{Sbm}(4)$ is represented by a (smooth) τ-vertical vector field Y tangent to $\mathscr{P}$ and defined at points of C_4. A vector $\widehat{Y}$ tangent to Sol at $C_4 \in \text{Sol}$ is represented by a τ-vertical vector field Y tangent to $\mathscr{P}$, defined at points of C_4, which satisfies the linearized version of field equation (2.1) (cf. Refs. 3, 12 and Appendix A). We denote the tangent bundles of Sbm(4), Sol by T(Sbm(4)) and T(Sol), respectively. In the framework presented in Refs. 3, 12 the notions of vector fields, differential forms, and their exterior derivatives can be defined in a natural way.

Remark: Throughout this paper we consider smooth (i.e., C^∞) sections of bundles, smooth vector fields, etc. However, we can choose another topology in these spaces—e.g., C^k (k-times differentiability) or H^s (Sobolev spaces). The precise choice of the topology is necessary if one intends to investigate the structure of the set of solutions deeper (cf. Refs. 9, 46).

In Refs. 1–3, 11–15, 41, 42, and 47 several examples of classical field theories, their multisymplectic bundles and the Hamilton-Cartan 4-forms were given, including the Einstein theory of gravity and its generalizations.

The symplectic 2-form Ω on the space Sol is defined in the following way.[3,4,12] Let $C_4 \in \text{Sol}$, $\widehat{Y}_1$, $\widehat{Y}_2 \in T_{C_4}(\text{Sol})$ and Y_1, Y_2 be τ-vertical vector fields defined on C_4 which represent $\widehat{Y}_1$, $\widehat{Y}_2$, respectively. Let $\sigma \subset M$ be a three-dimensional surface in M such that $M = \mathbb{R} \times \sigma$. We define

$$\begin{aligned}\Omega(\sigma)(\widehat{Y}_1, \widehat{Y}_2) &= \tfrac{1}{2}\int_{F(\sigma)} (Y_1 \wedge Y_2) \lrcorner d\Theta_{\text{H-C}} \\ &= \tfrac{1}{2}\int_{F(\sigma)} (Y_2 \lrcorner Y_1 \lrcorner d\Theta_{\text{H-C}}),\end{aligned} \tag{2.2}$$

where F is a section of $\mathscr{P}$ such that $C_4 = F(M)$. It was proved in Ref. 3 that the integral in (2.2) does not depend on the choice of σ in M and therefore (2.2) defines a differential 2-form Ω on Sol.

Remark: We must, of course, single out a class of three-dimensional surfaces σ in M for which the integral (2.2) is to be computed. One possible, reasonable set of axioms was presented in Ref. 1. If we discuss a particular theory, then the class of admissible surfaces σ in M has to be contained in the set of Cauchy surfaces, i.e., in the set of such three-dimensional surfaces in M for which the initial value problem is well posed.

We can generalize the definition of $\Omega(\sigma)$ for $C_4 \in \text{Sbm}(4)$, $\widehat{Y}_1$, $\widehat{Y}_2 \in T_{C_4}(\text{Sbm}(4))$. However, in this case the value of the integral depends essentially on the choice of $\sigma \subset M$ and we have the family $(\Omega(\sigma))_{\sigma\subset M}$ of differential 2-forms on Sbm(4). The form Ω on Sol is closed, i.e., $d\Omega = 0$ (cf. Refs. 3, 12). We are able to prove that all the forms $\Omega(\sigma)$ on Sbm(4) are also closed.

An essential feature of theories of gravity is that they are invariant with respect to the action of the diffeomorphism group of spacetime—that is, we have an action A in the bundle $\mathscr{P}$

$$(\text{Diff}\,M, \mathscr{P}) \ni (\Phi, p) \to A(\Phi)(p) \in \mathscr{P}, \tag{2.3}$$

and that for every $\Phi \in \text{Diff}\,M$

$$A^*(\Phi)\Theta_{\text{H-C}} = \Theta_{\text{H-C}} \tag{2.4}$$

It follows from (2.4) that the action A generates an action $\widehat{A}$ in the space Sol (see Sec. 6).

The action (2.3) induces the mapping

$$C^\infty(TM) \ni Z \to dA\,(\text{id})Z \in C^\infty(T\mathscr{P}) \tag{2.5}$$

where dA is the derivative of (2.3) with respect to the first variable, id is the identity in Diff M, and the vector field Z on M is generated by a one-parameter family of diffeomorphisms

$$\begin{aligned}&(\Phi_t)_{t\in\mathbb{R}}, \quad \Phi_0 = \text{id}, \\ &Z(\mathbf{x}) = \frac{d}{dt}\Phi_t(\mathbf{x})\Big|_{t=0}, \quad \mathbf{x}\in M.\end{aligned} \tag{2.6}$$

We have a globally defined vector field on $\mathscr{P}$

$$X_Z = dA\,(\text{id})Z. \tag{2.7}$$

For every section F of $\mathscr{P}$ the family $(\Phi_t)_{t\in\mathbb{R}}$ gives us a one-parameter family $(F_t)_{t\in\mathbb{R}}$ of sections of $\mathscr{P}$

$$\begin{aligned}&F_t(\mathbf{x}) = A(\Phi_t^{-1})F(\Phi_t(\mathbf{x})), \\ &\mathbf{x}\in M, \quad F_0 = F.\end{aligned} \tag{2.8}$$

Let

$$Y(F(\mathbf{x})) = \frac{d}{dt}F_t(\mathbf{x})\Big|_{t=0}, \quad \mathbf{x}\in M. \tag{2.9}$$

Y is a τ-vertical vector field defined on $C_4 = F(M)$ and tangent to $\mathscr{P}$. If the section F satisfies field equation (2.1), then the field Y represents a vector $\widehat{Y}\in T_{C_4}(\mathrm{Sol})$.

Proposition 1: If X_Z is the vector field on $\mathscr{P}$ defined by (2.7), $C_4 = F(M)$, and Y is defined by (2.9), then

$$X_Z(F(\mathbf{x})) = dF\cdot Z(\mathbf{x}) - Y(F(\mathbf{x})), \quad \mathbf{x}\in M. \tag{2.10}$$

Definition: The Hamilton 3-form ν_Z on $\mathscr{P}$ (corresponding to the vector field Z) is given by

$$\nu_Z = -X_Z \rfloor \Theta_{\mathrm{H\text{-}C}} \tag{2.11}$$

Definition: The energy–momentum function E_Z on Sol is given by

$$E_Z(C_4) = \int_{F(\sigma)} \nu_Z, \quad C_4\in\mathrm{Sol}, \tag{2.12}$$

where $\sigma\subset M$ is a three-dimensional submanifold (surface) in M. (We recall that σ is always a compact manifold without boundary.)

By virtue of (2.4) we have

$$\mathscr{L}_{X_Z}\Theta_{\mathrm{H\text{-}C}} = d(X_Z\rfloor\Theta_{\mathrm{H\text{-}C}}) + X_Z\rfloor d\Theta_{\mathrm{H\text{-}C}} = 0. \tag{2.13}$$

Using field equation (2.1), we get

$$d((X_Z\rfloor\Theta_{\mathrm{H\text{-}C}})|C_4) = 0, \quad C_4\in\mathrm{Sol}, \tag{2.14}$$

and therefore the integral (2.12) does not depend on the choice of $\sigma\subset M$ (for homotopically equivalent surfaces in M).

Proposition 2: Let $\widehat{V}\in T_{C_4}$ (Sol) and V be the τ-vertical vector field on C_4 representing $\widehat{V}$. Then

$$dE_Z\widehat{V} = \int_{F(\sigma)} V\rfloor d\nu_Z. \tag{2.15}$$

Proof: It is easy to see that

$$dE_Z\widehat{V} = \int_{F(\sigma)} \mathscr{L}_{\widetilde{V}}\nu_Z,$$

where $\widetilde{V}$ is an extension of V on a neighborhood of C_4 in $\mathscr{P}$. Using the properties of the Lie derivative [cf. (2.13)] and integrating by parts, we get (2.15).

Theorem 1: Let $\widehat{V}$ be an arbitrary vector tangent to Sol at $C_4 = F(M)$ and $\widehat{Y}$ the vector tangent to Sol defined by means of (2.9). Then

$$dE_Z\widehat{V} = -\Omega(\widehat{Y}\wedge\widehat{V}) = -2\Omega(\widehat{Y},\widehat{V}). \tag{2.16}$$

Proof: Let V be the τ-vertical vector field on C_4 representing $\widehat{V}$. From the invariance property (2.4) we have

$$0 = \mathscr{L}_{X_Z}\Theta_{\mathrm{H\text{-}C}} = d(X_Z\rfloor\Theta_{\mathrm{H\text{-}C}}) + X_Z\rfloor d\Theta_{\mathrm{H\text{-}C}}.$$

Contracting with V, we get

$$(V\rfloor d(X_Z\rfloor\Theta_{\mathrm{H\text{-}C}}))\,|\,C_4 + (V\rfloor X_Z\rfloor d\Theta_{\mathrm{H\text{-}C}})\,|\,C_4 = 0$$

The decomposition (2.10) enables us to write

$$-(V\rfloor d\nu_Z)\,|\,C_4 + (V\rfloor dFZ\rfloor d\Theta_{\mathrm{H\text{-}C}})\,|\,C_4 - (V\rfloor Y\rfloor d\Theta_{\mathrm{H\text{-}C}})\,|\,C_4 = 0. \tag{2.17}$$

But dFZ is tangent to C_4, C_4 satisfies (2.1), and therefore the second term in (2.17) vanishes. We see that (2.16) follows from (2.2), (2.15), and (2.17).

We remember that we have the family $(\Omega(\sigma))_{\sigma\subset M}$ of closed differential 2-forms on Sbm(4). In a similar way formula (2.12) defines a family of functions $(E_Z(\sigma))_{\sigma\subset M}$ on Sbm(4). Formulas (2.15) and (2.17) are valid also for $\widehat{V}\in T_{C_4}$(Sbm(4)). We have

Theorem 1′: Let $C_4\in\mathrm{Sol}$, $\widehat{V}\in T_{C_4}$(Sbm(4)) be an arbitrary tangent vector and $\widehat{Y}\in T_{C_4}$(Sol) be the vector defined by (2.9). Then

$$dE_Z(\sigma)\widehat{V} = -\Omega(\sigma)(\widehat{Y}\wedge\widehat{V}) = -2\Omega(\sigma)(\widehat{Y},\widehat{V}). \tag{2.18}$$

We show later that for theories of gravity $E_Z(\sigma)\equiv 0$ on Sol [but $E_Z(\sigma)\not\equiv 0$ on Sbm(4)] and that $\widehat{Y}$ belongs to the gauge distribution (degeneracy distribution) of Ω. These facts imply that formula (2.16) is trivial. On the other hand, Eq. (2.18) is not trivial and can be taken as the basic equation for the dynamical picture (Sec. 5).

3. THE BUNDLE OF INITIAL DATA AND THE TIME EVOLUTION

In this section we show how to pass from the static description of classical field theories, as presented in Sec. 2, to the evolution picture. In the static approach a state of the system is a solution of field equation (2.1); that is to say, a state is a field of geometric quantities on spacetime. For the coupled system of the gravitational and matter fields we have a metric tensor $\mathbf{g} = (g_{\mu\nu})$, an affine (linear) connection $\mathbf{\Gamma} = (\Gamma^{\lambda}_{\mu\nu})$ and a tensor matter field $\boldsymbol{\phi} = (\phi^{\alpha_1\cdots\alpha_k}_{\beta_1\cdots\beta_s})$ on M.

In order to describe the evolution problem, we assume that a slicing of spacetime into a family of three-dimensional surfaces $(\sigma_t)_{t\in\mathbb{R}}$ and a method of passing from one surface to another are given. This can be accomplished in the following way. Let σ be a three-dimensional submanifold in M such that M is diffeomorphic to the product $\mathbb{R}\times\sigma$. Let $(\Phi_t)_{t\in\mathbb{R}}$ be a one-parameter subgroup to diffeomorphisms of M (i.e., $\Phi_0 = \mathrm{id}$, $\Phi_{t+s} = \Phi_t\circ\Phi_s$, $\Phi_t^{-1} = \Phi_{-t}$), which satisfies the conditions

$$\begin{aligned} &\text{(i)}\quad \cup_{t\in\mathbb{R}}\Phi_t(\sigma) = M,\\ &\text{(ii)}\quad \Phi_{t_1}(\sigma)\cap\Phi_{t_2}(\sigma) = \emptyset \quad \text{for } t_1\neq t_2,\\ &\text{(iii)}\quad \text{the vector field } Z(\mathbf{x}) = d/dt\,\Phi_t(\mathbf{x}) \text{ is transversal to every submanifold } \sigma_t = \Phi_t(\sigma). \end{aligned} \tag{3.1}$$

Remark: A construction of the subgroup $(\Phi_t)_{t\in\mathbb{R}}$ by means of the exponential mapping exp: $C^\infty(TM)\to\mathrm{Diff}\,M$ was discussed in Refs. 48, 49.

By means of the diffeomorphisms $(\Phi_t)_{t\in\mathbb{R}}$ we transport geometrical quantities from points of M lying beyond σ onto σ. If F is a section of $\mathscr{P}$, then the formula

$$f_t(\mathbf{x}) = A(\Phi_t^{-1})(F(\Phi_t(\mathbf{x}))), \quad \mathbf{x}\in\sigma, \tag{3.2}$$

gives us a one-parameter family of sections of $\mathscr{P}$ over σ. Let $\tau_\sigma:\mathscr{P}(\sigma)\to\sigma$ be the restriction of the bundle $\tau:\mathscr{P}\to M$ to the submanifold $\sigma\subset M$.

Definition: A section $f:\sigma\to\mathscr{P}(\sigma)$ of the bundle $\mathscr{P}(\sigma)$ is called an admissible initial data if there exists a section $F:M\to\mathscr{P}$ such that

$$\begin{aligned} &\text{(i)}\quad F(\mathbf{x}) = f(\mathbf{x}), \quad \text{for } \mathbf{x}\in\sigma,\\ &\text{(ii)}\quad F \text{ satisfies field equation (2.1).} \end{aligned} \tag{3.3}$$

A one-parameter family $(f_t)_{t\in\mathbb{R}}$ of sections of $\mathscr{P}(\sigma)$ is called an evolution of the initial admissible data f if

(i) $f_0 = f$,

(ii) the section F of $\mathscr{P}$ given by

$$F(\mathbf{y}) = A(\Phi_t)(f_t(\Phi_t^{-1}(\mathbf{y}))), \quad \mathbf{y}\in M, \tag{3.4}$$

satisfies field equation (2.1).

The set of all admissible initial data on σ is denoted by (Id) (σ). For every section $f\in$(Id)(σ) there is a corresponding three-dimensional submanifold $c_3 = f(\sigma) \subset \mathscr{P}(\sigma)$. We call it an admissible initial surface and the corresponding space of three-dimensional submanifolds of $\mathscr{P}(\sigma)$ is denoted by (Is) (σ).

In general we have no 1–1 correspondence between elements of (Is) (σ) and Sol. In theories with gauge (electrodynamics, gravity, Yang–Mills fields) to every $f\in$(Is) (σ) there correspond many possible solutions of (2.1).

Let $t\rightarrow f_t$ be an evolution of an admissible initial data f. We define

$$'Y(f(\mathbf{x})) = \frac{d}{dt}(f_t(\mathbf{x}))\bigg|_{t=0}, \quad \mathbf{x}\in\sigma. \tag{3.5}$$

$'Y$ is a τ_σ-vertical vector field tangent to $\mathscr{P}(\sigma)$ and defined at points of the three-dimensional surface $c_3 = f(\sigma)$.

We see from (2.8) and (3.2) that the vector field $'Y$ is the restriction of the vector field Y [defined by (2.9)] to the submanifold $c_3 \subset C_4 = F(M) \subset \mathscr{P}$. $'Y$ gives us the initial values of Y on c_3. The vector field $'Y$ defines a vector $'\widehat{Y}$ tangent to (Is) (σ). We call $'\widehat{Y}$ the vector of evolution; in the literature $'\widehat{Y}$ is often called the Hamiltonian vector.[5]

We have a 1–1 correspondence between tangent spaces T_{c_3}((Is) (σ)) and T_{C_4} (Sol), respectively. In fact, if F is a section of $\mathscr{P}$ satisfying (2.1) and Y is defined on C_4 by means of (2.9), then the vector field $'Y$ on c_3 is defined by means of (3.2) and (3.5). Conversely, if $'Y$ is a vector field on c_3 and $t\rightarrow f_t$ is an evolution such that (3.5) holds, then we are able to construct the vector field Y on C_4 by means of (2.8), (2.9), and (3.4).

Formulas (2.2) and (2.12) define the symplectic form and the energy–momentum function on (Is) (σ). The independence of the definitions of the choice of σ means that these quantities are maintained in "time." Evolutions are symplectomorphisms, and the energy–momentum function is a preserved quantity. We can reformulate Theorems 1, 1′ in the space (Is) (σ). The vector of evolution $'\widehat{Y}$ satisfies the Hamilton equations (2.16) and (2.18). This problem is thoroughly investigated in Sec. 5. We shall prove there that admissible initial data have to satisfy some constraint equations. These equations together with the Hamilton equation (2.18) yield the complete characterization of the space of initial data and their evolutions.

Remark: In order to reformulate Theorem 1′ for (Is)(σ), we have to embed the space (Is) (σ) in the space Sbm_σ (3). This space consists of such three-dimensional submanifolds of $\mathscr{P}(\sigma)$ which are the images of sections of $\mathscr{P}(\sigma)$. We have the following natural definition: A vector $'\widehat{Y}$ tangent to Sbm_σ (3) at $c_3\in\mathrm{Sbm}_\sigma$ (3) is represented by a τ_σ-vertical vector field $'Y$ tangent to $\mathscr{P}(\sigma)$ and defined at points of c_3. We show later that a vector $'\widehat{Y}$ tangent to Sbm_σ (3) at $c_3\in$(Is) (σ) is tangent also to (Is) (σ) if and only if the components of the vector field $'Y$ satisfy the linearized version of the constraint equations (4.21).

In this paper we deal with geometrical quantities on M and with geometrical quantities on σ. We describe now the relations between these notions.

Elements of the fibre of the bundle $\mathscr{P}$ over $\mathbf{x}\in M$ are geometric 4-quantities (quantities tangent/cotangent to M). Elements of the fibre of $\mathscr{P}(\sigma)$ over $\mathbf{x}\in\sigma$ are also 4-quantities tangent/cotangent to M at $\mathbf{x}$. We would like to attach to every 4-quantity at $\mathbf{x}$ a family of 3-quantities tangent/cotangent to σ and to describe fibres of $\mathscr{P}(\sigma)$ by means of these 3-objects. Such a procedure is called the 1 +3 decomposition of geometrical quantities (objects) on M. We recall briefly this construction (see Ref. 14 and Appendix B).

Let $M \cong \mathbb{R}\times\sigma$, (x^k) be local coordinates on σ, x^0 be the coordinate in $\mathbb{R}$. Let

$$\sigma_t = \{\mathbf{x}\in M\, x^0 = t\}. \tag{3.6}$$

Let $\mathbf{g} = (g_{\mu\nu})$ be a Lorentz metric on M such that σ_t are spacelike surfaces for $\mathbf{g}$. Let $M\ni\mathbf{x}\rightarrow\mathbf{n}(\mathbf{x})$ be the field of vectors orthonormal to the slicing (i.e., $\mathbf{n}\cdot\mathbf{n} = -1$ and $\mathbf{n}$ is orthogonal to every σ_t).

We decompose vectors tangent to M at $\mathbf{x}\in\sigma_t$ into the normal part (parallel to $\mathbf{n}$) and the tangential part (tangent to σ_t). Similar constructions can be performed for covectors and arbitrary tensors. In Appendix B we define such decompositions by means of the "bar" operation, which commutes with the contractions of tensors and with the covariant differentiation. The bar operation applied to an affine connection $\mathbf{\Gamma} = (\Gamma^\lambda_{\mu\nu})$ on M gives rise to several geometrical objects on σ_t. Four of them $(\bar{\Gamma}^0_{ij}, \bar{\Gamma}^s_{k0}, \bar{\Gamma}^0_{k0}, \bar{\Gamma}^k_{ij})$ are specially interesting. The first three groups give us the fundamental forms of the embedding $i: \sigma_t \rightarrow M$ [14]; the quantities $\bar{\Gamma}^k_{ij}$ define a connection $\bar{\mathbf{\Gamma}}$ on σ_t induced by the connection $\mathbf{\Gamma}$ on M. By means of the 1 +3 decomposition (the bar operation) we redefine local coordinates in fibres of the bundle $\mathscr{P}(\sigma)$ and work with geometrical 3-objects on σ_t.

Remark: In the present section we have defined two kinds of slicings of spacetime into sets of three-dimensional surfaces. In a general case the dynamical slicing (3.1) does not coincide with the coordinate slicing (3.6). The most interesting case is when they, however, coincide. We will assume such a situation in Sec. 5.

4. THE SYMPLECTIC 2-FORM AND THE ENERGY–MOMENTUM FUNCTION FOR THE INTERACTING GRAVITATIONAL AND MATTER FIELDS

The system is described by a metric tensor $\mathbf{g} = (g_{\mu\nu})$, an affine (nonsymmetric) connection $\mathbf{\Gamma} = (\Gamma^\lambda_{\mu\nu})$, and a tensor matter field $\boldsymbol{\phi} = \phi^A = (\phi^{\alpha_1\cdots\alpha_k}_{\beta_1\cdots\beta_l})$ on M. The interaction between the geometry and the matter is given by the Lagrangian

$$L = (1/16\pi)R + L_{\mathrm{mat}}, \tag{4.1}$$

where R is the Ricci scalar built up from $\mathbf{g}$ and $\mathbf{\Gamma}$ and

$$L_{\text{mat}} = L_{\text{mat}}(g_{\mu\nu}, \Gamma^{\lambda}_{\mu\nu}, \phi^A, \partial_\lambda \phi^A)$$

is the Lagrangian of the matter.

The multisymplectic bundle $\tau:\mathscr{P}\to M$ is defined by local coordinates $(x^\lambda; g_{\mu\nu}; \Gamma^{\lambda}_{\mu\nu}; \phi^A; g_{\mu\nu,\tau}; \Gamma^{\lambda}_{\mu\nu,\tau}; \phi^A{}_{,\tau})$ and their transformation properties with respect to a change of local coordinates in spacetime M.

The transformation properties of $g_{\mu\nu}$, $\Gamma^{\lambda}_{\mu\nu}$, and ϕ^A are well known; the quantities $g_{\mu\nu,\tau}$, $\Gamma^{\lambda}_{\mu\nu,\tau}$, and $\phi^A{}_{,\tau}$ have the transformation properties of first partial derivatives of $g_{\mu\nu}$, $\Gamma^{\lambda}_{\mu\nu}$, and ϕ^A, respectively.

The Hamilton–Cartan 4-form $\Theta_{\text{H-C}}$ on $\mathscr{P}$ is given by

$$\begin{aligned}
\Theta_{\text{H-C}} = (1/16\pi)\Big[\sqrt{-g}(g^{\alpha\beta}\delta^\tau_\lambda - g^{\tau\beta}\delta^\alpha_\lambda)\\
\times dx^0\wedge\cdots\wedge \underbrace{d\Gamma^{\lambda}_{\alpha\beta}}_{\tau}\wedge\cdots\wedge dx^3\\
+ g^{\mu\nu}(\Gamma^{\tau}_{\nu\mu}\Gamma^{\alpha}_{\alpha\tau} - \Gamma^{\tau}_{\alpha\mu}\Gamma^{\alpha}_{\nu\tau})\\
\times\sqrt{-g}dx^0\wedge dx^1\wedge dx^2\wedge dx^3\Big]\\
+\sqrt{-g}p^\tau_A dx^0\wedge\cdots\wedge \underbrace{d\phi^A}_{\tau}\wedge\cdots\wedge dx^3\\
-(p^\tau_A\phi^A{}_{,\tau} - L_{\text{mat}})\sqrt{-g}dx^0\wedge dx^1\wedge dx^2\wedge dx^3.
\end{aligned}\tag{4.2}$$

In the definition of the bundle $\mathscr{P}$ the variables $g_{\mu\nu,\tau}$, $\Gamma^{\lambda}_{\mu\nu,\tau}$, and $\phi^A{}_{,\tau}$ are completely independent of $g_{\mu\nu}$, $\Gamma^{\lambda}_{\mu\nu}$, and ϕ^A. However, only such sections of $\mathscr{P}$ for which

$$g_{\mu\nu,\tau} = \partial_\tau g_{\mu\nu},\quad \Gamma^{\lambda}_{\mu\nu,\tau} = \partial_\tau\Gamma^{\lambda}_{\mu\nu},\quad \phi^A{}_{,\tau} = \partial_\tau\phi^A. \tag{4.3}$$

are interesting. From now on we assume that the space Sbm(4) consists of these (smooth) sections of $\mathscr{P}$ for which relations (4.3) are satisfied.

Proposition 3[14,15]: Field equation (2.1) for a section F of $\mathscr{P}$ [satisfying (4.3)]

$$x^\alpha \to F(x^\alpha) = (x^\alpha; g_{\mu\nu}(x^\alpha); \Gamma^{\lambda}_{\mu\nu}(x^\alpha); \phi^A(x^\alpha)\cdots)$$

reads

$$(\text{Eq. mat})_A = \frac{\partial L_{\text{mat}}}{\partial\phi^A} - (\sqrt{-g})^{-1}\partial_\lambda(\sqrt{-g}p^\lambda_A) = 0, \tag{4.4}$$

$$(\text{Eq. I})^{\mu\nu} = G^{\mu\nu} - 8\pi\,{}_{\text{sym}}T^{\mu\nu} = 0, \tag{4.5a}$$

$$(\text{Eq. II})^{\mu\nu}_\lambda = c^{\mu\nu}_\lambda + 16\pi s^{\mu\nu}_\lambda = 0, \tag{4.5b}$$

where $G^{\mu\nu}$, ${}_{\text{sym}}T^{\mu\nu}$, $c^{\mu\nu}_\lambda$, $s^{\mu\nu}_\lambda$, p^λ_A are the Einstein tensor, the symmetric stress–energy tensor of the matter, the hypermomentum tensors of the gravitational and matter fields, the 4-momentum tensor of the matter, respectively (see Appendix E).

Equations (4.5a) have the form of the Einstein equations, equations (4.5b) give us algebraic relations between the connection $\Gamma^{\lambda}_{\mu\nu}$ and the pseudo-Riemannian connection $\gamma^\lambda_{\mu\nu} = \{{}^{\lambda}_{\nu\mu}\}$ on M (cf. Refs. 14, 15 and Appendix E). Therefore system (4.4)–(4.5) yields a generalization of the Einstein equations for cases of tensor fields minimally coupled to gravity. It is not a unique conceivable generalization. Another possibility is to take into account only connections compatible with metrics on spacetime. Such a theory is called the Einstein–Cartan theory of gravity.[26,50,51] However, the variational principle based on a metric $g_{\mu\nu}$ and a connection $\Gamma^{\lambda}_{\mu\nu}$ gives rise to nonmetric compatible solutions of (4.5b). In order to get the Einstein–Cartan theory, we have to take fields of tetrads and anholonomic components of a connection as independent gravitational variables.[26,50–52] The third possible generalization of the Einstein theory was recently proposed by Kijowski.[38,39] His very ingenious variational principle is based only on symmetric (holonomic) components $\Gamma^{\lambda}_{\mu\nu}$ of a connection; the metric $g^{\mu\nu}$ is defined as the conjugate variable by means of the Legendre transformation. However, field equations in this theory are equivalent to those obtained from the Lagrangian (4.1) (in the case of a symmetric connection) by means of our variational principle.

We give now the diagonal formula for the symplectic 2-form $\Omega(\sigma)$ on Sbm(4). The special case, the diagonal expression for Ω on Sol, has been presented in Ref. 14.

Let σ be one of the surfaces of the slicing (3.6), e.g., $\sigma = \sigma_0$. Let $\mathbf{g} = (g_{\mu\nu})$ be a metric on M such that σ is spacelike for $\mathbf{g}$. Let $\boldsymbol{\Gamma} = (\Gamma^{\lambda}_{\mu\nu})$ be an affine (nonsymmetric) connection on M, $\boldsymbol{\phi} = (\phi^A) = (\phi^{\alpha_1\cdots\alpha_k}_{\beta_1\cdots\beta_s})$ be a tensor field on M and $p^\lambda_A = (p^{\lambda\,\alpha_1\cdots\alpha_k}_{\beta_1\cdots\beta_s}) = \partial L_{\text{mat}}/\partial(\partial_\lambda\phi^A)$ be the 4-momentum of $\boldsymbol{\phi}$. The metric $\mathbf{g}$ induces the metric $\bar{\mathbf{g}} = (g_{ij})$ on σ; we denote by $\bar g^{ij}$ the elements of the matrix inverse to the matrix (g_{ij}) and $\bar g = \det g_{ij}$. Let $\bar\Gamma^{\lambda}_{\mu\nu}$ be the bar components of the connection $\boldsymbol{\Gamma}$ on M, $\bar\phi^A$ and $\bar p^\lambda_A$ be the bar components of the tensor fields ϕ^A and p^λ_A, respectively. Let N and N^k be the lapse function and the shift vector of the slicing (3.6) (see Appendix B).

The symplectic variables on Sbm(4) are

$$\begin{aligned}
\Pi^{ij} = \tfrac{1}{2}\sqrt{\bar g}((\bar\Gamma^0_{ab} + \bar\Gamma^0_{ba})\bar g^{ai}\bar g^{bj}\\
- \bar g^{ij}(\bar\Gamma^0_{ab}\bar g^{ab} + \bar\Gamma^s_{s0})),\quad g_{ij}
\end{aligned}\tag{4.6}$$

$$\bar{\mathscr{P}}^{\beta_1\cdots\beta_k}_{\alpha_1\cdots\alpha_s} = \sqrt{\bar g}\bar p^{0\beta_1\cdots\beta_k}_{\alpha_1\cdots\alpha_s},\quad \bar\phi^{\alpha_1\cdots\alpha_k}_{\beta_1\cdots\beta_s}, \tag{4.7}$$

$$\begin{aligned}
\mathscr{M} = -(1/N)\sqrt{\bar g}(\bar c^{00}_0 + 16\pi\bar s^{00}_0),\quad N,\\
\mathscr{M}_k = -(1/N)\sqrt{\bar g}(\bar c^{00}_k + 16\pi\bar s^{00}_k),\quad N^k.
\end{aligned}\tag{4.8}$$

Remarks: (i) The geometrical meaning of the gravitational momenta Π^{ij} has been explained in Ref. 14 (see also Appendix B). (ii) We call Π^{ij}, g_{ij}, $\bar{\mathscr{P}}_A$, $\bar\phi^A$ the canonical variables of the theory. In Sec. 5 we derive the equations which govern their evolutions.

Proposition 4: If $C_4\in$Sbm(4), $\hat Y_1$, $\hat Y_2\in T_{C_4}$(Smb (4)), then

$$\begin{aligned}
\Omega(\sigma)(\hat Y_1,\hat Y_2)\\
= (1/32\pi)\int_{F(\sigma)}(\delta_1\Pi^{ij}\delta_2 g_{ij} - \delta_2\Pi^{ij}\delta_1 g_{ij})dx^1\wedge dx^2\wedge dx^3\\
+\tfrac{1}{2}\int_{F(\sigma)}(\delta_1\bar{\mathscr{P}}_A\delta_2\bar\phi^A - \delta_2\bar{\mathscr{P}}_A\delta_1\bar\phi^A)dx^1\wedge dx^2\wedge dx^3\\
+(1/32\pi)\int_{F(\sigma)}(\delta_1\mathscr{M}\delta_2 N - \delta_2\mathscr{M}\delta_1 N + \delta_1\mathscr{M}_k\delta_2 N^k\\
-\delta_2\mathscr{M}_k\delta_1 N^k)dx^1\wedge dx^2\wedge dx^3,
\end{aligned}\tag{4.9}$$

where

$$\delta_a\Pi^{ij},\quad \delta_a g_{ij},\quad \delta_a\bar{\mathscr{P}}_A,\quad \delta_a\bar\phi^A,$$
$$\delta_a\mathscr{M},\quad \delta_a N,\quad \delta_a\mathscr{M}_k,\quad \delta_a N^k,\quad a = 1, 2,$$

are the components of the vector fields Y_a representing $\widehat{Y}_a$, expressed in the symplectic variables.

In the special case, when $C_4 \in \mathrm{Sol}$, $\widehat{Y}_1, \widehat{Y}_2 \in T_{C_4}(\mathrm{Sol})$ we have, by virtue of (4.5b), $\mathscr{M} = 0$, $\mathscr{M}_k = 0$, $\delta\mathscr{M} = 0$, $\delta\mathscr{M}_k = 0$ and formula (4.9) reduces to that given in Ref. 14.

Remark: In order to get (4.9) from (2.2) we have neglected some integrals of 3-divergencies on σ (exact 3-forms). The terms which are to be omitted were given in Ref. 14.

We discuss now the energy–momentum formula.

Proposition 5: If $Z = \delta x^\lambda(x^\tau)\cdot\partial/\partial x^\lambda$ is a vector field on M, then

$$\begin{aligned} X_Z = {} & \delta x^\lambda\, \partial/\partial x^\lambda + \delta g_{\mu\nu}\partial/\partial g_{\mu\nu} + \delta\Gamma^{\lambda}_{\mu\nu}\, \partial/\partial\Gamma^{\lambda}_{\mu\nu} \\ & + \delta\phi^A\, \partial/\partial\phi^A + \delta g_{\mu\nu,\tau}\, \partial/\partial g_{\mu\nu,\tau} + \delta\Gamma^{\lambda}_{\mu\nu,\tau}\, \partial/\partial\Gamma^{\lambda}_{\mu\nu,\tau} \\ & + \delta\phi^A{}_{,\tau}\, \partial/\partial\phi^A{}_{,\tau}, \end{aligned}$$

where

$$\begin{aligned} \delta g_{\mu\nu} &= -(g_{\mu\epsilon}\partial_\nu\delta x^\epsilon + g_{\nu\epsilon}\partial_\mu\delta x^\epsilon), \\ \delta\Gamma^{\lambda}_{\mu\nu} &= -(\partial_\mu\partial_\nu\delta x^\lambda + \Gamma^{\lambda}_{\mu\epsilon}\partial_\nu\delta x^\epsilon + \Gamma^{\lambda}_{\epsilon\nu}\partial_\mu\delta x^\epsilon - \Gamma^{\epsilon}_{\mu\nu}\partial_\epsilon x^\lambda), \\ \delta\phi^{\alpha_1\cdots\alpha_k}_{\beta_1\cdots\beta_s} &= \sum_{n=1}^{k}\phi^{\alpha_1\cdots\epsilon_n\cdots\alpha_k}_{\beta_1\cdots\cdots\beta_s}\partial_{\epsilon_n}\delta x^{\alpha_n} - \sum_{m=1}^{s}\phi^{\alpha_1\cdots\cdots\alpha_k}_{\beta_1\cdots\tau_m\cdots\beta_s}\partial_{\beta_m}\delta x^{\tau_m}, \end{aligned} \tag{4.10}$$

$$\begin{aligned} \delta g_{\mu\nu,\tau} &= \partial_\tau\delta g_{\mu\nu} - \partial_\tau\delta x^\epsilon g_{\mu\nu,\epsilon}, \\ \delta\Gamma^{\lambda}_{\mu\nu,\tau} &= \partial_\tau\delta\Gamma^{\lambda}_{\mu\nu} - \partial_\tau\delta x^\epsilon\Gamma^{\lambda}_{\mu\nu,\epsilon}, \\ \delta\phi^A{}_{,\tau} &= \partial_\tau\delta\phi^A - \partial_\tau\delta x^\epsilon\phi^A{}_{,\epsilon}, \end{aligned}$$

Proposition 6: If F is a section of $\mathscr{P}$ and $C_4 = F(M)$, then

$$\begin{aligned} \nu_Z|C_4 &= -(X_Z \lrcorner\, \Theta_{\mathrm{H-C}})|C_4 \\ &= \sum_{\lambda=0}^{3}(-1)^\lambda e^\lambda\sqrt{-g}\,dx^0\wedge\cdots_{\hat\lambda}\cdots\wedge dx^3 \end{aligned} \tag{4.11}$$

$$\begin{aligned} e^\lambda &= e^\lambda(Z) \\ &= -\big[(1/16\pi)(RZ^\lambda - g^{\alpha\beta}\mathscr{L}_Z\Gamma^{\lambda}_{\alpha\beta} + g^{\lambda\beta}\mathscr{L}_Z\Gamma^{\alpha}_{\alpha\beta}) \\ &\quad + L_{\mathrm{mat}}Z^\lambda - p^\lambda_A\mathscr{L}_Z\phi^A\big], \end{aligned}$$

where $Z^\lambda = \delta x^\lambda(x^\tau)$ and the corresponding Lie derivatives are (cf. Ref. 15):

$$\begin{aligned} \mathscr{L}_Z\Gamma^{\lambda}_{\mu\nu} &= D_\mu D_\nu Z^\lambda + R^\lambda{}_{\nu\tau\mu}Z^\tau + D_\mu(Z^\tau Q^{\lambda}_{\tau\nu}), \\ \mathscr{L}_Z\phi^{\alpha_1\cdots\alpha_k}_{\beta_1\cdots\beta_s} &= Z^\tau\nabla_\tau\phi^{\alpha_1\cdots\alpha_k}_{\beta_1\cdots\beta_s} \\ &\quad - \nabla_\tau Z^{\alpha_1}\phi^{\tau\alpha_2\cdots\alpha_k}_{\beta_1\cdots\beta_s} - \cdots - \nabla_\tau Z^{\alpha_k}\phi^{\alpha_1\cdots\alpha_{k-1}\tau}_{\beta_1\cdots\cdots\beta_s} \\ &\quad + \nabla_{\beta_1}Z^\epsilon\phi^{\alpha_1\cdots\alpha_k}_{\epsilon\beta_2\cdots\beta_s} + \cdots + \nabla_{\beta_s}Z^\epsilon\phi^{\alpha_1\cdots\cdots\alpha_k}_{\beta_1\cdots\beta_{s-1}\epsilon} \end{aligned} \tag{4.12}$$

(D_μ, ∇_μ are the covariant derivatives with respect to the connections $\Gamma^{\lambda}_{\mu\nu}$ and $\gamma^\lambda_{\mu\nu} = \{^{\lambda}_{\mu\nu}\}$, respectively.

We have from (4.11), (4.12), and (E12), (E13), (E7).

Proposition 7:

$$\begin{aligned} e^\lambda(Z) = {} & (1/8\pi)\big[(H^\lambda{}_\tau - 8\pi\,{}_{\mathrm{can}}T^\lambda{}_\tau)Z^\tau \\ & + 8\pi(\nabla_\omega a^{\omega\lambda}_\tau + a^{\lambda\omega}_\epsilon r^\epsilon_{\tau\omega})Z^\tau\big] \\ & - (1/16\pi)(c^{\lambda\omega}_\epsilon + 16\pi s^{\lambda\omega}_\epsilon)(\nabla_\omega Z^\epsilon + r^\epsilon_{\tau\omega}Z^\tau) \\ & + (1/8\pi)\nabla_\tau b^{\tau\lambda} \end{aligned} \tag{4.13}$$

where the skew-symmetric tensor $b^{\tau\lambda}$ is defined by

$$\begin{aligned} b^{\alpha\beta} = 8\pi\big[& -a^{\alpha\beta}_\tau Z^\tau + (1/16\pi)(g^{\alpha\tau}D_\tau Z^\beta - g^{\tau\beta}D_\tau Z^\alpha) \\ & + (1/16\pi)Z^\tau(g^{\alpha\sigma}Q^{\beta}_{\tau\sigma} - g^{\beta\sigma}Q^{\alpha}_{\tau\sigma})\big] \end{aligned} \tag{4.14}$$

(tensors $H^\lambda{}_\tau$, $a^{\alpha\beta}_\tau$ and the canonical stress–energy tensor ${}_{\mathrm{can}}T^\lambda{}_\tau$ are defined in Appendix E).

The energy–momentum formula (4.13) is the sum of three terms. The second of them is expressed by the left sides of field equations (4.5b), the third is the Riemannian divergence of a skew-symmetric tensor. We show in Appendix D that the first term in (4.13) can be also expressed by the left sides of field equations (4.4) and (4.5a, b). This result gives us relations between our energy–momentum formula and those presented by Komar[7] and Kijowski[38,39] (see also the paper by Trautman[53]). For $C_4 \in \mathrm{Sbm}(4)$, $C_4 = F(M)$ we have from (2.12), (3.6), and (4.11)

$$E_Z(\sigma) = \int_{F(\sigma)}\sqrt{-g}\,e^0(Z)\,dx^1\wedge dx^2\wedge dx^3 \tag{4.15}$$

Proposition 8:

$$\begin{aligned} \sqrt{-g}\,e^0(Z) &= \sqrt{\bar g}\,\bar e^0(Z) \\ &= (1/16\pi)(\mathrm{Constr.}_0\bar Z^0 + \mathrm{Constr}_p\bar Z^p) \\ &\quad + (1/16\pi)(\mathscr{M}\partial_0\bar Z^0 + \mathscr{M}_p\partial_0\bar Z^p) \\ &\quad - (1/16\pi)\sqrt{\bar g}\,\bar\nabla_s\big[(\overline{\mathrm{Eq.\ II}})^{0s}_{0}\bar Z^0\big] \\ &\quad - (1/16\pi)\sqrt{\bar g}\,\bar\nabla_s\big[(\overline{\mathrm{Eq.\ II}})^{00}_{k}\bar g^{ks}\bar Z^0\big] \\ &\quad - (1/16\pi)\sqrt{\bar g}\,\bar\nabla_s\big[(\overline{\mathrm{Eq.\ II}})^{0s}_{k}\bar Z^k\big] \\ &\quad + (1/8\pi)\sqrt{\bar g}\,\bar\nabla_s\bar b^{s0} \\ &\quad + (1/16\pi)\mathscr{M}(-N^s\partial_s\bar Z^0 + \bar Z^s\partial_s N) \\ &\quad + (1/16\pi)\mathscr{M}_q(-N^s\partial_s\bar Z^q + \bar Z^s\partial_s N^q \\ &\quad - \bar g^{qu}N^2\partial_u(\bar Z^0/N)) \end{aligned} \tag{4.16}$$

where $\bar\nabla_k$ is the Riemannian covariant derivative on σ (cf. Appendix B) and

$$\begin{aligned} \mathrm{Constr}_0 = {} & -\sqrt{\bar g}\,{}^{(3)}R + (1/\sqrt{\bar g})(\Pi^{pq}\Pi_{pq} - \tfrac{1}{2}(\mathrm{tr}\Pi)^2) \\ & - (1/\sqrt{\bar g})\{(\Pi^{pq} - \pi^{pq})(\Pi_{pq} - \pi_{pq}) \\ & - \tfrac{1}{2}[\mathrm{tr}(\Pi - \pi)]^2\} - \bar\nabla_p(\sqrt{\bar g}\,\bar z^{0p}_0) \\ & - \bar\nabla_p(\sqrt{\bar g}\,\bar z^{00}_s\bar g^{sp}) + (1/\sqrt{\bar g})(\pi_{pq} - \tfrac{1}{2}g_{pq}\mathrm{tr}\pi) \\ & \times(\sqrt{\bar g}\,\bar z^{0p}_s\bar g^{sq}) - 16\pi\sqrt{\bar g}(L_{\mathrm{mat}} - (1/\sqrt{\bar g}) \\ & \times\bar{\mathscr{P}}^{\beta_1\cdots\beta_s}_{\alpha_1\cdots\alpha_k}\,\overline{D_0\phi^{\alpha_1\cdots\alpha_k}_{\beta_1\cdots\beta_s}}) \\ & + \sqrt{\bar g}\,\bar z^{0\lambda}_\tau\bar r^\tau_{0\lambda} - \tfrac{1}{2}\sqrt{\bar g}\,\bar c^{\alpha\beta}_\lambda\bar r^\lambda_{\alpha\beta}, \end{aligned} \tag{4.17a}$$

$$\begin{aligned} \mathrm{Constr}_r = {} & -2\,\bar\nabla_p\Pi^p{}_r - \bar\nabla_p(\sqrt{\bar g}\,\bar z^{0p}_r) \\ & + 16\pi\,\bar{\mathscr{P}}^{\beta_1\cdots\beta_s}_{\alpha_1\cdots\alpha_k}\bar\nabla_r\bar\phi^{\alpha_1\cdots\alpha_k}_{\beta_1\cdots\beta_s}. \end{aligned} \tag{4.17b}$$

Remark: ${}^{(3)}R$ is the scalar curvature of the metric $\bar{\mathbf{g}} = (g_{ij})$ on σ.

Among eight terms which appear in (4.16) four are 3-divergences and therefore are not important for integral formulas. In the special but the most important case when the dynamical slicing (3.1) coincides with the coordinate slicing (3.6), we have

$$\bar Z^0 = N, \quad \bar Z^p = N^p \tag{4.18}$$

and the last two terms in (4.16) vanish.

We define the reduced energy–momentum 3-density:

$$\mathscr{E}_{\mathrm{red}} = (1/16\pi)(\mathrm{Constr}_0 \bar{Z}^0 + \mathrm{Constr}_p \bar{Z}^p) + (1/16\pi)(\mathscr{M}\partial_0 \bar{Z}^0 + \mathscr{M}_p \partial_0 \bar{Z}^p). \tag{4.19}$$

We have from (4.15), (4.16), and (4.19)

$$E_Z(\sigma) = \int_{F(\sigma)} \mathscr{E}_{\mathrm{red}}\, dx^1 \wedge dx^2 \wedge dx^3, \tag{4.20}$$

The quantities Constr_0, Constr_p (a scalar density and a covector density on σ) are called the left-hand sides of constraint equations (or simply constraints). They play a fundamental role in the Hamiltonian formulation of the theory.

Remark: By means of (E12) we can express $\sqrt{g}\, \bar{z}_\tau^{0\lambda}$ by the canonical variables $\bar{\mathscr{P}}_A$, $\bar{\phi}^A$. Therefore Constr_p can be expressed by canonical variables Π^{pq}, g_{pq}, $\bar{\mathscr{P}}_A$, $\bar{\phi}^A$ and their spatial derivatives. For Constr_0 the situation is more complicated. We show in Sec. 5 that under some additional assumptions Constr_0 can be also expressed by means of canonical variables and their spatial derivatives.

The following result explains the name "constraints."

Proposition 9: If field equations (4.4), (4.5a, b) are satisfied, then

$$\mathrm{Constr}_0 = 0, \quad \mathrm{Constr}_p = 0. \tag{4.21}$$

The proof of Proposition 9 is given in Appendix D.

We show later that Eqs. (4.21) yield some relations (constraints) among initial values of canonical variables and that the evolution maintains these relations. We have from (4.19)–(4.21)

Proposition 10: If $C_4 \in \mathrm{Sol}$, then, for every σ, $E(\sigma) = 0$.

By virtue of Proposition 10 we see that from Eq. (2.16) we can only conclude that the vector of evolution $\hat{Y}$ belongs to the degeneracy subspace of Ω (see Sec. 6). This result does not help us too much. We are not able to determine from (2.16) the components of Y. Therefore we have to compute the derivatives of $E(\sigma)$ in directions tangent to Sbm(4) (they do not vanish) and to apply formula (2.18). This problem is solved in the next section.

5. HAMILTON EQUATION FOR THE COUPLED GRAVITATIONAL AND MATTER FIELDS

In Secs. 2 and 3 we have presented in outline the Hamiltonian formulation for classical field theories. We have proved that the variational principle and Euler–Lagrange equation (2.1) [cf. (4.4)–(4.5)] give rise to Hamilton equations (2.16), (2.18). We show that the converse statement is also true and that Hamilton equation (2.18) is equivalent to the system (4.4)–(4.5): instead of the Lagrangian (4.1) and Eqs. (4.4)–(4.5) we postulate the energy–momentum function $E(\sigma)$-(4.20), the symplectic 2-form $\Omega(\sigma)$-(4.9), and the Hamilton equation (2.18). Our main result reads

Theorem 2: Let f be a section of the bundle $\mathscr{P}(\sigma)$ such that constraint equations

$$\mathrm{Constr}_0 = 0, \quad \mathrm{Constr}_p = 0 \tag{5.1}$$

hold. Let $'\hat{Y}$ be a vector tangent to $\mathrm{Sbm}_\sigma(3)$ at $c_3 = f(\sigma)$ such that for every vector $'\hat{V} \in T_{c_3}(\mathrm{Sbm}_\sigma(3))$

$$dE(\sigma)\cdot{}'\hat{V} = -2\Omega(\sigma)('\hat{Y}, '\hat{V}) \quad [\text{cf. } (2.18)]. \tag{5.2}$$

Let

$$'Y = \delta\Pi^{pq}\, \partial/\partial\Pi^{pq} + \delta g_{pq}\, \partial/\partial g_{pq} + \delta\bar{\mathscr{P}}_A\, \partial/\partial\bar{\mathscr{P}}_A + \delta\bar{\phi}^A\, \partial/\partial\bar{\phi}^A + \delta\mathscr{M}\, \partial/\partial\mathscr{M} + \delta N\, \partial/\partial N + \delta\mathscr{M}_k\, \partial/\partial\mathscr{M}_k + \delta N^k\, \partial/\partial N^k + \cdots \tag{5.3}$$

be the τ_σ-vertical vector field on c_3 representing $'\hat{Y}$. Then the evolution equations for symplectic variables read

$$\partial_0\Pi^{pq} = \delta\Pi^{pq}, \quad \partial_0 g_{pq} = \delta g_{pq}, \quad \partial_0\bar{\mathscr{P}}_A = \delta\bar{\mathscr{P}}_A, \quad \partial_0\bar{\phi}^A = \delta\bar{\phi}^A, \tag{5.4a}$$

$$\partial_0\mathscr{M} = \delta\mathscr{M}, \quad \partial_0 N = \delta N, \quad \partial_0\mathscr{M}_k = \delta\mathscr{M}_k, \quad \partial_0 N^k = \delta N^k, \tag{5.4b}$$

and the system (5.1)–(5.4) is equivalent to (4.4)–(4.5).

Remarks: (i) From (5.2) we are able to compute only these components of $'Y$ which enter into the symplectic 2-form $\Omega(\sigma)$, i.e.,

$$\delta\Pi^{pq}, \quad \delta g_{pq}, \quad \delta\bar{\mathscr{P}}_A, \quad \delta\bar{\phi}^A, \quad \delta\mathscr{M}, \quad \delta N, \quad \delta\mathscr{M}_k, \quad \delta N^k.$$

(ii) We show later that the vector $'\hat{Y}$ is not only tangent to $\mathrm{Sbm}_\sigma(3)$ but also to $(\mathrm{Is})(\sigma)$—its components satisfy the linearized version of constraint equations (5.1) (see Proposition 14).

(iii) Hamilton equation (5.2) gives rise to relations (5.4) only on σ, i.e., for $x^0 = 0$. Therefore, in order to get (5.4) at all instants of time, we have to postulate constraint equations on every σ and to solve (5.2) also for every σ. We show further that this requirement can be reduced (Proposition 11).

Proof of Theorem 2: (I) Let us observe that in fibers of the bundle $\mathscr{P}(\sigma)$ we have the following independent variables:

$$g_{pq}, N, N^k; \quad \text{their spatial derivatives}; \tag{5.5a}$$

$$\bar{\gamma}^0_{pq} = (1/\sqrt{g})(\pi_{pq} - \tfrac{1}{2} g_{pq}\,\mathrm{tr}\,\pi), \partial_0 N, \partial_0 N^k; \tag{5.5b}$$

$$\Pi^{pq}, \mathscr{M}, \mathscr{M}_k; \quad \text{their spatial derivatives}; \tag{5.5c}$$

$$\tfrac{1}{2}(\bar{\Gamma}^0_{ab} - \bar{\Gamma}^0_{ba}), (\bar{\Gamma}^s_{k0} - \tfrac{1}{3}\delta^s_k \bar{\Gamma}^p_{p0}), \bar{\Gamma}^\lambda_{0\nu}, \bar{\Gamma}^k_{ab}; \quad \text{their spatial derivatives}; \tag{5.5d}$$

$$\partial_0 \bar{\Gamma}^\lambda_{\mu\nu}; \tag{5.5e}$$

$$\bar{\phi}^A; \quad \text{their spatial derivatives}; \tag{5.5f}$$

$$\bar{\mathscr{P}}_A; \tag{5.5g}$$

Remarks: (i) Variables $\tfrac{1}{2}(\bar{\Gamma}^0_{ab} + \bar{\Gamma}^0_{ba})$, $\bar{\Gamma}^p_{p0}$, $\bar{\Gamma}^0_{k0}$ can be expressed by the independent variables (5.5) [cf. (C1)].

(ii) We assume throughout this paper that there are no relations among the canonical momenta $\bar{\mathscr{P}}_A$ of the matter field. Degenerate cases, e.g., the Maxwell electrodynamics, can be treated in a similar way (cf. Ref. 13).

(II) Let

$$'V = \delta g_{pq}\, \partial/\partial g_{pq} + \delta N \partial/\partial N + \delta N^k\, \partial/\partial N^k + \cdots, \tag{5.6}$$

where δg_{pq}, δN, δN^k, etc. are arbitrary variations of the independent variables (5.5), be a vector field on $c_3 = f(\sigma)$ representing $'\hat{V} \in T(\mathrm{Sbm}_\sigma(3))$. We have from (4.19)–(4.20).

$$dE(\sigma)'\hat{V} = (1/16\pi)\int_{f(\sigma)} [\delta(\mathrm{Constr}_0)N + \delta(\mathrm{Constr}_p)N^p + \delta\mathscr{M}\partial_0 N + \delta\mathscr{M}_k\partial_0 N^k + \mathscr{M}\delta(\partial_0 N) + \mathscr{M}_k\delta(\partial_0 N^k)]dx^1\wedge dx^2\wedge dx^3. \tag{5.7}$$

We recall that $N = \overline{Z}^0$, $N^k = \overline{Z}^k$ [cf. (4.18)] and the equations (5.1) are satisfied. The formulas for $\delta(\mathrm{Constr}_\mu)$ are given in Appendix C.

We see from (C3) that the integrand in (5.7) contains variations $\delta\overline{\Gamma}^\lambda_{0\nu}$, $\delta(\overline{\Gamma}^s_{k0} - \frac{1}{3}\delta^s_k\overline{\Gamma}^p_{p0})$, $\delta\overline{\Gamma}^k_{pq}$, $\delta(\overline{\Gamma}^0_{ab} - \overline{\Gamma}^0_{ba})$. These variations do not appear in (4.9), and therefore all terms containing them have to vanish. It gives rise to the set of equations

$$\begin{aligned}
&(\overline{\text{Eq. II}})^{0\nu}_\lambda = 0, \quad (\overline{\text{Eq. II}})^{ab}_0 - (\overline{\text{Eq. II}})^{ba}_0 = 0,\\
&(\overline{\text{Eq. II}})^{s0}_k = 0 \text{ for } s\neq k,\\
&(\overline{\text{Eq. II}})^{10}_1 = (\overline{\text{Eq. II}})^{20}_2 = (\overline{\text{Eq. II}})^{30}_3,\\
&(\overline{\text{Eq. II}})^{pq}_s + \delta^p_s(\overline{\text{Eq. II}})^{q0}_0 = 0.
\end{aligned} \tag{5.8}$$

We have from (5.8)

$$\mathscr{M} = 0, \quad \mathscr{M}_k = 0, \tag{5.9}$$

and therefore the terms with $\delta(\partial_0 N)$, $\delta(\partial_0 N^k)$ vanish from (5.7).

(III) Let lin.symp.var. (linearized symplectic variables) be the vector space consisting of systems

$$\mathscr{X} = (\delta\Pi^{pq}, \delta g_{pq}, \delta\overline{\mathscr{P}}_A, \delta\bar{\phi}^A, \delta\mathscr{M}, \delta N, \delta\mathscr{M}_k, \delta N^k),$$

where components of $\mathscr{X}$ define corresponding tensor fields or tensor densities on σ. For every $'\hat{V}\in T(\mathrm{Sbm}_\sigma(3))$ there exists a unique $\mathscr{V}\in$lin.symp.var. [cf. (5.6)].

We know that equations (5.8) are satisfied and therefore [cf. (C2)–(C5)] the left sides of linearized constraints define a linear differential operator

$$P: \text{lin.symp.var.} \rightarrow C^\infty(\sigma, \mathrm{den})\times C^\infty(\sigma, \mathrm{den}T^*(\sigma)),$$

where $C^\infty(\sigma, \mathrm{den})$ is the vector space of (smooth) scalar densities on σ and $C^\infty(\sigma, \mathrm{den}T^*(\sigma))$ is the vector space of (smooth) covector densities on σ.

$$\begin{aligned}
&P(\mathscr{X}) = (\underline{s}, \underline{u}_k)\in C^\infty(\sigma, \mathrm{den})\times C^\infty(\sigma, \mathrm{den}T^*(\sigma))\\
&\underline{s} = \delta(\mathrm{Constr}_0), \quad \underline{u}_k = \delta(\mathrm{Constr}_k)
\end{aligned} \tag{5.10}$$

[see (C2)–(C5)]. The space lin.symp.var. has a natural scalar product. If

$$\mathscr{X}_j\in\text{lin.symp.var.}, \quad j = 1, 2, \quad \mathscr{X}_j = (\underset{j}{\delta\Pi}^{pq}, \underset{j}{\delta g}_{pq}, \cdots), \tag{5.11}$$

then

$$\begin{aligned}
g(\sigma)(\mathscr{X}_1, \mathscr{X}_2) = (1/16\pi)\int_\sigma\Big[&(1/\sqrt{\bar{g}})\underset{1}{\delta\Pi}^{pq}\underset{2}{\delta\Pi}^{ab}g_{pa}g_{qb} + \sqrt{\bar{g}}\underset{1}{\delta g}_{pq}\underset{2}{\delta g}_{ab}\bar{g}^{pa}\bar{g}^{qb} + (1/\sqrt{\bar{g}})\underset{1}{\delta\mathscr{M}}\underset{2}{\delta\mathscr{M}} + \sqrt{\bar{g}}\underset{1}{\delta N}\underset{2}{\delta N}\\
&+ (1/\sqrt{\bar{g}})\underset{1}{\delta\mathscr{M}}_k\underset{2}{\delta\mathscr{M}}_s\bar{g}^{ks} + \sqrt{\bar{g}}\underset{1}{\delta N}^k\underset{2}{\delta N}^s g_{ks}\Big]dx^1\wedge dx^2\wedge dx^3\\
&+\int_\sigma\Big[(1/\sqrt{\bar{g}})\underset{1}{\delta\overline{\mathscr{P}}}_A\bar{\bar{g}}^{AB}\underset{2}{\delta\overline{\mathscr{P}}}_B + \sqrt{\bar{g}}\underset{1}{\delta\bar{\phi}}^A\bar{\bar{g}}_{AB}\underset{2}{\delta\bar{\phi}}^B\Big]dx^1\wedge dx^2\wedge dx^3,
\end{aligned} \tag{5.12}$$

where

$$\begin{aligned}
&\bar{\bar{g}}_{AB} = \bar{\bar{g}}_{\alpha_1\mu_1}\cdots\bar{\bar{g}}_{\alpha_k\mu_k}\bar{\bar{g}}^{\beta_1\nu_1}\cdots\bar{\bar{g}}^{\beta_s\nu_s},\\
&\bar{\bar{g}}_{00} = 1, \quad \bar{\bar{g}}_{0k} = \bar{\bar{g}}_{k0} = 0, \quad \bar{\bar{g}}_{ks} = g_{ks},\\
&\bar{\bar{g}}^{00} = 1, \quad \bar{\bar{g}}^{0k} = g^{k0} = 0, \quad \bar{\bar{g}}^{ks} = \bar{g}^{ks}.
\end{aligned} \tag{5.13}$$

We define the symplectic operator J: lin.symp.var.$\rightarrow$lin.symp.var:

$$\begin{aligned}
&J(\delta\Pi^{pq}, \delta g^{pq}, \delta\overline{\mathscr{P}}_A, \delta\bar{\phi}^A, \delta\mathscr{M}, \delta N, \delta\mathscr{M}_k, \delta N^k)\\
&= (\sqrt{\bar{g}}\,\bar{g}^{pa}\bar{g}^{qb}\delta g_{ab}, -(1/\sqrt{\bar{g}})g_{pa}g_{qb}\delta\Pi^{ab}, \sqrt{\bar{g}}\,\bar{\bar{g}}_{AB}\delta\bar{\phi}^B, -(1/\sqrt{\bar{g}})\bar{\bar{g}}^{AB}\delta\overline{\mathscr{P}}_B, \sqrt{\bar{g}}\,\delta N, -(1/\sqrt{\bar{g}})\delta\mathscr{M},\\
&\quad\times\sqrt{\bar{g}}\,g_{ks}\delta N^s, -(1/\sqrt{\bar{g}})\delta\mathscr{M}_s\bar{g}^{ks}).
\end{aligned} \tag{5.14}$$

We have

$$J^2 = -\,\mathrm{id}. \tag{5.15}$$

Let $'\hat{X}_1, '\hat{X}_2\in T(\mathrm{Sbm}_\sigma(3))$ and $\mathscr{X}_1$, $\mathscr{X}_2$ be the corresponding elements in lin.symp.var. We have from (4.9) and (5.12)

$$2\Omega(\sigma)('\hat{X}_1, '\hat{X}_2) = g(\sigma)(\mathscr{X}_1, J\mathscr{X}_2) = -g(\sigma)(J\mathscr{X}_1, \mathscr{X}_2). \tag{5.16}$$

Let $(q, v^k)\in C^\infty(\sigma, \mathbb{R})\times C^\infty(\sigma, T(\sigma))$: q is a scalar function on σ and v^k is a (smooth) vector field on σ. We have the natural pairing

$$\langle(q, v^k)\,|\,(\underline{s}, \underline{u}_p)\rangle = (1/16\pi)\int_\sigma(q\underline{s} + v^k\underline{u}_k)dx^1\wedge dx^2\wedge dx^3. \tag{5.17}$$

By means of the scalar product (5.12) and the pairing (5.17) we define the adjoint operator

$$\begin{aligned}
&P^*: C^\infty(\sigma, \mathbb{R})\times C^\infty(\sigma, T(\sigma))\rightarrow\text{lin.symp.var.}\\
&g(\sigma)(P^*(q, v^k), \mathscr{X}) = \langle(q, v^k)\,|\,P\mathscr{X}\rangle.
\end{aligned}$$

The explicit formulas for P^* are given in Appendix C. From (5.7) and (5.9) we get

$$dE(\sigma)'\hat{V} = \langle(N, N^k)\,|\,P\mathscr{V}\rangle + g(\sigma)(\mathscr{Z}, \mathscr{V}) \tag{5.18}$$

where $\mathscr{V}$ is the element in lin.symp.var. corresponding to $'\hat{V}$ and

$$\mathscr{Z} = (0, 0, 0, 0, \sqrt{\bar{g}}\partial_0 N, 0, \sqrt{\bar{g}}\,g_{ks}\partial_0 N^s, 0). \tag{5.19}$$

We have from (5.18)

$$dE(\sigma)'\widehat{V} = g(\sigma)(J(-JP^*(N, N^k) - J\mathscr{Z}), \mathscr{V}), \tag{5.20}$$

and by virtue of (5.16) the solution of (5.2) is

$$\mathscr{Y} = -(JP^*(N, N^k) + J\mathscr{Z}). \tag{5.21}$$

$\mathscr{Y}$ determines the components $(\delta\Pi^{pq}, \delta g_{pq}, \cdots)$ of the right sides of equations (5.4).

(IV) We have from (5.14) and (5.19)

$$-J\mathscr{Z} = (0, 0, 0, 0, 0, \partial_0 N, 0, \partial_0 N^k). \tag{5.22}$$

From (C1), (C3), and (C6) we compute δg_{pq}, δN, δN^k components of $\mathscr{Y}$. They read

$$\begin{aligned}\delta g_{pq} &= \overline{\nabla}_p N_q + \underline{\overline{\nabla}_q N_p} + (2N/\sqrt{\bar{g}})(\pi_{pq} \underline{- \tfrac{1}{2} g_{pq}} \mathrm{tr}\pi) \\ &\quad + N\underline{[-(}\,\mathrm{Eq.\ II}\,)_0^{(ab)} g_{ap} g_{bq} + \tfrac{1}{2} g_{pq}(\,\mathrm{Eq.\ II}\,)_0^{(ab)} g_{ab} \\ &\quad + \tfrac{1}{6}(\,\mathrm{Eq.\ II}\,)_s^{so} g_{pq}\,],\end{aligned} \tag{5.23}$$

$$\delta N = \partial_0 N + N[\tfrac{1}{4}N(\,\overline{\mathrm{Eq.\ II}}\,)_0^{(ab)} g_{ab} - \tfrac{1}{12}N(\,\overline{\mathrm{Eq.\ II}}\,)_s^{so}], \tag{5.24}$$

$$\delta N^k = \partial_0 N^k + N\,[N(\overline{\,\mathrm{Eq.\ II}\,})_0^{k0}]. \tag{5.25}$$

However, we know from (5.4b) that

$$\partial_0 g_{pq} = \delta g_{pq}, \quad \partial_0 N = \delta N, \quad \partial_0 N^k = \delta N^k.$$

Therefore, it follows from (5.23)–(5.25) and (E22) that

$$\overline{(\,\mathrm{Eq.\ II}\,)}_0^{k0} = 0, \quad \overline{(\,\mathrm{Eq.\ II}\,)}_s^{s0} = 0, \quad \overline{(\,\mathrm{Eq.\ II}\,)}_0^{(ab)} = 0 \tag{5.26}$$

Relations (5.26) together with (5.8) give rise to

$$\overline{(\,\mathrm{Eq.\ II}\,)}_\lambda^{\mu\nu} = 0, \tag{5.27}$$

and thus Eqs. (4.5b) hold.

(V) From now on we assume that Eqs. (5.27) hold. Explicit formulas for $-JP^*$ in this special case are given in Appendix C. We get from (C6) and (5.4)

$$\partial_0 \mathscr{M} = 0, \quad \partial_0 \mathscr{M}_k = 0. \tag{5.28}$$

This formula is in perfect agreement with (5.27) [cf. (4.8)]. We see also from (C6) that equation

$$\partial_0 \bar{\phi}^A = \delta \bar{\phi}^A \tag{5.29}$$

are simply the identity. Equations

$$\partial_0 \overline{\mathscr{P}}_A = \delta \overline{\mathscr{P}}_A \tag{5.30}$$

are exactly the E–L matter field equations (4.4).

Now we explain the meaning of dynamical equations for the gravitational momenta

$$\partial_0 \Pi^{pq} = \delta \Pi^{pq}. \tag{5.31}$$

It has been proved in Ref. 14 that this system is equivalent to

$$\begin{aligned}&\sqrt{\bar{g}}({}_{\mathrm{sym}}R_{pq} - 8\pi({}_{\mathrm{sym}}T_{pq} - \tfrac{1}{2}g_{pq}\mathrm{tr}\,{}_{\mathrm{sym}}T))\bar{g}^{pi}\bar{g}^{qj} \\ &\quad - \sqrt{\bar{g}}\bar{g}^{ij}({}_{\mathrm{sym}}R_{ab} - 8\pi({}_{\mathrm{sym}}T_{ab} - \tfrac{1}{2}g_{ab}\mathrm{tr}\,{}_{\mathrm{sym}}T))\bar{g}^{ab} \\ &\quad = 0.\end{aligned} \tag{5.31'}$$

Remark: In order to prove the equivalence between (5.31) and (5.31′) we have to use equations (5.27) and (5.30).

(VI) Relations (D1) give rise to the following result: If Eqs. (5.27) and (5.30) hold, then

$$\begin{aligned}\mathrm{Constr}_0 &= 2\sqrt{\bar{g}}(\overline{G}^0{}_0 - 8\pi\,{}_{\mathrm{sym}}\overline{T}^0{}_0) \\ &= 2\sqrt{\bar{g}}[(G^0{}_0 - 8\pi\,{}_{\mathrm{sym}}T^0{}_0) \\ &\quad - N^r(G^0{}_r - 8\pi\,{}_{\mathrm{sym}}T^0{}_r)],\end{aligned} \tag{5.32a}$$

$$\begin{aligned}\mathrm{Constr}_p &= 2\sqrt{\bar{g}}(\overline{G}^0{}_p - 8\pi\,{}_{\mathrm{sym}}\overline{T}^0{}_p) \\ &= 2\sqrt{\bar{g}}N(G^0{}_p - 8\pi\,{}_{\mathrm{sym}}T^0{}_p).\end{aligned} \tag{5.32b}$$

These equations show that the constraint equations (5.1) are equivalent to

$$G^0{}_\lambda - 8\pi\,{}_{\mathrm{sym}}T^0{}_\lambda = 0. \tag{5.33}$$

It is also easy to prove that Eqs. (5.31′) and (5.33) are equivalent to (4.5a). The proof of Theorem 2 is now complete.

Let us observe that if constraint equations (5.1) do not hold, we get a nonconsistent system. In fact we would have then the additional term $(1/16\pi)(\mathrm{Constr}_0\delta N + \mathrm{Constr}_p\delta N^p$ in (5.7) and instead of (5.28) we would have

$$\partial_0 \mathscr{M} = -(\mathrm{Constr}_0), \quad \partial_0 \mathscr{M}_p = -(\mathrm{Constr}_p), \tag{5.28'}$$

which would be in contradiction with (5.27). [In this way we have proved that the constraint equations (5.1) give the necessary and sufficient conditions for the solvability of (5.2).]

We have pointed out in Sec. 4 that Constr_p can be expressed by canonical variables and their spatial derivatives. In a general case this is not true for Constr_0. However, if Eqs. (4.5b) are satisfied, then we see from (C2)–(C5) that $\delta(\mathrm{Constr}_0)$ can be expressed by variations of canonical variables and their spatial derivatives only.

Let us assume that Eqs. (4.5b) and (4.4) are satisfied. We know that Constr_μ are now related to the Einstein tensor by (5.32) and the Einstein tensor satisfies the contracted Bianchi identities (cf. Refs. 14, 15).

$$\nabla_\lambda(G^\lambda{}_\mu - 8\pi\,{}_{\mathrm{sym}}T^\lambda{}_\mu) = 0. \tag{5.34}$$

All these facts give rise to the following:

Proposition 11 (see Refs. 14 and 15): If Eqs. (4.4), (4.5b), (5.31′) are satisfied for every $x^0 = 0$, then constraint equations (5.1) are satisfied for all x^0.

6. THE DEGENERACY DISTRIBUTION OF THE SYMPLECTIC FORM AND THE ACTION OF DIFF M IN THE SPACE OF SOLUTIONS

The left action A of Diff M in $\mathscr{P}$ generates a right action $\widehat{A}$ in the space Sol:

$$(\mathrm{Diff}\,M \times \mathrm{Sol}) \ni (\Phi, C_4) \rightarrow \widehat{A}(\Phi)(C_4) \in \mathrm{Sol}. \tag{6.1}$$

If $C_4 \in \mathrm{Sol}$ is the image of a section $F: M \rightarrow \mathscr{P}$, then $\widehat{A}(\Phi)(C_4)$ is the image of the section

$$M \ni \mathbf{x} \rightarrow A(\Phi^{-1})(F(\Phi(\mathbf{x}))) \in \mathscr{P}. \tag{6.2}$$

A one-parameter family of diffeomorphisms of M $(\Phi_t)_{t \in \mathbb{R}}$ defines the vector field on Sol:

$$\widehat{Y}(C_4) = \frac{d}{dt}\widehat{A}(\Phi_t)(C_4)\Big|_{t=0} = d\widehat{A}(C_4)\cdot U, \tag{6.3}$$

where $U(\mathbf{x}) = (d/dt)\Phi_t(\mathbf{x})|_{t=0}$; $\mathbf{x} \in M$ is a vector field on M.

Proposition 12: The vector $Y(C_4)$ is represented by the τ-vertical vector field Y

$$Y = -(X_U - dFU) \tag{6.4}$$

In local coordinates we have

$$U = \delta x^\lambda \, \partial/\partial x^\lambda,$$
$$\begin{aligned} Y = {} & \delta g_{\mu\nu}\, \partial/\partial g_{\mu\nu} + \delta\Gamma^\lambda_{\mu\nu}\, \partial/\partial\Gamma^\lambda_{\mu\nu} + \delta\phi^A\, \partial/\partial\phi^A \\ & + \delta g_{\mu\nu,\tau}\, \partial/\partial g_{\mu\nu,\tau} + \delta\Gamma^\lambda_{\mu\nu,\tau}\, \partial/\partial\Gamma^\lambda_{\mu\nu,\tau} \\ & + \delta\phi^A{}_{,\tau}\, \partial/\partial\phi^A{}_{,\tau}, \end{aligned}$$
$$\delta g_{\mu\nu} = \nabla_\mu \delta x_\nu + \nabla_\nu \delta x_\mu,$$
$$\delta\Gamma^\lambda_{\mu\nu} = D_\mu D_\nu \delta x^\lambda + R^\lambda{}_{\nu\tau\mu}\delta x^\tau + D_\mu(\delta x^\tau Q^\lambda_{\tau\nu})$$
$$\begin{aligned} \delta\phi^A = {} & \delta x^\tau \nabla_\tau \phi^A - \nabla_\tau \delta x^{\alpha_1} \phi^{\tau\alpha_2\cdots\alpha_k}_{\beta_1\cdots\beta_l} - \cdots \\ & - \nabla_\tau \delta x^{\alpha_k} \phi^{\alpha_1\cdots\alpha_{k-1}\tau}_{\beta_1\cdots\cdots\beta_l} \\ & + \nabla_{\beta_1}\delta x^\tau \phi^{\alpha_1\cdots\alpha_k}_{\tau\beta_2\cdots\beta_l} + \cdots + \nabla_{\beta_l}\delta x^\tau \phi^{\alpha_1\cdots\cdots\alpha_k}_{\beta_1\cdots\beta_{l-1}\tau}, \end{aligned}$$
$$\delta g_{\mu\nu,\tau} = \partial_\tau \delta g_{\mu\nu} \quad \delta\Gamma^\lambda_{\mu\nu,\tau} = \partial_\tau \delta\Gamma^\lambda_{\mu\nu} \quad \delta\phi^\alpha{}_{,\tau} = \partial_\tau \delta\phi^A. \tag{6.5}$$

Proposition 13: Let $\alpha^k = \overline{\delta x^k} = \delta x^k + N\delta x^0$, $\beta = \overline{\delta x^0} = N\delta x^0$ be the tangential and the normal components of the vector field U at points of σ. Then the components $(\delta\Pi^{pq}, \delta g_{pq}, \delta\overline{\mathscr{P}}_A, \delta\overline{\phi}^A)$ of Y are given by formula (C6) with $s = \beta$, $u^k = \alpha^k$; the components δN, δN^k are

$$\begin{aligned} \delta N &= \partial_0\beta - N^s\overline{\nabla}_s\beta + \alpha^s\overline{\nabla}_s N, \\ \delta N^k &= \partial_0\alpha^k + \alpha^s\overline{\nabla}_s N^k - N^s\overline{\nabla}_s\alpha^k + \beta\overline{\nabla}^k N - N\overline{\nabla}^k\beta. \end{aligned} \tag{6.6}$$

Remarks: (i) In order to get Proposition 13 from (6.5), we have used the canonical form of field equations (5.4).

(ii) For special cases of the Einstein and Einstein–Maxwell theories the correspodning formulas have been presented in Refs. 11–13.

(iii) It follows from (6.6) that by means of the appropriate choice of α^k, β we can get arbitrary values of δN, δN^k, $\partial_0\delta N$, $\partial_0\delta N^k$ on σ (on c_3).

Let $C_4 \in \mathrm{Sol}$. By means of (3.2) we can split C_4 into a one-parameter family of initial surfaces $({}_t c_3)_{t\in\mathbb{R}}$. We have proved in Sec. 5 that on every initial surface ${}_t c_3$ we are able to determine [from the Hamilton equation (5.2)] components $(\delta\Pi^{pq}, \delta g_{pq}, \delta\overline{\mathscr{P}}_A, \delta\overline{\phi}^A)$ of the vector field $'Y$ representing the vector of evolution $'\widehat{Y}$. Therefore we can construct a τ-vertical vector field Y on C_4. We show now that Y represents a vector $\widehat{Y}$ tangent to Sol and that this vector coincides with that given by the action $\widehat{A}$ in Sol. We can expect this result if we recall the considerations in Sec. 2 which gave rise to Theorems 1 and 1′. We prove also that $\widehat{Y}$ belongs to the degeneracy subspace of the symplectic 2-form Ω on Sol. Also this fact is understandable by virtue of (2.16) and the fact that the energy–momentum function E vanishes on Sol.

For $C_4 \in \mathrm{Sol}$ we define the vector space lin.can.var. (linearized canonical variables) which consists of systems $\mathscr{X} = (\delta\Pi^{pq}, \delta g_{pq}, \delta\mathscr{P}_A, \delta\overline{\phi}^A)$. The components of $\mathscr{X}$ represent families of geometrical 3-quantities on σ depending on the parameter x^0. We assume that they satisfy the linearized constraint equations

$$\delta(\mathrm{Constr}_0) = 0, \quad \delta(\mathrm{Constr}_p) = 0 \tag{6.7}$$

(thus $\mathscr{X} \in \ker P$).

Remark: We recall that for $C_4 \in \mathrm{Sol}$ the linearized constraints depend only on $\delta\Pi^{pq}$, δg_{pq}, $\delta\overline{\mathscr{P}}_A$, $\delta\overline{\phi}^A$. It is clear that every vector $\widehat{Y}$ tangent to Sol at C_4 defines an element $\mathscr{Y} \in$ lin.can.var.

Proposition 14: If $\widehat{Y} = d\widehat{A}(C_4)U$ [cf. (6.3)], then

(i) the corresponding element $\mathscr{Y} \in$ lin.can.var. is equal to

$$\mathscr{Y} = -JP^*(\overline{U}^0, \overline{U}^k) \quad \text{[see (C6)]}, \tag{6.8}$$

(ii) $\operatorname{im} JP^* \subset \ker P$, (6.9)

(iii) for every vector $\widehat{X}$ tangent to Sol at C_4

$$\Omega(\widehat{Y}, \widehat{X}) = 0 \tag{6.10}$$

Statement (i) follows directly from Proposition 13; (ii) is the consequence of the fact that $\widehat{Y} \in T_{C_4}(\mathrm{Sol})$; (iii) follows from (5.16) and the properties of adjoint operators.

Remark: It is possible to prove (ii) directly substituting (C6) into (C2)–(C5) but the calculations are rather long.

Corollary: The vector of evolution $'\widehat{Y}$ is determined by the vector $\widehat{Y} = d\widehat{A}(C_4)Z$, where Z is the vector field on M given by the dynamical slicing (3.1), and the isomorphism between $T_{C_4}(\mathrm{Sol})$ and $T_{c_3}((\mathrm{Is})(\sigma))$.

The degeneracy subspace W_{C_4} (the gauge subspace) of the symplectic 2-form Ω is

$$W_{C_4} = \{\widehat{Y} \in T_{C_4}(\mathrm{Sol}) \colon \Omega(\widehat{Y}, \widehat{X}) = 0 \text{ for every } \widehat{X} \in T_{C_4}(\mathrm{Sol})\}. \tag{6.11}$$

We know from (6.10) that $\operatorname{im} d\widehat{A}(C_4) \subset W_{C_4}$. The question arises whether $\operatorname{im} d\widehat{A}(C_4) = W_{C_4}$. We have the following:

Proposition 15: (i) If the matter field Lagrangian L_{mat} is regular, that is, if the momenta $\overline{\mathscr{P}}_A$ are in a 1–1 correspondence with x^0 derivatives of ϕ^A

(ii) and if Eqs. (4.5b) determine $\Gamma^\lambda_{\mu\nu}$ as function of

$$g_{\mu\nu}, \partial_\lambda g_{\mu\nu}, \phi^A, \partial_\lambda\phi^A,$$

then

$$W_{C_4} = \overline{\operatorname{im} d\widehat{A}(C_4)}, \tag{6.12}$$

where the closure is taken in the C^∞-topology of sections.

Remark: For the proof we need Remark (iii) of Proposition 13. We think that it is possible to prove that $\operatorname{im} d\widehat{A}(C_4)$ is a closed subspace, i.e.,

$$W_C = \operatorname{im} d\widehat{A}(C_4). \tag{6.12'}$$

For the Einstein theory (6.12′) has been proved independently by Fischer–Marsden–Moncrief[8,34] and the present author [11,12] by means of the theory of elliptic differential operators. It seems that the methods presented in those papers can be also adopted for more general cases.

Proposition 16: If (6.12′) holds, then

$$\ker P = (\ker P \cap \ker(PJ)) \oplus \operatorname{im} JP^* \tag{6.13}$$

For the Einstein theory the decomposition (6.13) has been proved by Moncrief[34] and, independently, by the present author.[11,12] The subspace $\ker P \cap \ker(PJ)$ describes the genuine degrees of freedom (in the tangent space), the subspace $\operatorname{im} JP^*$ describes the gauge directions.

Remarks: (i) Examples of regular matter field Lagrangians (cf. Ref. 15):

the covector field $L_{\mathrm{mat}} = \frac{1}{2}(g^{\mu\nu}g^{\alpha\beta}D_\mu\phi_\alpha D_\nu\phi_\beta - m^2\phi^\alpha\phi_\alpha)$;

the Fermi electrodynamics

$$L_{\mathrm{mat}} = -(1/16\pi)F^{\mu\nu}F_{\mu\nu} - (1/8\pi)(g^{\mu\nu}D_\mu A_\nu)^2,$$
$$F_{\mu\nu} = D_\mu A_\nu - D_\nu A_\mu;$$

(ii) If $L_{\mathrm{mat}} = 0$, the we have from (E7) $\Gamma^\lambda_{\mu\nu} = \{^{\lambda}_{\mu\nu}\}$

$-\chi_\mu \delta_\nu^\lambda$, where χ_μ is an arbitrary function on M. Therefore we have an additional degeneracy of Ω connected with the transformation $\Gamma^\lambda_{\mu\nu} \to {}'\Gamma^\lambda_{\mu\nu} = \Gamma^\lambda_{\mu\nu} - \chi_\mu \delta^\lambda_\nu$ (in the Einstein case $\Gamma^\lambda_{\mu\nu} = \Gamma^\lambda_{\nu\mu}$ and $\chi_\mu = 0$).

(iii) Example of a nonregular Lagrangian: the Einstein–Maxwell theory $L_{\rm mat} = -(1/16\pi) f_{\mu\nu} f^{\mu\nu}, f_{\mu\nu} = \partial_\mu A_\nu - \partial_\nu A_\mu$. We have $\overline{\mathscr{P}}_0 = 0$, $\overline{\mathscr{P}}_A = (0, \mathscr{D}^k)$, $\bar\phi^A = (\bar A_0, \bar A_k)$ and the additional constraint equation $\overline{\nabla}_k \mathscr{D}^k = 0$ on σ. In this case we have only three dynamical equations for momenta $\partial_0 \mathscr{D}^k = (\mathrm{rot}\mathscr{H})^k$ (cf. Ref. 13). The gauge subspace W is determined by the action of the semidirect product of Diff M and of the gradient gauge group of electrodynamics.

7. THE HAMILTON–JACOBI RELATIONS

In classical mechanics we can view the action integral

$$S = \int_{t_0}^{t} L\,(x^i(\tau), \dot x^i(\tau), \tau) d\tau$$

as a function of final values of coordinates q^i in the configuration space and a final instant of time t. (Initial values of coordinates $\mathring{q}^i$ and an initial instant of time t_0 are kept fixed). We have

$$S = S(q^i, t)$$

and

$$\frac{\partial S}{\partial q^i} = p_i, \quad \frac{\partial S}{\partial t} = -H(p_i, q^i, t)$$

(cf. Ref. 16, Chap. 21). The above formulas give us the Hamilton–Jacobi equation in classical mechanics. In the present section we deal with a similar problem in field theory.

Let F be a section of the bundle $\mathscr{P}$, $C_4 = F(M)$, $\mathscr{D}$ be a compact four-dimensional domain in M and $\partial\mathscr{D}$ be its boundary. Let $(\Psi_s)_{s\in\mathbf{R}}$ be a one-parameter family of diffeomorphisms of $\mathscr{P}$ (in general not preserving fibres of $\mathscr{P}$). This family generates a family of four-dimensional submanifolds $(\underset{s}{C}_4)_{s\in\mathbf{R}}$ in $\mathscr{P}$ and a family of domains $(\mathscr{D}_s)_{s\in\mathbf{R}}$ in M: $\underset{s}{C}_4 = \Psi_s(C_4)$; $\mathscr{D}_s = (\tau\circ\Psi_s\circ F)(\mathscr{D})$. We compute an infinitesimal change of the action integral

$$S(\underset{s}{C}_4) = \int_{F(\mathscr{D}_s)} \Theta_{\text{H-C}}. \tag{7.1}$$

Let V be the vector field in $\mathscr{P}$ defined by the family $(\Psi_s)_{s\in\mathbf{R}}$

$$V(\mathbf{p}) = \frac{d}{ds}\Psi_s(\mathbf{p})\Big|_{s=0}. \tag{7.2}$$

We have

$$\frac{d}{ds} S(\underset{s}{C}_4)\Big|_{s=0} = \int_{F(\mathscr{D})} \mathscr{L}_V \Theta_{\text{H-C}} = \int_{F(\mathscr{D})} (V \lrcorner d\Theta_{\text{H-C}} + d(V\lrcorner\Theta_{\text{H-C}}). \tag{7.3}$$

Proposition 17: If F satisfies field equation (2.1), then

$$\frac{d}{ds} S(\underset{s}{C}_4)\Big|_{s=0} = \int_{F(\partial\mathscr{D})} V\lrcorner\Theta_{\text{H-C}}. \tag{7.4}$$

Let U be the projection of V onto M

$$U = \tau_* V. \tag{7.5}$$

In local coordinates

$$V = \delta x^\lambda\, \partial/\partial x^\lambda + \delta g_{\mu\nu}\, \partial/\partial g_{\mu\nu} + \delta\Gamma^\lambda_{\mu\nu}\, \partial/\partial\Gamma^\lambda_{\mu\nu} + \delta\phi^A\, \partial/\partial\phi^A + \cdots, \tag{7.6}$$

$$U = \delta x^\lambda\, \partial/\partial x^\lambda.$$

Let X_U be the vector field in $\mathscr{P}$ defined by (2.7) and (4.10). In local coordinates

$$V - X_U = \hat\delta g_{\mu\nu}\, \partial/\partial g_{\mu\nu} + \hat\delta\Gamma^\lambda_{\mu\nu}\, \partial/\partial\Gamma^\lambda_{\mu\nu} + \hat\delta\phi^A\, \partial/\partial\phi^A + \cdots, \tag{7.7}$$

where

$$\begin{aligned} \hat\delta g_{\mu\nu} &= \delta g_{\mu\nu} - \delta x^\tau \partial_\tau g_{\mu\nu} + \mathscr{L}_U g_{\mu\nu}, \\ \hat\delta\Gamma^\lambda_{\mu\nu} &= \delta\Gamma^\lambda_{\mu\nu} - \delta x^\tau \partial_\tau \Gamma^\lambda_{\mu\nu} + \mathscr{L}_U \Gamma^\lambda_{\mu\nu}, \\ \hat\delta\phi^A &= \delta\phi^A - \delta x^\tau \partial_\tau \phi^A + \mathscr{L}_U \phi^A, \text{ etc.} \end{aligned} \tag{7.8}$$

[the corresponding Lie derivatives are given in (4.12)].

We explain now why quantities $\hat\delta g_{\mu\nu}$, $\hat\delta\Gamma^\lambda_{\mu\nu}$, $\hat\delta\phi^A$, etc. are important in our considerations. In previous sections we have dealt with geometrical quantities on a given three-dimensional surface σ in spacetime. This surface was an element of the coordinate slicing (3.6). If we perform now the action of the family $(\Psi_s)_{s\in\mathbf{R}}$ in $\mathscr{P}$ we get a family $\tilde\sigma_s$ of three-dimensional surfaces in M. However, these surfaces are not consistent with the slicing (3.6). Therefore, formulas (B2b) and (B11) for the metric $\mathbf{g}$ on $\tilde\sigma_s$, for the fundamental forms of embedding $\mathbf{K}$, $\mathbf{S}$, for the canonical components of field potentials $\bar\phi^A$ and their momenta $\overline{\mathscr{P}}_A$, for the lapse function N and the shift vector N^k are not valid in this case. We must compute these quantities by means of their geometric definitions (cf. 12, 14, 16, 54). In the first order approximation (for small $s = \epsilon$) we have on $\tilde\sigma_\epsilon$

$$\begin{aligned} &{}_\epsilon\bar g_{pq} \cong \bar g_{pq} + \epsilon\hat\delta\bar g_{pq}, \quad {}_\epsilon K_{pq} \cong K_{pq} + \epsilon\hat\delta K_{pq}, \\ &{}_\epsilon S_{pq} \cong S_{pq} + \epsilon\hat\delta S_{pq}, \\ &{}_\epsilon\bar\phi^A \cong \bar\phi^A + \epsilon\hat\delta\bar\phi^A, \quad {}_\epsilon\overline{\mathscr{P}}_A \cong \overline{\mathscr{P}}_A + \epsilon\hat\delta\overline{\mathscr{P}}_A, \\ &{}_\epsilon N \cong N + \epsilon\hat\delta N, \quad {}_\epsilon N^k \cong N^k + \epsilon\hat\delta N^k \end{aligned} \tag{7.9}$$

and $\hat\delta\bar g_{pq}$, $\hat\delta K_{pq}$, $\hat\delta S_{pq}$,, etc. can be expressed by means of (7.8) (in particular, $\hat\delta\bar g_{pq} = \hat\delta g_{pq}$). We omit here these complicated formulas.

Let us suppose that $\partial\mathscr{D}$ consists of two three-dimensional surfaces σ_0 and σ and that V vanishes on σ_0. We have from (7.4)

$$\begin{aligned} \frac{d}{ds} S(\underset{s}{C}_4)\Big|_{s=0} &= \int_{F(\sigma)} X_U \lrcorner\Theta_{\text{H-C}} + \int_{F(\sigma)} (V - X_U)\lrcorner\Theta_{\text{H-C}} \\ &= -E_U(\sigma) + \int_{F(\sigma)} [-(1/16\pi) g_{pq} \hat\delta\Pi^{pq} + \overline{\mathscr{P}}_A \hat\delta\bar\phi^A \\ &\quad - (1/32\pi)\hat\delta(\sqrt{\bar g}(S_{pq} - K_{pq})\bar g^{pq})] dx^1\wedge dx^2\wedge dx^3 \\ &\quad + (1/16\pi)\int_{F(\sigma)} \sqrt{\bar g}\,[\overline{\nabla}_p(\hat\delta N^p/N) + \overline{\nabla}_p((N^p/N)\partial_0\delta x^0 \\ &\quad + (1/N)\partial_0 \delta x^p)] dx^1\wedge dx^2\wedge dx^3. \end{aligned} \tag{7.10}$$

But C_4 satisfies field equations, and therefore, by virtue of

Proposition 10, $E_U(\sigma) = 0$. The last integral can be also omitted and we get

Proposition 18:

$$\frac{d}{ds}S(C_4)\Big|_{s=0} = \int_{F(\sigma)} \big[-(1/16\pi) g_{pq}\hat{\delta}\Pi^{pq} + \overline{\mathscr{P}}_A \hat{\delta}\bar{\phi}^A - (1/32\pi)\hat{\delta}(\sqrt{\bar{g}}(S_{pq} - K_{pq})\bar{g}^{pq})\big] dx^1 \wedge dx^2 \wedge dx^3. \tag{7.11}$$

We have expected such a result; only the term $-(1/32\pi)\hat{\delta}(\sqrt{\bar{g}}(S_{pq} - K_{pq})\bar{g}^{pq})$ is a bit surprising. Therefore, in order to get a correct formula for the generating function we substract from $\Theta_{\text{H-C}}$ the exact 4-form (cf. Ref. 14)—the proof of Theorem 1)

$$\underset{3}{\Theta} = -(1/32\pi) d(\sqrt{\bar{g}}(S_{pq} - K_{pq})\bar{g}^{pq}) dx^1 \wedge dx^2 \wedge dx^3. \tag{7.12}$$

The generating functional S_1 is

$$S_1(C_4) = \int_{F(\mathscr{S})} (\Theta_{\text{H-C}} - \underset{3}{\Theta}) \tag{7.13}$$

and

$$\frac{\delta S_1}{\hat{\delta}\Pi^{pq}} = -\frac{1}{16\pi} g_{pq}, \quad \frac{\delta S_1}{\hat{\delta}\bar{\phi}^A} = \overline{\mathscr{P}}_A. \tag{7.14}$$

Formulas (7.14) generalize those of the Einstein theory.[12,16,54]

8. CONCLUDING REMARKS

(i) In the present paper we have defined the energy–momentum (Hamiltonian) 3-form ν_Z and proved that the corresponding energy function E determines Hamilton equation (2.18) which is equivalent to the variational Euler–Lagrange equation (2.1). A similar problem was earlier investigated by Kondracki.[47] His paper outlines, however, only how to pass from E–L equation to Hamilton equation in field theories in flat Minkowski spacetime. The solution of the inverse problem presented in Ref. 47 seems to be incorrect.

Our formula for the energy momentum 3-form ν_Z, (4.13), is a generalization of Komar's formula[7] (cf. Ref. 53) and Appendix D). An elegant derivation of Komar's formula given in Ref. 55 explains to us why our result generalizes that of Komar. Recently Kijowski and Tulczyjew[38] have presented their own ingenious approach to the notion of energy–momentum in classical field theories. Kijowski[39] investigated the problem in a theory of gravity (with a symmetric, metric noncompatible connection) and derived the formula which (almost) coincides with ours (cf. Appendix D). At first glance the approach presented in Ref. 39 seems very different from ours. However, the discussion of the problem in Ref. 38 reveals deep relations between our 3-form ν_Z and their original definition of energy-momentum.

(ii) The theory developed in this paper works only for spatially closed spacetimes, that is, M is diffeomorphic to $\mathbb{R}\times\sigma$, where σ is a compact three-dimensional manifold without boundary. For a noncompact σ we have to take into account boundary integrals in the formula for the symplectic 2-form Ω, (4.9), in the definition of the energy–momentum function E, (4.15)–(4.20), and in the formula for JP^*, (C6). A detailed discussion on these questions will be presented in the next paper. We shall show their relations between our approach and those developed earlier by ADM,[10] DeWitt,[56] Regge–Teitelboim,[22,57] Murchadha–York,[29] and Choquet–Bruhat, Fischer, and Marsden[58] for spatially noncompact spacetimes.

(iii) In the present paper matter is described by a finite representation of the group of general coordinate transformations (the local GL (4, $\mathbb{R}$) group). Therefore we have only tensor matter fields. In order to describe SL(2, $\mathbb{C}$) spinor matter fields (the Dirac field), one has to work in the framework of the tetrad formulation of gravity (gauge formulation) and with the local Lorentz group as the transformation group for fields of tetrad.[26,51,52] In this case the connection on M is always metric compatible, i.e., $D_\lambda g_{\mu\nu} = 0$. One of our next papers[52] is devoted to problems of the Hamiltonian formulation of theories of gravity in the language of tetrad fields. Spinor matter fields can be also discussed in this schema.

Remark: Recently Ne'eman and Šijački have discovered some nontensorial representations of GL(4, $\mathbb{R}$). Corresponding bandors (generalized spinors) are to be related to strong interactions.[59]

ACKNOWLEDGMENTS

I would like to thank Professor F. W. Hehl for his kind hospitality at the Institute for Theoretical Physics, University of Cologne, where this paper has been written in its final form. I am also very indebted to the Humboldt Foundation for the award of the fellowship in 1979/80. My special thanks are due to Dr. S. Krebs for her help during the preparation of the manuscript.

APPENDIX A: LINEARIZED FIELD EQUATIONS

Let F be a section of $\mathscr{P}$ satisfying (2.1) and $C_4 = F(M)$. A vector field Y on C_4 (tangent to $\mathscr{P}$) represents a vector $\hat{Y} \in T(\text{Sol})$ if and only if for any extension $\widetilde{Y}$ of Y onto a neighborhood $\mathscr{U}$ of C_4 in $\mathscr{P}$ and for every vector field $\widetilde{X}$ in $\mathscr{U}$ we have (cf. Refs. 3, 12)

$$(\mathscr{L}_{\widetilde{Y}}(\widetilde{X} \lrcorner d\Theta_{\text{H-C}})) \,|\, C_4 = 0. \tag{A1}$$

Using properties of Lie derivatives and (2.1), we see that (A1) is equivalent to

$$(\widetilde{X} \lrcorner d(\widetilde{Y} \lrcorner d\Theta_{\text{H-C}})) \,|\, C_4 = 0 \tag{A2}$$

or

$$(\widetilde{X} \lrcorner d(\widetilde{Y} \lrcorner d\Theta_{\text{H-C}}))(\widetilde{Z}_0, \widetilde{Z}_1, \widetilde{Z}_2, \widetilde{Z}_3) = 0 \quad \text{on } C_4,$$

where $\widetilde{Z}_\lambda$ are vector fields given in $\mathscr{U} \subset \mathscr{P}$ and tangent to C_4 at points of C_4.

On the other hand, we have from (A1)

$$(\widetilde{Y} \lrcorner d(\widetilde{X} \lrcorner d\Theta_{\text{H-C}}) + d(\widetilde{Y} \lrcorner \widetilde{X} \lrcorner d\Theta_{\text{H-C}})) \,|\, C_4 = 0 \tag{A3}$$

or equivalently (cf. Refs. 3, 12)

$$\widetilde{Y}((\widetilde{X} \lrcorner d\Theta_{\text{H-C}})(\widetilde{Z}_0, \widetilde{Z}_1, \widetilde{Z}_2, \widetilde{Z}_3)) + \sum_{\lambda=0}^{3} (-1)^{\lambda+1} (\widetilde{X} \lrcorner d\Theta_{\text{H-C}})([\widetilde{Y}, \widetilde{Z}_\lambda], \widetilde{Z}_0, \ldots\hat{\lambda}\ldots, \widetilde{Z}_3) = 0 \quad \text{on } C_4.$$

We have two equivalent formulations of linearized field equations for components of Y. We see from (A2) that these equations do not involve partial derivatives of $\widetilde{X}, \widetilde{Z}$ and from (A3) we see that they depend only on values of $\widetilde{Y}$ on C_4 and derivatives of $\widetilde{Y}$ in directions tangent to C_4.

Formulas (A2)–(A3) imply also the following facts:

(i) We can consider only τ-vertical vector fields $\widetilde{X}, \widetilde{Y}$.

(ii) Condition (A1) does not depend on prolongations of $\widetilde{X}, \widetilde{Y}$ beyond the submanifold C_4 in $\mathscr{P}$.

(iii) In a local coordinate system in $\mathscr{P}$ we get system of linear partial differential equations for vertical components of Y.

APPENDIX B: THE 1 +3 DECOMPOSITION OF GEOMETRICAL OBJECTS IN SPACETIME

In Ref. 14 we have defined the "bar" operation for geometrical objects in spacetime M. This operation gives us decompositions of tensor fields on M into normal and tangential components to a given three-dimensional surface σ in M. We present now a somewhat different approach, which works also beyond σ in M. Let $\sigma_t = \{ \mathbf{x} \in M : x^0 = t \}$ be a slicing of M in a family of three-dimensional submanifolds. Let $\mathbf{g} = (g_{\mu\nu})$ be a Lorentz metric on M such that all surfaces σ_t are spacelike with respect to $\mathbf{g}$. The lapse function and the shift vector field are defined by

$$N = (-g^{00})^{-1/2}, \quad N^k = \bar{g}^{ks} g_{s0},$$

where $[\bar{g}^{ks}]$ is the inverse matrix of $[g_{ks}]$. We define the matrices $[A^{\bar{\mu}}_{\ \nu}]$, $[B^{\ \nu}_{\bar{\mu}}]$, where $\mathbf{B} = \mathbf{A}^{-1}$:

$$A^{\bar{0}}_{\ 0} = N, \quad A^{\bar{k}}_{\ 0} = N^k, \quad A^{\bar{0}}_{\ s} = 0, \quad A^{\bar{k}}_{\ s} = \delta^k_s, \tag{B1}$$

$$B^{\ 0}_{\bar{0}} = 1/N, \quad B^{\ k}_{\bar{0}} = -N^k/N, \quad B^{\ 0}_{\bar{s}} = 0, \quad B^{\ k}_{\bar{s}} = \delta^k_s,$$

and $N = N(x^0, x^s)$, $N^k = N^k(x^0, x^s)$. For a tensor field $\boldsymbol{\phi}$ on M we define

$$\bar{\phi}^{\alpha_1 \cdots \alpha_k}_{\ \beta_1 \cdots \beta_s} = A^{\bar{\alpha}_1}_{\ \mu_1} \cdots A^{\bar{\alpha}_k}_{\ \mu_k} B^{\ \nu_1}_{\bar{\beta}_1} \cdots B^{\ \nu_s}_{\bar{\beta}_s} \phi^{\mu_1 \cdots \mu_k}_{\ \nu_1 \cdots \nu_s}. \tag{B2}$$

Geometrically $\bar{\phi}^A(t, x^k)$ are the normal and tangential components of $\boldsymbol{\phi}$ on the submanifold σ_t; e.g., for a vector v^α we have

$$\bar{v}^0 = N v^0 = -\mathbf{v}\cdot\mathbf{n}, \quad \bar{v}^k = v^k + N^k v^0 = v^k + (\mathbf{v}\cdot\mathbf{n}) n^k, \tag{B2a}$$

where $\mathbf{n} = (n^\lambda) = (1/N, -N^k/N)$ is the unit vector field normal to the slicing σ_t. Definition (B2) coincides for $t = 0$ with that given in Ref. 14. For a given tensor field $\boldsymbol{\phi}$ on M, (B2) gives a family of tensor fields on σ_t. The valency of $\bar{\phi}^{\alpha_1 \cdots \alpha_k}_{\ \beta_1 \cdots \beta_s}$ is determined by nonzero indices; the index 0 refers to scalars. For the metric tensor $\mathbf{g}$ we have

$$\bar{g}_{00} = -1, \quad \bar{g}_{0k} = \bar{g}_{k0} = 0, \quad \bar{g}_{pq} = g_{pq}, \tag{B2b}$$

$$\bar{g}^{00} = -1, \quad \bar{g}^{0k} = \bar{g}^{k0} = 0, \quad \bar{g}^{pq} = [g_{pq}]^{-1}.$$

Other examples (cf. Ref. 14)

$$\bar{a}_{pq} = a_{pq}, \quad \bar{b}^{0}_{\ 0} = b^0_{\ 0} - N^s b^0_{\ s}, \quad \bar{b}^0_{\ k} = N b^0_{\ k}, \cdots. \tag{B2c}$$

Remark: We can also treat $\bar{\phi}^A$ as a family of tensor fields on $\sigma = \sigma_0$ parametrized by x^0. At points $\mathbf{x} \in \sigma_t$ quantities $\bar{\phi}^A$ coincide with corresponding components of $\boldsymbol{\phi}$ in the normal coordinate system on σ_t. However, we cannot treat $\bar{\phi}^A$ as components of $\boldsymbol{\phi}$ in some coordinate system in M because equations

$$\frac{\partial x^{\bar{\mu}}}{\partial x^\nu} = A^{\bar{\mu}}_{\ \nu}$$

are not integrable.

We define formal differential operators

$$\bar{\partial}_\mu = B^{\ \alpha}_{\bar{\mu}} \partial_\alpha \tag{B3}$$

and formal coefficients $\bar{\Gamma}^\lambda_{\mu\nu}$, where $\mathbf{\Gamma} = (\Gamma^\lambda_{\mu\nu})$ is an affine connection on M

$$\bar{\Gamma}^\lambda_{\mu\nu} = A^{\bar{\lambda}}_{\ \tau} B^{\ \alpha}_{\bar{\mu}} B^{\ \beta}_{\bar{\nu}} \Gamma^\tau_{\alpha\beta} + A^{\bar{\lambda}}_{\ \tau} \bar{\partial}_\mu B^{\ \tau}_{\bar{\nu}}, \tag{B4}$$

$$\Gamma^\tau_{\alpha\beta} = B^{\ \tau}_{\bar{\lambda}} A^{\bar{\mu}}_{\ \alpha} A^{\bar{\nu}}_{\ \beta} \bar{\Gamma}^\lambda_{\mu\nu} + B^{\ \tau}_{\bar{\lambda}} \partial_\alpha A^{\bar{\lambda}}_{\ \beta}.$$

For the Riemannian connection $\gamma^\lambda_{\mu\nu} = \{^{\ \lambda}_{\mu\nu}\}$ in M, we have

$$\bar{\gamma}^0_{00} = 0, \quad \bar{\gamma}^0_{k0} = 0,$$
$$\bar{\gamma}^p_{00} = (1/N) \bar{\nabla}^p N, \quad \bar{\gamma}^s_{k0} = N \bar{g}^{su} \gamma^0_{uk}, \tag{B5}$$

$$\bar{\gamma}^0_{0k} = (1/N) \bar{\nabla}_k N, \quad \bar{\gamma}^0_{ks} = N \gamma^0_{ks},$$
$$\bar{\gamma}^s_{0k} = N \bar{g}^{su} \gamma^0_{uk} + (1/N) \partial_k N^s, \quad \bar{\gamma}^r_{ks} = \gamma^r_{ks} + N^r \gamma^0_{ks}.$$

Remarks: (i) $\bar{\gamma}^\lambda_{\mu\nu}$ is not a symmetric quantity.

(ii) $\bar{\gamma}^\lambda_{\mu\nu}$ on σ_t do not coincide with the corresponding components of $\boldsymbol{\gamma}$ in the normal coordinate system on σ_t [cf. Ref. 14 formulas (A.2.8-9)]. It is only true for the elements of the second column of (B 5).

(iii) Property (ii) holds also for an arbitrary connection $\mathbf{\Gamma}$.

(iv) Because of (iii), the change of the notation in comparison with Ref. 14 does not affect the important formulas of that paper.

(v) Quantities $\bar{\gamma}^r_{ks}$ are coefficients of the Riemannian connection given by the metric $\bar{\mathbf{g}} = (g_{ij})$ on σ_t (cf. E19); all other quantities (with exception of $\bar{\gamma}^s_{0k}$) have transformation properties of 3-tensor fields on σ_t.

For the covariant derivative of a tensor field $\boldsymbol{\phi}$ on M we have according to (B2)

$$\overline{D_\lambda \phi}^{\alpha_1 \cdots \alpha_k}_{\ \beta_1 \cdots \beta_s} = B^{\ \tau}_{\bar{\lambda}} A^{\bar{\alpha}_1}_{\ \mu_1} \cdots A^{\bar{\alpha}_k}_{\ \mu_k} B^{\ \nu_1}_{\bar{\beta}_1} \cdots B^{\ \nu_s}_{\bar{\beta}_s} D_\tau \phi^{\mu_1 \cdots \mu_k}_{\ \nu_1 \cdots \nu_s} \tag{B6}$$

and

$$\begin{aligned}\overline{D_\lambda \phi}^{\alpha_1 \cdots \alpha_k}_{\ \beta_1 \cdots \beta_s} = {}& \bar{\partial}_\lambda \bar{\phi}^{\alpha_1 \cdots \alpha_k}_{\ \beta_1 \cdots \beta_s} + \bar{\Gamma}^{\alpha_1}_{\lambda\tau} \bar{\phi}^{\tau \alpha_2 \cdots \alpha_k}_{\ \beta_1 \cdots \beta_s} + \cdots + \bar{\Gamma}^{\alpha_k}_{\lambda\tau} \bar{\phi}^{\alpha_1 \cdots \alpha_k}_{\ \beta_1 \cdots \beta_s}{}^{,\tau} \\ & - \bar{\Gamma}^\epsilon_{\lambda\beta_1} \bar{\phi}^{\alpha_1 \cdots \alpha_k}_{\ \epsilon\beta_2 \cdots \beta_s} - \cdots - \bar{\Gamma}^\epsilon_{\lambda\beta_s} \bar{\phi}^{\alpha_1 \cdots \cdots \alpha_k}_{\ \beta_1 \cdots \beta_s , \epsilon}.\end{aligned} \tag{B7}$$

We have the following formulas for the Riemann tensor, the difference tensor, and torsion:

$$\begin{aligned}\bar{R}^\alpha_{\ \beta\mu\nu} &= A^{\bar{\alpha}}_{\ \tau} B^{\ \sigma}_{\bar{\beta}} B^{\ \omega}_{\bar{\mu}} B^{\ \epsilon}_{\bar{\nu}} R^\tau_{\ \sigma\omega\epsilon} \\ &= \bar{\partial}_\mu \bar{\Gamma}^\alpha_{\nu\beta} - \bar{\partial}_\nu \bar{\Gamma}^\alpha_{\mu\beta} + \bar{\Gamma}^\omega_{\nu\beta} \bar{\Gamma}^\alpha_{\mu\omega} - \bar{\Gamma}^\omega_{\mu\beta} \bar{\Gamma}^\alpha_{\nu\omega} \\ &\quad - (\bar{\partial}_\mu B^{\ \epsilon}_{\bar{\nu}} - \bar{\partial}_\nu B^{\ \epsilon}_{\bar{\mu}}) A^{\bar{\omega}}_{\ \epsilon} \bar{\Gamma}^\alpha_{\omega\beta},\end{aligned} \tag{B8}$$

$$\bar{r}^\lambda_{\mu\nu} = \bar{\Gamma}^\lambda_{\mu\nu} - \bar{\gamma}^\lambda_{\mu\nu}, \tag{B9}$$

$$\bar{Q}^\lambda_{\mu\nu} = \bar{r}^\lambda_{\mu\nu} - \bar{r}^\lambda_{\nu\mu} = \bar{\Gamma}^\lambda_{\mu\nu} - \bar{\Gamma}^\lambda_{\nu\mu} - A^{\bar{\lambda}}_{\ \tau} (\bar{\partial}_\mu B^{\ \tau}_{\bar{\nu}} - \bar{\partial}_\nu B^{\ \tau}_{\bar{\mu}}). \tag{B10}$$

The bar operation commutes with the contraction of tensors and with their covariant differentiations.

The fundamental forms of the embedding $i:\sigma \to M$ can be expressed by means of the bar operation (cf. Ref. 14)

$$K_{ij} = -\bar{\Gamma}^{s}_{i0} g_{sj}, \quad S_{ij} = -\bar{\Gamma}^{0}_{ij}; \tag{B11}$$

the covariant derivative $\bar{\mathbf{D}}$ for a vector field Y^k on σ is given by

$$\bar{D}_s Y^k = \partial_s Y^k + \bar{\Gamma}^{k}_{sp} Y^p, \tag{B12}$$

that is, $\bar{\mathbf{\Gamma}} = (\bar{\Gamma}^{k}_{sp})$ is the connection induced on σ by $\mathbf{\Gamma}$. The gravitational canonical momenta are [see (4.6)]

$$\Pi^{pq} = -\tfrac{1}{2}\sqrt{\bar{g}}\,[(S_{ab} + S_{ba})\bar{g}^{ap}\bar{g}^{bq} - \bar{g}^{pq}(S_{ab}\bar{g}^{ab} + K_{ab}\bar{g}^{ab})]. \tag{B13}$$

APPENDIX C: THE LINEARIZED CONSTRAINTS—OPERATOR P

It follows from (4.6), (4.8), (E7), (E13) that

$$\begin{aligned}
&\tfrac{1}{2}\sqrt{\bar{g}}(\bar{\Gamma}^{0}_{pq} + \bar{\Gamma}^{0}_{qp}) = (\Pi_{pq} - \tfrac{1}{2}g_{pq}\operatorname{tr}\Pi) + \tfrac{1}{4}g_{pq}(\sqrt{\bar{g}}\bar{z}^{00}_{0} - N\mathscr{M}),\\
&\sqrt{\bar{g}}\bar{\Gamma}^{p}_{p0} = -\tfrac{1}{4}(\sqrt{\bar{g}}\bar{z}^{00}_{0} - N\mathscr{M} + 2\operatorname{tr}\Pi),\\
&\sqrt{\bar{g}}\bar{\Gamma}^{0}_{k0} = \sqrt{\bar{g}}(\bar{\Gamma}^{p}_{pk} - \bar{\gamma}^{p}_{pk}) + \sqrt{\bar{g}}\bar{z}^{00}_{k} - N\mathscr{M}_k.
\end{aligned} \tag{C1}$$

Remark: The terms $\sqrt{\bar{g}}\,\bar{z}^{00}_{\lambda}$ can be expressed by $\bar{\phi}^A$, $\bar{\mathscr{P}}_A$, (E12), and $\bar{\gamma}^{p}_{pk}$ can be expressed by g_{ab} and their spatial derivatives (E19). If

$$\underset{\sim}{s} = \delta(\mathrm{Constr}_0), \tag{C2}$$

then

$$\begin{aligned}
\underset{\sim}{s} = {} & \sqrt{\bar{g}}({}^{(3)}R^{ab} - \tfrac{1}{2}\bar{g}^{ab}\,{}^{(3)}R)\delta g_{ab} - \sqrt{\bar{g}}\bar{\nabla}^a\bar{\nabla}^b\delta g_{ab} + \sqrt{\bar{g}}\bar{g}^{ab}\bar{\nabla}^k\bar{\nabla}_k\delta g_{ab} + (2/\sqrt{\bar{g}})\delta\Pi^{pq}(\pi_{pq} - \tfrac{1}{2}g_{pq}\operatorname{tr}\pi) - (1/2\sqrt{\bar{g}})(\pi_{pq}\pi^{pq}\\
& - \tfrac{1}{2}(\operatorname{tr}\pi)^2)\bar{g}^{ab}\delta g_{ab} + (2/\sqrt{\bar{g}})(\pi^{p}{}_{b}\pi^{bq} - \tfrac{1}{2}\pi^{pq}\operatorname{tr}\pi)\delta g_{pq} - \bar{\nabla}_p\delta(\sqrt{\bar{g}}\bar{z}^{0p}_{0}) - \bar{\nabla}_p\delta(\sqrt{\bar{g}}\bar{z}^{00}_{s})\bar{g}^{sp} - \bar{\nabla}_p(\sqrt{\bar{g}}\bar{z}^{00}_{s})\delta\bar{g}^{sp} - \sqrt{\bar{g}}\bar{z}^{00}_{s}\bar{\nabla}_p\delta\bar{g}^{sp}\\
& + \bar{\gamma}^{0}_{pq}\delta(\sqrt{\bar{g}}\bar{z}^{0p}_{s})\bar{g}^{sp} + \delta(\sqrt{\bar{g}}\bar{z}^{0\lambda}_{\tau})\bar{r}^{\tau}_{0\lambda} + \bar{\gamma}^{0}_{pq}(\sqrt{\bar{g}}\bar{z}^{0p}_{s})\delta\bar{g}^{sq} + \bar{\gamma}^{0}_{pq}(\sqrt{\bar{g}}\bar{z}^{p0}_{s})\delta\bar{g}^{sq} - \tfrac{1}{4}\sqrt{\bar{g}}\bar{g}^{ab}\delta g_{ab}\bar{c}^{\alpha\beta}_{\lambda}\bar{r}^{\lambda}_{\alpha\beta} + \tfrac{1}{2}\sqrt{\bar{g}}(-\bar{c}^{\alpha\beta}_{p}\bar{r}^{p}_{\alpha\beta}\\
& + \bar{c}^{\alpha s}_{\tau}\bar{r}^{\tau}_{\alpha p} + \bar{c}^{s\alpha}_{\tau}\bar{r}^{\tau}_{p\alpha})\bar{g}^{pu}\delta g_{su} - 8\pi\sqrt{\bar{g}}\,{}_{\mathrm{sym}}T_{pq}\bar{g}^{pa}\bar{g}^{qb}\delta g_{ab} - 16\pi\sqrt{\bar{g}}\bar{j}_A\delta\bar{\phi}^A + 16\pi\delta\bar{\mathscr{P}}_A\,\overline{D_0\phi^A} - 16\pi\sqrt{\bar{g}}\bar{p}^{r\beta_1\cdots\beta_s}_{\alpha_1\cdots\alpha_k}\cdot\bar{\nabla}_r\delta\bar{\phi}^{\alpha_1\cdots\alpha_k}_{\beta_1\cdots\beta_s}\\
& + \tfrac{1}{2}\sqrt{\bar{g}}\bar{z}^{kp}_{q}\bar{g}^{qa}(\bar{\nabla}_k\delta g_{pa} + \bar{\nabla}_p\delta g_{ka} - \bar{\nabla}_a\delta g_{kp}) - 16\pi\sqrt{\bar{g}}(\bar{p}^{a\nu_1\cdots\nu_s}_{\epsilon\mu_2\cdots\mu_k}\delta\bar{\phi}^{\tau\mu_2\cdots\mu_k}_{\nu_1\cdots\nu_s} + \cdots + \bar{p}^{a\nu_1\cdots\nu_s}_{\mu_1\cdots\mu_{k-1}\epsilon}\delta\bar{\phi}^{\mu_1\cdots\mu_{k-1}\tau}_{\nu_1\cdots\nu_s}\\
& - \bar{p}^{a\tau\nu_2\cdots\nu_s}_{\mu_1\cdots\mu_k}\delta\bar{\phi}^{\mu_1\cdots\mu_k}_{\epsilon\nu_2\cdots\nu_s} - \cdots - \bar{p}^{a\nu_1\cdots\nu_{s-1}\tau}_{\mu_1\cdots\mu_k}\delta\bar{\phi}^{\mu_1\cdots\mu_k}_{\nu_1\cdots\nu_{s-1}\epsilon})\bar{r}^{\epsilon}_{a\tau} - 16\pi\sqrt{\bar{g}}(\bar{p}^{a\nu_1\cdots\nu_s}_{0\mu_2\cdots\mu_k}\delta\bar{\phi}^{q\mu_2\cdots\mu_k}_{\nu_1\cdots\nu_s} + \cdots + \bar{p}^{a\nu_1\cdots\nu_s}_{\mu_1\cdots\mu_{k-1}0}\delta\bar{\phi}^{\mu_1\cdots\mu_{k-1}q}_{\nu_1\cdots\nu_s})\bar{\gamma}^{0}_{aq}\\
& - 16\pi\sqrt{\bar{g}}(\bar{p}^{a\nu_1\cdots\nu_s}_{r\mu_2\cdots\mu_k}\delta\bar{\phi}^{0\mu_2\cdots\mu_k}_{\nu_1\cdots\nu_s} + \cdots + \bar{p}^{a\nu_1\cdots\nu_s}_{\mu_1\cdots\mu_{k-1}r}\delta\bar{\phi}^{\mu_1\cdots\mu_{k-1}0}_{\nu_1\cdots\nu_s})\bar{\gamma}^{r}_{a0} + 16\pi\sqrt{\bar{g}}(\bar{p}^{a0\nu_2\cdots\nu_s}_{\mu_1\cdots\mu_k}\delta\bar{\phi}^{\mu_1\cdots\mu_k}_{r\nu_2\cdots\nu_s} + \cdots + \bar{p}^{a\nu_1\cdots\nu_{s-1}0}_{\mu_1\cdots\mu_k}\delta\bar{\phi}^{\mu_1\cdots\mu_k}_{\nu_1\cdots\nu_{s-1}r})\bar{\gamma}^{r}_{a0}\\
& + 16\pi\sqrt{\bar{g}}(\bar{p}^{aq\nu_2\cdots\nu_s}_{\mu_1\cdots\mu_k}\delta\bar{\phi}^{\mu_1\cdots\mu_k}_{0\nu_2\cdots\nu_s} + \cdots + \bar{p}^{a\nu_1\cdots\nu_{s-1}q}_{\mu_1\cdots\mu_k}\delta\bar{\phi}^{\mu_1\cdots\mu_k}_{\nu_1\cdots\nu_{s-1}0})\bar{\gamma}^{0}_{aq} - \sqrt{\bar{g}}(\overline{\text{Eq. II}})^{[ab]}_{0}\delta\bar{\Gamma}^{0}_{[ab]} - \sqrt{\bar{g}}(\overline{\text{Eq. II}})^{k0}_{s}(\delta\bar{\Gamma}^{s}_{k0} - \tfrac{1}{3}\delta^{s}_{k}\delta\bar{\Gamma}^{c}_{c0})\\
& - \sqrt{\bar{g}}(\overline{\text{Eq. II}})^{0\mu}_{\lambda}\delta\bar{\Gamma}^{\lambda}_{0\mu} - \sqrt{\bar{g}}[(\overline{\text{Eq. II}})^{pq}_{s} + \delta^{p}_{s}(\overline{\text{Eq. II}})^{q0}_{0}]\cdot\delta\bar{\Gamma}^{s}_{pq} - \sqrt{\bar{g}}(\overline{\text{Eq. II}})^{(ab)}_{0}\delta\bar{\Gamma}^{0}_{(ab)} - \tfrac{1}{3}\sqrt{\bar{g}}(\overline{\text{Eq. II}})^{k0}_{k}\delta\bar{\Gamma}^{c}_{c0}\\
& - \sqrt{\bar{g}}(\overline{\text{Eq. II}})^{k0}_{0}\delta(\bar{\Gamma}^{0}_{k0} - \bar{\Gamma}^{p}_{pk}) + \sqrt{\bar{g}}(\overline{\text{Eq. II}})^{0b}_{0}\delta[(1/N)\partial_b N] + \tfrac{1}{2}\sqrt{\bar{g}}(\overline{\text{Eq. II}})^{ab}_{k}\bar{g}^{ks}(\bar{\nabla}_a\delta g_{sb} + \bar{\nabla}_b\delta g_{sa} - \bar{\nabla}_s\delta g_{ab})\\
& + \sqrt{\bar{g}}(\overline{\text{Eq. II}})^{0b}_{k}\{\delta\bar{g}^{ku}\bar{\gamma}^{0}_{ub} + \delta[(1/N)\partial_b N^k]\} + \sqrt{\bar{g}}(\overline{\text{Eq. II}})^{00}_{k}\delta[(1/N)\bar{g}^{ku}\partial_u N] + \sqrt{\bar{g}}(\overline{\text{Eq. II}})^{a0}_{k}\delta\bar{g}^{ku}\bar{\gamma}^{0}_{au}.
\end{aligned} \tag{C3}$$

Remark: ${}^{(3)}R^{pq}$, ${}^{(3)}R$ are the Ricci tensor and the scalar curvature of the metric $\bar{\mathbf{g}} = (g_{ij})$ on σ. If

$$\underset{\sim}{u}_k = \delta(\mathrm{Constr}_k), \tag{C4}$$

then

$$\underset{\sim}{u}_k = -2\bar{\nabla}_p(\delta g_{ks}\Pi^{ps}) + \Pi^{ab}\bar{\nabla}_k\delta g_{ab} - 2g_{ks}\bar{\nabla}_p\delta\Pi^{ps} - \bar{\nabla}_p(\delta(\sqrt{\bar{g}}\bar{z}^{0p}_{k})) + 16\pi\delta\bar{\mathscr{P}}_A\bar{\nabla}_k\bar{\phi}^A + 16\pi\bar{\mathscr{P}}_A\bar{\nabla}_k\delta\bar{\phi}^A \tag{C5}$$

Lemma: If

$$(\text{Eq. II})^{\mu\nu}_{\lambda} = 0, \quad \mathscr{Y} = -JP^*(s, u^k), \quad \mathscr{Y} = (\delta\Pi^{pq}, \delta g_{pq}, \delta\bar{\mathscr{P}}_A, \delta\bar{\phi}^A, \delta\mathscr{M}, \delta N, \delta\mathscr{M}_k, \delta N^k),$$

then

$$\begin{aligned}
\delta\Pi^{pq} = {} & \bar{\nabla}_r(\Pi^{pq}u^r) - \Pi^{rq}\bar{\nabla}_r u^p - \Pi^{rp}\bar{\nabla}_r u^q - \sqrt{\bar{g}}({}^{(3)}R^{pq} - \tfrac{1}{2}\bar{g}^{pq}\,{}^{(3)}R)s + \sqrt{\bar{g}}\bar{\nabla}^p\bar{\nabla}^q s\\
& - \bar{g}^{pq}\sqrt{\bar{g}}\bar{\nabla}^k\bar{\nabla}_k s - (2/\sqrt{\bar{g}})(\pi^{p}{}_{a}\pi^{aq} - \tfrac{1}{2}\pi^{pq}\operatorname{tr}\pi)s + (1/2\sqrt{\bar{g}})\bar{g}^{pq}(\pi_{ab}\pi^{ab} - \tfrac{1}{2}(\operatorname{tr}\pi)^2)s + \tfrac{1}{2}\sqrt{\bar{g}}\bar{z}^{00}_{r}\bar{g}^{rq}\bar{\nabla}^p s + \tfrac{1}{2}\sqrt{\bar{g}}\bar{z}^{00}_{r}\bar{g}^{rp}\bar{\nabla}^q s\\
& + \tfrac{1}{4}\sqrt{\bar{g}}\bar{c}^{\alpha\beta}_{\lambda}\bar{r}^{\lambda}_{\alpha\beta}\bar{g}^{pq}s + 8\pi\,{}_{\mathrm{sym}}T_{ab}\bar{g}^{ap}\bar{g}^{bq}\sqrt{\bar{g}}s + \tfrac{1}{2}(\bar{z}^{r0}_{u} + \bar{z}^{0r}_{u})\bar{g}^{up}(\pi^{q}{}_{r} - \tfrac{1}{2}\delta^{q}_{r}\operatorname{tr}\pi)s + \tfrac{1}{2}(\bar{z}^{r0}_{u} + \bar{z}^{0r}_{u})\bar{g}^{uq}(\pi^{p}{}_{r} - \tfrac{1}{2}\delta^{p}_{r}\operatorname{tr}\pi)s\\
& + \tfrac{1}{4}\sqrt{\bar{g}}[\bar{\nabla}_r(\bar{z}^{rq}_{u}s)\bar{g}^{up} + \bar{\nabla}_r(\bar{z}^{rp}_{u}s)\bar{g}^{uq} + \bar{\nabla}_r(\bar{z}^{qr}_{u}s)\bar{g}^{up} + \bar{\nabla}_r(\bar{z}^{pr}_{u}s)\bar{g}^{uq}] - \tfrac{1}{4}\sqrt{\bar{g}}\bar{\nabla}^r(\bar{z}^{pq}_{r}s + \bar{z}^{qp}_{r}s) + \tfrac{1}{4}\sqrt{\bar{g}}(\bar{c}^{\alpha\beta}_{u}\bar{r}^{q}_{\alpha\beta}\bar{g}^{up}\\
& + \bar{c}^{\alpha\beta}_{u}\bar{r}^{p}_{\alpha\beta}\bar{g}^{uq})s - \tfrac{1}{4}\sqrt{\bar{g}}(\bar{c}^{\epsilon q}_{\tau}\bar{r}^{\tau}_{\epsilon u}\bar{g}^{up} + \bar{c}^{\epsilon p}_{\tau}\bar{r}^{\tau}_{\epsilon u}\bar{g}^{uq} + \bar{c}^{q\epsilon}_{\tau}\bar{r}^{\tau}_{u\epsilon}\bar{g}^{up} + \bar{c}^{p\epsilon}_{\tau}\bar{r}^{\tau}_{u\epsilon}\bar{g}^{uq})s,
\end{aligned} \tag{C6a}$$

$$\delta g_{pq} = \bar{\nabla}_p u_q + \bar{\nabla}_q u_p + (2/\sqrt{\bar{g}})(\pi_{pq} - \tfrac{1}{2}g_{pq}\operatorname{tr}\pi)s, \tag{C6b}$$

$$\begin{aligned}
\delta \overline{\mathscr{P}}{}^{\beta_1\cdots\beta_s}_{\alpha_1\cdots\alpha_k} &= \overline{\nabla}_c(u^c \overline{\mathscr{P}}{}^{\beta_1\cdots\beta_s}_{\alpha_1\cdots\alpha_k}) + \sqrt{\overline{g}}\overline{J}{}^{\beta_1\cdots\beta_s}_{\alpha_1\cdots\alpha_k} s - \sqrt{\overline{g}}\overline{\nabla}_c(s\cdot\overline{p}{}^{c\beta_1\cdots\beta_s}_{\alpha_1\cdots\alpha_k}) + s\sqrt{\overline{g}}(\overline{r}{}^{\epsilon}_{\tau\alpha_1}\overline{p}{}^{\tau\beta_1\cdots\beta_s}_{\epsilon\cdots\alpha_k} + \cdots + \overline{r}{}^{\epsilon}_{\tau\alpha_k}\overline{p}{}^{\tau\beta_1\cdots\beta_s}_{\alpha_1\cdots\epsilon} \\
&\quad - \overline{r}{}^{\beta_1}_{\tau\lambda}\overline{p}{}^{\tau\lambda\cdots\beta_s}_{\alpha_1\cdots\alpha_k} - \cdots - \overline{r}{}^{\beta_s}_{\tau\lambda}\overline{p}{}^{\tau\beta_1\cdots\lambda}_{\alpha_1\cdots\alpha_k}) + s\sqrt{\overline{g}}(- \delta^{\beta_1}_0\overline{\gamma}{}^{0}_{ur}\overline{p}{}^{ur\cdots\beta_s}_{\alpha_1\cdots\alpha_k} - \delta^{\beta_1}_q\overline{\gamma}{}^{q}_{u0}\overline{p}{}^{u0\cdots\beta_s}_{\alpha_1\cdots\alpha_k} - \cdots - \delta^{\beta_s}_0\overline{\gamma}{}^{0}_{ur}\overline{p}{}^{u\beta_1\cdots r}_{\alpha_1\cdots\alpha_k} \\
&\quad - \delta^{\beta_s}_q\overline{\gamma}{}^{q}_{u0}\overline{p}{}^{u\beta_1\cdots 0}_{\alpha_1\cdots\alpha_k}) + s\sqrt{\overline{g}}(\delta^0_{\alpha_1}\overline{\gamma}{}^{r}_{u0}\overline{p}{}^{u\beta_1\cdots\beta_s}_{r\cdots\alpha_k} + \delta^i_{\alpha_1}\overline{\gamma}{}^{0}_{ui}\overline{p}{}^{u\beta_1\cdots\beta_s}_{0\cdots\alpha_k} + \cdots + \delta^0_{\alpha_k}\overline{\gamma}{}^{r}_{u0}\overline{p}{}^{u\beta_1\cdots\beta_s}_{\alpha_1\cdots r} + \delta^{\,i}_{\alpha_k}\overline{\gamma}{}^{0}_{ui}\overline{p}{}^{u\beta_1\cdots\beta_s}_{\alpha_1\cdots 0}) \\
&\quad + s\sqrt{\overline{g}}(- \delta^{\beta_1}_r\overline{\gamma}{}^{r}_{j0}\overline{p}{}^{0j\cdots\beta_s}_{\alpha_1\cdots\alpha_k} - \cdots - \delta^{\beta_s}_r\overline{\gamma}{}^{r}_{j0}\overline{p}{}^{0\beta_1\cdots j}_{\alpha_1\cdots\alpha_k} + \delta^r_{\alpha_1}\overline{\gamma}{}^{i}_{r0}\overline{p}{}^{0\beta_1\cdots\beta_s}_{i\cdots\alpha_k} + \cdots + \delta^r_{\alpha_k}\overline{\gamma}{}^{i}_{r0}\overline{p}{}^{0\beta_1\cdots\beta_s}_{\alpha_1\cdots i}) \\
&\quad + \sqrt{\overline{g}}(- \delta^{\beta_1}_r\overline{\nabla}_j u^r\overline{p}{}^{0j\cdots\beta_s}_{\alpha_1\cdots\alpha_k} - \cdots - \delta^{\beta_s}_r\overline{\nabla}_j u^r\overline{p}{}^{0\beta_1\cdots j}_{\alpha_1\cdots\alpha_k} + \delta^r_{\alpha_1}\overline{\nabla}_r u^i\overline{p}{}^{0\beta_1\cdots\beta_s}_{i\cdots\alpha_k} + \cdots + \delta^r_{\alpha_k}\overline{\nabla}_r u^i\overline{p}{}^{0\beta_1\cdots\beta_s}_{\alpha_1\cdots i}) \\
&\quad + \sqrt{\overline{g}}(- \delta^{\beta_1}_0\partial_j s\overline{p}{}^{0j\cdots\beta_s}_{\alpha_1\cdots\alpha_k} - \cdots - \delta^{\beta_s}_0\partial_j s\,\overline{p}{}^{0\beta_1\cdots j}_{\alpha_1\cdots\alpha_k} + \delta^r_{\alpha_1}\partial_r s\overline{p}{}^{0\beta_1\cdots\beta_s}_{0\cdots\alpha_k} + \cdots + \delta^r_{\alpha_k}\partial_r s\,\overline{p}{}^{0\beta_1\cdots\beta_s}_{\alpha_1\cdots 0}) \\
&\quad + \sqrt{\overline{g}}(- \delta^{\beta_1}_r\overline{\nabla}{}^r s\overline{p}{}^{00\cdots\beta_s}_{\alpha_1\cdots\alpha_k} - \cdots - \delta^{\beta_s}_r\overline{\nabla}{}^r s\,\overline{p}{}^{0\beta_1\cdots 0}_{\alpha_1\cdots\alpha_k} + \delta^0_{\alpha_1}\overline{\nabla}{}^r s\,\overline{p}{}^{0\beta_1\cdots\beta_s}_{r\cdots\alpha_k} + \cdots + \delta^0_{\alpha_k}\overline{\nabla}{}^r s\,\overline{p}{}^{0\beta_1\cdots\beta_s}_{\alpha_1\cdots r}),
\end{aligned} \tag{C6c}$$

$$\begin{aligned}
\delta\overline{\phi}{}^{\alpha_1\cdots\alpha_k}_{\beta_1\cdots\beta_s} &= \overline{D_0\phi{}^{\alpha_1\cdots\alpha_k}_{\beta_1\cdots\beta_s}}s + u^k\overline{\nabla}_k\overline{\phi}{}^{\alpha_1\cdots\alpha_k}_{\beta_1\cdots\beta_s} + (- \overline{r}{}^{\alpha_1}_{0\tau}\overline{\phi}{}^{\tau\cdots\alpha_k}_{\beta_1\cdots\beta_s} - \cdots - \overline{r}{}^{\alpha_k}_{0\tau}\overline{\phi}{}^{\alpha_1\cdots\tau}_{\beta_1\cdots\beta_s} + \overline{r}{}^{\lambda}_{0\beta_1}\overline{\phi}{}^{\alpha_1\cdots\alpha_k}_{\lambda\cdots\beta_s} + \cdots + \overline{r}{}^{\lambda}_{0\beta_s}\overline{\phi}{}^{\alpha_1\cdots\alpha_k}_{\beta_1\cdots\lambda})s \\
&\quad - \delta^{\alpha_1}_0\partial_r s\,\overline{\phi}{}^{r\cdots\alpha_k}_{\beta_1\cdots\beta_s} - \cdots - \delta^{\alpha_k}_0\partial_r s\,\overline{\phi}{}^{\alpha_1\cdots r}_{\beta_1\cdots\beta_s} + \delta^r_{\beta_1}\partial_r s\,\overline{\phi}{}^{\alpha_1\cdots\alpha_k}_{0\cdots\beta_s} + \cdots + \delta^{\,r}_{\beta_s}\partial_r s\,\overline{\phi}{}^{\alpha_1\cdots\alpha_k}_{\beta_1\cdots 0} - \delta^{\alpha_1}_j\overline{\nabla}_r u^j\overline{\phi}{}^{r\cdots\alpha_k}_{\beta_1\cdots\beta_s} - \cdots \\
&\quad - \delta^{\alpha_k}_j\overline{\nabla}_r u^j\overline{\phi}{}^{\alpha_1\cdots r}_{\beta_1\cdots\beta_s} + \delta^r_{\beta_1}\overline{\nabla}_r u^i\overline{\phi}{}^{\alpha_1\cdots\alpha_k}_{i\cdots\beta_s} + \cdots + \delta^r_{\beta_s}\overline{\nabla}_r u^i\overline{\phi}{}^{\alpha_1\cdots\alpha_k}_{\beta_1\cdots i} + (- \delta^{\alpha_1}_j\overline{\gamma}{}^{j}_{r0}\overline{\phi}{}^{r\cdots\alpha_k}_{\beta_1\cdots\beta_s} - \cdots - \delta^{\alpha_k}_j\overline{\gamma}{}^{j}_{r0}\overline{\phi}{}^{\alpha_1\cdots r}_{\beta_1\cdots\beta_s})s \\
&\quad + (\delta^i_{\beta_1}\overline{\gamma}{}^{r}_{i0}\overline{\phi}{}^{\alpha_1\cdots\alpha_k}_{r\cdots\beta_s} + \cdots + \delta^i_{\beta_s}\overline{\gamma}{}^{r}_{i0}\overline{\phi}{}^{\alpha_1\cdots\alpha_k}_{\beta_1\cdots r})s + (- \delta^{\alpha_1}_r\overline{\phi}{}^{0\cdots\alpha_k}_{\beta_1\cdots\beta_s} - \cdots - \delta^{\alpha_k}_r\overline{\phi}{}^{\alpha_1\cdots 0}_{\beta_1\cdots\beta_s} + \delta^0_{\beta_1}\overline{\phi}{}^{\alpha_1\cdots\alpha_k}_{r\cdots\beta_s} + \cdots + \delta^0_{\beta_s}\overline{\phi}{}^{\alpha_1\cdots\alpha_k}_{\beta_1\cdots r})\overline{\nabla}{}^r s.
\end{aligned} \tag{C6d}$$

$$\delta\mathscr{M} = 0, \quad \delta\mathscr{M}_k = 0, \quad \delta N = 0, \quad \delta N^k = 0. \tag{C6e}$$

APPENDIX D

We discuss relations between our energy–momentum formula (4.13) and those given in the literature. We need the following result, which follows from the Belinfante–Rosenfeld identities for matter field Lagrangians (cf. Ref. 15) and (E15)–(E16).

Lemma: If Eqs. (4.4) and (4.5b) are satisfied, then

$$\begin{aligned} G^{\lambda}{}_{\mu} - 8\pi\, {}_{\rm sym}T^{\lambda}{}_{\mu} &= H^{\lambda}{}_{\mu} - 8\pi\, {}_{\rm can}T^{\lambda}{}_{\mu} + 8\pi(\nabla_{\omega}a^{\omega\lambda}_{\mu} + a^{\lambda\omega}_{\tau}r^{\tau}_{\mu\omega}) \end{aligned} \tag{D1}$$

We get from (4.13) and (D1)

$$e^{\lambda}(Z) = (1/8\pi)(G^{\lambda}{}_{\tau} - 8\pi\, {}_{\rm sym}T^{\lambda}{}_{\tau})Z^{\tau} + (1/8\pi)\nabla_{\tau}b^{\tau\lambda}. \tag{D2}$$

The above formula was given by Komar[7] for the Einstein theory (without matter). Recently Kijowski[38,39] has presented an interesting, nonstandard approach to the energy problem. He considered a theory with vanishing torsion (symmetric connection) and obtained a formula which (almost) coincides with (D2). Nonessential small differences in the divergence term $(1/8\pi)\nabla_{\tau}b^{\tau\lambda}$ between Kijowski's and our formulas follow from the fact that the Belinfante–Rosenfeld identities for symmetric theories do not coincide with those given in Ref. 15.

APPENDIX E

The Riemann tensor:

$$R^{\beta}{}_{\alpha\mu\nu} = \partial_{\mu}\Gamma^{\beta}_{\nu\alpha} - \partial_{\nu}\Gamma^{\beta}_{\mu\alpha} + \Gamma^{\tau}_{\nu\alpha}\Gamma^{\beta}_{\mu\tau} - \Gamma^{\tau}_{\mu\alpha}\Gamma^{\beta}_{\nu\tau}. \tag{E1}$$

The Ricci tensor, symmetric Ricci tensor, curvature scalar:

$$R_{\alpha\nu} = R^{\beta}{}_{\alpha\beta\nu}, \quad {}_{\rm sym}R_{\alpha\nu} = \tfrac{1}{2}(R_{\alpha\nu} + R_{\nu\alpha}), \quad R = g^{\alpha\beta}R_{\alpha\beta}.$$

The Einstein tensor:

$$G^{\mu\nu} = {}_{\rm sym}R^{\mu\nu} - \tfrac{1}{2}g^{\mu\nu}R. \tag{E2}$$

The H-tensor:

$$H^{\mu}{}_{\nu} = \tfrac{1}{2}(R^{\mu}{}_{\nu} + g^{\alpha\beta}R^{\mu}{}_{\alpha\nu\beta}) - \tfrac{1}{2}\delta^{\mu}_{\nu}R. \tag{E3}$$

The torsion tensor:

$$Q^{\lambda}_{\mu\nu} = \Gamma^{\lambda}_{\mu\nu} - \Gamma^{\lambda}_{\nu\mu}. \tag{E4}$$

The (pseudo-) Riemannian connection of the metric **g** on M:

$$\gamma^{\lambda}_{\mu\nu} = \left\{{\lambda \atop \mu\nu}\right\} = \tfrac{1}{2}g^{\lambda\tau}(\partial_{\mu}g_{\nu\tau} + \partial_{\nu}g_{\mu\tau} - \partial_{\tau}g_{\mu\nu}). \tag{E5}$$

The defect tensor (the difference tensor):

$$r^{\lambda}_{\mu\nu} = \Gamma^{\lambda}_{\mu\nu} - \gamma^{\lambda}_{\mu\nu}, \tag{E6}$$

$$c^{\mu\nu}_{\lambda} = g^{\mu\nu}r^{\tau}_{\tau\lambda} + g^{\alpha\beta}r^{\nu}_{\alpha\beta}\delta^{\mu}_{\lambda} - g^{\tau\nu}r^{\mu}_{\tau\lambda} - g^{\tau\mu}r^{\nu}_{\lambda\tau} \tag{E7}$$

(the inverse formula to (E7) was given in Ref. 14).

The symmetric stress energy tensor:

$${}_{\rm sym}T^{\mu\nu} = 2\frac{\partial L_{\rm mat}}{\partial g_{\mu\nu}} + g^{\mu\nu}L_{\rm mat}. \tag{E8}$$

The canonical stress–energy tensor:

$${}_{\rm can}T^{\mu}{}_{\lambda} = \delta^{\mu}_{\lambda}L_{\rm mat} - \frac{\partial L_{\rm mat}}{\partial(\partial_{\mu}\phi^{A})}D_{\lambda}\phi^{A}. \tag{E9}$$

The hypermomentum tensor of matter field:

$$s^{\mu\nu}_{\lambda} = \frac{\partial L_{\rm mat}}{\partial \Gamma^{\lambda}_{\mu\nu}}. \tag{E10}$$

4-momentum of matter:

$$p^{\tau}_{A} = \frac{\partial L_{\rm mat}}{\partial(\partial_{\tau}\phi^{A})}, \tag{E11}$$

$$\begin{aligned} z^{\mu\nu}_{\lambda} = &- 16\pi(p^{\mu\beta_1\cdots\beta_s}_{\lambda\cdots\alpha_k}\phi^{\,\nu\cdots\alpha_k}_{\beta_1\cdots\beta_s} + \cdots + p^{\mu\beta_1\cdots\beta_s}_{\alpha_1\cdots\lambda}\phi^{\,\alpha_1\cdots\nu}_{\beta_1\cdots\beta_s} \\ &- p^{\mu\nu\cdots\beta_s}_{\alpha_1\cdots\alpha_k}\phi^{\,\alpha_1\cdots\alpha_k}_{\lambda\cdots\beta_s} - \cdots - p^{\mu\beta_1\cdots\nu}_{\alpha_1\cdots\alpha_k}\phi^{\,\alpha_1\cdots\alpha_k}_{\beta_1\cdots\lambda}) \end{aligned} \tag{E12}$$

Invariance properties of $L_{\rm mat}$ imply

$$L_{\mathrm{mat}}(\phi^A, \partial_\lambda \phi^A, g_{\mu\nu}, \Gamma^\lambda_{\ \mu\nu}) = \widetilde{L}_{\mathrm{mat}}(\phi^A, D_\lambda \phi^A, g_{\mu\nu}, Q^\lambda_{\ \mu\nu}).$$

Let $a^{\mu\nu}_\lambda = 2\,\partial \widetilde{L}_{\mathrm{mat}}/\partial Q^\lambda_{\ \mu\nu}$. We have

$$s^{\mu\nu}_\lambda + (1/16\pi) z^{\mu\nu}_\lambda = a^{\mu\nu}_\lambda. \qquad \text{(E13)}$$

The current of matter field

$$j_A = \frac{\partial \widetilde{L}_{\mathrm{mat}}}{\partial \phi^A}. \qquad \text{(E14)}$$

The Belinfante–Rosenfeld identities (cf. Ref. 15).

If matter field equations (4.4) are satisfied, then

$${}_{\mathrm{can}}T^\lambda_{\ \mu} - {}_{\mathrm{sym}}T^\lambda_{\ \mu} + D_\tau s^{\tau\lambda}_\mu - r^\tau_{\tau\omega} s^{\omega\lambda}_\mu - \nabla_\tau a^{\tau\lambda}_\mu - a^{\lambda\omega}_\tau r^\tau_{\mu\omega} = 0, \qquad \text{(E15)}$$

$$D_\mu({}_{\mathrm{can}}T^\mu_{\ \nu}) + {}_{\mathrm{sym}}T^\alpha_{\ \beta} r^\beta_{\nu\alpha} - {}_{\mathrm{can}}T^\mu_{\ \nu} r^\tau_{\tau\mu} - {}_{\mathrm{can}}T^\mu_{\ \lambda} Q^\lambda_{\ \nu\mu}$$
$$- s^{\alpha\beta}_\tau R^{\ \tau}_{\ \ \beta\nu\alpha} + a^{\alpha\beta}_\tau R^{\ \tau}_{\ \ \beta\nu\alpha} - \tfrac{1}{2} a^{\alpha\beta}_\tau D_\nu Q^{\ \tau}_{\ \ \alpha\beta} = 0. \qquad \text{(E16)}$$

The bar operation for metrics on M:

$$N = (-g^{00})^{-1/2}, \quad N_k = g_{0k}, \quad N^k = \bar{g}^{ks} N_s, \qquad \text{(E17)}$$

$$g^{pq} = \bar{g}^{pq} - (N^p N^q / N^2), \quad g_{00} = -N^2 + N^s N_s,$$
$$g^{0p} = N^p / N^2,$$

$$g = \det g_{\mu\nu}, \quad \bar{g} = \det g_{ij}, \quad \sqrt{-\bar{g}} = N\sqrt{\bar{g}}, \qquad \text{(E18)}$$

$$\bar{\gamma}^k_{pq} = \tfrac{1}{2}\bar{g}^{ks}(\partial_p g_{sq} + \partial_p g_{sp} - \partial_s g_{pq}) \quad [\text{cf. (B5)}], \qquad \text{(E19)}$$

$$\pi^{ij} = \sqrt{\bar{g}}\,[\bar{\gamma}^0_{pq} - g_{pq}(\bar{\gamma}^0_{ab}\bar{g}^{ab})]\bar{g}^{pi}\bar{g}^{qj},$$

$$\bar{\gamma}^0_{pq} = (1/\sqrt{\bar{g}})(\pi_{pq} - \tfrac{1}{2} g_{pq}\,\mathrm{tr}\pi), \quad \mathrm{tr}\pi = g_{ab}\pi^{ab}, \qquad \text{(E20)}$$

$$\partial_\lambda g_{\mu\nu} = g_{\mu\alpha}\gamma^\alpha_{\lambda\nu} + g_{\nu\alpha}\gamma^\alpha_{\lambda\mu}, \qquad \text{(E21)}$$

$$\partial_0 g_{ij} = \bar{\nabla}_i N_j + \bar{\nabla}_j N_i + 2N\bar{\gamma}^0_{ij}. \qquad \text{(E22)}$$

[1] J. Kijowski, "A Finite Dimensional Canonical Formalism in the Classical Field Theory," Comm. Math. Phys. **30**, 99–128 (1973).

[2] H. Goldschmidt and S. Sternberg, "The Hamilton-Cartan Formalism in the Calculus of Variations," Ann. Inst. Fourier **23**, 203–67 (1973).

[3] J. Kijowski and W. Szczyrba, "A Canonical Structure for Classical Field Theories," Comm. Math. Phys. **46**, 183–206 (1976).

[4] V. Guillemin and S. Sternberg, *Geometric Asymptotics*, Math. Survey, Vol. 14 (AMS, Providence, R. I., 1977).

[5] R. Abraham and J. Marsden, *Foundations of Mechanics* (Benjamin, New York, 1978).

[6] V. Arnold, *Mathematical Methods of Classical Mechanics* (Springer-Verlag, Berlin-Heidelberg-New York, 1978).

[7] A. Komar, "Covariant Conservation Laws in General Relativity," Phys. Rev. **113**, 934–6 (1959).

[8] A. Fischer and J. Marsden, "Topics in the Dynamics of General Relativity," in *Isolated Gravitating Systems in General Relativity, Proceedings of the International School of Physics "Enrico Fermi" Course* 67, edited by J. Ehlers (North-Holland, Amsterdam, New York, 1979); "The Initial Value Problem and the Dynamical Formulation of General Relavitvty", in *General Relativity—An Einstein Centenary Survey*, edited by S. W. Hawking and W. Israel (Cambridge U. P., Cambridge 1979).

[9] J. Marsden, "Geometric Methods in Mathematical Physics," Lectures at NSF-CBMS Conferences, Lowell, Mass., 1979.

[10] R. Arnowitt, S. Deser, and C. Misner, "The Dynamics of General Relativity," in *Gravitation: An introduction to current research*, edited by L. Witten (Wiley, New York, 1962).

[11] W. Szczyrba, "A Symplectic Structure on the Set of Einstein Metrics," Comm. Math. Phys. **51**, 163–82 (1976).

[12] W. Szczyrba, "On the Geometric Structure of the Set of Solutions of Einstein Equations," Dissertationes Math. (Rozprawy Mat.) **150**, 1–83 (1977).

[13] W. Szczyrba, "A Symplectic Structure for the Einstein–Maxwell Field," Rep. Math. Phys. **12**, 169–91 (1977).

[14] W. Szczyrba, "The Canonical Variables, the Symplectic Structure and the Initial Value Formulation of the Generalized Einstein–Cartan Theory of Gravity," Comm. Math. Phys. **60**, 215–32 (1978).

[15] W. Szczyrba, "Field Equations and Contracted Bianchi Identities in the Generalized Einstein–Cartan theory," Lett. Math. Phys. **2**, 265–74 (1978).

[16] C. W. Misner, K. S. Thorne, and J. A. Wheeler, *Gravitation* (Freeman, San Francisco, 1973).

[17] P. A. M. Dirac, "Generalized Hamiltonian Dynamics," Proc. Roy. Soc. Lond. **A 246**, 326–333 (1958); "The Theory of Gravitation in Hamiltonian Form," Proc. Roy. Soc. Lond. **A 246**, 333–46 (1958); "Fixation of Coordinates in the Hamiltonian Theory of Gravitation," Phys. rev. **114**, 924–30 (1959); "Interacting Gravitational and Spinor Fields," in *Recent Developments in General Relativity* (Pergamon–PWN, Oxford–Warszawa, 1962).

[18] P. G. Bergmann, "The General Theory of Relativity," in *Encyclopedia of Physics*, edited by S. Flügge *Vol.* 4. *Principles of Electrodynamics and Relativity*, (Springer-Verlag, Berlin, 1955).

[19] L. Faddeev, "Symplectic Structure and Quantisation of the Einstein Gravitation Theory," in *Actes du Congres Intern. Math. Nice* (1970), Vol. 3, pp. 35–39. L. Faddeev and V. Popov, "A Covariant quantization of the Gravitational Field," Usp. Fiz. Nauk **111**, 427–50 (1973).

[20] K. Kuchař, "Geometrodynamics regained: A Lagrangian approach," J. Math. Phys. **15**, 708–15 (1974); "Geometry of Hypersurface. I–IV J. Math. Phys. **17**, 777–91, 792–800, 801–20 (1976); **18**, 1589–97 (1977).

[21] S. Hojman, K. Kuchař, and C. Teitelboim, "Geometrodynamics Regained," Ann. Phys. **96**, 88–135 (1976).

[22] A. Hanson, T. Regge, and C. Teitelboim, *Constrained Hamiltonian Systems* (Accademia Nazionale dei Lincei, Rom, 1976), No. 22.

[23] J. E. Nelson and C. Teitelboim, "Hamiltonian Formulation of the Theory of Interacting Gravitational and Electron Fields," Ann. Phys. **116**, 86–104 (1978).

[24] C. Teitelboim, "The Hamiltonian Structure of Spacetime," in *General Relativity and Gravitation*, edited by A. Held (Plenum, New York, 1980).

[25] M. Pilati, "The Canonical Formulation of Supergravity," Nucl. Phys. **B 132**, 138–54 (1978).

[26] J. M. Nester, "Canonical Formalism and the ECSK Theory," Ph. D. thesis (University of Maryland, 1977).

[27] J. A. Isenberg and J. M. Nester, "The Effect of Gravitational Interaction on Classical Fields. A Hamilton-Dirac Analysis," Ann. Phys. **107**, 56–81 (1977).

[28] J. A. Isenberg and J. M. Nester, "Canonical Gravity", in *General Relativity and Gravitation*, edited by A. Held (Plenum, New York, 1980).

[29] N. Ō. Murchadha and J. W. York Jr., "Gravitational Energy," Phys. Rev. D. **10**, 2345–57 (1974).

[30] J. W. York Jr., "Kinematics and Dynamics of General Relativity," in *Sources of Gravitational Radiation, Proceedings of the Battelle Seattle Workshop*, edited by L. L. Smarr (Cambridge U. P., Cambridge, 1979).

[31] A. Śniatycki, "Dirac Brackets in Geometric Dynamics," Ann. Inst. H. Poincaré **20**, 365–72 (1974).

[32] W. M. Tulczyjew, "Hamiltonian Systems, Lagrangian Systems and the Legendre transformation," Symp. Math. **14**, 247–58 (1974).

[33] M. J. Gotay, J. M. Nester, and G. Hinds, "Presymplectic manifolds and the Dirac–Bergmann theory of constraints," J. Math. Phys. **19**, 2388–99 (1978).

[34] V. Moncrief, "Decomposition of gravitational perturbations," J. Math. Phys. **16**, 1556–60 (1975).

[35] V. Moncrief, "Spacetime symmetries and linearization stability of the Einstein equations. I, II," J. Math. Phys. **16**, 493–498 (1975); **17**, 1893–1902 (1976).

[36] J. Arms, "Linearization Stability of the Einstein–Maxwell system," J. Math. Phys. **18**, 830–33 (1977); "Linearization stability of gravitational and gauge fields," J. Math. Phys. **20**, 443–53 (1979).

[37] M. Francaviglia, "Applications of Infinite Dimensional Differential Geometry to General Relativity," Nuovo Cimento **1** (7), 1–54 (1978).

[38] J. Kijowski and W. M. Tulczyjew, *A Symplectic Framework for Field Theories*, Lecture Notes in Physics 107 (Springer-Verlag, Berlin 1979).

[39] J. Kijowski, "On a New Variational Principle in General Relativity and the Energy of the Gravitational Field". Gen. Relativity Gravitation **9**, 857–78 (1978).

[40] P. Dedecker, "Calcul des Variations, Formes Differentieles et Champs Geodesiques," in Coll. Intern. Geometr. Different. Strasbourg 1953.

[41] W. Szczyrba, "Lagrangian Formalism in the Classical Field Theory," Ann. Pol. Math. **32**, 145–85 (1976).

[42] P. L. Garciá, "Reducibility of the Symplectic Structure of Classical

Fields with Gauge Symmetry," in *Lecture Notes in Mathematics* 570 (Springer-Verlag, Berlin-Heidelberg-New York, 1977); P. L. Garciá and A. Pérez-Rendon, "Reducibility of the Symplectic Structure of Minimal Interactions," in *Lecture Notes in Mathematics* 676 (Springer, Berlin-Heidelberg-New York, 1978), pp. 409–33.
[43]J. Eells, "A Setting For Global Analysis," Bull. Am. Math. Soc. **72**, 751–807 (1966).
[44]J. Kijowski, "Existence of Differentiable Structure in the Set of Submanifolds," Studia Math. **33**, 93–108 (1969).
[45]A. Fischer and J. Marsden, "Linearization Stability of Non-linear Partial Differential Equations," Proc. Symp. Pure Math. (AMS) **27**, 219–263 (1975).
[46]A. Fischer, J. Marsden, and V. Moncrief, "The Structure of the Space of Solutions of Einstein Equations. I. One Killing Field," preprint (University of California, 1979).
[47]W. Kondracki, "Geometrization of the Classical Field Theory and its Application in Yang–Mills Field Theory," Rep. Math. Phys. **16**, 9–67, (1979).
[48]D. Ebin and J. Marsden, "Group of Diffeomorphisms and the Motion of an Incompressible Fluid," Ann. Math. **92**, 102–63 (1970).
[49]J. Marsden, D. Ebin, and A. Fischer, "Diffeomorphism Groups, Hydrodynamics and General Relativity," in Proceedings of the 13th Biennial Seminar of the Canadian Mathematical Congress 1972.
[50]A. Trautman, "On the Structure of the Einstein–Cartan Equations," Symp. Math. **12**, 139–62 (1973).
[51]F. W. Hehl, P. von der Heyde, G. D. Kerlick, and J. M. Nester, "General Relativity with Spin and Trosion. Foundations and Prospects," Rev. Mod. Phys. **48**, 393–416 (1976).
[52]W. Szczyrba, "Contracted Bianchi Identities and Conservation Laws in Poincaré Gauge Theories of Gravity," in *Proceedings of the International School of Cosmology and Gravitation Erice* 1979, edited by P. Bergmann and V. De Sabbata (Plenum, New York, 1980); "The Hamiltonian Dynamics for Gauge Theories of Gravity," preprint (University of Cologne, 1980).
[53]A. Trautman, "Conservation Laws in General Relativity," in *Gravitation: An Introduction to Current Research*, edited by L. Witten (Wiley, New York, 1962).
[54]J. A. Wheeler, "Geometrodynamics and the Issue of the Final State," in *Relativity, Groups and Topology*, edited by B. DeWitt and C. DeWitt (Gordon and Breach, New York, 1964).
[55]W. R. Davis, *Classical Theory of Particles and Fields and the Theory of Relativity* (Gordon and Breach, New York, 1970).
[56]B. DeWitt, "Quantum Theory of Gravity, I. The Canonical Theory," Phys. Rev. **160**, 1113–48 (1967).
[57]T. Regge and C. Teitelboim, "Role of Surface Integrals in the Hamiltonian Formulation of General Relativity," Ann. Phys. **88**, 286–318 (1974).
[58]Y. Choquet-Bruhat, A. Fischer, and J. Marsden, "Maximal Hypersurfaces and Positivity of Mass," in *Isolated Gravitating Systems in General Relativity—Proceedings of the International School of Physics "Enrico Fermi"*, Course 67 edited by J. Ehlers (North-Holland, Amsterdam, 1979).
[59]Y. Ne'eman and D. Šijački, "Unified Affine Gauge Theory of Gravity and Strong Interactions with Finite and Infinite GL (4, R) Spinor Fields," Ann. Phys. **120**, 292–315 (1979).

Path integration for time-dependent metrics

Bruce Nelson
Institute for Theoretical Physics, University of California, Santa Barbara, California 93106
Benny Sheeks
Massachusetts Institute of Technology, Lincoln Laboratory, Lexington, Massachusetts 02173

(Received 7 August 1979; accepted for publication 20 June 1980)

We define path integrals for systems with time-dependent metrics in terms of prodistributions and discuss the relation of the path integral quantization to the Schrödinger one. We work an example that displays the elegance and utility of the prodistribution definition. We also discuss how our definition is particularly suited for changing integration variables.

PACS numbers: 03.65.Ca

INTRODUCTION

Cécile DeWitt–Morette has introduced a definition of path integrals in terms of projective systems of distributions.[1–5] Her definition takes full advantage of the vector-space nature of the space of paths and of the Gaussian nature of the free particle (and other simple) integrals. Her definition gives the same results as Feynman's in those cases where his definition can be used; however, hers is much more versatile, applying in more cases and being very suitable for studying general problems of integration such as the effects of changing variables. In particular her definition yields directly the "normalization factors" that are so difficult to obtain using the time-slicing approach. The extension of her prodistribution definition to systems with curved configuration spaces (i.e., with spatial-dependent metrics) has been presented in Ref. 5. Here we extend the definition to include systems with a time-dependent (but spatially flat) metrics. (The case of time-dependent metrics in one dimension has been considered in Ref. 6. The case of arbitrary dimensions is contained implicitly in Ref. 4, however the true nature of the time-dependent metric is perhaps obscurred by the additional complications of phase space. Also, unlike Ref. 6, our philosophy is not to construct path integral solutions of a given Schrödinger equation, but rather to quantize a given classical system directly in terms of its Lagrangian.) As an example, we quantize the system of a particle subjected to a random classical force, where the effects of friction are also treated as classical. This problem, although simple, illustrates the power and elegance of the prodistribution approach to path integration.

Unlike Feynman who considers all continuous paths that go from some initial point, a, to some final point b, in a given time interval $T \equiv [t_b, t_a]$, we begin with the space of paths which all end at the origin and begin anywhere. This space, called X_+, forms a vector space. Since many of the paths are continuous, but not differentiable, we cannot write expressions like dq/dt and give them meaning for all paths. We do, however, want to construct an object (a prodistribution) that corresponds to the Dx of Feynman plus the exponential of the free particle part of the Lagrangian of the system under consideration.

THE PRODISTRIBUTION

In the same way that a normalized Gaussian measure (or distribution) γ on R^n is most simply described in terms of its covariance G^{ij}, and infinite-dimensional Gaussian promeasure or prodistribution can be defined by its covariance $G(t, s)$. In the finite dimensional case

$$d\gamma(x) \equiv dx^1 \cdots dx^n (2\pi i)^{-n/2} (\det G^{ij})^{-1/2} \exp(\tfrac{1}{2} i x^i G^{-1}_{ij} x^j).$$

Two important properties of γ are its moments and its Fourier transform:

$$\int d\gamma(x) x^i x^j = i G^{ij}$$

and

$$(\mathscr{F}\gamma)(y) \equiv \int d\gamma(x) e^{-ix\cdot y} = e^{-\frac{1}{2} i y_i G^{ij} y_j}.$$

Note that the Fourier transform is much simpler than γ itself. More importantly, under linear mappings of R^n (either onto R^n again or onto R^m for $m < n$) the transformation of γ is very simple whereas that of γ is in general quite complicated. It is this property of γ that makes it, rather than γ, the object of study in a system of projective distributions for an infinite-dimensional space (such as the space of paths). In defining path integrals for quantum systems, the physical input is that the prodistribution should contain all the information about the kinetic energy part of the Lagrangian. [It is in fact possible to incorporate much more of the Lagrangian into the prodistribution. However, we always begin with the simplest prodistribution and incorporate other parts of the Lagrangian by a change of integration variables (See Ref. 5, pp. 271–284 for a full discussion of this procedure.)]

The Gaussian prodistribution appropriate for quantizing a system with a time-dependent metric is defined by the covariance $G_+(t, s)$ which is a Green function of the small disturbance (or Jacobi) operator of the free particle part of the system's Lagrangian. (The definition of G_+ and a discussion of its properties are given below.) Then, to quantize a Lagrangian

$$L = \tfrac{1}{2} m g_{\alpha\beta}(t) \dot{q}^\alpha(t) \dot{q}^\beta(t) + \dot{q}^\alpha(t) A_\alpha[q(t), t] - V[q(t)]. \tag{1}$$

We define a wave function $\psi(b, t_b)$ in terms of its initial value

$$\psi(a, t_a) \equiv \psi_0(a)$$

by the path integral

$$\psi(b, t_b) \equiv \int_{X_+} d\gamma_+(x) \exp\left\{\frac{i}{\hbar}\int_T [A_\alpha(q(t), t)]\, dq^\alpha(t) - V[q(t)\, dt]\right\} \psi_0[q(t_a)]. \tag{2}$$

γ_+ is the prodistribution defined by G_+. X_+ is the space of all paths in R^n (parameterized by $t \in [t_a, t_b]$) that end at 0. The path q is $b + \mu x$. Thus, we add b to all paths x so that they end at the point of interest, b. The paths are scaled by the factor $\mu = \sqrt{\hbar/m}$ because our prodistribution is defined for $\hbar = m = 1$. (It would have been possible to include the $m/\hbar$ in the definition of γ_+, however, it is more convenient for doing the semiclassical expansion of a path integral to have the factors of $\hbar$ explicitly displayed in the integrand instead of in γ). For the velocity dependent potential, we write $\int A \cdot dq$ instead of $\int A \cdot \dot{q}\, dt$ because q is a stochastic variable and $\dot{q}$ really has no meaning. Equation (2) is the generalized Feynman–Kac formula. Perhaps a justification for it will come below when we show that ψ satisfies a Schrödinger equation

$$i\hbar\frac{\partial\psi}{\partial t} = \widehat{H}\psi,$$

where $\widehat{H}$ is the quantum Hamiltonian formed from L (1) by the symmetric factor ordering. The propagator $K(b, t_b; a, t_a)$ is given by the path integral (2) with $\psi_0(a)$ given by $\delta[q(t_a) - a]$.

We now construct the covariance G_+ of γ_+ from the classical Lagrangian L given by Eq. (1). We consider a classical system with a time-dependent metric $g_{\alpha\beta}(t)$. We require g to be flat, in the sense that for each t, the curvature tensor formed from g is zero. For example, in Cartesian coordinates, g might take the simple form

$$g_{\alpha\beta}(t) = f(t)\delta_{\alpha\beta}.$$

More generally, g might not vary the same in all directions. Thus, we might have a g given by the line-element

$$ds^2 = f_1(t)(dx^1)^2 + f_2(t)(dx^2)^2 + \cdots.$$

In order to construct the prodistribution appropriate to this system, we consider the covariance $G_+(t, s)$ that is a Green function of the free particle part of L. Thus,

$$L_0 \equiv \tfrac{1}{2}\dot{q}(t)\cdot g(t)\cdot\dot{q}(t),$$

(where the dots denote contraction of indices). The action is

$$S_0 = \int_T dt\, L_0$$

and its first and second variations are

$$S_0'[x] = x(t)\cdot g(t)\cdot\dot{q}(t)\big|_{t_a}^{t_b} + \int_T dt\,[-x(t)\cdot g(t)\cdot\ddot{q}(t) - x(t)\cdot\dot{g}(t)\cdot\dot{q}(t)]$$

and

$$S_0''[x, y] = x(t)\cdot g(t)\cdot\dot{y}(t)\big|_{t_a}^{t_b} + \int_T x(t)\cdot[\mathscr{J}(t)\cdot y(t)]\, dt,$$

where $\mathscr{J}_{\alpha\beta}$ is the small disturbance (or Jacobi) operator, given by

$$\mathscr{J}_{\alpha\beta}(t) = -g_{\alpha\beta}(t)\nabla_t^2 - \dot{g}_{\alpha\beta}(t)\nabla_t.$$

The Green function G_+ is defined by

$$\mathscr{J}_{\alpha\beta}(t)G_+^{\beta\gamma}(t, s) = \delta(t - s)\delta_\alpha^\gamma,$$

with

$$G_+(t = t_b, s) = 0$$

and

$$\frac{d}{dt}G_+(t = t_a, s) = 0.$$

G_+ can be written conveniently in terms of the Jacobi matrices: Let $K^{\beta\gamma}(t, s)$ and $J^{\beta\gamma}(t, s)$ be the solutions to

$$\mathscr{J}_{\alpha\beta}(t)K^{\beta\gamma}(t, s) = 0,$$

with

$$K^{\alpha\beta}(s, s) = g^{\alpha\beta}(s), \tag{3a}$$

and

$$\frac{d}{dt}K^{\alpha\beta}(t = s, s) = 0,$$

and

$$\mathscr{J}_{\alpha\beta}(t)J^{\beta\gamma}(t, s) = 0,$$

with

$$J^{\beta\alpha}(t, t) = 0, \tag{3b}$$

$$\frac{d}{dt}J^{\beta\gamma}(t = s, s) = g^{\beta\gamma}(s).$$

Then G_+ is given by

$$G_+(t, s) = \Theta(s - t)\cdot K(t, t_a)\cdot N(t_a, t_b)\cdot J(t_b, s) - \Theta(t - s)J(t, t_b)\cdot\widetilde{N}(t_b, t_a)\cdot\widetilde{K}(t_a, s), \tag{3c}$$

where N is the inverse of K:

$$K^{\alpha\beta}(t, s)N_{\beta\gamma}(s, t) = \delta_\gamma^\alpha$$

and

$$\widetilde{K}^{\alpha\beta}(t, s) \equiv K^{\beta\alpha}(s, t).$$

In practice J and K are much easier to calculate than is G_+ directly. Also, in general discussions it is convenient to have an explicit representation of G_+.

In order to discuss the Schrödinger equation satisfied by ψ defined by (2), we follow the discussion of Ref. 7 (p. 76). For simplicity, we first consider the case of $V = A = 0$ in order to bring out the effects of the time-dependent metric in γ_+. We then show how the A and V terms affect the results. Thus we have

$$\psi(b, t_b) = \int_{X_+} d\gamma_+(x)\psi[b + \mu x(t_a), t_a].$$

The integrand in this case is a simple cylindrical function (see Ref. 5, pp. 259–271 for a full discussion of the techniques of integrating cylindrical functions with Gaussian prodistributions) and reduces immediately to an integral over R^n (the configuration space):

$$\psi(b, t_b) = \int \frac{dx}{(2\pi i)^{n/2}|\det G|^{1/2}} \exp(\tfrac{1}{2}ix^iG_{ij}^{-1}x^j) \times \psi(b + \mu x, t_a).$$

Here $G \equiv G(t_a, t_a) = K(t_a, t_a) \cdot N(t_a, t_b) \cdot J(t_b, t_a)$. We now let $t_b = t_a + \epsilon$, for ϵ small. We expand $(b, t_a + \epsilon)$ in powers of ϵ and $\psi(b + \mu x, t_a)$ in powers of μ. Also, J, N, and K can be expanded in powers of ϵ. Using the Jacobi equation, we have

$$K(t+\epsilon, t) = g^{-1}(t) + O(\epsilon^2),$$

$$N(t, t+\epsilon) = g(t) + O(\epsilon^2),$$

and

$$J(t+\epsilon, \epsilon) = \epsilon g^{-1}(t) + O(\epsilon^2).$$

Thus

$$G = \epsilon g^{-1}(t) + O(\epsilon^2).$$

The expansion of ψ is then

$$\psi(b, t_b) = \psi(b, t_a) + \epsilon \frac{\partial \psi}{\partial t} + O(\epsilon^2)$$
$$= \int_{x_+} \frac{dx}{(2\pi i)^{n/2} |\det G|^{1/2}} e^{(i/2) x\, G\, x}$$
$$\times \left[\psi(b, t_a) + \mu x \cdot \nabla \psi(b, t_a) + \frac{\mu^2}{2} x \cdot x \cdot \nabla\nabla \psi(b, t_a) + O(X^3) \right].$$

The integrals are standard Gaussian integrals. Let

$$d\gamma^{(n)} \equiv \frac{dx^n}{(2\pi i)^{n/2} |\det G|^{1/2}} e^{\frac{1}{2} i x \cdot G^{-1} \cdot x}.$$

Then

$$\int d\gamma^{(n)}(x) = 1,$$

$$\int d\gamma^{(n)}(x) x^\alpha = 0,$$

$$\int d\gamma^{(n)}(x) x^\alpha x^\beta = i G^{\alpha\beta} = i\epsilon g^{\alpha\beta}(t) + O(\epsilon^2),$$

$$\int d\gamma^{(n)}(x) x^\alpha x^\beta x^\gamma x^\delta = O(\epsilon^2).$$

Thus

$$\epsilon \frac{\partial \psi(b, t)}{\partial t} = \frac{\mu^2}{2} (i\epsilon g^{\alpha\beta}(t)) \nabla_\alpha \nabla_\beta \psi(b, t) + O(\epsilon^2)$$

or, as $\epsilon \to 0$,

$$i\hbar \frac{\partial \psi}{\partial t} = \frac{-\hbar^2}{2} g^{\alpha\beta}(t) \nabla_\alpha \nabla_\beta \psi(b, t) \equiv \hat{H}(t) \psi \tag{4}$$

(since $\mu \equiv \sqrt{\hbar/m}$). $\hat{H}(t)$ is just the quantum operator formed from $L_0 = (m/2) g_{\alpha\beta}(t) \dot{q}^\alpha(t) \dot{q}^\beta(t)$ in the usual way.

In order to include the A and V terms of L, we approximate $\int_{t_a}^{t_a+\epsilon} V[b + \mu x(t)] dt$ by $\epsilon V(b) + O(\epsilon^2)$ and

$$\int_{t_a}^{t_a+\epsilon} \mu\, dx^\alpha(t) A_\alpha [b + \mu x(t)]$$

by

$$-\mu x^\alpha(t_a) A_\alpha \left(b + \frac{\mu x(t_a)}{2} \right).$$

(the midpoint rule).

The exponential

$$e^{(i/\hbar)\int (A \cdot dq - V dt)}$$

is expanded up to first order in ϵ, second order in x, and the terms integrated as above. The result is, as expected,

$$i\hbar \frac{\partial \psi}{\partial t} = \frac{-\hbar^2}{2m} g^{\alpha\beta}(t) [\partial_\alpha - i A_\alpha(b)][\partial_\beta - i A_\beta(b)] \psi + V(b) \psi \tag{5}$$

for ψ defined by (2).

EXAMPLE

As an example of a system that has a time-dependent metric, we consider a particle subject to a frictional force and a random fluctuating force f. The classical equation of motion is

$$\ddot{q} = -\beta \dot{q} + \alpha f(t). \tag{6}$$

For simplicity, let f be white noise [i.e., given by a Gaussian measure with the simple covariance $\delta(t - s)$]. This equation has been studied often (See Ref. 8, for example.) If the friction and the random force are due to some background system at temperature T, then the equipartition theorem requires that as $t \to \infty$, $\langle \frac{1}{2} \dot{q}^2 \rangle \to \frac{1}{2} kT$. This requires a relationship between the magnitude of the damping and the magnitude of the associated fluctuating force:

$$\beta = \alpha^2 / kT. \tag{7}$$

A review of the quantization of (6) in terms of the Schrödinger equation can be found in Ref. 9. (See also Ref. 10.) In order to quantize this system using path integrals, we note that (6) can be derived from a Lagrangian with a time-dependent metric:

$$L = \tfrac{1}{2} e^{\beta t} \dot{q}^2(t) + \alpha e^{\beta t} f(t) q(t). \tag{8}$$

(Note that we could just as well work the problem in several dimensions with different friction constants in the different directions.)

One could consider this to be a particle without friction viewed on a movie that runs slower and slower, according to the factor $e^{-\beta t}$. For this system

$$\mathscr{J}(t) = -e^{\beta t} \nabla_t^2 - \beta e^{\beta t} \nabla_t,$$

$$J(t, s) = \frac{1}{\beta} (e^{-\beta s} - e^{-\beta t})$$

$$K(t, s) = e^{-\beta s},$$

and

$$G_+(t, s) = \frac{1}{\beta} \left(-e^{-\beta t_b} + e^{-\max(t_b - t,\, t_b - s)} \right) \tag{9}$$

where $\max(t, s)$ is the larger of t and s.

First, we construct the propagator for L. This illustrates the use of our γ_+ for time-dependent metrics and the power and elegance of the prodistribution method for evaluating path integrals.

Let the initial wave function be simply

$$\psi_0(a) = e^{ip \cdot a / \hbar}.$$

(We take $t_a = 0$.) The evolution of ψ is given by the generalized Feynman–Kac formula (2):

$$\psi(b, t_b) = \int_{X_+} d\gamma_+(x)\exp\left\{-\frac{i}{\hbar}\int_0^{t_b} ds e^{\beta s}\alpha f[s(b+\mu x(s))]\right\}$$
$$\times \exp\left\{\frac{i}{\hbar}[b+\mu x(0)]p\right\}.$$

γ_+ is defined by the covariance G_+ given by (9). The parts of the integrand independent of x can be taken out, leaving

$$\psi(b, t_b) = \exp\left[-\frac{i\alpha b}{\hbar}\int_T ds e^{\beta s} f(s)\right]\exp(\frac{i}{\hbar}bp)I,$$

where I is the simple path integral:

$$I \equiv \int_{X_+} d\gamma_+(x)\exp\left[-\frac{i\alpha\mu}{\hbar}\int_T ds e^{\beta s} f(s)x(s)\right]\exp\left[\frac{i}{\hbar}x(0)p\right].$$

I is a simple cylindrical integral: Consider the linear map P

$$P{:}X_+ \to R^1$$

defined by

$$x \to U \equiv \alpha\int_T ds[e^{\beta s} f(s)x(s)] + x(0) \equiv \langle v, x\rangle,$$

[where v is the measure (in X_+^*) given by

$$dv(s) = \alpha e^{\beta s} f(s) ds + \delta_0(s).]$$

The prodistribution $d\gamma_+$ is mapped to $d\gamma^1$, given by

$$d\gamma^1(U) = \frac{dU}{[2\pi i G(t_b)]^{\frac{1}{2}}}\exp\left[\frac{i}{2}G^{-1}(t_b)U^2\right],$$

where

$$\begin{aligned} G(t_b) &\equiv \int\int dv(t)dv(s)G_+(t, s) \\ &= G_+(0, 0) + 2p\alpha\int_T e^{\beta s} f(s) G(0, s)\, ds \\ &\quad + \alpha^2\int_T\int_T e^{\beta r}e^{\beta s} f(r)f(s)G_+(r, s)\, dr\, ds. \end{aligned} \tag{10}$$

Thus I becomes a simple integral over R^1:

$$I = \int_{R^1} d\gamma^1(U)\exp(\frac{i\mu U}{\hbar}) = \exp\left[-\frac{iG(t_b)}{2\hbar}\right].$$

The wave function ψ is now completely evaluated:

$$\psi(b, t_b) = \exp\frac{i}{\hbar}\left[bp + \alpha b\int_T e^{\beta s} f(s) ds - \frac{G(t_b)}{2}\right].$$

$G(t_b)$ is given by (10), or explicitly as

$$G(t) = \frac{p^2}{\beta}(1 - e^{-\beta t}) + \frac{2p\alpha}{\beta}\int_0^t e^{\beta s} f(s)(e^{-\beta s} - e^{-\beta t})\, ds$$
$$+ \frac{2\alpha^2}{\beta}\int_0^t dr\int_0^r ds\, e^{\beta r}e^{\beta s} f(r)f(s)(e^{-\beta r} - e^{-\beta t}).$$

CHANGE OF VARIABLES

It is possible to make a linear change of integration variables in the integrals for γ_+, just as in the case of prodistributions defined for time-independent metrics. The results are essentially the same as given on pp. 271–282 of Ref. 5. The difference is that here, the Jacobian of the transformation is in terms of the Jacobi matrix K defined by (3a). Thus, consider the linear transformation

$$M{:}X_+ \to X_+$$

given by

$$y(t) \to x(t) = y(t) + \int_0^t \dot{\bar{K}}(t, 0)\cdot\bar{N}(0, r)y(r)\, dr,$$

where $\bar{K}$ is itself some Jacobi matrix, $d\gamma_+(y)$ will be mapped to $d\gamma_+(x)$ times the exponential of some quadratic terms. The Jacobian of the transformation is

$$\operatorname{Det} M = \left|\frac{\det \bar{K}(t, 0)}{\det K(t, 0)}\right|^{\frac{1}{2}}.$$

These linear transformations are used to incorporate the quadratic terms of the semiclassical expansion of a path integral into a larger Gaussian prodistribution. Thus, we always begin with the simple Feynman–Kac formula (2), and generate new and improved prodistributions by changes of variables. The Jacobians of these transformations are usually the functions of time that are sometimes left unevaluated in other approaches to path integration.

As pointed out by the referee, although still the wavefunction ψ satisfies a Schrödinger equation with a self-adjoint Hamiltonian [as, for example, Eq. (4)] if the normalization of the wave function is taken to be the natural invariant one with the measure being $\sqrt{g}dx$:

$$\int \psi^*(x, t)\psi(x, t)\sqrt{g(t)}\, dx,$$

then the norm will not in general be preserved due to the time dependence in $\sqrt{g}$.

[1] C. DeWitt-Morette, "Feynman's Path Integral, Definition without Limiting Procedure," Commun. Math. Phys. **28**, 47–67 (1972).

[2] C. DeWitt-Morette, "Feynman Path Integrals: I. Linear and Affine Techniques; II. The Feynman–Green Function," Commun. Math. Phys. **37**, 63–81 (1974).

[3] C. DeWitt-Morette, "The Semi-Classical Expansion," Ann. Phys. **97**, 367–399 (1976).

[4] C. DeWitt-Morette, A. Maheshwari, and B. Nelson, "Path Integration in Phase Space," Gen. Rel. Grav. **8**, 581–593 (1977).

[5] C. DeWitt-Morette, A. Maheshwari, and B. Nelson, "Path Integration in Non-relativistic Quantum Mechanics," Phys. Rep. **50**, 255–372 (1979).

[6] M. Mizrahi, "Phase Space Path Integrals without Limiting Procedure," J. Math. Phys. **19**, 298–307 (1978).

[7] R. P. Feynman and A. R. Hibbs, *Quantum Mechanics and Path Integrals* (McGraw–Hill, New York, 1965).

[8] G. E. Uhlenbeck and L. S. Ornstein, "On the Theory of Brownian Motion," Phys. Rev. **36**, 823–841 (1930).

[9] J. Messer, "Friction in Quantum Mechanics," Acta Phys. Austriaca **50**, 75–91 (1979).

[10] I. R. Senitzsky, "Dissipation in Quantum Mechanics. The Harmonic Oscillator," Phys. Rev. **119**, 670–9 (1960) and **124**, 642–8 (1961).

Dynamical symmetries of the Schrödinger equation

Thomas P. Mitchell
University of California, Santa Barbara, Santa Barbara, California 93106

(Received 12 May 1980; accepted for publication 20 June 1980)

The conditions are derived under which a symmetry operator quadratic in the momenta exists for a spin-zero structureless point charge interacting with an externally applied uniform magnetic field in the presence of a potential field. The conditions apply to the possible forms of the potential. The explicit form of the symmetry operator in general is constructed and some particular examples are examined.

PACS numbers: 03.65.Fd

1. INTRODUCTION

Geometrical symmetry groups involving point transformations of the coordinates alone are the source of the first applications of group theory in quantum mechanics. The group operators are linear expressions in the momenta effecting displacements in space and time. As is well known, the fundamental conservation laws follow from the invariance of the dynamical equations under these operations. In fact, the transformation operators themselves represent conserved quantities. For example, the nonrelativistic Hamiltonian of a point particle moving in a Coulomb potential field commutes with all of the operators of the three-dimensional rotation group SO_3. However, Fock[1] showed that a higher symmetry group viz. SO_4 exists, each of whose operators also commutes with the Hamiltonian and thus explained the degeneracy of the energy levels of the hydrogen atom. The group SO_4 contains, of course, the purely geometrical SO_3 as a subgroup. It was the first example of a nongeometrical symmetry group. The nongeometrical, or more generally dynamical, groups depend for their existence on particular types of interactions rather than on assumed properties of space and time. They reflect the existence of more subtle invariances involving simultaneous transformations of coordinates and momenta. A dynamical group can determine the energy spectrum, the degeneracies of its eigenvalues, and all of the quantum numbers of a quantum mechanical system. Accordingly, it is a matter of considerable practical and theoretical interest to demonstrate the existence of dynamical groups wherever possible.

This paper establishes the conditions under which a higher order (quadratic) symmetry operator exists for a point charge interacting simultaneously with a constant uniform magnetic field and an arbitrary potential field. The approach adopted is to determine the explicit form of an operator $\hat{L}$ quadratic in the momenta which commutes with the Hamiltonian. The latter is written in a gauge-independent formulation in order to maintain the generality of the results.

A study of dynamical symmetry groups for the interaction with a potential field alone has been carried out in both nonrelativistic classical and quantum mechanics by Winternitz *et al.*[2] and by Makarov *et al.*[3] Yanagawa and Moriya[4] have constructed in the quantum mechanical case a realization of a Lie Group using first order infinitesimal operators, and a special choice of gauge, with only the magnetic field present. The Lie algebras associated with the classical relativistic motion of a charge in the presence of both the magnetic field and a potential have been presented by Mitchell.[5]

2. DETERMINATION OF THE SYMMETRY OPERATOR

The quantum mechanical nonrelativistic Hamiltonian of a spin-zero structureless point charge e of mass m moving in an externally applied electromagnetic field described by the potentials $(\mathbf{A},\phi)$ is

$$\widehat{H}=\frac{\hat{p}^2}{2m}-\frac{e}{2mc}(\widehat{A}\cdot\hat{p}+\hat{p}\cdot\widehat{A})+\frac{e^2}{2mc^2}A^2+e\phi. \quad (1)$$

This expression can be rewritten, while making allowance for the presence of any additional scalar potential $V(\mathbf{r})$, in the form

$$\widehat{H}=\frac{\hat{p}^2}{2m}-\frac{e}{mc}\widehat{A}\cdot\hat{p}+\frac{ie\hbar}{2mc}\nabla\cdot\mathbf{A}+\frac{e^2}{2mc^2}\mathbf{A}^2+U \quad (2)$$

on using the relationship

$$\hat{p}\cdot\widehat{A}-\widehat{A}\cdot\hat{p}=-i\hbar\nabla\cdot\mathbf{A}, \quad (3)$$

with

$$U=e\phi+V. \quad (4)$$

If consideration is now confined to the time-independent two-dimensional case $A=A(x,y)$, $U=U(x,y)$, the usual Cartesian coordinates being represented by (x,y,z), then $[\hat{p}_z,\widehat{H}]=0$. The commutation of $\hat{p}_z$ and $\widehat{H}$ implies translational invariance in the direction of the z-axis with the concomitant conservation of the corresponding component of the linear momentum which will be denoted by p_{z0}. It is convenient for the purposes of the following analysis to introduce complex pairs of coordinates, momenta and magnetic potentials defined by

$$x_1=x+iy,\quad A_1=A_x-iA_y$$
$$x_2=x-iy,\quad A_2=A_x+iA_y \quad (5)$$

$$\hat{p}_1=\hat{p}_x-i\hat{p}_y=\frac{2\hbar}{i}\frac{\partial}{\partial x_1}, \quad (6)$$

$$\hat{p}_2=\hat{p}_x+i\hat{p}_y=\frac{2\hbar}{i}\frac{\partial}{\partial x_2}. \quad (7)$$

Using these definitions one can write the Hamiltonian (2) as

$$\widehat{H}=\frac{\hat{p}_1\hat{p}_2}{2m}-\frac{\mu}{\hbar}(A_1\hat{p}_2+A_2\hat{p}_1)+i\mu\left(\frac{\partial A_1}{\partial x_2}+\frac{\partial A_2}{\partial x_1}\right)+\frac{2m\mu^2}{\hbar^2}A_1A_2+U(x_1,x_2)+\text{const}$$

$$= \frac{\hat p_1 \hat p_2}{2m} - \frac{\mu}{\hbar}(A_1\hat p_2 + A_2 \hat p_1) + W(x_1,x_2), \tag{8}$$

in which $\mu = e\hbar/2mc$. Although the magnetic field of interest from this point on is a uniform constant field of strength B in the z-direction, specification of explicit forms of A_1 and A_2 will not be made in Eq. (8) other than to note that in general

$$\frac{\partial A_1}{\partial x_2} = \frac{\partial A_2}{\partial x_1} - iB. \tag{9}$$

In this way a gauge invariant formulation will be maintained throughout the work. The quadratic operator $\hat L$ is now taken in the form

$$\hat L = f_1 \hat p_1^2 + f_2 \hat p_2^2 + f_3 \hat p_1 \hat p_2 + f_4 \hat p_1 + f_5 \hat p_2 + f_6, \tag{10}$$

where the six functions $f_i(x_1,x_2)$, $i = 1,...,6$ are arbitrary. This is a symmmetry operator if it commutes with the Hamiltonian. The commutation condition can be expressed, on expanding the bracket $[\hat L, \hat H]$, by equating to zero the coefficients of the various powers of the momenta $\hat p_1$ and $\hat p_2$. The expansion is carried out by using Eqs. (6) and (7) in conjunction with the relations

$$[f,\hat p_i^2] = 4i\hbar \frac{\partial f}{\partial x_i}\hat p_i + 4\hbar^2 \frac{\partial^2 f}{\partial x_i^2}; \quad i = 1,2, \tag{11}$$

$$[f,\hat p_1 \hat p_2] = 2i\hbar \frac{\partial f}{\partial x_2}\hat p_1 + 2i\hbar \frac{\partial f}{\partial x_1}\hat p_2 + 4\hbar^2 \frac{\partial^2 f}{\partial x_1 \partial x_2}, \tag{12}$$

which are valid for any $f(x_1,x_2)$. In summary, it is found that the ten conditions are as follows:

$$\frac{\partial f_1}{\partial x_2} = \frac{\partial f_2}{\partial x_1} = 0, \tag{13}$$

$$\frac{\partial f_1}{\partial x_1} + \frac{\partial f_3}{\partial x_2} = \frac{\partial f_2}{\partial x_2} + \frac{\partial f_3}{\partial x_1} = 0, \tag{14}$$

$$\frac{\hbar}{2m\mu}\frac{\partial f_4}{\partial x_2} - A_2 \frac{\partial f_1}{\partial x_1} + 2f_1 \frac{\partial A_2}{\partial x_1} + f_3 \frac{\partial A_2}{\partial x_2} = 0, \tag{15}$$

$$\frac{\hbar}{2m\mu}\frac{\partial f_5}{\partial x_1} - A_1 \frac{\partial f_2}{\partial x_2} + 2f_2 \frac{\partial A_1}{\partial x_2} + f_3 \frac{\partial A_1}{\partial x_1} = 0, \tag{16}$$

$$\begin{aligned} &\frac{\hbar}{2m\mu}\left(\frac{\partial f_4}{\partial x_1} + \frac{\partial f_5}{\partial x_2}\right) - \frac{i\hbar^2}{m\mu}\frac{\partial^2 f_3}{\partial x_1 \partial x_2} + \frac{\partial}{\partial x_1}(A_1 f_1) \\ &+ f_1 \frac{\partial A_1}{\partial x_1} + \frac{\partial}{\partial x_2}(A_2 f_2) + f_2 \frac{\partial A_2}{\partial x_2} \\ &\qquad + f_3\left(\frac{\partial A_1}{\partial x_2} + \frac{\partial A_2}{\partial x_1}\right) = 0, \end{aligned} \tag{17}$$

$$\begin{aligned} &\frac{\hbar}{2\mu}\frac{\partial f_6}{\partial x_1} + f_5 \frac{\partial A_1}{\partial x_2} + f_4 \frac{\partial A_1}{\partial x_1} - A_2 \frac{\partial f_5}{\partial x_1} \\ &\qquad - A_1 \frac{\partial f_5}{\partial x_2} - \frac{i\hbar^2}{m\mu}\frac{\partial^2 f_5}{\partial x_1 \partial x_2} - \frac{f_3}{\mu}\frac{\partial U}{\partial x_1} \\ &\qquad -2i\hbar f_3 \frac{\partial^2 A_1}{\partial x_1 \partial x_2} - 2i\hbar f_2 \frac{\partial^2 A_1}{\partial x_2^2} \\ &\qquad - \frac{2\hbar}{\mu} f_2 \frac{\partial U}{\partial x_2} - 2i\hbar f_1 \frac{\partial^2 A_1}{\partial x_1^2} = 0, \end{aligned} \tag{18}$$

$$\begin{aligned} &\frac{\hbar}{2\mu}\frac{\partial f_6}{\partial x_2} + f_5 \frac{\partial A_2}{\partial x_2} + f_4 \frac{\partial A_2}{\partial x_1} - A_2 \frac{\partial f_4}{\partial x_1} \\ &\qquad - A_1 \frac{\partial f_4}{\partial x_2} - \frac{i\hbar^2}{m\mu}\frac{\partial^2 f_4}{\partial x_1 \partial x_2} - \frac{f_3}{\mu}\frac{\partial U}{\partial x_2} \\ &\qquad -2i\hbar f_3 \frac{\partial^2 A_2}{\partial x_1 \partial x_2} - 2i\hbar f_2 \frac{\partial^2 A_2}{\partial x_2^2} \\ &\qquad - \frac{2\hbar}{\mu} f_1 \frac{\partial U}{\partial x_1} - 2i\hbar f_1 \frac{\partial^2 A_2}{\partial x_2^2} = 0, \end{aligned} \tag{19}$$

$$\begin{aligned} &\frac{\partial^2 f_6}{\partial x_1 \partial x_2} - i\mu m A_2 \frac{\partial f_6}{\partial x_1} - i\mu m A_1 \frac{\partial f_6}{\partial x_2} - \frac{im}{\hbar} f_5 \frac{\partial W}{\partial x_2} \\ &\qquad - \frac{im}{\hbar} f_4 \frac{\partial W}{\partial x_1} - 2mf_3 \frac{\partial^2 W}{\partial x_1 \partial x_2} \\ &\qquad - 2mf_2 \frac{\partial^2 W}{\partial x_2^2} - 2mf_1 \frac{\partial^2 W}{\partial x_1^2} = 0. \end{aligned} \tag{20}$$

The first three functions f_1, f_2, and f_3 are easily found from Eqs. (13) and (14) to be

$$f_1 = \tfrac{1}{2}a_1 x_1^2 + b_1 x_1 + c_1 \tag{21}$$

$$f_2 = \tfrac{1}{2}a_1 x_2^2 + b_2 x_2 + c_2 \tag{22}$$

$$f_3 = -a_1 x_1 x_2 - b_2 x_1 - b_1 x_2 + c_3, \tag{23}$$

where $a_1, b_1, b_2, c_1, c_2, c_3$ are arbitrary constants. On combining equations (14) and (15) and utilizing Eq. (9), the fourth function is determined to be

$$f_4 = (-2m\mu/\hbar)(2A_1 f_1 + 2iBx_2 f_1 + A_2 f_3 - \psi_1). \tag{24}$$

Similarly on making use of Eqs. (14), (16), and (9), one finds the fifth function in the form

$$f_5 = (-2m\mu/\hbar)(2A_2 f_2 - 2iBx_1 f_2 + A_1 f_3 - \psi_2). \tag{25}$$

In Eqs. (24) and (25) the two functions

$$\psi_1 = -iBb_2 x_1^2 + \gamma_1 x_1 + \delta_1, \tag{26}$$

$$\psi_2 = iBb_1 x_2^2 + \gamma_2 x_2 + \delta_2, \tag{27}$$

in which δ_1, δ_2 are arbitrary constants while

$$\gamma_1 + \gamma_2 = -i(\hbar^2/m\mu)a_1 \tag{28}$$

have been introduced for conciseness.

The last function, f_6, in the representation of $\hat L$ is determined from Eqs. (18) and (19) to be

$$\begin{aligned} f_6 = {} & 4\frac{\mu^2 m^2}{\hbar^2}\left(A_1^2 + 2iBx_2 A_1 + \frac{i\hbar^2}{\mu m}\frac{\partial A_1}{\partial x_1} - B^2 x_2^2\right) f_1 \\ & + 4\frac{\mu^2 m^2}{\hbar^2}\left(A_2^2 - 2iBx_1 A_2 + \frac{i\hbar^2}{\mu m}\frac{\partial A_2}{\partial x_2} - B^2 x_1^2\right) f_2 \\ & + 4\frac{\mu^2 m^2}{\hbar^2}\left[A_1 A_2 + \frac{1}{2}\frac{i\hbar^2}{\mu m}\left(\frac{\partial A_1}{\partial x_2} + \frac{\partial A_2}{\partial x_1}\right)\right] f_3 \\ & - 4\frac{\mu^2 m^2}{\hbar^2}(\psi_1 A_1 + \psi_2 A_2) \\ & - 4i\frac{\mu^2 m^2}{\hbar^2} B\left[\left(\gamma_1 + \frac{i\hbar^2}{2\mu m}a_1\right)x_1 x_2 - \left(\delta_2 + \frac{i\hbar^2}{2\mu m}b_2\right)\right. \\ & \left. \times x_1 + \left(\delta_1 + \frac{i\hbar^2}{2\mu m}b_1\right)x_2\right] + mN(x_1,x_2), \end{aligned} \tag{29}$$

where the function $N(x_1,x_2)$ is defined by

$$\begin{aligned} N(x_1,x_2) = {} & 2\int\left(f_3 \frac{\partial U}{\partial x_1} + 2f_2 \frac{\partial U}{\partial x_2}\right)dx_1 \\ & + 2\int\left[2f_1 \frac{\partial U}{\partial x_1} + f_3 \frac{\partial U}{\partial x_2}\right. \end{aligned}$$

$$-\frac{\partial}{\partial x_2}\int\left(f_3\frac{\partial U}{\partial x_1}+2f_2\frac{\partial U}{\partial x_2}\right)dx_1\Bigg]\,dx_2. \tag{30}$$

Equations (18) and (19) also require, however, that the consistency requirement $\partial^2 f_6/\partial x_1\partial x_2=\partial^2 f_6/\partial x_2\partial x_1$ be satisfied. This requirement can be reduced to the form

$$2f_1\frac{\partial^2 U}{\partial x_1^2}-2f_2\frac{\partial^2 U}{\partial x_2^2}+3\frac{df_1}{dx_1}\frac{\partial U}{\partial x_1}-3\frac{df_2}{dx_2}\frac{\partial U}{\partial x_2}=0. \tag{31}$$

Finally the established expressions for the six functions f_i, show that the tenth condition Eq. (20), is tantamount to

$$\left(\frac{df_1}{dx_1}-\frac{2im\mu}{\hbar^2}\psi_1-\frac{4\mu m}{\hbar^2}Bx_2f_1\right)\frac{\partial U}{\partial x_1}+\left(\frac{df_2}{dx_2}-\frac{2i\mu m}{\hbar^2}\psi_2+\frac{4\mu m}{\hbar^2}Bx_1f_2\right)\frac{\partial U}{\partial x_2}=0. \tag{32}$$

The general solution of the latter equation is

$$U=U\Bigg[\frac{2\mu B}{\hbar^2}x_2^2f_1+\frac{2\mu B}{\hbar^2}x_1^2f_2-a_1\frac{\mu B}{\hbar^2}x_1^2x_2^2-\left(a_1-\frac{2i\mu m}{\hbar^2}\gamma_1\right)x_1x_2+\left(b_2-\frac{2i\mu m}{\hbar^2}\delta_2\right)x_1-\left(b_1-\frac{2i\mu m}{\hbar^2}\delta_1\right)x_2\Bigg], \tag{33}$$

the functional dependence on the argument being arbitrary. Accordingly, it can be asserted that the operator $\widehat{L}$ is a symmetry operator of the Hamiltonian $\widehat{H}$ for those potentials of the form given by Eq. (33) which satisfy equation (31). The operator being quadratic should of course contain the Hamiltonian itself. It is easily seen on using Eqs. (23) through (30) that the coefficient of c_3 in $\widehat{L}$ is in fact equal to $2m\widehat{H}$, omitting the additive constant term in p_{z0}^2.

Equation (32) can also be satisfied by setting the coefficients of the partial derivatives separately equal to zero. This choice leads to $f_1=f_2=\psi_1=\psi_2=0$, $f_3=c_3$, the operator $\widehat{L}$ becoming proportional to the Hamiltonian, and hence is not of interest. However in the absence of the magnetic field this choice is not so restrictive[2] and potential functions other than those possessing the geometrical symmetry in Eq. (33) satisfy equations (31) and (32). The presence of the magnetic field effectively removes this possibility. The application of the foregoing results to three particular field geometries will now be briefly examined.

3. POTENTIAL FIELD CONFIGURATIONS

A. Axially symmetric field

The potential is a function of the argument x_1x_2. Equation (31) is identically satisfied and the form prescribed by Eq. (33) is met if $b_1=b_2=c_1=c_2=\delta_1=\delta_2=0$. The quadratic symmetry operator is expressible in the form

$$\widehat{L}=\frac{a_1}{2}\widehat{L}_1^2+i\hbar\left(a_1-2i\frac{\mu m}{\hbar^2}\gamma_1\right)\widehat{L}_1 \tag{34}$$

in which the first order operator

$$\widehat{L}_1=x_1\left(\hat{p}_1-\frac{e}{c}A_1\right)-x_2\left(\hat{p}_2-\frac{e}{c}A_2\right)-i\frac{eB}{c}x_1x_2 \tag{35}$$

also commutes with the Hamiltonian.

This result applies in particular to the isotropic oscillator.

B. Nonisotropic oscillator

The potential in this case is

$$U(x_1,x_2)=\tfrac{1}{8}[(k_1-k_2)x_1^2+2(k_1+k_2)x_1x_2+(k_1-k_2)x_2^2] \tag{36}$$

the unequal oscillator constants being represented by k_1 and k_2. It is a member of the class defined by Eq. (33) and satisfies Eq. (31) if $a_1=b_1=b_2=\delta_1=\delta_2=0$ and

$$c_1=c_2=\frac{\hbar^2}{16\mu B}(k_1-k_2), \tag{37}$$

$$\gamma_1=-\gamma_2=\frac{-i\hbar^2}{8\mu m}(k_1+k_2). \tag{38}$$

The second-order operator can be now written without difficulty. It is noteworthy that no first order symmetry operator exists for the nonisotropic oscillator, because by Eq. (37) $c_1\neq 0$, even though a second-order one does.

C. Magnetic field alone

Here $U=0$ and Eqs. (31) and (32) are identically satisfied without restrictions on the functions f_i $(i=1,...,6)$. The full quadratic operator contains within it three linear operators found by letting $f_1=f_2=f_3=0$ and consequently

$$f_4=\frac{2\mu m}{\hbar}(\gamma_1x_1+\delta_1), \tag{39}$$

$$f_5=\frac{2\mu m}{\hbar}(\gamma_2x_2+\delta_2), \tag{40}$$

$$f_6=-4\frac{\mu^2m^2}{\hbar^2}[(\gamma_1x_1+\delta_1)A_1+(\gamma_2x_2+\delta_2)A_2]-4i\frac{\mu^2m^2}{\hbar^2}B(\gamma_1x_1x_2-\delta_2x_1+\delta_1x_2) \tag{41}$$

with $\gamma_1+\gamma_2=0$ from Eq. (28). On setting the independent coefficients $\gamma_1,\delta_1,\delta_2$ equal to zero in pairs one finds the three operators

$$\widehat{L}_1=\hat{p}_1-\frac{2\mu m}{\hbar}A_1-2i\frac{\mu mB}{\hbar}x_2, \tag{42}$$

$$\widehat{L}_2=\hat{p}_2-\frac{2\mu m}{\hbar}A_2+2i\frac{\mu mB}{\hbar}x_1, \tag{43}$$

$$\widehat{L}_3=x_1(\hat{p}_1-(e/c)A_1)-x_2(\hat{p}_2-(e/c)A_2)-2i\frac{\mu mB}{\hbar}x_1x_2. \tag{44}$$

Their commutators are

$$[\widehat{L}_1,\widehat{L}_2]=4\mu mB\hat{1}, \tag{45}$$

$$[\widehat{L}_2,\widehat{L}_3]=2i\hbar\widehat{L}_2, \tag{46}$$

$$[\widehat{L}_3,\widehat{L}_1]=2i\hbar\widehat{L}_1, \tag{47}$$

and of course each commutes with the Hamiltonian. The expressions for these operators are the gauge-independent forms of the operators utilized by Yanagawa and Moriya[4] to construct a realization of the four-dimensional Lie algebra with the basis $\widehat{L}_1$, $\widehat{L}_2$, $\widehat{L}_3$, $\hat{1}$. These authors' choice of gauge corresponds to $A_1 = -(iB/2)x_2$.

[1]V. A. Fock, Z. Physik. **98**, 145–54 (1935).

[2]P. Winternitz, Ya. A. Smorodinskii, M. Uhlir, and I. Fris, Yad. Fiz. **4**, 625–35 (1966) [Sov. J. Nucl. Phys. **8**, 444–50 (1967)].

[3]A. A. Makarov, J. A. Smorodinsky, Kh. Valiev, and P. Winternitz, Nuovo Cimento A **52**, 1061–84 (1967).

[4]S. Yanagawa and T. Moriya, J. Math. Phys. **11**, 3244 (1970).

[5]T. P. Mitchell, J. Math. Phys. **20**, 382–84 (1979).

Exponential perturbations of the harmonic oscillator

Marco Maioli[a),b)]
Centre de Physique Theorique 2 et Centre de Physique Theorique C.N.R.S., Marseille, France

(Received 10 June 1980; accepted for publication 20 March 1981)

The operators $H(\beta) = p^2 + x^2 + \beta e^{rx}(r\in\mathbb{R}\setminus\{0\})$ in $L^2(\mathbb{R})$ are studied. The spectrum is discrete for $|\arg\beta| < \pi$, the eigenvalues admit an asymptotic expansion as $|\beta| \to 0$, and they have no Bender–Wu type singularities in the analytic continuation to any punctured sector of a logarithmic Riemann surface. For $\beta' < 0$, $H(\beta')$ defines a symmetric operator with deficiency indices (1,1), and all its self-adjoint extensions have discrete spectrum; however, any eigenvalue of $H(\beta')$, when continued to $\beta' < 0$, can be interpreted as a resonance of the problem.

PACS numbers: 03.65.Ge

I. INTRODUCTION

The object of this paper is any Hamiltonian of the form $H(\beta) = p^2 + x^2 + \beta e^{rx}(r\in\mathbb{R}\setminus\{0\})$, viewed as a quantum-dynamic operator in the Hilbert space $L^2(\mathbb{R})$. Though some similar perturbation problems have been widely studied in recent years (see Ref. 1 for a review) no treatment seems to exist for nonpolynomial perturbations of the harmonic oscillator which misbehave at infinity. In such cases, for example, it is not clear what kind of singularity can be expected for the eigenvalues at $\beta = 0$: here, for small $|\beta|$, we have analyticity on an arbitrary sector of a logarithmic Riemann surface, as we shall see. Notice that the perturbation coefficients give a stronger divergence than in the examples of perturbation series recently treated (Refs. 1 and 2): the interest of this research is mainly in view of a rigorous study of such singular perturbation series, as well as for the anharmonic oscillator. Besides, the one-dimensional potential $x^2 + \beta' e^{rx}(\beta' < 0)$ is not complete (see e.g. Ref. 3), and all self-adjoint extensions of the symmetric operator $H(\beta')$ have discrete spectrum, in spite of the shape of such potentials, for which the barrier penetration would rather suggest a continuous spectrum along $(-\infty, +\infty)$, and "unstable states" or resonances: a "pathology" which is also verified in other noncomplete potentials (see Ref. 4). These self-adjoint extensions are not very significant, and we shall rather study the analytic continuation to $\beta' < 0$ of the eigenvalues of $H(\beta)$, $|\arg\beta| < \pi$. The consideration of such continued eigenvalues is common in physical literature (see, e.g., Ref. 5, p. 1622, where the authors consider $p^2 + x^2 + \beta' x^4, \beta' < 0$) when they are found to be nonreal: then their real part in interpreted as the energy of the unstable state, and the imaginary part is assumed to be exponentially related to the lifetime (see also Ref. 1, XII. 6). In some important cases, it has been proved that such a continuation can be obtained as a second sheet pole of matrix elements of the resolvent of the given self-adjoint Schrödinger operator (and this way the most correct definition of resonance can be given; see Ref. 1, XII. 6).

Now, let us consider the analytic continuation to $\beta' < 0$ of the eigenvalues of $H(\beta)$, $|\arg\beta| < \pi$. First we can show that they are related to the self-adjoint extensions of $H(\beta')$ through a spectral concentration phenomenon (see Refs. 6, or 1, XII. 5). Then our question is whether they can be obtained as second sheet poles of "resolvent" matrix elements, taking into account that the resolvent of any self-adjoint extension of $H(\beta')$ cannot even have a second sheet, because each spectrum is discrete. To answer this question we can use the notion of generalized resolvent of a given symmetric operator. Recall that if A is a symmetric operator in Hilbert space $\mathscr{H}$ with deficiency indices (n,m) [so that $(A - E)^{-1}$ is well defined on the range of $(A - E)$ for Im $E > 0$] a generalized resolvent of A is a bounded, densely defined extension of $(A - E)^{-1}$ which can be expressed as $P^+(B^+ - E)^{-1}$, where $P^+ : \mathscr{H}^+ \to \mathscr{H}$ is the orthogonal projection onto $\mathscr{H}$ from some larger Hilbert space $\mathscr{H}^+ \supset \mathscr{H}$; and B^+ is a self-adjoint extension of A acting in $\mathscr{H}^+$ (for a detailed discussion see Ref. 7). In these terms, the result is the existence of a unique generalized resolvent $\mathscr{R}(E)$ of the symmetric operator $H(\beta')$, such that its scalar products, on some dense set in $L^2(\mathbb{R})$, admit second sheet poles; and these poles are exactly the above eigenvalues of $H(\beta)$ when continued to $\beta' < 0$. Thus we can say that our problems admit a natural notion of resonance, although all self-adjoint extensions have a discrete spectrum, by simply replacing their resolvents by $\mathscr{R}(E)$. At this point, we can give $\mathscr{R}(E)$ a physical characterization, as a limit of resolvents, corresponding to problems with potentials locally approximating our one.

In Sec. II the perturbation theory of $H(\beta)$, $|\arg\beta| < \pi$, is discussed, by introducing the translation $x \to x + r^{-1}\log(\beta^{-1})$, and by referring to Ref. 8 whenever possible. As a result, the existence and asymptotic expansion of the eigenvalues as $|\beta| \to 0$ and the absence of Bender–Wu type singularities are shown, and these properties are extended to N-dimensional cases. In Sec. III the Hamiltonian $H(\beta')$, $\beta' < 0$, is considered: it is proved that the real part of any eigenvalue of $H(\beta)$, when continued to $\beta' < 0$, is a pseudoeigenvalue of any self-adjoint extension of $H(\beta')$; it is also proved that any such continued eigenvalue is a second sheet pole of the generalized resolvent $\mathscr{R}(E)$. In the Appendix $\mathscr{R}(E)$ is also characterized as a weak limit of the resolvents of certain Schrödinger operators, whose potentials locally approximate $x^2 + \beta' e^{rx}$.

[a)]Partially supported by G.N.F.M., C.N.R.
[b)]Permanent address: Istituto Matematico, Università di Modena, 41100 Modena, Italy.

II. PERTURBATION THEORY

In the following, we mean by p the differential operator $(-id/dx)$, acting in $L^2(\mathbb{R})$ and unitarily equivalent to the multiplication operator by p through the Fourier transform. The other symbols of operator theory are also taken in the usual meaning. The natural logarithm of w is denoted by $\ln(w)$. First consider the differential expression $H(\alpha,\beta)=p^2+x^2+\alpha x+\beta e^{rx}$, $r\in\mathbb{R}\setminus\{0\}$, in order to prove the above mentioned results.

Lemma II.1: The operator $H(0,1)=p^2+x^2+e^{rx}$ is self-adjoint on the domain $D=D(p^2+x^2)\cap D(e^{rx})$. $H(0,1)$ has compact resolvent and, for all $u\in D$,

$$\|(p^2+x^2)u\|^2+\tfrac{1}{2}\|e^{rx}u\|^2\leqslant\|(p^2+x^2+e^{rx})u\|^2+b\,\|u\|^2. \quad \text{(II1)}$$

Proof: By standard arguments, $H(0,1)$ is self-adjoint on D with compact resolvent if (II1) is proved. To this end, it is enough to remark that, for some constant b,

$$\begin{aligned}(p^2+x^2+e^{rx})^2&=(p^2+x^2)^2+e^{2rx}+e^{rx}(p^2+x^2)\\&\quad+(p^2+x^2)e^{rx}\\&\geqslant(p^2+x^2)^2+e^{2rx}+p(-ire^{rx}+e^{rx}p)+e^{rx}p^2\\&=(p^2+x^2)^2+e^{2rx}-r^2e^{rx}+2pe^{rx}p\\&\geqslant(p^2+x^2)^2+\tfrac{1}{2}e^{2rx}-b.\end{aligned}$$

Taking expectation values (for $u\in D$), the lemma is proved.

By the same methods, the following property can be seen.

Lemma II.2: The maximal mulstiplication operator by x is $H(0,1)$—relatively bounded with an arbitrarily small relative bound. On the basis of Lemmas II.1 and II.2, by exactly repeating the arguments of Ref. 8, pp. 83–84, one can prove the following theorem.

Theorem II.3: For any complex α, and $\beta>0$, $H(\alpha,\beta)=p^2+x^2+\alpha x+\beta e^{rx}$ is defined as a closed operator on the domain D, with compact resolvents which are analytic in α. Moreover, $[H(\alpha_0,\beta)-z]^{-1}$ is analytic in β for β in a neighborhood of any positive number. All eigenvalues $E(\alpha,\beta)$ of $H(\alpha,\beta)$ are nondegenerate and analytic in both variables in a neighborhood of any real α_0 and positive β_0.

These facts allow labeling of the eigenvalues by an ordering: $E_n(\alpha,\beta)$, $n=0,1,2,\cdots$. To study them in more detail, let us introduce a convenient translation.

Theorem II.4: Let $n=0,1,2,\cdots$. For $\beta>0$, $\lambda>0$ and wherever an analytic continuation is possible,

$$E_n(0,\beta)=(r^{-1}\ln\lambda)^2+E_n(2r^{-1}\ln\lambda,\beta\lambda). \quad \text{(II2)}$$

Proof: Let us define $[U(\lambda)f](x)=f(x+r^{-1}\ln\lambda)$. Then $U(\lambda)$ is a group of unitary translations for $\lambda>0$, and

$$U(\lambda)(p^2+x^2+\beta e^{rx})U(\lambda)^{-1}=p^2+x^2+(r^{-1}\ln\lambda)^2+2(r^{-1}\ln\lambda)x+\beta\lambda e^{rx},$$

which proves (II2).

Remark: As a consequence of Theroem II.4, for $\lambda=\beta^{-1}$ we have

$$E_n(0,\beta)=(r^{-1}\ln\beta^{-1})^2+E_n(2r^{-1}\ln\beta^{-1},1), \quad \text{(II3)}$$

which will be useful later and allows us to state the following:

Corollary II.5: Let $n=0,1,\cdots$. Then $E_n(0,\beta)$ has a convergent expansion in powers of $\ln\beta$ near $\beta=1$.

Lemma II.6: Let β be complex with $\mathrm{Im}\beta\neq0$, and let $X=(x+r^{-1}\ln(m))$, $m\in\mathbb{N}$. Then, for all $v\in D$, the following estimate holds:

$$\begin{aligned}&a\|(p^2+X^2)v\|^2+a|\beta|^2\|e^{rX}v\|^2\\&\leqslant\|(p^2+X^2+\beta e^{rX})v\|^2+b\,\|v\|^2\end{aligned} \quad \text{(II4)}$$

for some a, b, m depending on β.

Proof: As quadratic forms on $D\otimes D$, we have

$$\begin{aligned}&(p^2+X^2+\bar\beta e^{rX})(p^2+X^2+\beta e^{rX})=(p^2+X^2+m\bar\beta e^{rx})(p^2+X^2+m\beta e^{rx})=(p^2+X^2)^2+m^2|\beta|^2e^{2rx}\\&+m(\mathrm{Re}\beta)[(p^2+X^2)e^{rx}+e^{rx}(p^2+X^2)]+im(\mathrm{Im}\beta)(p^2e^{rx}-e^{rx}p^2)=|\beta^{-1}\mathrm{Re}\beta|(p^2+X^2\pm m|\beta|e^{rx})\\&+(1-|\beta^{-1}\mathrm{Re}\beta|)[(p^2+X^2)^2+m^2|\beta|^2e^{2rx}]+mr(\mathrm{Im}\beta)(pe^{rx}+e^{rx})\geqslant\quad(\text{with }\ 2a=1-|\beta^{-1}\mathrm{Re}\beta|)\\&2a[(p^2+X^2)^2+|\beta|^2m^2e^{2rx}]+|mr\mathrm{Im}\beta|(p\pm e^{rx})^2-|mr\mathrm{Im}\beta|(p^2+e^{2rx})\\&\geqslant a[(p^2+X^2)^2+|\beta|^2e^{2rX}]-b+[a(p^2+X^2)^2-|mr\mathrm{Im}\beta|p^2+b]+[am^2|\beta|^2e^{2rx}-|mr\mathrm{Im}\beta|e^{2rx}].\end{aligned}$$

By a suitable choice of m and $b=b(m)$, the last two terms can be made positive, whence (II4).

Theorem II.7: $H(0,\beta)=p^2+x^2+\beta e^{rx}$, defined on $D=D(p^2+x^2)\cap D(e^{rx})$, is an analytic family of type A with compact resolvent for $|\arg\beta|<\pi$.

Proof: Let $U_m v(x)=v(x-r^{-1}\ln(m))$: then the unitary translation U_m leaves unchanged the domain D, so that any $u\in D$ can be written as $u=U_m v$, where $v\in D$. In the notation of the preceding lemma, $\|(p^2+X^2+\beta e^{rX})v\|=\|U_m^{-1}(p^2+x^2+\beta e^{rx})U_m v\|=\|(p^2+x^2+\beta e^{rx})u\|$, and $H(0,\beta)$ satisfies a quadratic estimate similar to (II4). Thus, as a consequence of this estimate, exactly as in Ref. 8, Theorem II.9.2, the statement is verified.

Theorem II.8: $H(\alpha,0)=p^2+x^2+\alpha x$, $\alpha\in\mathbb{C}$, defined on $D(p^2+x^2)$, is an entire family of type A with compact resolvent and analytic eigenvalues $E_n(\alpha,0)=2n+1-\alpha^2/4$.

Proof: If $\alpha\in\mathbb{R}$, $H(\alpha,0)$ has compact resolvent by a translation $x\to x-\alpha/2$; moreover, the resolvents are analytic in a neighborhood of any real α, since x is infinitesimally small with respect to $H(\alpha,0)$. For $\mathrm{Im}\alpha\neq0$, again as quadratic forms we have:

$$\begin{aligned}&(p^2+x^2+\bar\alpha x)(p^2+x^2+\alpha x)=p^4+|\alpha|^2x^2+2\mathrm{Re}\alpha x^3-2\\&+2px^2p+\mathrm{Re}\alpha(p^2x+xp^2)+i\mathrm{Im}\alpha[p^2,x]\\&\geqslant(1-|\alpha^{-1}\mathrm{Re}\alpha|)(p^4+|\alpha|^2x^2)+|\alpha^{-1}\mathrm{Re}\alpha|(p^2\pm|\alpha|x)^2\\&+x^4+2\mathrm{Re}\alpha x^3-2+\mathrm{Im}\alpha p.\end{aligned}$$

As in Lemma II.6, one can find $a,b>0$ such that $ap^4+ax^4\leqslant(p^2+x^2+\bar\alpha x)(p^2+x^2+\alpha x)+b$. This implies that $H(\alpha,0)$ is closed on $D(p^2)\cap D(x^2)$ and that x is small with respect to $H(\alpha,0)$, whence the assertion about $H(\alpha,0)$. Then, by the above translation $E_n(\alpha,0) = 2n + 1 - \alpha^2/4$.

Theorem II.9: Let $\alpha\in\mathbb{C}$, $|\arg\beta| < \pi$; then $H(\alpha,\beta) = p^2 + x^2 + \alpha x + \beta e^{rx}$, defined on D, is an analytic family of type A with compact resolvent. For any purely imaginary α,

$$|\beta|\,\|e^{rx}u\| \leqslant a\|H(\alpha,|\beta|)u\| + b\,\|u\| \qquad \text{(II5)}$$

for all $u\in D$, with a,b independent of $\alpha,|\beta|$.

Proof: For $\alpha\in\mathbb{R}$, $p^2 + x^2 + \alpha x + \beta e^{rx}$ is unitarily equivalent to $p^2 + x^2 - \alpha^2/4 + \beta e^{-r\alpha/2}e^{rx}$ by a translation $x\to x - \alpha/2$; so it is closed on D with compact resolvents by Theorem II.7. On the other hand, by the quadratic estimate (II4), x is infinitesimally small with respect to $p^2 + x^2 + \beta e^{rx}$ when $\mathrm{Im}\beta \neq 0$. So, by Theorem II.3, $H(\alpha,\beta)$ is an entire family of type A for any fixed β in the cut plane. Since it has compact resolvents for real α, the property extends to the whole region of analyticity. In order to show (II5) it is enough to consider, as quadratic forms on $D \otimes D$,

$$\begin{aligned}
&(p^2 + x^2 + \bar{\alpha}x + |\beta|e^{rx})(p^2 + x^2 + \alpha x + |\beta|e^{rx}) \\
&\quad = (p^2 + x^2 + \bar{\alpha}x)(p^2 + x^2 + \alpha x) + |\beta|^2 e^{2rx} \\
&\qquad + |\beta|[e^{rx}(p^2 + x^2) + (p^2 + x^2)e^{rx}] \\
&\quad \geqslant |\beta|^2 e^{2rx} + |\beta|(p^2 e^{rx} + e^{rx}p^2),
\end{aligned}$$

and to proceed in analogy with the preceding estimates. So the assertion is proved.

The above propositions allow us to get a norm resolvent convergence of $H(\alpha,\beta)$ as $\beta\to 0$; it is known that an analogous convergence holds for every x^{2m}- perturbation of the harmonic oscillator ($m\in\mathbb{N}$, Ref. 8), as well as βx^{2m+1}- perturbations (for nonreal β only, Ref. 4). Let $H(\alpha,|\beta|)$ be given as in Theorem II.9, and let $H(\alpha,0)$ be defined by Theorem II.8.

Theorem II.10: If $\mathrm{Re}\alpha = 0$ and $d > 0$ is given,

$$\|[H(\alpha,|\beta|) - E]^{-1} - [H(\alpha,0) - E]^{-1}\| \to 0 \quad \text{as} \quad |\beta| \to 0$$

for some $E\in\mathbb{C}$ and uniformly for $|\alpha| < d$.

Proof: If $\mathrm{Re}\alpha = 0$, the union of the numerical ranges of $H(\alpha,|\beta|)$ is contained in a half-plane; so there are $E\in\mathbb{C}$ and $c > 0$ such that $\|[H(\alpha,|\beta|) - E]^{-1}\| < c$. We have $\Delta \equiv [H(\alpha,|\beta|) - E]^{-1}\beta - [H(\alpha,0) - E]^{-1} = -[H(\alpha,0) - E]^{-1}(1 + |x|)(\ln|\beta|^{-1/2|r|})^{-1}$ $(\ln|\beta|^{-1/2|r|})(1 + |x|)^{-1}|\beta|e^{rx}[H(\alpha,|\beta|) - E]^{-1}$. Now, $(\ln|\beta|^{1/2r})^2(1 + |x|)^{-2}|\beta|^2 e^{2rx} \leqslant |\beta|^2 e^{2rx} + 1$. Indeed, it is enough to show such inequality for $(1 + |x|) < \ln|\beta|^{-1/2|r|}$. For such values of x, $(\ln|\beta|^{-1/2|r|})^2|\beta|^2 e^{2|rx|}$ $\leqslant (\ln|\beta|^{-1/2|r|})^2|\beta| \leqslant 1$, uniformly for $0 < |\beta| < \beta_0$, for some $\beta_0 > 0$. By this inequality and by (II5) we have $(\ln|\beta|^{-1/2|r|})(1 + |x|)^{-1}|\beta|e^{rx}[H(\alpha,|\beta|) - E]^{-1}$ bounded uniformly over α and β, $\mathrm{Re}\alpha = 0$ and $0 < |\beta| < \beta_0$. On the other hand, if d is fixed $(1 + |x|)$ is relatively bounded with respect to $H(\alpha,0)$ uniformly for $|\alpha| < d$ (see, e.g., the proof of Theorem II.8). This implies that $\|\Delta\| \to 0$ as $|\beta| \to 0$, at least as $(-\ln|\beta|^{1/2|r|})^{-1}$.

Theorem II.11: Let $\mathrm{Re}\alpha = 0$, $d > 0$, and let $E_n(\alpha,0)$ be the nth eigenvalue of $H(\alpha,0)$. Then, for small $|\beta|$, there is exactly one eigenvalue $E_n(\alpha,|\beta|)$ of $H(\alpha,|\beta|)$ near $E_n(\alpha,0)$. As $|\beta| \to 0$, one has $E_n(\alpha,|\beta|) \to E_n(\alpha,0)$ uniformly for $|\alpha| < d$.

Proof: A direct consequence of the norm resolvent convergence of Theorem II.10.

Theorem II.12: There is a $B > 0$ such that for $|\beta| < B$, $|\arg\beta| < \pi$, $H(0,\beta)$ has exactly one eigenvalue near $2n + 1$. Such eigenvalues are analytic functions of β for $|\arg\beta| < \pi$, $|\beta| < B$, and admit an analytic continuation across the real axis, on a logarithmic Riemann surface, to any sector $\{\beta: 0 < |\beta| < B, |\arg\beta| < \theta\}$, $\theta > \pi$, $\widetilde{B} = \widetilde{B}(\theta)$...

Proof: For $|\arg\beta| < \pi$, $|\beta|$ small, by setting $\lambda = \exp(-i\arg\beta)$ in (II2),

$$E_n(0,\beta) = (-ir^{-1}\arg\beta)^2 + E_n(-2ir^{-1}\arg\beta,|\beta|),$$

so that $E_n(0,\beta)$ tends to $2n + 1$ as $\beta\to 0$. By the same reason, it admits an analytic continuation to any sector as above.

Theorem II.13: Let $E(0,\beta)$ denote an arbitrary eigenvalue of $H(0,\beta)$, $|\arg\beta| < \pi$, $|\beta| < B$. Then the Rayleigh–Schrödinger formal series is uniformly asymptotic, as $\beta\to 0$, to the function $E(0,\beta)$ in any sector $|\arg\beta| < \theta$, $\theta \geqslant \pi$.

Proof: Since $E(0,\beta) = \alpha^2 + E(2\alpha,|\beta|)$, where $\beta = |\beta|e^{-r\alpha}$, it suffices to see that $E(2\alpha,|\beta|)$ admits asymptotic expansion uniformly for $|\alpha| < |r^{-1}\theta|$, when α is purely imaginary. Let $S(\mathbb{R})$ be the Schwartz space and let $S_1 \subset S(\mathbb{R})$ be the subset of functions that decrease at infinity faster than any inverse power of $\cosh(x)$. Since every unperturbed eigenvector $\psi(2\alpha,0)$ of $H(2\alpha,0)$ belongs to S_1, setting $V = e^{rx}$, $H_\beta = H(2\alpha,|\beta|)$, the formal equality,

$$\begin{aligned}
(H_\beta - E)^{-1} = {} & \sum_{k=0}^{N}(-\beta)^k(H_0 - E)^{-1}[V(H_0 - E)^{-1}]^k \\
& + (-\beta)^{N+1}(H_\beta - E)^{-1}[V(H_0 - E)^{-1}]^{N+1},
\end{aligned}$$

holds on $\psi(2\alpha,0)$. Moreover, if $E(2\alpha,0)$ is the unperturbed eigenvalue and $\psi\in S_1$, one can prove that $[V(H_0 - E)^{-1}]^N\psi$ is continuous in E on $\{E: |E - E(2\alpha,0)| = \epsilon\}$, for small ϵ. So, by the norm resolvent convergence of Theorem II.10, one can apply the arguments of Ref. 1, Theorem XII.14, and the assertion is proved.

It is clear that all the preceding arguments do not depend on the dimension. This fact allows extending the results of this section to any N-dimensional case (compare with Ref. 8, Theorem II.2.1).

Theorem II.14: Let

$$H_0 = \sum_{k=1}^{N}(p_k^2 + w_k^2 q_k^{\,2}) \quad \text{and} \quad V = \exp\Big(\sum_{k=1}^{N} a_k q_k\Big),$$

where $w_k, a_k \in \mathbb{R}\setminus\{0\}$ for all k; let $E_n(0,\beta)$ be an eigenvalue of $H_0 + \beta V$. Then in any sector $|\arg\beta| < \theta$, $\theta > 0$, $E_n(0,\beta)$ is analytic for small $|\beta|$ and the Rayleigh–Schrödinger series is asymptotic in the sector.

By a reasonable expectation about the perturbation coefficients, the Rayleigh–Schrödinger series is not convergent. In Ref. 2 the authors estimate the perturbation coefficients by the method of Lipatov, and the result is that the divergence is faster than $(km)!$ for any $k\in\mathbb{N}$. Thus, from now on, we shall make the assumption that $\beta = 0$ is a singular point for the eigenvalues. Then, since $E_n(0,\beta) \to E_n(0,0)$ as $\beta\to 0$, the negative half-axis must be a cut, and the eigenvalues admit an analytic continuation across the cut, for small $|\beta|$, to arbitrary sectors of a logarithmic Riemann surface. Notice the complete absence of Bender–Wu type singularities (see Ref. 1 for related references).

III. SPECTRAL ANALYSIS

As already mentioned in the Introduction, the potential $x^2 + \gamma e^{rx}$, $\gamma < 0$, does not give rise to a continuous spectrum. This behavior is not very exceptional,[4] and is intuitively related to noncompleteness of these potentials. The following proof is also analogous to the corresponding one in Ref. 4.

Lemma III.1: Let $H(\gamma) = p^2 + x^2 + \gamma e^{rx}$ ($\gamma < 0$) be the symmetric operator defined on D. Its closure $\bar{H}(\gamma)$ has deficiency indices (1,1), and admits infinitely many self-adjoint extensions, which have discrete spectrum.

Proof: Only the last assertion has to be verified (for the remaining ones, see, e.g., Ref. 9). For $\gamma < 0$, $r > 0$, let $u_a(x)$ be a solution of the differential equation $-y'' + x^2 y + \gamma e^{rx} y = iy$ such that $\lim_{x\to+\infty} [u_a(x)\bar{u}_a'(x) - \bar{u}_a'(x)\bar{u}_2(x)] = 0$ and let $v_a(x)$ be a (linearly independent) solution which is L^2 at $-\infty$. Then it is well known that the Green function

$$G_a(x,y;i) = \begin{cases} W(a)^{-1} v_a(x) u_a(y), & -\infty < x \leqslant y < +\infty, \\ W(a)^{-1} u_a(x) v_a(y), & -\infty < y \leqslant x < +\infty, \end{cases}$$

$W(a)$ being the Wronskian of u_a and v_a, specifies the integral kernel of $(H_a - i)^{-1}$, where H_a is a self-adjoint extension of $H(\gamma)$. Through standard "WKB type" estimates (again Ref. 9 can be seen), one easily finds the following asymptotic behaviors:

$$|u_a(x)| \sim |x|^{-(1/2)} e^{(1/2)x^2} (x \to -\infty)),$$
$$|v_a(x)| \sim |x|^{-1/2} e^{-(1/2)x^2} (x \to -\infty),$$
$$|u_a(x)| \sim e^{-rx/4} (x \to +\infty),$$
$$|v_a(x)| \sim e^{-rx/4} \quad (x \to +\infty),$$

so that an easy computation yields

$$\int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} |G(x,y;i)|^2 dx\, dy < -\infty.$$

Hence $(H_a - i)^{-1}$ is Hilbert–Schmidt, and H_a has a discrete spectrum. Since all self-adjoint extensions have the same essential spectrum (again Ref. 9) the assertion is proven for $r > 0$. For $r < 0$ the proof is analogous.

From now on, let $H_a(\gamma)$ be any fixed self-adjoint extension of $H(\gamma)$, $\gamma < 0$. Let $H(\beta)$ be the operator family of Theorem II.7 and let $E(\beta)$ be any eigenvalue of $H(\beta)$, $|\arg\beta| < \pi$. The following theorem is a consequence of the results of Sec. II.

Theorem III.2: Let $|\beta|$ be small, $|\arg\beta| < \pi$. Then the analytic continuation of $E(\beta)$ to the negative half-axis is such that

$$\lim_{(\arg\beta)\to\pi-0} E(\beta)$$

is not identically zero.

Proof: Call $E(\alpha,\beta)$ any eigenvalue of $p^2 + x^2 + \alpha x + \beta e^{rx}$ (defined on D), so that $E(0,\beta) = E(\beta)$. Set

$$\lim_{(\arg\beta)\to\pi-0} E(\beta) = E(|\beta| e^{i\pi})$$

and

$$\lim_{(\arg\beta)\to\pi+0} E(\beta) = E(|\beta| e^{-i\pi}).$$

By analyticaly continuing (II3), we get

$$E(|\beta| e^{i\pi}) = (r^{-1}\ln|\beta| + r^{-1} i\pi)^2 + E(-2r^{-1}\ln|\beta| - 2r^{-1} i\pi, 1)j,$$
$$E(|\beta| e^{-i\pi}) = (r^{-1}\ln|\beta| - r^{-1} i\pi)^2 + E(-2r^{-1}\ln|\beta| + 2r^{-1} i\pi, 1).$$

So $\bar{E}(|\beta| e^{i\pi}) = E(|\beta| e^{-i\pi})$ and

$$(2i) \lim_{(\arg\beta)\to\pi-0} \operatorname{Im} E(\beta)$$

gives the difference across the cut. Since $E(\beta)$ is not analytic near $\beta = 0$, this value is not identically zero.

Remark: Let $\gamma < 0$; then the perturbation series $\Sigma a_k \gamma^k$ has real partial sums because every a_k is real. thus if $\operatorname{Im} E(\gamma)$ is the above defined limit, $|\gamma|^{-k} \operatorname{Im} E(\gamma) \to 0$ as $\gamma \to 0$, for any $k \in \mathbb{N}$; and $\operatorname{Re} E(\gamma)$ has $\Sigma a_k \gamma^k$ as an asymptotic expansion to all orders in γ. We can now specify the relation between such continued eigenvalues and the self-adjoint extensions of $H(\gamma)$, $\gamma < 0$.

Theorem III.3: Let $\gamma < 0$. As $\gamma \to 0$, the spectrum of any self-adjoint extension $H_a(\gamma)$ of $H(\gamma)$ is asymptotically concentrated to all orders in γ near the eigenvalues of the harmonic oscillator H_0. If $E(\beta)$ is an eigenvalue of $H(\beta)$, $|\arg\beta| < \pi$, the real part of its analytic continuation to $\gamma < 0$ is a pseudoeigenvalue (to all orders of γ) of any self-adjoint extension $H_a(\gamma)$.

Proof: For $\gamma < 0$, $H_a(\gamma)$ converges strongly in the generalized sense to H_0 as $\gamma \to 0$; in fact, for every $u \in C_0^\infty(\mathbb{R})$, $H_a(\gamma) u \to H_0 u$ as $\gamma \to 0$, and $C_0^\infty(\mathbb{R})$ is a core of H_0; so Corollary VIII.1.6 of Ref. 6 can be applied. Thus, by the standard criteria of the spectral concentration (Ref. 1, Theorem VIII.5.4 and Remark VIII.5.6; see also Ref. 1, XII.5), if E_0 is an eigenvalue of H_0 and I is an open interval such that $\bar{I} \cap \sigma(H_0) = \{E_0\}$, for all positive integers N, there is a function $f^{(N)}(\gamma)$ obeying $|\gamma|^{-N} f^{(N)}(\gamma) \to 0$ as $\gamma \to 0$, such that the part of the spectrum of $H_a(\gamma)$ in I is asymptotically concentrated in the interval

$$\Big(\sum_0^N a_k \gamma^k - f^{(N)}(\gamma), \sum_0^N a_k \gamma^k + f^{(N)}(\gamma)\Big),$$

where $\Sigma a_k \gamma^k$ is the Rayleigh–Schrödinger expansion. Since $\operatorname{Re} E(\gamma)$ admits this series as an asymptotic expansion (Remark after Theorem III.2), it satisfies the same property as $\gamma \to 0$. Hence, it is a pseudoeigenvalue of $H_a(\gamma)$, to all orders in γ (see Ref. 6, Remark VIII.5.5).

The second result involves the notion of a generalized resolvent of a given symmetric operator. The eigenvalues of $H(\beta)$, in the limit $\gamma < 0$, are proved to be second sheet poles of a uniquely determined generalized resolvent of the symmetric operator $H(\gamma)$. Recall that an operator-valued function $F(t): \mathbb{R} \to B(X)$ [X is a complex Hilbert space, $B(X)$ the space of all bounded operators on X] is a spectral function if (a) $F(-\infty) = 0$, $F(+\infty) = I$, (b) $F(t-0) = F(t)$, $t \in \mathbb{R}$ (c) for $s > t$, $F(s) - F(t) \geqslant 0$. An operator-valued function $\mathscr{R}(E): \mathbb{C} \setminus \mathbb{R} \to B(X)$ is a generalized resolvent of a symmetric operator $A: X \to X$, if it can be represented, in the weak sense, as

$$\mathscr{R}(E) = \int_{-\infty}^{+\infty} (t-E)^{-1} d\,F(t),$$

where $F(t)$ is a spectral function, and $\mathscr{R}(E) \supset (A-E)^{-1}$. If $F(t)$ is an orthogonal spectral family, $\mathscr{R}(E)$ is the resolvent of a self-adjoint extension of A, (for a more detailed discussion, see Ref. 7).

For this proof, the analytic translation methods are needed (compare with Ref. 10). Set $U(\theta)f(x) = f(x + r^{-1}\theta)$. Then, for real θ,

$$H(\beta,\theta) = U(\theta)H(\beta)U(\theta)^{-1} = p^2 + (x + r^{-1}\theta)^2 + \beta e^{\theta} e^{rx} \tag{III1}$$

is unitarily equivalent to $H(\beta)$. As θ is complex, by the results of Sec. II, $H(\beta,\theta)$ is an analytic family of operators with compact resolvents for $|\arg(\beta e^{\theta})| < \pi$, when defined on $D = D(p^2 + x^2) \cap D(re^{rx})$. $U(\theta)$ too. In particular, the eigenvalues of $H(\beta,\theta)$ do not depend on θ by the analytic translation property, and when $|\arg\beta| < \pi$, they coincide with the eigenvalues of $H(\beta)$ because for real θ the translation is a unitary operation. Let $\gamma < 0$, $\theta = \theta_1 + i\theta_2$, $0 < \theta_2 < \pi/2$. Let $H(\gamma,\theta_1) = p^2 + (x + r^{-1}\theta_1)^2 + \gamma e^{\theta_1} e^{rx}$ [unitarily equivalent to $H(\gamma)$]; and let $H(\gamma,\theta)$ be the family of compact resolvent operators defined on D by (III1).

Lemma III.4: Let $\gamma < 0$, $0 < \theta_2 < \pi/2$. Then the numerical range of $H(\gamma,\theta)$ is contained in the half-plane $\mathrm{Re}z \geqslant (\cot\theta_2)\mathrm{Im}z - c$, where the constant c is independent of θ_2.

Proof: Let $z = \langle H(\gamma,\theta)u,u\rangle, \|u\| = 1$. Then $\mathrm{Re}z = \langle p^2u,u\rangle + \langle (x + r^{-1}\theta_1)^2 u,u\rangle - r^{-2}\theta_2^2\langle u,u\rangle + \gamma e^{\theta_1}(\cos\theta_2)\langle e^{rx}u,u\rangle$.
$\mathrm{Im}z = 2r^{-1}\theta_2\langle (x + r^{-1}\theta_1)u,u\rangle + \gamma e^{\theta_1}(\sin\theta_2)\langle e^{rx}u\, u\rangle$.

Hence, for some constant c, independent of θ_2,

$$\begin{aligned}
\mathrm{Re}z &\geqslant \langle (x + r^{-1}\theta_1)^2 u,u\rangle - r^{-2}\theta_2^2\langle u,u\rangle + \gamma e^{\theta_1}(\cos\theta_2)\langle e^{rx}u,u\rangle \\
&\geqslant (\cot\theta_2)\,(2r^{-1}\theta_2)\langle (x + r^{-1}\theta_1)u,u\rangle \\
&\quad + \gamma e^{\theta_1}(\cot\theta_2)\,(\sin\theta_2)\langle e^{rx}u,u\rangle - c \\
&= (\cot\theta_2)\mathrm{Im}z - c.
\end{aligned}$$

This proves the lemma.

Let us define $H_c(\gamma,\theta) = H(\gamma,\theta) + c$, c as in Lemma III.4, so that its numerical range lies in the half-plane $\mathrm{Re}z \geqslant (\cot\theta_2)\mathrm{Im}z$. Similarly, let $H_c(\gamma) = H(\gamma) + c$.

Lemma III.5: Let $\mathrm{Im}z > 0$, $0 < \mathrm{Im}\theta < \pi/2$. Then $[e^{-\theta}H_c(\gamma,\theta) - z]^{-1}$ converges strongly to $[\overline{H}_c(\gamma) - z]^{-1}$ on $R(\overline{H}_c(\gamma) - i) = L^2 - N_i$ as $\theta \to 0$ [where N_i is the deficiency subspace of $H_c(\gamma)$], uniformly on compacts in z.

Proof: First remark that, by Lemma III.4, the compact resolvent $[e^{-\theta}H_c(\gamma,\theta) - z]^{-1}$ is uniformly bounded by $(\mathrm{Im}z)^{-1}$ for $0 < \mathrm{Im}\theta < \pi/2$. In addition, $[\overline{H}_c(\gamma) - z]^{-1}$ acting on $R(\overline{H}_c(\gamma) - z)$, $\mathrm{Im}z > 0$, is bounded. Now, for $u \in R(H_c(\gamma) - i)$ we can write $[H_c(\gamma,\theta) - z]^{-1}u - [H_c(\gamma) - z]^{-1}u = [H_c(\gamma,\theta) - z]^{-1} [H_c(\gamma,\theta) - H_c(\gamma)]\,[H_c(\gamma) - z]^{-1}u \to 0$ as $\theta \to 0$, because $H_c(\gamma,\theta)v \to H_c(\gamma)v$ for $v \in D$ as $\theta \to 0$. Now, $R(H_c(\gamma) - i)$ is a dense set in $R(\overline{H}_c(\gamma) - i)$ because D is by definition a core of $H_c(\gamma)$. Hence, by the uniform boundedness, the convergence takes place on the whole of $R(\overline{H}_c(\gamma) - i)$, with the stated uniformity.

Lemma III.6: Let $\gamma < 0$ be fixed, $\mathrm{Im}z > 0$ and $0 < \mathrm{Im}\theta < \pi/2$. Then the resolvent $S_\theta(z) = [e^{-\theta}H_c(\gamma,\theta) - z]^{-1}$ weakly converges, as $\theta \to 0$, to some operator $S(z)$, which is bounded and analytic for $\mathrm{Im}z > 0$.

Proof: By the same remark of the proof given above, $S_\theta(z)$ is uniformly bounded by $(\mathrm{Im}z)^{-1}$. Moreover, if Q is a dense set of translation analytic vectors, then for $u, v \in Q$, $u(\theta)(x) = u(x + r^{-1}\theta)$, every scalar product

$$\langle [e^{-\theta}H_c(\gamma,\theta + \eta) - z]^{-1}u(\eta),v(\bar\eta)\rangle \tag{III2}$$

is an analytic function of η in the strip $-\mathrm{Im}\theta < \mathrm{Im}\eta < \pi/2 - \mathrm{Im}\theta$. Since it does not depend on the real part of θ, it is also constant in the whole strip and coincides with $\langle S_\theta(z)u,v\rangle$. If we fix η in the strip, with $\mathrm{Im}\eta > 0$, (III2) converges as $\theta \to 0$, because $[e^{-\theta}H_c(\gamma,\theta + \eta) - z]^{-1}$ converges in norm to $[H_c(\gamma,\eta) - z]^{-1}$. Then, for every $u,v \in Q$,

$$\lim_{\theta \to 0}\langle S_\theta(z)u,v\rangle = \langle [H_c(\gamma,\eta) - z]^{-1}u(\eta),v(\bar\eta)\rangle, \tag{III3}$$

where the right hand side does not depend on η by the translation analyticity still. Since u,v can vary in a dense set, and by the uniform boundedness, (III3) means that $S_\theta(z)$ satisfies a weak Cauchy condition as $\theta \to 0$. Thus $S_\theta(z)$ weakly converges to some bounded operator $S(z)$, which is so defined for all z in $\mathrm{Im}z > 0$. Again, the convergence is uniform on compact subsets of the upper half-plane, so that every function $\langle S(z)u,v\rangle$ is analytic for $\mathrm{Im}z > 0$.

Remark: By (III3),

$$\langle S(z)u,v\rangle = \langle [H_c(\gamma,\theta) - z]^{-1}u(\theta),v(\bar\theta)\rangle \tag{III4}$$

identically in the strip $0 < \mathrm{Im}\theta < \pi/2$, when $u,v \in Q$. On the other hand, by definition of $S_\theta(z)$, at least for all z such that $\mathrm{Re}z = 0$, $\mathrm{Im}z > 0$, $[e^{-\theta}H_c(\gamma,\theta) - e^{-\theta}z]^{-1} = S_\theta(e^{-\theta}z)$ (by keeping, e.g., $0 < \mathrm{Im}\theta < \pi/4$; see again Lemma III.4) whence, by (III4)

$$\langle S(z)u,v\rangle = \langle e^{-\theta}S_\theta(e^{-\theta}z)u(\theta),v(\bar\theta)\rangle. \tag{III5}$$

The identities (III4) and (III5) are useful for the following proof.

Theorem III.7: Let $\gamma < 0$ ($|\gamma|$ small) and let $H(\gamma)$ be defined on D. If Q is a dense set of translation analytic vectors, there is a uniquely determined generalized resolvent $\mathscr{R}(E)$ of the symmetric operator $H(\gamma)$ such that any function

$$f_u(E) = \langle \mathscr{R}(E)u,u\rangle, u \in Q \tag{III6}$$

a priori analytic for $\mathrm{Im}E > 0$, has a meromorphic continuation to the lower half-plane $\mathrm{Im}E \leqslant 0$. The set of singularities $\{E\,|f_u$ has a pole at E for some $u \in Q\}$ coincides with $\sigma(H(\gamma,\theta))$, $0 < \mathrm{Im}\theta < \pi/2$.

Proof: It is convenient to show the analogous statement for the symmetric opertor $H_c(\gamma) = H(\gamma) + c$ (c being the positive constant of Lemma III.4); then Theorem III.7 will follow by replacing $z \to E = z - c$, $S(z) \to \mathscr{R}(E) = S(z - c)$. Take $S(z)$ as in the preceding Lemma. For $u \in Q$, $0 < \mathrm{Im}\theta < \pi/2$, from (III4) we have

$$\langle S(z)u,u\rangle = \langle [H_c(\gamma,\theta) - z]^{-1}u(\theta),u(\bar\theta)\rangle. \tag{III7}$$

By Lemma III.6 the left hand side is *a priori* analytic for $\mathrm{Im}z > 0$. Since the right hand side is meromorphic in the whole z plane with poles given by the eigenvalues of $H_c(\gamma,\theta)$ (which exist at least for small $|\gamma|$), the left hand side admits poles in the continuation to $\mathrm{Im}z \leqslant 0$, so we have only to prove that $S(z)$ is a generalized resolvent of the symmetric operator $H_c(\gamma)$.

First, it is clear that $S(z)$ is an extension of the resolvent $[H_c(\gamma)-z]^{-1}$. Indeed, by Lemma III.5, $[H_c(\gamma)-z]^{-1}$ is the strong limit, as $\theta\to 0$, of $S_\theta(z)\upharpoonright(L^2(\mathbb{R})-N_i)$ [N_i being the deficiency subspace of $H_c(\gamma)$]. So, by Lemma III.6, $S(z)$ extends $[H_c(\gamma)-z]^{-1}$ as a weak limit of $S_\theta(z)$ on the whole of $L^2(\mathbb{R})$.

In order to prove that $S(z)$ is a generalized resolvent in the sense above specified, we can use a result of Dolph[11,12] about maximal closed operators in a Hilbert space with numerical range in a half-plane; from that, we know that $S_\theta(z)=[e^{-\theta}H_c(\gamma,\theta)-z]^{-1}$ is a generalized resolvent, i.e., there exists a spectral function $F_\theta(t)$ such that, for $u,v\in L^2(\mathbb{R})$, $\operatorname{Im}z>0$,

$$\langle S_\theta(z)u,v\rangle=\int_{-\infty}^{+\infty}(t-z)^{-1}d\,\langle F_\theta(t)u,v\rangle.$$

In particular, by setting $u=v$, θ fixed, we see that the function $\langle S_\theta(z)u,u\rangle$ can be expressed as an integral where the measure is given by a nondecreasing function of bounded variation, $\rho(t)=\langle F_\theta(t)u,u\rangle$. So (see, e.g., Ref. 7, Theorem 59.3) $\langle S_\theta(z)u,u\rangle$ has a nonnegative imaginary part in the half-plane $\operatorname{Im}z>0$. Besides, setting $z=x+iy$, $\langle S_\theta(iy)u,u\rangle\leqslant\langle u,u\rangle y^{-1}$ since $\|S_\theta(z)\|\leqslant y^{-1}$. These properties are carried into the weak limit as $\theta\to 0$; so $\langle S(z)u,y\rangle$, which is analytic for $\operatorname{Im}z>0$ by Lemma III.6, has a nonnegative imaginary part and satisfies the inequality $\langle S(iy)u,u\rangle\geqslant\langle u,u\rangle y^{-1}$.

Using this inequality and repeating the argument of Ref. 7, Appendix I.4, we obtain a representation of the operator $S(z)$ in the form

$$\langle S(z)u,v\rangle=\int_{-\infty}^{+\infty}(t-z)^{-1}d\,\langle F(t)u,v\rangle. \qquad \text{(III8)}$$

Here too, $F(t)$ is a nondecreasing left-continuous operator function which tends to zero as $t\to-\infty$, and satisfies the condition $\langle F(t)u,u\rangle\leqslant\langle u,u\rangle$. Moreover, $F(t)$ has a strong limit as $t\to+\infty$, as an increasing, uniformly bounded family. So, to complete the proof, it remains to show that $\langle F(t)u,u\rangle\to\langle u,u\rangle$ as $t\to+\infty$, for every u in a dense set. By the representation (III8), this is equivalent to the convergence $(iy)\langle S(iy)u,u\rangle\to-\langle u,u\rangle$ as $y\to+\infty$. This in turn is true for $u\in Q$ since by (III5)

$$\lim_{y\to+\infty}(iy)\langle S(iy)u,u\rangle=\lim_{y\to+\infty}(e^{-\theta}iy)\int_{-\infty}^{+\infty}(t-e^{-\theta}iy)^{-1}\times d\,\langle F_\theta(t)u(\theta),u(\bar\theta)\rangle,$$

where the right hand side is equal to $-\langle u(\theta),u(\bar\theta)\rangle$ because $F_\theta(+\infty)=I$. Since $\langle u(\theta),u(\bar\theta)\rangle=\langle u,u\rangle$ by the translation analyticity, $F(+\infty)=I$. Since we have already remarked that $S(z)$ extends the resolvent of the symmetric operator $H_c(\gamma)$, $S(z)$ is a generalized resolvent of $H_c(\gamma)$ and the theorem is proved.

Remark 1: The analogous statement for odd anharmonic oscillators $p^2+x^2+\gamma x^{2m+1}$, $\gamma\in\mathbb{R}\setminus\{0\}$, has been proved in Ref. 4, by means of the well-known connection between symmetric operators cyclically generated and the classical moment problem. Modulo some technical details, that procedure could be applied here and vice versa. The proof exposed here emphasizes the existence of a bounded extension of $[H(\gamma)-E]^{-1}$, $\operatorname{Im}E>0$ (which admits a meromorphic continuation to $\operatorname{Im}E<0$) independently of its interpretation as a generalized resolvent. Explicitly, it is the weak limit, as $\theta\to 0$, of $[H(\gamma,\theta)-E]^{-1}$, which is compact when $0<\operatorname{Im}\theta<\pi/2$.

Remark 2: Exactly as in Ref. 4, Therorem III.7 implies that the analytic continuation of $\langle\mathscr{R}(E)u,y\rangle$ along any path crossing the real axis is not single valued, i.e., the real axis is a cut for the generalized resolvent $\mathscr{R}(E)$, which is *a priori* analytic for $\operatorname{Im}E>0$ and for $\operatorname{Im}E<0$. In this sense it admits "second sheet poles", which can be interpreted as resonances of the problem.

APPENDIX

The purpose of this Appendix is to show that the generalized resolvent $\mathscr{R}(E)$, introduced in Theorem III.7, is determined by Hamiltonians whose potentials locally approximate $x^2+\gamma e^{rx}$. Let us consider, for fixed $r\neq 0$, $\gamma<0$ and small, $0<\epsilon<1$, the bounded potentials

$$V_\epsilon(x,\gamma)=(2x^2+2\gamma e^{rx})\,(\epsilon^2x^4+\epsilon^2\gamma^2e^{2rx}+4)^{-1/2} \qquad \text{(A1)}$$

If $T_\epsilon(\gamma)=p^2+V_\epsilon(x,\gamma)$ is the self-adjoint operator defined on $D(p^2)$, it is well known (see, e.g., Ref. 1) that its continuous spectrum is the half-line $[-2\epsilon^{-1},+\infty]$ and it tends to cover the whole real axis as $\epsilon\to 0$. We wish to prove that $[T_\epsilon(\gamma)-E]^{-1}$ for $\operatorname{Im}E>0$ is weakly convergent as $\epsilon\to 0$ and, among all bounded extensions of $[H(\gamma)-E]^{-1}$, the limit is just the generalized resolvent $\mathscr{R}(E)$. We define $T_\epsilon(\gamma,\theta)=p^2+V_\epsilon(x+r^{-1}\theta,\gamma)$, at least for $|\operatorname{Im}\theta|<\pi/4$, on the domain of $T_\epsilon(\gamma)$.

*Lemma A.*1: $T_\epsilon(\gamma,\theta)$ is an analytic family of type A in the whole strip $|\operatorname{Im}\theta|<\pi/4$. If $0\leqslant\operatorname{Im}\theta<\pi/4$ and if V is the union, over $\epsilon\in(0,1)$, of the numerical ranges of $T_\epsilon(\gamma,\theta)+2c$ (where c is the constant of Lemma III.4), V is contained in the region $\{E\in C:\ -\pi\leqslant\arg(E)\leqslant\operatorname{Im}\theta\}$.

Proof: Since $V_\epsilon(x+r^{-1}\theta,\gamma)$ is a bounded holomorphic family for θ in the strip, the first assertion is immediate (Ref. 6, Problem VII.1.2). For the other one, it is enough to see that $V_\epsilon(x+r^{-1}\theta,\gamma)+2c$ lies in the stated region when $\operatorname{Re}\theta=0$, by unitary equivalence with respect to different values of $\operatorname{Re}\theta$. We have

$$\begin{aligned}&V_\epsilon(x+r^{-1}\theta,\gamma)+2c\\&=(2(x+r^{-1}\theta)^2+2\gamma e^\theta e^{rx}+2c)\rho(x,\theta)^{-1}e^{-i\varphi(x,\theta)}\\&\quad+(2c-2c\,\rho(x,\theta)^{-1}e^{-i\varphi(x,\theta)}),\end{aligned}$$

where $\rho(x,\theta)>2$ and $0\leqslant\varphi(x,\theta)\leqslant\operatorname{Im}\theta$ for all $x\in\mathbb{R}$, $0\leqslant\operatorname{Im}\theta<\pi/4$. By Lemma III.4 the first term is contained in the desired region, taking into account the bounds over ρ and φ; the argument of the second term does not exceed $\operatorname{Im}\theta$. So the lemma is proved.

*Lemma A.*2: Let γ,ϵ be given as above and $0<\operatorname{Im}\theta<\pi/4$. Then $[T_\epsilon(\gamma,\theta)+2c-E]^{-1}$ strongly converges to $[H(\gamma,\theta)+2c-E]^{-1}$ as $\epsilon\to 0$, for E such that $\operatorname{Im}\theta<\arg(E)<\pi$.

Proof: By Lemma A.1, every E with $\operatorname{Im}\theta<\arg(E)<\pi$ is a regular point for $T_\epsilon(\gamma,\theta)$ and $\|[T_\epsilon(\gamma,\theta)+2c-E]^{-1}\|\leqslant k$, where $k^{-1}=\operatorname{dist}(E,\bar V)$. Besides, if $0<\operatorname{Im}\theta<\pi/4$, E is in the resolvent set of $H(\gamma,\theta)$ by the results of paragraph 3. Finally, for $u\in C_0^\infty(\mathbb{R})$, $T_\epsilon(\gamma,\theta)u\to H(\gamma,\theta)u$ as $\epsilon\to 0$, and $C_0^\infty(\mathbb{R})$ is a core of $H(\gamma,\theta)$.

Then, by Theorem VIII.1.5 of Ref. 6, the strong resolvent convergence holds.

Theorem A.3: For $\mathrm{Im}E > 0$, $[T_\epsilon(\gamma) - E]^{-1}$ is weakly convergent, as $\epsilon \to 0$, to the generalized resolvent $\mathscr{R}(E)$, defined by Theorem III.7, of the symmetric operator $H(\gamma)$.

Proof: Let u,v belong to a dense set of translation analytic vectors for $|\mathrm{Im}\theta| < \pi/4$. By Lemma A.1 and the usual analyticity arguments $\langle [T_\epsilon(\gamma,\theta) + 2c - E]^{-1} u(\theta), v(\bar{\theta})\rangle$ is constant for $|\mathrm{Im}\theta| < \pi/4$ and hence it is equal to $\langle [T_\epsilon(\gamma) + 2c - E]^{-1} u,v\rangle$. For $\mathrm{Im}\theta < \arg(E) < \pi$, by Lemma A.2 any one of these scalar products converges to $\langle [H(\gamma,\theta) + 2c - E]^{-1} u(\theta), v(\bar{\theta})\rangle$ as $\epsilon \to 0$. By (III4) (and by the proof of Theorem III.7) this in turn coincides with $\langle \mathscr{R}(E - 2c)u,v\rangle$. So, for u,v in a dense set we have $\langle [T_\epsilon(\gamma) + 2c - E]^{-1} u,v\rangle \to \langle \mathscr{R}(E - 2c)u,v\rangle$ as $\epsilon \to 0$. Since $T_\epsilon(\gamma)$ is self-adjoint, its resolvent is uniformly bounded with respect to ϵ, so the weak convergence follows for $\mathrm{Im}\theta < \arg(E) < \pi$. By letting $\mathrm{Im}\theta$ be arbitrarily small, the convergence extends to any E with $\mathrm{Im}E > 0$.

ACKNOWLEDGMENTS

It is a pleasure to thank Professors S. Graffi, V. Grecchi, and A. Grossmann for useful suggestions, and Professors F. Bentosela and T. T. Wu for their interest in this work. I am also grateful for hospitality to C. P. T. of Marseille, C.N.R.S., where this research has been completed.

[1]M. Reed and B. Simon, *Methods of Modern Mathematical Physics*, Vol. IV (Academic, New York, 1978).

[2]A. D. Doglov and V. S. Popov, Phys. Lett. B **79**, 403 (1978).

[3]M. Reed and B. Simon, *Methods of Modern Mathematical Physics*, Vol. II (Academic, New York, 1975).

[4]E. Caliceti, S. Graffi and M. Maioli, Commun. Math. Phys. **75**, 51 (1980).

[5]C. M. Bender and T. T. Wu, Phys. Rev. D **7**, 1620 (1973).

[6]T. Kato, *Perturbation Theory for Linear Operators* (Springer, Berlin, 1966).

[7]N. I. Akhiezer and I. M. Glasman, *Theory of Linear Operators in Hilbert Space*, Vol. II (Ungar, New York, 1963).

[8]B. Simon, Ann. Phys. (N.Y.) **58**, 76 (1970).

[9]N. A. Naimark, *Linear Differential Operators*, Part II (Harrap, 1964).

[10]J. Avron and I. Herbst, Commun. Math. Phys. **52**, 239 (1977).

[11]C. L. Dolph, Bull. Am. Math. Soc. **67**, 1 (1961).

[12]C. L. Dolph, "Positive Real Resolvents and Linear Passive Hilbert Systems," Ann. Acad. Sci. Fenn. (1963).

Symmetry of time-dependent Schrödinger equations. I. A Classification of time-dependent potentials by their maximal kinematical algebras

D. Rodney Truax
Department of Chemistry, University of Calgary, Calgary, Alberta, T2N 1N4, Canada

(Received 5 December 1980; accepted for publication 10 April 1981)

Potentials for the time-dependent Schrödinger equation $[-\frac{1}{2}\partial_{xx} + V(x,t)]\Psi(x,t) = i\partial_t \Psi(x,t)$ are classified according to their space–time or kinematical algebras in a search for exactly solvable time-dependent models. In addition, it is shown that their dynamical algebras are isomorphic to their kinematical algebras on the solution space of the Schrödinger equation.

PACS numbers: 03.65.Ge

1. INTRODUCTION

The quantum mechanics of time-dependent systems has considerable importance for our understanding of decaying or oscillatory systems. The usual approach to solving time-dependent Schrödinger equations has been time-dependent perturbation theory.[1] Although this method will probably remain the primary computational method, much could be gained from the study of exactly solvable, time-dependent models. A classification of potentially, exactly solvable Schrödinger equations with time-dependent interactions seems, then, to be appropriate. In this regard, symmetry, particularly the space–time or kinematical symmetries,[2] admitted by these equations would be a powerful investigative tool.

In this paper, we begin with a linear Schrödinger equation with an arbitrary time-dependent potential $V(x,t)$,

$$\mathscr{H}\Psi(x,t) = [-\tfrac{1}{2}\partial_{xx} + V(x,t)]\Psi(x,t) = i\partial_t \Psi(x,t), \tag{1.1}$$

where our analysis is limited to a single spatial dimension. Extensions to higher dimensions would seem to be straightforward. In Sec. 2, we determine the general restrictions on the form of the interaction, $V(x,t)$, essential for the existence of kinematical symmetry. Then we show that both the time-dependent oscillator and linear potential $g_2(t)x^2 + g_1(t)x$ have the Schrödinger algebra $\mathscr{S}_1 = sl(2,\mathbb{R})\Box w_1$ as their kinematical algebra, analogous to[2] the time-independent case $\omega^2x^2/2 + kx$. Here g_2 and g_1 are arbitrary functions of time. The algebras $sl(2,\mathbb{R})$ and w_1 are the Lie algebras of the two-dimensional, real special linear group and the Heisenberg–Weyl algebra in one dimension, respectively. Using these results, we calculate in Sec. 3, the form of the time-dependent potentials consistent with the subalgebras of the Schrödinger algebra, $\mathscr{S}_1$. Our analysis employs the subalgebra decompositions of $\mathscr{S}_1$ obtained by Boyer *et al.*,[3] for their calculations on nonlinear, time-dependent Schrödinger equations. In Sec. 4, we explore the possibility of broadening our classification using dynamical symmetries. However, no new interactions or symmetries are uncovered which implies that, at least for one spatial dimension, it is sufficient to compute the kinematical algebra only.

2. THE OSCILLATOR AND LINEAR POTENTIALS

Let us rewrite Eq. (1.1) in a more convenient form

$$Q\Psi(x,t) = [\partial_{xx} + 2i\partial_t - 2V(x,t)]\Psi(x,t) = 0. \tag{2.1}$$

We wish to find those space–time symmetries which transform solutions of (2.1) into solutions. Generators of these symmetries have the form[2,4]

$$L = A(x,t)\partial_x + B(x,t)\partial_t + C(x,t), \tag{2.2}$$

where L must satisfy the operator equation

$$[Q,L] = \lambda(x,t)Q \tag{2.3}$$

and $\lambda(x,t)$ is an arbitrary function of its arguments.[2,4] Let $\mathscr{G}$ be the collection of all such generators L. Then, if L,L' are two members of $\mathscr{G}$, the $[L,L']$ is a member of $\mathscr{G}$ also, where $[L,L'] = LL' - L'L$, the usual commutator bracket. $\mathscr{G}$ is then a Lie algebra[4] called the kinematical algebra.

By substituting (2.1) and (2.2) into (2.3) and identifying the coefficients of corresponding derivatives in the resulting equation, we obtain the following set of coupled partial differential equations for the coefficients $A(x,t)$, $B(x,t)$, $C(x,t)$, and $\lambda(x,t)$:

$$A_x = 0,\quad B_x = \lambda/2, \tag{2.4}$$

$$\begin{aligned} &A_{xx} + 2iA_t = 2i\lambda,\\ &B_{xx} + 2iB_t + 2C_x = 0,\\ &AV_t + BV_x + \lambda V = -iC_t - \tfrac{1}{2}C_{xx}, \end{aligned} \tag{2.5}$$

where $A_x = \partial A/\partial x$ and $A_{xx} = \partial^2 A/\partial x^2$, etc. Solving the system (2.4) yields the following forms for A, B, and C:

$$A(x,t) = A(t), \tag{2.6a}$$

$$B(x,t) = \tfrac{1}{2}\dot{A}x + b(t), \tag{2.6b}$$

$$C(x,t) = -\tfrac{1}{4}i\ddot{A}x^2 - i\dot{b}x + c(t), \tag{2.6c}$$

where for the moment the functions $A(t)$, $b(t)$, and $c(t)$ are arbitrary functions of time and $\dot{A} = dA/dt$.

Substitution of (2.6) into (2.5) generates a first-order partial differential equation for the potential $V(x,t)$,

$$AV_t + (\tfrac{1}{2}\dot{A}x + b)V_x + \dot{A}V = -\tfrac{1}{4}\dddot{A}x^2 - \ddot{b}x - i\dot{c} + \tfrac{1}{4}i\ddot{A}. \tag{2.7}$$

The general solution to this partial differential equation has the form

 0022-2488/81/091959-06$01.00

$$V(x,t) = \bar{V}(x,t) + g_2(t)x^2 + g_1(t)x + g_0(t), \tag{2.8}$$

where $g_0(t)$, $g_1(t)$, and $g_2(t)$ are arbitrary functions of time and $\bar{V}(x,t)$ is solution to the homogeneous first-order linear partial differential equation

$$A\bar{V}_t + (\tfrac{1}{2}\dot{A}x + b)\bar{V}_x + \dot{A}\bar{V} = 0. \tag{2.9}$$

Now, if we substitute (2.8) into (2.7), we obtain a system of ordinary differential equations for $A(t)$, $b(t)$, and $c(t)$:

$$\dddot{A} + 8g_2(t)\dot{A} + 4\dot{g}_2(t)A = 0, \tag{2.10a}$$

$$\ddot{b} + 2g_2(t)b = -\tfrac{3}{2}g_1(t)\dot{A} - \dot{g}_1(t)A, \tag{2.10b}$$

$$\dot{c} = \tfrac{1}{4}\ddot{A} + i[g_0(t)\dot{A} + \dot{g}_0(t)A] + ibg_1(t). \tag{2.10c}$$

Next we shall consider the case where

$$V(x,t) = g_2(t)x^2 + g_1(t)x + g_0(t), \tag{2.11}$$

that is, where $\bar{V}(x,t)$ vanishes and is the trivial solution to (2.9). We wish to find a set of linearly independent solutions to the system of ordinary differential equations (2.10) such that closure of the symmetry algebra $\mathscr{G}$ is guaranteed. If we assume that φ_1, φ_2, and φ_3 are three linearly independent solutions of the ordinary differential equation (2.10a), then the general solution is a linear combination,

$$A(t) = \sum_{j=1}^{3} \beta_j \varphi_j, \tag{2.12}$$

where the β_j are constants. For closure of the algebra $\mathscr{G}$,

$$\psi_{jk} = \begin{vmatrix} \varphi_j & \varphi_k \\ \dot{\varphi}_j & \dot{\varphi}_k \end{vmatrix} \tag{2.13}$$

must also be a solution of (2.10a). That this is indeed the case is shown in Theorem 1 of the Appendix. Furthermore, by Theorem 2 of the Appendix, if χ_1 and χ_2 are two linearly independent, nontrivial solutions of the homogeneous equation

$$\ddot{b} + 2g_2(t)b = 0, \tag{2.14}$$

then we can choose

$$\varphi_1 = (1/\alpha)\chi_1^2, \quad \varphi_2 = (1/\alpha)\chi_2^2, \quad \varphi_3 = (2/\alpha)\chi_1\chi_2, \tag{2.15}$$

where the Wronskian of solutions to (2.14), $W(\chi_1,\chi_2) = \alpha$, is a constant. Together Eqs. (2.13) and (2.15) imply (by Theorem 2) that

$$\psi_{12} = \varphi_3, \quad \psi_{31} = -2\varphi_1, \quad \psi_{32} = 2\varphi_2. \tag{2.16}$$

A general solution to Eq. (2.10b) can be written as follows[5]:

$$b(t) = \beta_4\chi_1 + \beta_5\chi_2 + \sum_{j=1}^{3}\beta_j\left\{\frac{1}{\alpha}\left[\chi_1\int^t \chi_2(\tfrac{3}{2}g_1\dot{\varphi}_j + \dot{g}_1\varphi_j) - \chi_2\int^t \chi_1(\tfrac{3}{2}g_1\dot{\varphi}_j + \dot{g}_1\varphi_k)\right]\right\}, \tag{2.17a}$$

$$= \beta_4\chi_1 + \beta_5\chi_2 + \sum_{j=1}^{3}\beta_j\{\mathscr{A}_j/\alpha\}, \tag{2.17b}$$

where

$$\begin{aligned} \mathscr{A}_1 &= -\chi_1\mathscr{C}_1, \\ \mathscr{A}_2 &= -\chi_2\mathscr{C}_2, \\ \mathscr{A}_3 &= -(\chi_1\mathscr{C}_2 + \chi_2\mathscr{C}_1), \end{aligned} \tag{2.18}$$

and

$$\mathscr{C}_\sigma = \int^t \chi_\sigma g_1, \quad \sigma = 1,2. \tag{2.19}$$

The $\mathscr{A}_j$ of (2.18) were obtained from a partial integration of the integral expressions of the third term in (2.17a) followed by a substitution of (2.15) for the φ_j.

With expressions (2.12) for $A(t)$ and (2.17b) for $b(t)$, Eq. (2.10c) can be integrated for $c(t)$:

$$\begin{aligned} c(t) &= \tfrac{1}{4}\dot{A} + ig_0A + i\int^t bg_1 + i\beta_6 \\ &= \sum_{j=1}^{3}\beta_j\left\{\tfrac{1}{4}\dot{\varphi}_j + ig_0\varphi_j + \frac{i}{\alpha}\mathscr{D}_j\right\} \\ &\quad + \beta_4\{i\mathscr{C}_1\} + \beta_5\{i\mathscr{C}_2\} + i\beta_6, \end{aligned} \tag{2.20}$$

where $\mathscr{D}_j$ is defined by

$$\mathscr{D}_j = \int^t g_1\mathscr{A}_j, \quad 1 \leqslant j \leqslant 3. \tag{2.21}$$

Integrating (2.21) by parts and substituting (2.18) for the $\mathscr{A}_j$ yields

$$\mathscr{D}_1 = -\tfrac{1}{2}\mathscr{C}_1^2, \quad \mathscr{D}_2 = -\tfrac{1}{2}\mathscr{C}_2^2, \quad \mathscr{D}_3 = -\mathscr{C}_1\mathscr{C}_2. \tag{2.22}$$

Expressions for the coefficients $B(x,t)$ and $C(x,t)$ can now be obtained by substitution of Eqs. (2.12), (2.17b), and (2.20) into (2.6b) and (2.6c). Thus,

$$B(x,t) = \sum_{j=1}^{3}\beta_j\left\{\tfrac{1}{2}\dot{\varphi}_jx + \frac{1}{\alpha}\mathscr{A}_j\right\} + \beta_4\{\chi_1\} + \beta_5\{\chi_2\} \tag{2.23}$$

and

$$\begin{aligned} C(x,t) = \sum_{j=1}^{3}\beta_j\Big\{ &-\frac{i}{4}\ddot{\varphi}_jx^2 - \frac{i}{\alpha}\dot{\mathscr{A}}_jx + \frac{1}{4}\dot{\varphi}_j \\ &+ ig_0\varphi_j + \frac{i}{\alpha}\mathscr{D}_j\Big\} + \beta_4\{-i\dot{\chi}_1x + i\mathscr{C}_1\} \\ &+ \beta_5\{-i\dot{\chi}_2 + i\mathscr{C}_2\} + i\beta_6. \end{aligned} \tag{2.24}$$

The symmetry generator L is given by

$$L = \sum_{j=1}^{6}\beta_jL_j,$$

where the $\{L_j, 1 \leqslant j \leqslant 6\}$ form a basis for the kinematic potential (2.11). The form of the generators of these kinematic symmetries, the L_j, can be obtained from the generator (2.2) and the coefficients (2.12), (2.23), and (2.24). They are

$$\begin{aligned} L_j &= \varphi_j\partial_t + \{\tfrac{1}{2}\dot{\varphi}_jx + (1/\alpha)\mathscr{A}_j\}\partial_x - \tfrac{1}{4}i\ddot{\varphi}_jx^2 \\ &\quad - (i/\alpha)\dot{\mathscr{A}}_jx + \tfrac{1}{4}\dot{\varphi}_j + ig_0\varphi_j + (i/\alpha)\mathscr{D}_j, \quad 1 \leqslant j \leqslant 3, \end{aligned} \tag{2.25a}$$

$$\begin{aligned} L_4 &= \chi_1\partial_x - i\dot{\chi}_1x + i\mathscr{C}_1, \\ L_5 &= \chi_2\partial_x - i\dot{\chi}_2 + i\mathscr{C}_2, \\ L_6 &= E = i. \end{aligned} \tag{2.25b}$$

To find the specific kinematical algebra we must compute the commutators of the generators of (2.25). For $1 \leqslant j, k \leqslant 3$, the nonzero commutators are

$$\begin{aligned} [L_j, L_k] &= \psi_{jk}\partial_t + \{\tfrac{1}{2}\dot{\psi}_{jk}x + (1/\alpha)\mathscr{A}_{jk}\}\partial_x - \tfrac{1}{4}i\ddot{\psi}_{jk}x^2 \\ &\quad - \tfrac{1}{4}i\dot{\mathscr{A}}_{jk}x + \tfrac{1}{4}\dot{\psi}_{jk} + ig_0\psi_{jk} + (i/\alpha)\mathscr{D}_{jk}, \end{aligned} \tag{2.26}$$

where ψ_{jk} is given by (2.13) and

$$\left.\begin{aligned}\mathscr{A}_{jk} &= \varphi_j \dot{\mathscr{A}}_k - \varphi_k \dot{\mathscr{A}}_j + \tfrac{1}{2}(\dot{\varphi}_k \mathscr{A}_j - \dot{\varphi}_j A_k)\\ \mathscr{D}_{jk} &= \varphi_j \dot{\mathscr{D}}_k - \varphi_k \dot{\mathscr{D}}_j - (1/\alpha)(\mathscr{A}_j \dot{\mathscr{A}}_k - \dot{\mathscr{A}}_j \mathscr{A}_k)\end{aligned}\right\} 1 \leqslant j, k \leqslant 3. \tag{2.27a}$$

After some algebra it is possible to show that

$$\mathscr{A}_{12} = \mathscr{A}_3, \quad \mathscr{A}_{31} = -2\mathscr{A}_1, \quad \mathscr{A}_{32} = 2\mathscr{A}_2$$
$$\mathscr{D}_{12} = \mathscr{D}_3, \quad \mathscr{D}_{31} = -2\mathscr{D}_1, \quad \mathscr{D}_{32} = 2\mathscr{D}_2 \tag{2.27b}$$

and thence by (2.16) that

$$[L_1,L_2] = L_3, \quad [L_3,L_1] = -2L, \quad [L_3,L_2] = 2L_2. \tag{2.28}$$

The generators $\{L_1,L_2,L_3\}$ with the commutators (2.28) form a $sl(2,\mathbb{R})$ algebra.[5] Because

$$[L_4,L_5] = -\alpha E,$$

we will find it convenient to define the generators

$$B_1 = -(1/\sqrt{\alpha})(\chi_1 \partial_x - i\dot{\chi}_1 x + i\mathscr{C}_1),$$
$$B_2 = (1/\sqrt{\alpha})(\chi_2 \partial_x - i\dot{\chi}_2 x + i\mathscr{C}_2), \tag{2.29}$$

with commutator

$$[B_1,B_2] = E. \tag{2.30}$$

The generators $\{B_1,B_2,E\}$ with commutator (2.30) form a Heisenberg–Weyl algebra w_1 in one dimension. Again, after some calculation we have the commutators

$$\begin{aligned}&[L_1,B_1] = 0, && [L_2,B_1] = B_2, && [L_3,B_1] = -B_1,\\ &[L_1,B_2] = -B_1, && [L_2,B_2] = 0, && [L_3,B_2] = B_2,\end{aligned} \tag{2.31}$$

which implies that the maximal kinematical algebra for the time-dependent harmonic oscillator and linear potential is the semidirect sum $sl(2,\mathbb{R})\square w_1$, the Schrödinger algebra $\mathscr{S}_1$.

A number of special cases are mentioned in Table I. The free particle, the attractive and repulsive oscillators, and linear potential [$g_1(t) = \kappa$, a constant] have been discussed by Boyer.[2] The solutions to $\ddot{b} + 2g_2 b = 0$, where $g_2 = a/t^2$, is included for three different values of the constant a.

The interpretation of the generators of $sl(2,\mathbb{R})$ for the free particle case as a space–time dilation, a conformal symmetry, and time translation[2–4] is apparent from the action of the corresponding group elements[3,4] on an arbitrary function of x,t. Similarly, the Heisenberg algebra consists of operators generating a spatial translation and a Galilean boost.[2–4] The specific interpretation of the generators will depend upon the form of the φ_j chosen and will vary with the choice of g_2.

It is perhaps surprising that the time-dependent linear potential does not break the symmetry of the oscillator term. That this is in fact the case when $\bar{V}(x,t)$ of (2.8) is nonvanishing will be demonstrated in the next section.

TABLE I. Solutions to Eq. (2.14) for different values of $g_2(t)$.

$g_2(t)$	χ_1	χ_2	*Remarks*
0	1	t	Free particle[a]
$\omega^2/_2$	$(1/\sqrt{\omega})\cos\omega t$	$(1/\sqrt{\omega})\sin\omega t$	Attractive Oscillator[a]
$-\omega^2/_2$	$(1/\sqrt{\omega})\cosh\omega t$	$(1/\sqrt{\omega})\sinh\omega t$	Repulsive Oscillator[a]
$-\dfrac{n(n-1)}{2t^2}$	t^n	$-\dfrac{1}{(2n-1)t^{n-1}}$	$n\geqslant 2$, n integer
$1/8t^2$	$t^{1/2}$	$t^{1/2}\ln t$	$t>0$
$3/4t^2$	$t^{-1/2}$	$t^{3/2}$	$t>0$

[a]See Boyer, Ref. 2.

3. CLASSIFICATION OF POTENTIALS

In this section we shall find the potential $\bar{V}(x,t)$, solution to (2.9), which reduces the symmetry of $V(x,t) = \bar{V}(x,t) + g_2(t)x^2 + g_1(t)x + g_0(t)$ from $\mathscr{S}_1$ to a subalgebra of $\mathscr{S}_1$. Thus we must compute the form of $\bar{V}(x,t)$ which satisfies the first-order partial differential equation

$$A\bar{V}_t + B\bar{V}_x + \dot{A}\bar{V} = 0 \tag{3.1}$$

and is consistent with the subalgebra $\mathscr{K}$ of $\mathscr{S}_1$ under consideration. The first-order equation (3.1) can be solved by integrating the subsidiary conditions

$$\frac{dt}{A} = \frac{dx}{B} = -\frac{d\bar{V}}{\dot{A}\bar{V}}, \tag{3.2}$$

where A and B are given, in general, by (2.12) and (2.23), respectively.

The subalgebras, or rather conjugacy classes of subalgebras where conjugacy is with respect to the Schrödinger group S_1, have been determined by Boyer *et al.*,[3] employing standard methods. Since condition (3.1) for $\bar{V}$ has no reference to the coefficient $C(x,t)$:

(i) It is unnecessary to consider one-dimensional subalgebras of the type $\{X + aE\}$ where X is a generator for $\mathscr{S}_1$ or higher-dimensional subalgebras of $\mathscr{S}_1$ containing elements of this form.[6]

(ii) Subalgebras containing the central element E can be excluded.[7] The remaining subalgebras determined for by Boyer *et al.*,[3] $\mathscr{S}_1$ can be found in Table II.

For multidimensional algebras, the number of conditions of the type (3.1) for $\bar{V}$ will equal the dimensionality of the subalgebra, and $\bar{V}(x,t)$ must be consistent with each of these conditions. Subalgebras and their associated interactions, $V(x,t)$, are listed in Table III for the two cases $g_1(t) = 0$ and $g_1(t) \neq 0$. Below a few examples are worked for the former case; the potentials for the latter follow analogously.

(a) Any subalgebra containing B_1 or B_2 as an element will have $\bar{V}_x = 0$ as a minimal condition. Since $\bar{V}$ is then a function of t only it can be absorbed into the $g_0(t)$ term in (2.8)

TABLE II. Proper subalgebras of $\mathscr{S}_1$ classified under S_1.

	Dimension of the subalgebra $\mathscr{K}$ 1	2	3
Subalgebra $\mathscr{K}$	$\{L_1\}$	$\{L_3,L_1\}$	$\{L_1,L_2,L_3\}$
	$\{L_1 + L_2\}$		
	$\{L_3\}$	$\{L_3,B_1\}$	
	$\{B_1\}$		
	$\{L_1 + B_2\}$		

TABLE III. Potentials and their associated kinematical algebras.

Kinematic algebra	Potential $V(x,t)$
$sl(2,\mathbb{R})\square w_1$	$g_2(t)x^2 + g_1(t)x + g_0(t)$
$sl(2,\mathbb{R})$	$\kappa/x^2 + g_2(t)x^2$
$\{L_1 + L_2\}$	$g_2(t)x^2 + \frac{1}{\varphi_1+\varphi_2} f\left(\frac{x^2}{\varphi_1+\varphi_2}\right)$
	$g_2(t)x^2 + g_1(t)x + \frac{1}{\varphi_1+\varphi_2} f\left(\frac{x}{(\varphi_1+\varphi_2)^{1/2}} - \frac{1}{\alpha}\int^t \frac{\mathscr{A}_1+\mathscr{A}_2}{(\varphi_1+\varphi_2)^{3/2}}\right)$
$\{L_j\}, j = 1$ or 3	$g_2(t)x^2 + \frac{1}{\varphi_j} f\left(\frac{x^2}{\varphi_j}\right)$
	$g_2(t)x^2 + g_1(t)x + \frac{1}{\varphi_j} f\left(\frac{x}{\varphi_j^{1/2}} - \frac{1}{\alpha}\int^t \frac{\mathscr{A}_1}{\varphi_j^{3/2}}\right)$
$\{L_1 + B_2\}$	$g_2(t)x^2 + \frac{1}{\varphi_1} f\left(\frac{x}{\varphi_1^{1/2}} - \frac{1}{\sqrt{\alpha}}\int^t \frac{\chi_2}{\varphi_1^{3/2}}\right)$
	$g_2(t)x^2 + g_1(t)x + \frac{1}{\varphi_1} f\left(\frac{x}{\varphi_1^{1/2}} - \frac{1}{\alpha}\int^t \frac{\mathscr{A}_1}{\varphi_1^{3/2}} - \frac{1}{\sqrt{\alpha}}\int^t \frac{\chi_2}{\varphi_1^{3/2}}\right)$

and these cases are excluded from Table III. Furthermore, for $\mathscr{S}_1$ and any subalgebras of dimension 3,4, or 5 containing B_1 or B_2 we require $\bar{V}=0$.

(b) $\{L_1 + B_2\}$. Here we have

$$A = \varphi_1, \quad B = \tfrac{1}{2}\dot{\varphi}_1 x + \chi_2/\sqrt{\alpha},$$

and the subsidiary conditions are

$$\frac{dt}{\varphi_1} = \frac{dx}{\dot{\varphi}_1 x/2 + \chi_2/\sqrt{\alpha}} = \frac{-d\bar{V}}{\dot{\varphi}_1 \bar{V}},$$

which may be integrated to give

$$\bar{V}(x,t) = \frac{1}{\varphi_1} f\left(\frac{x}{\varphi_1^{1/2}} - \frac{1}{\sqrt{\alpha}}\int^t \frac{\chi_2}{\varphi_1^{3/2}}\right),$$

where f is an arbitrary function of its argument.

(c) $\{L_1, L_2, L_3\} = sl(2,\mathbb{R})$. We have three conditions on the potential $\bar{V}(x,t)$:

$$\varphi_j \bar{V}_t + \tfrac{1}{2}\dot{\varphi}_j x \bar{V}_x + \dot{\varphi}_j \bar{V} = 0, \quad 1 \leqslant j \leqslant 3.$$

Eliminating the $\bar{V}_t$ dependence leaves us with the single condition

$$\tfrac{1}{2}x\bar{V}_x + \bar{V} = 0,$$

which has κ/x^2 as its solution. Since κ/x^2 has $sl(2,\mathbb{R})$ as its kinematic algebra, it will also be an invariant interaction for the two-and one-dimensional subalgebras of $sl(2,\mathbb{R})$. The generators in this case will have the form

$$L_j = \varphi_j \partial_t + \tfrac{1}{2}\dot{\varphi}_j x \partial_x - \tfrac{1}{4}i\ddot{\varphi}_j x^2 + \tfrac{1}{4}\dot{\varphi}_j, \quad 1 \leqslant j \leqslant 3,$$

where φ_j are solutions of (2.10a). Even though the Heisenberg–Weyl algebra is not a symmetry algebra of κ/x^2, we can still define the differential equation (2.14) and use its solutions to construct the φ_j as in (2.15) so that they have the property (2.16). This is the situation in the time-independent problem where the kinematical algebras for $V=0$ and $V=\kappa/x^2$ are realized by the same set of three operators,[2] and similarly for $V=\omega^2x^2/2$ and $V=\omega^2x^2/2+\kappa/x^2$.

For the one-dimensional problem, the existence of kinematical symmetry is intimately related to separation of variables.[4] For each of the algebras listed in Table III, it is possible to partition its elements into orbits of operators. To each of the nontrivial orbits[4] it is possible to associate a separable coordinate system. For example, each one-dimensional lgebra corresponds directly to an orbit and so separates irone coordinate system. For cases with $sl(2,\mathbb{R})$ and $\mathscr{S}_1$ symmtry, separation occurs in more than one coordinate systemDetails for the one-dimensional free-particle Schrödinger equation have been worked out by Kalnins and Miller[8] and Iiller.[4] Thus in the sense of separation of variable, if the separated ordinary differential equations can be solved,hen the Hamiltonians with potentials found in Table III ccrespond to exactly solvable models.

4. DYNAMICAL SYMMETRIES

In this section we explore the relationship betwee the kinematical algebras for the quantum mechanical systeis in Table III and their dynamical algebras, including consants of the motion.[9] We do this with a view to extending our lt of solvable models in Table III. It is possible to find dynaiical symmetries for the Schrödinger equation (2.1) by compting those generators

$$S = F(x,t)\partial_{xx} + G(x,t)\partial_x + H(x,t) \tag{4.1}$$

which satisfy the commutator relation[4,9]

$$[S,Q] = 0, \tag{4.2}$$

where Q is given by (2.1). The S operators will be constats of the motion for (2.1) since they satisfy the relation[10]

$$\frac{dS}{dt} = \frac{\partial S}{\partial t} + i[\mathscr{H},S] = 0, \tag{4.3}$$

where $\mathscr{H}$ is the Hamiltonian specified in (1.1). The exression (4.3) vanishes because of (4.2).

Substitution of (2.1) and (4.1) into (4.2) yields theet of coupled partial differential equations

$$F_x = 0, \tag{4.4a}$$

$$F_{xx} + 2G_x + 2iF_t = 0, \tag{4.4b}$$

$$G_{xx} + 2H_x + 2iG_t + 4FV_x = 0, \tag{4.5}$$

$$H_{xx} + 2iH_t + 2FV_{xx} + 2GV_x = 0. \tag{4.6}$$

Equations (4.4) imply that

$$G(x,t) = -2i[\tfrac{1}{2}\dot{F}x + g(t)], \tag{4.7}$$

where F is a function of t only and $-2ig(t)$ is an arbitrary function of time resulting from the integration of (4.4b) with respect to x. Furthermore, (4.5) can be simplified by (4.7) and integrated to give

$$H(x,t) = -(\tfrac{1}{2}\ddot{F}x^2 + 2FV + 2\dot{g}x) + 2ih(t), \tag{4.8}$$

where $2ih(t)$ is an arbitrary function of time. When (4.7) and (4.8) are substituted into (4.6), we obtain a first-order partial differential equation for V,

$$FV_t + [\tfrac{1}{2}\dot{F}x + g(t)]V_x + \dot{F}V = -\tfrac{1}{4}\dddot{F}x^2 - \ddot{g}x + i\dot{h} + \tfrac{1}{4}i\ddot{F}. \tag{4.9}$$

The treatment of (4.9) proceeds analgously to that of (2.7) to which it is essentially equivalent. Again we partition V into two parts: a solution $g_2(t)x^2 + g_1(t)x + g_0(t)$ to the inhomogeneous equation and $\overline{V}(x,t)$ and a solution to the homogeneous equation

$$F\overline{V}_t + [\tfrac{1}{2}\dot{F}x + g(t)]\overline{V}_x + \dot{F}\overline{V} = 0, \tag{4.10}$$

which is identical in form to Eq. (2.9). The solution to the inhomogeneous equation yields the following differential equations for F, g, and h:

$$\dddot{F} + 8g_2(t)\dot{F} + 4\dot{g}_2(t)F = 0, \tag{4.11a}$$

$$\ddot{g} + 2g_2(t)g = (\tfrac{3}{2}g_1(t)\dot{F} - \dot{g}_1(t)F, \tag{4.11b}$$

$$\dot{h} = -\tfrac{1}{4}\ddot{F} - i(\dot{g}_0F + g_0\dot{F}) - ig_1(t)g. \tag{4.11c}$$

If we restrict ourselves for the moment to $V(x,t) = g_2(t)x^2 + g_1(t)x + g_0(t)$, these equations (4.11a) and (4.11b) are identical to (2.10a) and (2.10b), respectively, and therefore have the same solutions

$$F(t) = \sum_{j=1}^{3}\beta_j\varphi_j \tag{4.12}$$

and

$$g(t) = \beta_4\chi_1 + \beta_5\chi_2 + \sum_{j=1}^{3}\beta\{\mathscr{A}_j/\alpha\}, \tag{4.13}$$

where χ_1 and χ_2 are solutions to the homogeneous equation $\ddot{g} + 2g_2(t)g = 0$. The specific form for $\varphi_j, 1\leqslant j\leqslant 3$ is given by (2.15) with property (2.16). Equation (4.11c) may now be integrated if we substitute for $F(t)$ and $g(t)$, (4.12) and (4.13) respectively. Hence

$$h(t) = \sum_{j=1}^{3}\beta_j\left\{-\tfrac{1}{4}\dot{\varphi}_j - ig_0\varphi_j - \frac{i}{\alpha}\mathscr{D}_j\right\} + \beta_4\{-i\mathscr{C}_1\} + \beta_5\{-i\mathscr{C}_2\} + \tfrac{1}{2}\beta_6. \tag{4.14}$$

Thus, for $G(x,t)$ and $H(x,t)$ we have

$$G(x,t) = \sum_{j=1}^{3}\beta_j\left\{-2i(\dot{\varphi}_j\frac{x}{2} + \mathscr{A}_j/\alpha\right\} + \beta_4\{-2i\chi_1\} + \beta 5\{-2i\chi_2\} \tag{4.15}$$

and

$$H(x,t) = \sum_{j=1}^{3}\beta_j\left\{-\tfrac{1}{2}(\ddot{\varphi}_j + 4g_2\varphi_j)x^2 - 2(g_1\varphi_j + \dot{\mathscr{A}}_j/\alpha)x - \frac{i}{2}\dot{\varphi}_j + \frac{2}{\alpha}\mathscr{D}_j\right\} + \beta_4\{-2\dot{\chi}_1x + 2\mathscr{C}_1\} + \beta_5\{-2\dot{\chi}_2x + 2\mathscr{C}_2\} + i\beta_6, \tag{4.16}$$

respectively.

The generator S_1 can be written as a linear combination

$$S = \sum_{j=1}^{3}\beta_jS_j, \tag{4.17}$$

where

$$S_j = \varphi_j\partial_{xx} - 2i(\tfrac{1}{2}\dot{\varphi}_jx + \mathscr{A}_j/\alpha)\partial_x - \tfrac{1}{2}(\ddot{\varphi}_j + 4g_2\varphi_j)x^2 - 2(g_1\varphi_J + \dot{\mathscr{A}}_j/\alpha)_x - \tfrac{1}{2}i\dot{\varphi}_j + (2/w)\mathscr{D}_j, \quad 1\leqslant j\leqslant 3, \tag{4.18}$$

$$S_4 = -2i(\chi_1\partial_x - i\dot{\chi}_1x + i\mathscr{C}_1),$$
$$S_5 = -2i(\chi_2\partial_x - i\dot{\chi}_2x + i\mathscr{C}_2), \tag{4.19}$$
$$S_6 = E = i.$$

Clearly, the algebra $\{S_4,S_5,S_6\}$ can be identified with the Heisenberg–Weyl algebra $\{B_1,B_2,E\}$ of Sec. 2. Because of (2.16) and (2.27) we have the commutators

$$[S_1,S_2] = S_3, \quad [S_3,S_1] = -2S_1, \quad [S_3,S_2] = 2S_2. \tag{4.20}$$

Thus, the algebra of $\{S_1,S_2,S_3\}$ with commutators (4.20) is $sl(2,\mathbb{R})$ and is isomophic to the algebra of $\{L_1,L_2,L_3\}$ with commutators (2.28). In addition,

$$\begin{aligned} &[S_1,B_1] = 0, && [S_2,B_1] = B_2, && [S_3,B_1] = -B_1,\\ &[S_1,B_2] = -B_1, && [S_2,B_2] = 0, && [S_3,B_2] = B_2, \end{aligned} \tag{4.21}$$

which implies that for the time-dependent oscillator and linear potential, the algebra of the constants of the motion $\{S_j, 1\leqslant j\leqslant 6\}$ is $sl(2,\mathbb{R})\square w_1$, isomorphic to the kinematic algebra. The isomorphism between the two algebras $\{S_1,S_2,S_3\}$ and $\{L_1,L_2,L_3\}$ occurs because of the mapping $S_j \rightarrow L_j$ on the solution space of (2.1).[4,9] In this case, $S_j + R(x,t)Q$ is also a symmetry of (2.1), where $R(x,t)$ is an arbitrary function. If we choose $R = -\varphi_j$, then $S_j - \varphi_jQ = 2iL_j, 1\leqslant j\leqslant 3$.

We mention that for the time-dependent oscillator and linear potentials the S_j are second degree polynomials in the nontrivial elements of the Heisenberg algebra

$$S_1 = B_1^2, \quad S_2 = B_2^2, \quad S_3 = -[B_1,B_2]_+, \tag{4.22}$$

where $[X,Y]_+ = XY + YX$.

It is clear from the identity of Eqs. (2.8) and (4.10) that the only potentials with dynamical algebras will be those of Table III. Furthermore, their dynamical algebras will be isomorphic to their kinematical algebras, related by the mapping described above. No new dynamical symmetries are obtained unlike some situations in higher dimensions.[9] This implies that, for one spatial dimension, we can discuss the model systems of Table III entirely in terms of their kinematical algebras.

In conclusion, we point out the interesting comparison between the dynamical algebra $sl(2,\mathbb{R})\square w_1$ for the time-dependent quantum oscillator and the algebra of the five invariants as computed by Leach[11] for its classical analog. Both the quantum and classical algebras have $sl(2,\mathbb{R})$ components, but their Heisenberg algebras are distinguished by the fact that in the former $[B_1,B_2] = E$ but in the latter $[B_1,B_2] = 0$. The identity E is not a component of the classical algebra.

APPENDIX

Lemma 1: Let $\varphi_1,\varphi_2,\ldots,\varphi_n$ be n linearly independent, nontrivial solutions to the homogeneous, linear ordinary differential equation

$$y^{(n)} + a_2(t)y^{(n-2)} + \cdots + a_{n-1}(t)y' + a_n(t) = 0,$$
$$n \geqslant 2, \quad a_1(t) = 0. \tag{A1}$$

Then the Wronskian of the solutions

$$W(\varphi_1,\varphi_2,\ldots,\varphi_n) = \begin{vmatrix} \varphi_1 & \varphi_2 & \cdots & \varphi_n \\ \dot{\varphi}_1 & \dot{\varphi}_2 & \cdots & \dot{\varphi}_n \\ \vdots & \vdots & & \vdots \\ \varphi_1^{(n-1)} & \varphi_2^{(n-1)} & \cdots & \varphi_n^{(n-1)} \end{vmatrix} \tag{A2}$$

is constant.

Proof: This follows directly from the fact that[5]

$$\frac{W'(\varphi_1,\varphi_2,\ldots,\varphi_n)}{W(\varphi_1,\varphi_2,\ldots,\varphi_n)} = -a_1(t) = 0.$$

Theorem 1: Let $\varphi_1,\varphi_2,\varphi_3$ be three linearly independent, nontrivial solutions for the homogeneous, linear ordinary differential equation

$$\dddot{A} + 2g(t)\dot{A} + \dot{g}(t)A = 0. \tag{A3}$$

Then,

$$\psi_{kl} = \begin{vmatrix} \varphi_k & \varphi_l \\ \dot{\varphi}_k & \dot{\varphi}_l \end{vmatrix} = \varphi_k\dot{\varphi}_l - \dot{\varphi}_k\varphi_l, \quad 1 \leqslant k < l \leqslant 3 \tag{A4}$$

is a nontrivial solution of (A3). Furthermore, the three functions $\psi_{12},\psi_{23},\psi_{13}$ are linearly independent.

Proof: It is straightforward to show from (A3) that for $1 \leqslant k < l \leqslant 3$,

$$\begin{vmatrix} \varphi_k & \varphi_l \\ \dddot{\varphi}_k & \dddot{\varphi}_l \end{vmatrix} = -2g\begin{vmatrix} \varphi_k & \varphi_l \\ \dot{\varphi}_k & \dot{\varphi}_l \end{vmatrix}, \tag{A5}$$

and

$$\begin{vmatrix} \dot{\varphi}_k & \dot{\varphi}_l \\ \dddot{\varphi}_k & \dddot{\varphi}_l \end{vmatrix} = \dot{g}\begin{vmatrix} \varphi_k & \varphi_l \\ \dot{\varphi}_k & \dot{\varphi}_l \end{vmatrix}. \tag{A6}$$

Hence, we have from (A4), (A5), and (A6),

$$\dot{\psi}_{kl} = \begin{vmatrix} \varphi_k & \varphi_l \\ \ddot{\varphi}_k & \ddot{\varphi}_l \end{vmatrix}, \tag{A7}$$

$$\dddot{\psi}_{kl} = -\dot{g}\begin{vmatrix} \varphi_k & \varphi_l \\ \dot{\varphi}_k & \dot{\varphi}_l \end{vmatrix} - 2g\begin{vmatrix} \varphi_k & \varphi_l \\ \ddot{\varphi}_k & \ddot{\varphi}_l \end{vmatrix}. \tag{A8}$$

Substitution of (A4), (A7), and (A8) into (A3) yields

$$\dddot{\psi}_{kl} + 2g(t)\dot{\psi}_{kl} + \dot{g}(t)\psi_{kl} = 0,$$

and so ψ_{kl} is a solution of (A3). If we assume that for $k \neq l$, ψ_{kl} is the trivial solution, then $\psi_{kl} = \varphi_k\dot{\varphi}_l - \dot{\varphi}_k\varphi_l = 0$ which implies that $\varphi_k = c\varphi_l$ c a constant. This is contrary to our assumption that φ_k and φ_l are linearly independent.

Now let us assume that the ψ_{kl} are linearly dependent, that is,

$$a_1\psi_{12} + a_2\psi_{13} + a_3\psi_{23} = 0, \tag{A9}$$

where a_1,a_2, and a_3 are not all zero. Now by Lemma 1, the Wronskian of the solutions φ_1,φ_2, and φ_3 is

$$W(\varphi_1,\varphi_2,\varphi_3) = \begin{vmatrix} \varphi_1 & \varphi_2 & \varphi_3 \\ \dot{\varphi}_1 & \dot{\varphi}_2 & \dot{\varphi}_3 \\ \ddot{\varphi}_1 & \ddot{\varphi}_2 & \ddot{\varphi}_3 \end{vmatrix} = c, \quad \text{a constant.} \tag{A10}$$

By the properties of determinants and elementary transformations of rows and columns we have

$$W(\varphi_1,\varphi_2,\varphi_3) = \frac{1}{\varphi_1}\begin{vmatrix} \psi_{12} & \psi_{13} \\ \dot{\psi}_{12} & \dot{\psi}_{13} \end{vmatrix} = -\frac{1}{\varphi_2}\begin{vmatrix} \psi_{12} & \psi_{23} \\ \dot{\psi}_{12} & \dot{\psi}_{23} \end{vmatrix}$$
$$= -\frac{1}{\varphi_3}\begin{vmatrix} \psi_{13} & \psi_{23} \\ \dot{\psi}_{13} & \dot{\psi}_{23} \end{vmatrix} = c. \tag{A11}$$

If $a_1 \neq 0$, we can substitute for ψ_{12} in $W(\varphi_1,\varphi_2,\varphi_3)$ in (A11):

$$W(\varphi_1,\varphi_2,\varphi_3) = \frac{a_3}{a_1\varphi_1}\begin{vmatrix} \psi_{13} & \psi_{23} \\ \dot{\psi}_{13} & \dot{\psi}_{23} \end{vmatrix} = \frac{1}{\varphi_3}\begin{vmatrix} \psi_{13} & \psi_{23} \\ \dot{\psi}_{13} & \dot{\psi}_{23} \end{vmatrix}.$$

Hence, $a_1\varphi_1 = a_3\varphi_3$, a contradiction. If $a_1 = 0$, then $a_2,a_3 \neq 0$ if the ψ_{jk} are linearly dependent. But then by (A11), $W(\varphi_1,\varphi_2,\varphi_3) = 0$ which implies that the $\varphi_1,\varphi_2,\varphi_3$ are linealy dependent contrary to hypothesis. Thus the ψ_{jk}, $1 \leqslant j < k \leqslant 3$ must be linearly independent.

Theorem 2: Let $\varphi_1,\varphi_2,\varphi_3$ be three linearly independent, nontrivial solutions to the homogeneous, linear ordinary differential equation

$$\dddot{A} + 8g_2(t)\dot{A} + 4\dot{g}_2(t)A = 0. \tag{A12}$$

Now let χ_1,χ_2 be any two linearly independent, nontrivial solutions to the homogeneous ordinary differential equaton

$$\ddot{b} + 2g_2(t)b = 0. \tag{A13}$$

Then the product $\chi_j\chi_k$ is a solution of the differential equation (A12) and furthermore we can identify three linearly independent solutions

$$\varphi_1 = (1/\alpha)\chi_1^2, \quad \varphi_2 = (1/\alpha)\chi_2^2, \quad \varphi_3 = (2/\alpha)\chi_1\chi_2 \tag{A14}$$

such that

$$\psi_{12} = \varphi_3, \quad \psi_{31} = -2\varphi_1, \quad \psi_{32} = 2\varphi_2, \tag{A15}$$

where the Wronskian $W(\chi_1,\chi_2) = \alpha$.

Proof: That $\chi_j\chi_k$ is a solution of (A12) can be shown by substitution into (A12) and using (A13). That the three soutions (A14) are linearly independent follows from Theorem 2. That Eqs. (A15) hold can be demonstrated by direct computation.

[1] See for example, A. Messiah, *Quantum Mechanics*, Vol. 2, 3rd ed. (Wiley, New York, 1961).
[2] C. P. Boyer, Helv. Phys. Acta **47**, 589 (1974).
[3] C. P. Boyer, R. T. Sharp, and P. Winternitz, J. Math. Phys. **17**, 1439 (1976).
[4] W. Miller, Jr., *Symmetry and Separation of Variables* (Addision-Wesly, Reading, Mass., 1977).
[5] E. A. Coddington, *An Introduction to Ordinary Differential Equations* (Prentice-Hall, Englewood Cliffs, N. J., 1961).
[6] If $\{X + aE\}$ is a kinematic algebra for V, then so is $\{X\}$.
[7] If $\mathscr{K}$ is a subalgebra of $\mathscr{S}_1$, E is not a member of $\mathscr{K}$, and $\mathscr{K}$ is a kinematical algebra for a Schrödinger equation, then so is $\mathscr{K} \cup \{E\}$.
[8] E. G. Kalnins and W. Miller, Jr., J. Math Phys. **15**, 1728 (1974).
[9] D. R. Truax, J. Math. Phys. **21**, 807 (1980), and references therein. The analysis in this paper is for three-dimensional Schrödinger equations with time-independent interactions.
[10] A. Messiah, *Quantum Mechanics*, Vol. 1, 3rd ed. (Wiley, New York, 1961).
[11] P. G. L. Leach, J. Math. Phys. **21**, 300 (1980).

Partition combinatorics and multiparticle scattering theory

K. L. Kowalski
Case Western Reserve University, Cleveland, Ohio 44106
Wayne N. Polyzou[a] and Edward F. Redish
Department of Physics and Astronomy, University of Maryland, College Park, Maryland 20742

(Received 12 September 1980; accepted for publication 20 January 1981)

The recently developed combinatoric methods for handling partition-labeled operators in N-particle scattering theory are studied from an abstract point of view. The relation of these methods to approaches of the cluster/cumulant type in many areas of mathematical physics is pointed out. The concept of connectedness is defined abstractly and the mathematical structure of the partition lattice is considered in detail. Many of the useful results of combinatoric scattering theory are shown to be natural expressions of properties of the partition lattice. The conditions on these results can then be stated with precision. A number of new operator theorems are also obtained by means of applying simple extensions and analogs of the known properties of the partition lattice.

PACS numbers: 03.65.Nk, 03.80. + r, 02.10.Eb

I. INTRODUCTION

Recent developments in time-independent nonrelativistic multiparticle scattering theory[1-7] have yielded a number of previously unfamiliar relations among operators labelled by channel or partition indices. It has since been observed that many of these relations follow directly from the lattice structure of the label set.[8,9] In this paper we explicate the abstract relations behind the practical ones which have been obtained, generalize them to obtain new relations, and clarify the assumptions and conditions inherent in some previous work. By so doing we are able to see the relation of the scattering theory structure to similar structures in other fields. In addition, these techniques provide powerful and concise tools for deriving and determining the structure of operator equations.

The goal of scattering theory is to construct transition or resolvent operators for an N-particle system given the interactions present in the fewer-particle systems and the boundary conditions.[10-13] One method of doing this which has been highly successful in the two-[14,15] and three-[12] particle problems has been through the numerical solution of compact kernel integral equations which incorporate both the dynamical input and sufficient boundary conditions to specify a unique solution . As is now well known, obtaining such integral equations (and in particular the proof that the kernel is compact) is decidedly nontrivial for three or more particles.[13,16,17] In most cases such proofs rely on restrictions on the behavior of the interparticle potentials and on the construction of a kernel such that some iterate is completely connected, i.e., involves every particle with every other by a chain of interactions. A completely connected operator constructed from resolvents and few-particle interactions is compact given sufficiently restrictive assumptions on the nature of the potentials.[16]

With the above result, connectivity becomes the central issue in the construction of well-posed integral equations for the N-particle problem. It is therefore important to have a sharp specification of what is meant by connectivity and which of its properties are needed for which results. A sharp definition should permit the development of a better understanding of the relation between compactness and the various possible notions of connectivity.

The underlying notion of connectivity is graph theoretical and depends only on the abstract concepts of lines and vertices.[18] In N-particle scattering theory we can use Weinberg graphs,[10] in which each particle is represented by a line, and each interaction by a line or blob connecting different groups of particles. If one puts the system in a box of volume L^3, in the limit of large volume the matrix elements of a given graph will be proportional to the volume raised to a power equal to the number of linked parts the graph contains.[19] This notion of connectivity can be restated in more convenient terms in the infinite volume limit in terms of the translational symmetry properties of the operators.[9] These connectivity properties are classified in terms of partitions: assignments of a way of grouping particles to operators. The set of partitions forms a lattice.[20,21] The relation of the lattice structure of the partition set and the operator product is the central feature required. We discuss the lattice structure of the partition set in Sec. II. This leads to an abstract definition of connectivity.

The classification of various N-particle operators by their connectivity leads to nontrivial combinatoric problems associated with operator decomposition theorems and product rules. This was clearly recognized by Weinberg[10] who pointed out that some of the same problems have been encountered, solved, and applied in a host of other mathematical and physical contexts.[22] Central among these are the cluster/cumulant expansions which appear in statistics,[23] the theory of functions of random variables,[24] the theory of fluids,[25-27] quantum field theory,[28-32] and S-matrix theory,[33] as well as in several other contexts.[6,7,34,35]

Since modern multiparticle scattering theory is concerned with the formulation of exact as well as with a wide variety of approximate integral equations,[5] one must be able to treat very general situations. As a result, standard combinatoric techniques are not adequate for the operator analy-

[a]Present address: Los Alamos Scientific Laboratory, Los Alamos, New Mexico 87545.

 0022-2488/81/091965-18$01.00

sis one encounters there. (In fact, some of the combinatorial questions which have arisen remain unanswered at present.[36]) This circumstance has led to the development of new methods for handling the combinatorics.[9]

Some of the most interesting techniques have been developed in conjunction with the generalization extension of Sloan's[37] four-particle equations to N particles.[38,39] These are based on the idea of distribution laws: intercluster potentials are divided up among the various partitions in different ways.[1–3,40–44] In particular, the distribution law found in Ref. 1 was accompanied by important combinatoric results which have led to further developments and applications, in particular with the possibility of deriving connected-kernel equations reflecting various physically motivated approximations.[4] We put these distribution laws in context by explicating and elaborating a number of results concerning the structure and properties of the partition lattice. In Sec. III we present these results emphasizing the mathematical lattice structure. Taking this viewpoint allows us to see symmetries in the structure and thereby to obtain new results analogous in structure to previously known ones. In particular, we observe that the duality structure of the partition lattice leads to results for the completely disconnected partition (the minimal element of the lattice) analogous to the ones known for the completely connected partition (the maximal element of the lattice).

In Sec. IV these abstract lattice-theoretic results are used to generate operator theorems. Many well-known theorems, such as the distribution laws mentioned above, are seen in this context not as the surprising accidents they once appeared, but rather as a natural and fundamental expression of the mathematical properties of the partitioning process. In addition to providing a new perspective, the lattice theory results allow us to obtain relations involving a new set of operators. These operators satisfy a cluster/cumulant type of expansion involving the standard coefficients, and distribution laws involving a different natural set of coefficients from those used previously.

Section V contains a brief summary of the paper.

II. PARTITIONS AND CONNECTIVITY

In this section we introduce the notion of partition and consider its structure as a mathematical lattice. The general concept of connectivity is then defined in an abstract sense. The specific connectivity arising in the case of N-particle scattering theory with potentials of compact support is discussed and relation with the structure of the translation group is presented. Finally, an application is considered in which a matrix notation is introduced. This notation is of considerable importance for later developments.

We begin by reviewing briefly the structure of the partition lattice. (For more extensive treatments the reader is referred to Refs. 20 and 21.) A partition of a set X of N objects[45] is a grouping of the N objects into K $(1\leqslant K\leqslant N)$ nonempty disjoint equivalence classes. In the N-particle case, the equivalence classes are referred to as *clusters*. We let lower case Latin letters, a, b, c,$\cdots$ denote partitions of X into $n_a, n_b, n_c, \cdots$ clusters. Exceptions to this notation are made in the case of the two partitions which are uniquely specified by their number of clusters. We denote the N-cluster partition by $\underline{0}$, and the 1-cluster partition by $\underline{1}$. Let $\mathscr{P}$ denote the set of all partitions of X.

Given two partitions a and b, one says that *a contains b* (denoted as $a\supseteq b$) if any two particles which are equivalent under the equivalence relation associated with b are necessarily equivalent under the relation associated with a. That is, each cluster of b is wholly contained within some cluster of a.

The binary relation $\supseteq$ on $\mathscr{P}\times\mathscr{P}$ can be shown to be a partial ordering on $\mathscr{P}$ in the sense that it is reflexive ($a\supseteq a$ $\forall a\in\mathscr{P}$), antisymmetric ($a\supseteq b$ and $b\supseteq a$ implies $a=b$), and transitive ($a\supseteq b$ and $b\supseteq c$ implies $a\supseteq c$). A lattice structure is defined on $\mathscr{P}$ by introducing two operations. The *join* or union) of two partitions a and b is the least upper bound of a and b with respect to $\supseteq$. It is the unique partition c satisfying: if $c\supseteq a$, $c\supseteq b$, and $\forall d\supseteq a,b$, then $d\supseteq c$. This is the partition c obtained by joining in the same cluster all particles which are in the same cluster of either a or b. The *meet* (or intersection) of the partitions a and b is the greatest lower bound of a and b with respect to $\supseteq$. It is the unique partition c satisfying $a\supseteq c$, $b\supseteq c$, and $\forall d\supseteq a,b$ then $c\supseteq d$. This is the partition obtained by requiring that those pairs which lie in different clusters in either a or b lie in different clusters in c. We denote the union of a and b by $a\cup b$ and the intersection by $a\cap b$.

It is easy to show these relations are commutative ($\forall a,b\in\mathscr{P}$, $a\cup b=b\cup a$, $a\cap b=b\cap a$), associative [$\forall a,b,c\in P$,$(a\cup b)\cup c=a\cup(b\cup c)$ and $(a\cap b)\cap c=a\cap(b\cap c)$], absorptive [$\forall a,b\in\mathscr{P}$,$a\cup(b\cap a)=a\cap(b\cup a)=a$], and idempotent ($\forall a\in\mathscr{P}$,$a\cup a=a\cap a=a$). This gives $(\mathscr{P},\cup,\cap)$ the abstract structure of a lattice. This lattice was introduced by Birkhoff and is called the *partition* or *Birkhoff lattice*.[20] It is nondistributive and semimodular. The semimodularity turns out to be an essential ingredient in studying the class of N-body equations labeled by chains of partitions.[13]

The concept of connectivity is introduced in order to specify what groups of particles are mutually associated in some way. We specify this abstractly by introducing the concept of connectivity structure.

Definition: A *connectivity structure* on an operator algebra $\mathscr{B}$ over $\mathscr{C}$ is a triple $(\mathscr{L},\mathscr{C},C)$ where $\mathscr{L}$ is a lattice, $\mathscr{C}$ is a linear subspace of $\mathscr{B}$, and C is a map $\mathscr{L}\times\mathscr{C}\rightarrow\mathscr{C}$ satisfying

$$\text{(i)}\,\forall A\in\mathscr{C}\ \sum_{a\in\mathscr{L}} C(a,A)=A. \tag{2.1}$$

(ii) If $\{a_1,\ldots,a_n\}$ are a set of distinct elements of $\mathscr{L}$ then the nonvanishing elements of the set $C(a_i,A)$, $i=1,\ldots,n$, are linearly independent for all $A\in\mathscr{C}$.

(iii) The functions $C(a,\cdot)$ are linear.

(iv) If $A,B\in\mathscr{C}$ then

$$C(c,C(a,A)C(b,B))=0 \tag{2.2}$$

unless $c=a\cup b$.

We refer to the map C as the *connectivity map*. For typographical simplicity we write the operator $C(a,A)$ as $[A]_a$. This operator is referred to as *the part of the operator A which* *has connectivity a*. Condition (i) says that the operators in $\mathscr{C}$ are those for which the operator is the sum of its parts of all possible connectivities. Equation (2.1) is referred to as the *cluster expansion*.[46] The condition (2.2) is particularly important as it guarantees the consistency of the lattice union and the operator product. We refer to this condition as the *connectivity product rule*. Here we only consider the case where $\mathscr{L}$ is the partition lattice for N distinguishable objects.

A prototype for this structure is obtained by considering the space $\mathscr{C}_0$ = sums of finite strings of operators which are (alternating) products of pair potentials of compact support with free resolvents. The map C is specified by defining it on any single string, $\mathscr{O}$. A unique partition $a_{\mathscr{O}}$ can be assigned to $\mathscr{O}$ by defining all mutually interacting particles as equivalent. The partition $a_{\mathscr{O}}$ is the one whose clusters are the equivalence classes under this relation. The operator C is then uniquely defined by

$$C(a,\mathscr{O}) = \mathscr{O}\delta(a,a_{\mathscr{O}}). \tag{2.3}$$

We refer to the connectivity structure so defined as *string connectivity*. As an example, consider the string $\mathscr{O} = V_{12}G_0V_{34}$ in the five-body problem (see Fig. 1). The connectivity assigned to this operator is $a_{\mathscr{O}} = (12)(34)(5)$.

We restrict ourselves to the space $\mathscr{C}_0$ even though we eventually will want to work with some larger set of operators $\overline{\mathscr{C}}_0 \supset \mathscr{C}_0$. What the largest space of operators on which a connectivity structure can be defined is an interesting question, but one beyond the scope of the present paper.

The physical connectivity structure defined above has a number of interesting properties. One of these is that if an operator A has string connectivity a, then its momentum space matrix elements have the form

$$\langle \mathbf{P}_1\cdots\mathbf{P}_N|\mathrm{A}|\mathbf{P}'_1\cdots\mathbf{P}'_N\rangle = A^{\mathrm{red}}(\mathbf{P}_1\cdots\mathbf{P}_N;\mathbf{P}'_1\cdots\mathbf{P}_N)$$
$$\times\prod_{i=1}^{n_a}\delta(\mathbf{P}(a,i)-\mathbf{P}'(a,i)), \tag{2.4}$$

where

$$\mathbf{P}(a,i) = \sum_{j\in a_i}\mathbf{P}_j \tag{2.5}$$

is the momentum of the center of mass of the i^{th} cluster a_i of a. The function A^{red} is only defined on the manifold $\{\mathbf{P}(a,i) = \mathbf{P}'(a,i)\}$, $i = 1,\ldots,n_a$, and possesses no delta function singularities. One should note that the statement (2.4) is not adequate in itself to provide a definition of connectivity. To say an operator A has a momentum conserving δ function in the momentum representation is equivalent to the statement that A is invariant under the unitary group (of translations) generated by the conserved momenta. If we restrict our attention to the space of bounded operators on N-body Hilbert space, $\mathscr{L}(\mathscr{H}_N,\mathscr{H}_N)$, then the set of operators A invariant under this group form a subspace of $\mathscr{L}(\mathscr{H}_N,\mathscr{H}_N)$. Since $\mathscr{L}(\mathscr{H}_N,\mathscr{H}_N)$ is not a Hilbert space, there does *not* exist a unique decomposition of an arbitrary element into a part contained in a given subspace and a part not contained in that subspace. This implies that additional constraints are needed to make this decomposition unique.

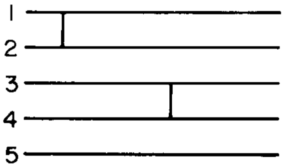


FIG. 1. Weinberg graph of the operator $V_{12}\,G_0\,V_{34}$ for $N = 5$. This graph has the connectivity (12) (34)(5).

The $3N$-parameter unitary group of translations, $\mathscr{U}$, gives a useful characterization of string connectivity as well as a constructive procedure for obtaining the cluster decomposition of a given operator. Define the group of unitary translation operators of the N-body system to be the distinct elements of the set of operators $\{U_a(\mathbf{y}_1,\ldots,\mathbf{y}_{n_a})|a\in\mathscr{L},\ \mathbf{y}_j\in R^3,\ j = 1,\ldots,n_a\}$, where

$$U_a(\{y_i\})\equiv U_a(\mathbf{y}_1,\ldots,\mathbf{y}_{n_a}) = \exp\Big[-i\sum_{j=1}^{n_a}\mathbf{P}(a,j)\cdot\mathbf{y}_j\Big].$$

The operator $U_a(\mathbf{y}_1,\ldots,\mathbf{y}_{n_a})$ translates the CM of the ith cluster of partition a by $\mathbf{y}_i$.

To each partition $a\in\mathscr{L}$, $\mathscr{U}$ has a subgroup $\mathscr{U}_a = \{U_a(\mathbf{y}_1,\ldots,\mathbf{y}_{n_a})|\mathbf{y}_i\in R^3,\ i = 1,\ldots,n_a\}$. The connections among the group and lattice properties are illustrated by the following propositions[8,9]:

(i) $\mathscr{U}_a$ is a subgroup of $\mathscr{U}_b$ iff $a\supseteq b$,

(ii) $\mathscr{U}_{a\cup b}$ is the largest group that is a subgroup of both $\mathscr{U}_a$ and $\mathscr{U}_b$,

(iii) $\mathscr{U}_{a\cap b}$ is the smallest group that contains both $\mathscr{U}_a$ and $\mathscr{U}_b$ as subgroups.

This implies that the lattice of subgroups of the translation group contains a sublattice isomorphic to the partition lattice.

The role of the lattice structure in many of the standard operator manipulations of N-particle scattering theory can be seen through the following example. If the operator A is an element of $\mathscr{C}$ then it admits a cluster decomposition

$$A = \sum_{a\in\mathscr{L}}[A]_a. \tag{2.1'}$$

For an arbitrary partition a we have the decomposition

$$A = A_a + A^a, \tag{2.6}$$

where we define the interior and exterior maps of $\mathscr{L}\times\mathscr{C}\rightarrow\mathscr{C}$, M_{I} and M_{E}, by

$$M_{\mathrm{I}}(a,A)\equiv A_a\equiv\sum_b\Delta_{ab}[A]_b, \tag{2.7}$$

$$M_{\mathrm{E}}(a,A)\equiv A^a\equiv\sum_b\bar{\Delta}_{ab}[A]_b. \tag{2.8}$$

We refer to A_a as the *a-interior part of A* and to A^a as the *a-exterior part of A*. The containment and noncontainment matrices Δ and $\bar{\Delta}$ are defined by

$$\Delta_{ab} = \begin{Bmatrix}1 & \text{if } a\supseteq b,\\ 0 & \text{otherwise}\end{Bmatrix}, \tag{2.9}$$

$$\bar{\Delta}_{ab} = 1-\Delta_{ab}. \tag{2.10}$$

The matrices Δ and $\bar{\Delta}$ play an essential role in constructing cluster and cumulant expansions and in generating various powerful sum rules. We consider their properties in detail in

Sec. III. We also introduce the general multi-index notation

$$A^{a_1\cdots a_n}_{b_1\cdots b_m} = \sum_{c\in\mathscr{S}} \Delta_{b_1c}\cdots\Delta_{b_mc}\bar{\Delta}_{a_1c}\cdots\bar{\Delta}_{a_nc}[A]_c. \tag{2.11}$$

If string connectivity is used then A_a is the part of A invariant under $\mathscr{U}_a$ while A^a is that part that satisfies[8,9]

$$U_a(\{y_i\})A^aU_a(\{y_i\})^\dagger \to 0,\ |\mathbf{y}_i|\to\infty\ \forall i.$$

The convergence is strong. These properties can be used to extend the definition of string connectivity. If we define

$$A_a = \lim_{\{|y_i|\}\to\infty} U_a(\{y_i\})AU_a(\{y_i\})^\dagger, \tag{2.12}$$

the parts of various connectivities $[A]_a$ may be obtained from the set of operators $\{A_a\}$ by the cumulant expansion discussed in Sec. IV below. Physically, this process corresponds to extracting the various disconnected parts of the operator by going to that region of configuration space where only the particles within the clusters of the partition a are close and the c.m.'s of the clusters are very far apart. The parts of subordinated connectivities, e.g., $b\subsetneq a$, are extracted from the asymptotic region of partition b. Once all the subordinately connected parts are determined and extracted from A_a, the part $[A]_a$ is what is left. It is clear that this definition of connectivity agrees with string connectivity on $\mathscr{C}_0$. We refer to this connectivity as *translation-invariance connectivity.*

We suppose that the N-particle Hamiltonian can be decomposed into a total kinetic energy H_0 and the sum of all interactions among the N particles, V:

$$H = H_0 + V. \tag{2.13}$$

In most cases of interest, V has a cluster decomposition

$$V = \sum_{a\in\mathscr{S}}[V]_a. \tag{2.14}$$

This corresponds to expressing V as sum of pair-wise, three-body, etc., terms. Thus,

$$[V]_a = 0 \quad \text{for } a\notin\mathscr{S}_{N-1}, \tag{2.15}$$

where $\mathscr{S}_{N-1}$ is the set of partitions $a\in\mathscr{S}$ that have only a single cluster with more than one particle. The partitions $a\in\mathscr{S}_{N-1}$ represent the classification of the subsystems of the N-particle system in the manner used by Weinberg.[10]

The potentials V_a and V^a represent the familiar channel (or internal) and residual (or external) interactions, respectively. The partition Hamiltonian corresponding to the partition a is

$$H_a = H_0 + V_a. \tag{2.16}$$

By virtue of (2.6)

$$H = H_a + V^a, \tag{2.17}$$

and we note H_0 is the 0 connected part of H.

The eigenstates of H_a on which all of the clusters of the partition a are bound are referred to as *maximally bound asymptotic states*, or *channel states.* If a partition Hamiltonian possesses such states the partition is referred to as *stable.* It is generally assumed (though proved only for a few cases) that if the potentials are short-ranged and not "pathological" that all of the unbound eigenstates of H look asymptotically like maximally bound states of some H_a. The transition amplitude for the scattering from one channel state corresponding to the partition a to another eigenstate $|\phi_b\rangle$ of H_b is given by the on-shell matrix element[47]

$$\langle\phi_b|T^{ba}_{(+)}|\phi_a\rangle, \tag{2.18}$$

where the transition operator $T^{ba}_{(+)}$ is

$$T^{ba}_{(+)} = V^b + V^bG(Z)V^a, \tag{2.19}$$

and

$$G(Z) = (Z-H)^{-1} \tag{2.20}$$

is the resolvent of the full Hamiltonian with parametric energy Z. The resolvent G can be expressed in terms of the partition resolvents

$$G_a(Z) = (Z-H_a)^{-1} \tag{2.21}$$

by means of the second resolvent relations

$$G(Z) - G_a(Z) = G_a(Z)V^aG(Z) \tag{2.22a}$$

$$= G(Z)V^aG_a(Z). \tag{2.22b}$$

The identification of $G_a(Z)$ as the a-invariant part of $G(Z)$ follows immediately from (2.21) and (2.22). It follows that

$$G^a(Z) = G_a(Z)V^aG(Z) \tag{2.23a}$$

$$= G(Z)V^aG_a(Z). \tag{2.23b}$$

Equations (2.22) can be rewritten in a matrix notation in the partition indices.[48] If we let $\hat{G}$ and $\bar{V}$ represent diagonal matrices with elements $G_a\delta_{ab}$ and $V^a\delta_{ab}$, respectively, then (2.22) becomes

$$G\mathscr{S} - \hat{G}\mathscr{S} = \hat{G}\bar{V}G\mathscr{S} = \hat{G}\bar{V}\mathscr{S}G \tag{2.24a}$$

$$= G\mathscr{S}\bar{V}\hat{G}. \tag{2.24b}$$

Here $\mathscr{S}$ denotes the matrix with elements $(\mathscr{S})_{ab} = 1$, and G is the diagonal matrix $(G)_{ab} = \delta_{ab}G(Z)$.

The matrix forms of Eqs. (2.24) immediately suggest the introduction of channel coupling schemes defined by the stipulation that[1,40–44]

$$\bar{V}\mathscr{S} = \mathscr{Y}\mathscr{S}. \tag{2.25}$$

Here, the elements $(\mathscr{Y})_{bc}$ of the matrix $\mathscr{Y}$ are not necessarily diagonal in the partition indices. Then, since (2.19) can be written as

$$T^{(+)} = \bar{V}\mathscr{S} + \bar{V}G\mathscr{S}\bar{V}, \tag{2.26}$$

we obtain, using (2.33a) and (2.24), the set of coupled integral equations (in matrix form)

$$T^{(+)} = \mathscr{Y}\hat{G}\mathscr{S}\hat{G}^{-1} + \mathscr{Y}\hat{G}T^{(+)}, \tag{2.27}$$

Several significant avenues of investigation have been pursued using (2.27) along with the requirement that $\mathscr{Y}$ be chosen so that the kernel of Eq. (2.27) becomes a connected operator after a finite number of iterations.[5]

In dealing with relations such as resolvent or scattering integral equations we must deal with products of operators. If we are given cluster decompositions of two operators, we would like to be able to construct the cluster decomposition of their product. Property (iv) of the definition of a connectivity structure implies that

$$[AB]_a = \sum_{\{(b,c)\in P\times P\,|\,b\cup c = a\}} [A]_b [B]_c. \tag{2.28}$$

III. LATTICE PROPERTIES

The structure of the partition lattice is expressed through the properties of the containment matrix Δ and its complement $\bar{\Delta}$ which provide an algebraic representation of the lattice order relations $a \supseteq b$ and $a \not\supseteq b$, respectively. These matrices enter into the treatment of connectivity structures through the interior and exterior maps (2.7) and (2.8). Upper and lower indexed operators are often the natural quantities to use in performing calculations while the parts of an operator having well-defined connectivities are the natural quantities for determining the compactness properties of the scattering equations. As a result, the properties of the matrices relating these sets of objects, Δ and $\bar{\Delta}$, play an important role in constructing scattering equations.

In this section we investigate the properties of matrices Δ and $\bar{\Delta}$ including the existence of inverses, sum rules, and product laws. These relations provide a concise representation of the lattice properties. The cases of the three- and four-particle partition lattices are explicated in the appendix as examples. The results contained in this section only depend on the structure of the partition lattice and do not rely on any of the properties of connectivity.

We begin by considering the matrix Δ. Most of its properties discussed below follow from the important fact that there exist unique elements $\underline{1}$ and $\underline{0}$ such that for every $a \in \mathscr{L}$, $\underline{1} \supseteq a \supseteq \underline{0}$. In order to express our results in matrix as well as indexed notation, we introduce the projectors

$$(P_{\underline{0}})_{ab} = \delta_{a\underline{0}}\delta_{ab}, \tag{3.1a}$$

$$(P_{\underline{1}})_{ab} = \delta_{a\underline{1}}\delta_{ab} \tag{3.1b}$$

and their complements

$$Q_{\underline{0}} = \underline{I} - P_{\underline{0}}, \tag{3.1c}$$

$$Q_{\underline{1}} = \underline{I} - P_{\underline{1}}. \tag{3.1d}$$

The fundamental properties therefore have the following representations:

Proposition 3.1:

(A) $\Delta P_{\underline{0}} = \mathscr{S} P_{\underline{0}}, \quad \Delta_{a\underline{0}} = 1$

(all partitions contain $\underline{0}$), (3.2a)

(B) $P_{\underline{0}}\Delta = P_{\underline{0}}, \qquad \Delta_{\underline{0}a} = \delta_{\underline{0}a}$

(only $\underline{0}$ is contained in or equal to $\underline{0}$), (3.2b)

(C) $\Delta P_{\underline{1}} = P_{\underline{1}}, \qquad \Delta_{a\underline{1}} = \delta_{a\underline{1}}$

(only $\underline{1}$ contains or is equal to $\underline{1}$), (3.2c)

(D) $P_{\underline{1}}\Delta = P_{\underline{1}}\mathscr{S}, \quad \Delta_{\underline{1}a} = 1$

(all partitions are contained in $\underline{1}$). (3.2d)

In the above and in what follows we represent our results in three ways: in matrix form, in indexed form, and in words. Proofs are generally simpler to carry out in matrix form, but the structure of the result is occasionally obscured by the large number of operators required to specify the allowed index set.[49]

We now construct Δ^{-1}. This allows us to invert expressions of the form (2.7) and corresponds to the representation of the cumulants in terms of moments of a distribution. Choose some definite ordering of the partitions, $>$ satisfying $a_n > b_m$ whenever $n > m$. We refer to this as the *standard ordering*. With this ordering, Δ is upper triangular and has ones runnning down the diagonal, so $\det\Delta = 1$, and therefore, Δ^{-1} exists on $\mathscr{L}$.

The inverse of the matrix Δ can be expressed in two convenient forms:

Theorem 3.1: The inverse of Δ on $\mathscr{L}$ is given by the forms

$$\Delta^{-1} = \sum_{k=0}^{N-1} \frac{(-1)^k}{k!} \left\{ \sum_{l=k}^{N-1} \frac{l!}{(l-k)!} \right\} \Delta^k, \tag{3.3a}$$

$$\Delta^{-1}_{ab} = \begin{cases} (-1)^{n_a} \prod_{i=1}^{n_a} (-1)^{n_{b_i}} (n_{b_i} - 1)!, & a \supseteq b, \\ 0, & a \not\supseteq b. \end{cases} \tag{3.3b}$$

In (3.3b), n_{b_i} is the number of clusters of b contained in the ith cluster of a.

The matrix elements of Δ^{-1} are referred to in scattering theory[50] as *anticluster coefficients* because of their role in inverting the cluster expansion. In lattice theory Δ^{-1} is referred to[51] as the *Möbius function* for the partial ordering $\supseteq$.

Proof of Theorem 3.1: The form (3.3a) can be obtained by writing

$$\Delta = I + \hat{\delta}. \tag{3.4}$$

From the definition of Δ it follows that $\hat{\delta}_{ab} = 0$ if a $a \not\supset b$ or $a = b$, so $(\hat{\delta}^p)_{ab} = 0$ unless $a \supset b$ for $p \geqslant 1$, where $\supset$ indicates proper containment. Successive iterations of the identity

$$\Delta^{-1} = I - \Delta^{-1}\hat{\delta}, \tag{3.5}$$

and the stated properties of $\hat{\delta}$ yield (3.3a).

Equation (3.3a) gives an algorithm for computing the anticluster coefficients in terms of powers of Δ. They may be obtained explicitly as follows. We first note that (3.3a) implies $\Delta^{-1}_{ab} = 0$ for $a \not\supseteq b$. To compute a specific form of Δ^{-1}_{ab} for $a \supseteq b$, we fix a and observe that the set of partitions $\mathscr{L}_a = \{b \,|\, a \supseteq b\}$ with the induced lattice structure is lattice isomorphic to the direct product of the n_a lattices associated with the cluster of a. Since Δ^{-1}_{ab} and Δ_{bc} are nonvanishing only for $b,c \in \mathscr{L}_a$, it suffices to compute Δ^{-1}_{ab} on $\mathscr{L}_a$. In the direct product representation, Δ_{bc} is mapped into $\otimes_{i=1}^{n_a} \Delta_{b_i c_i}$ where b_i and c_i are the partitions of the ith cluster of a induced by the refinements b and c, respectively. In the product representation this has the inverse $\otimes_{i=1}^{n_a} \Delta^{-1}_{d_i b_i}$ which induces an inverse on the set of matrices indexed by $\mathscr{L}_a$. Since a is a product of 1-cluster partitions in the product lattice,

$$\Delta^{-1}_{ab} = \otimes_{i=1}^{n_a} \Delta^{-1}_{1_i b_i}. \tag{3.6}$$

The elements $\Delta^{-1}_{1,b}$ are shown below [Eqs. (3.14a) and (3.25a)] to have the form $(-1)^{n_b+1}(n_b - 1)!$ and the result (3.3b) follows.■

We note some properties of Δ^{-1}. Equation (3.3a) implies directly that Δ^{-1} only has nonvanishing matrix elements in the same places where Δ has them. From (3.3b) we see that all the matrix elements of Δ^{-1} are integral and that all its diagonal elements are 1. The reader is invited to note these properties for the examples in the Appendix.

We next consider the construction of "sum rules" for Δ^{-1}. By this, we mean expressions arising from summing a row or column of the matrix. These results have been used with considerable elegance in the construction of connected-kernel scattering equations. How these enter can be seen by considering two types of objects.

A matrix W which satisfies one of the conditions

$$\mathscr{S}W=\mathscr{S} \tag{3.7a}$$

or

$$W\mathscr{S}=\mathscr{S} \tag{3.7b}$$

is referred to as[52] (right or left) *$\mathscr{S}$-invariant*. Matrices A and B which satisfy

$$A\mathscr{S}=B\mathscr{S} \tag{3.8a}$$

or

$$\mathscr{S}A=\mathscr{S}B \tag{3.8b}$$

are said to be (right or left) *$\mathscr{S}$-equivalent*. Through the appearance of the term $\mathscr{S}\bar{V}$ in the Eqs. (2.24) and (2.26), any $\mathscr{S}$-invariant quantity may be inserted before or after the $\mathscr{S}$, or $\bar{V}$ may be replaced by an $\mathscr{S}$-equivalent matrix. Either of these procedures lead to different forms for the coupling structure of the equations and therefore to a rich variety of equations for the scattering operators.[1,2,5] Both $\mathscr{S}$-invariant and $\mathscr{S}$-equivalent quantities appear throughout our subsequent development.

We note that the multiplication of a matrix by $\mathscr{S}$ on the right (left) has the effect of summing the matrix's columns (rows), viz.,

$$(M\mathscr{S})_{ab}=\sum_c M_{ac} \quad \text{(independent of } b\text{)}, \tag{3.9a}$$

$$(\mathscr{S}M)_{ab}=\sum_c M_{cb} \quad \text{(independent of } a\text{)}. \tag{3.9b}$$

$\mathscr{S}$ equivalence leads to the formulations of useful sum rules if one of the equivalent quantities is very simple.

Other aspects of out subsequent development are intimately related to $\mathscr{S}$ invariance and equivalence. Given any nonsingular matrix M one can define unique diagonal matrices $\mathscr{D}^{\mathrm{R,L}}$ by the stipulations that $\mathscr{D}^{\mathrm{R}}M$ and $M\mathscr{D}^{\mathrm{L}}$ be right and left $\mathscr{S}$-invariant, respectively:

$$\mathscr{S}\mathscr{D}^{\mathrm{R}}M=\mathscr{S}, \tag{3.10a}$$

$$M\mathscr{D}^{\mathrm{L}}\mathscr{S}=\mathscr{S}. \tag{3.10b}$$

$\mathscr{D}^{\mathrm{R}}$ is the diagonal matrix with the sum of the rows of M^{-1} as its entries while $\mathscr{D}^{\mathrm{L}}$ has the sum of the columns of M^{-1} as its entries. Equations (3.10) correlate diagonal matrices with sum rules (which may be confined to a subset of $\mathscr{L}$). Let $\{f\}_a$ be a set of partition-labelled elements of an operator algebra, and let

$$F\equiv\sum_a\{f\}_a, \tag{3.11a}$$

$$F_{\mathrm{R}}(a)=\sum_b M_{a,b}\{f\}_b, \tag{3.11b}$$

$$F_{\mathrm{L}}(a)=\sum_b\{f\}_b M_{b,a}. \tag{3.11c}$$

Then it follows that

$$F=\sum_a F_{\mathrm{R}}(a)\mathscr{D}^{\mathrm{R}}_a=\sum_a F_{\mathrm{L}}(a)\mathscr{D}^{\mathrm{L}}_a. \tag{3.12}$$

Specific realizations of these algebraic structures appear throughout Sec. III and IV.

Sum rules corresponding to the choice $M=\Delta$ may be constructed from (3.2a) and (3.2d). It follows immediately that

Lemma 3.1:

$$\text{(A)}\quad \Delta^{-1}\mathscr{S}P_{\underline{0}}=P_{\underline{0}} \text{ or } \sum_b\Delta^{-1}_{ab}=\delta_{a\underline{0}}, \tag{3.13a}$$

$$\text{(B)}\quad P_{\underline{1}}\mathscr{S}\Delta^{-1}=P_{\underline{1}} \text{ or } \sum_a\Delta^{-1}_{ab}=\delta_{\underline{1}b}. \tag{3.13b}$$

We now define the quantities

$$C_a\equiv-\Delta^{-1}_{\underline{1}a}, \tag{3.14a}$$

$$D_a\equiv-\Delta^{-1}_{a\underline{0}}. \tag{3.14b}$$

The sum rules (3.10) then imply the

Corollary:

$$\sum_a{}'\Delta^{-1}_{ab}=C_b+\delta_{\underline{1}b},\quad \sum_b{}^{*}\Delta^{-1}_{ab}=D_a+\delta_{a\underline{0}}, \tag{3.15a}$$

$$\sum_a{}^{*}\Delta^{-1}_{ab}=\delta_{\underline{1}b}-\delta_{\underline{0}b},\quad \sum_b{}'\Delta^{-1}_{ab}=\delta_{a\underline{0}}-\delta_{a\underline{1}}, \tag{3.15b}$$

$$\begin{aligned}&\sum_a{}'^{*}\Delta^{-1}_{ab}=C_b+\delta_{\underline{1}b}-\delta_{\underline{0}b},\\ &\sum_a{}'^{*}\Delta^{-1}_{ab}=D_a+\delta_{a\underline{0}}-\delta_{a\underline{1}},\end{aligned} \tag{3.15c}$$

where the prime and the star on the sums indicate sums over the subsets of $\mathscr{L}$:

$$\mathscr{L}'=\mathscr{L}-\{\underline{1}\}, \tag{3.16}$$

$$\mathscr{L}'^{*}=\mathscr{L}-\{\underline{0}\}, \tag{3.17}$$

respectively, and

$$\mathscr{L}'^{*}=\mathscr{L}-\{\underline{0}\}-\{\underline{1}\}, \tag{3.18}$$

with the corresponding sum Σ'^{*}.

The sum rules will be employed in terms of the results:

Theorem 3.2:

$$\text{(A)}\quad \sum_a{}'C_a\Delta_{ab}=1-\delta_{\underline{1}b}\equiv\bar{\delta}_{\underline{1}b}, \tag{3.19a}$$

$$\text{(B)}\quad \sum_b{}^{*}\Delta_{ab}D_b=1-\delta_{a\underline{0}}\equiv\bar{\delta}_{a\underline{0}}. \tag{3.19b}$$

Proof: Result (A) follows from the identity

$$\sum_b\Delta^{-1}_{ab}\Delta_{bc}=\delta_{ac}, \tag{3.20}$$

upon setting $a=1$ and using the properties $\Delta_{1c}=1$ and

$$C_{\underline{1}}=-1. \tag{3.21}$$

Result (B) is obtained starting from

$$\sum_b\Delta_{ab}\Delta^{-1}_{bc}=\delta_{ac}, \tag{3.22}$$

only now we set $a=\underline{0}$ and use the properties $\Delta_{a\underline{0}}=1$ and

$$D_{\underline{0}} = -1. \blacksquare \tag{3.23}$$

The equations of the theorem for $b = \underline{0}$ and $a = \underline{1}$, respectively, give the particularly useful results:

Corollary:

$$\sum_a{}' C_a = 1, \tag{3.24a}$$

$$\sum_b{}^{*} D_b = 1, \tag{3.24b}$$

$$\sum_a{}'^{*} C_a = 1 - C_{\underline{0}} = \sum_b{}'^{*} D_b. \tag{3.24c}$$

In obtaining (3.24c) we used the fact that $D_{\underline{1}} = C_{\underline{0}}$. (This important number is referred to[51] as the *Euler characteristic* of the lattice in the mathematics literature.)

The result (A) of Theorem 2 is well known in the scattering theory literature,[2,4,9] and possesses considerable power in simplifying results otherwise obtained at great effort. The result (B) is presented here for the first time (to our knowledge) and is only one aspect of a powerful duality which is a standard lattice property.[21] A number of the new results presented below are duals of previously known results. In this duality, completely connected ($\underline{1}$) is replaced by completely disconnected ($\underline{0}$), joins and meets interchange, and partitions interchange with the sublattices corresponding to the clusters of the partition.

The coefficients C_a and D_a are sufficiently important that we would like to have explicit expressions for them. They are given by the following

Proposition 3.2: If a is a partition into the n_a clusters $a_1, a_2, \ldots, a_{n_a}$ having, respectively, $n_{a_1}, n_{a_2} \cdots$ particles, then

$$\text{(A)} \quad C_a = (-1)^{n_a} (n_a - 1)!, \tag{3.25a}$$

$$\text{(B)} \quad D_a = (-1)^{n_a + 1} \prod_{i=1}^{n_a} C_{a_i} = (-1)^{N + n_a + 1} \prod_{i=1}^{n_a} (n_{a_i} - 1)!. \tag{3.25b}$$

Proof: The expression for C_a can be demonstrated by an indirect argument. We introduce S_n^m, the number of distinct m-cluster partitions which can be made from n distinguishable objects.[50] In the combinatoric literature, this is referred to as the *Stirling number of the second kind*.[53] It is easy to show using standard properties of S_n^m that

$$\sum_{m=2}^{N} (-1)^m (m-1)! S_k^m = 1, \quad k \geqslant 2. \tag{3.26}$$

Using the definition of S_n^m,

$$\sum_a \delta_{m n_a} \Delta_{ab} = S_{n_b}^m, \tag{3.27}$$

we can write (3.23) as

$$\sum_a{}' (-1)^{n_a} (n_a - 1)! \Delta_{ab} = 1, \quad \text{any } b \neq \underline{1}. \tag{3.28}$$

Equations (3.19a) are $B_N - 1$ equations for the $B_N - 1$ unknowns C_a, where B_N is the total number of possible partitions (Bell number). Since $\det \Delta = 1$ on $\mathscr{L}$ it is clear that they determine the C_a uniquely. Since (3.28) shows that the form (A) satisfies all these equations, the result follows. Part (B) of the proposition follows directly from the explicit form for Δ^{-1}, Eq. (3.3b). $\blacksquare$

We now consider the analogous properties for the complementary matrix $\bar{\Delta}$ defined by (2.10), which in matrix form is

$$\bar{\Delta} = \mathscr{S} - \Delta. \tag{2.10}$$

The fundamental properties of Δ given by Proposition 1 imply corresponding results for $\bar{\Delta}$ simply by using (2.10), viz.,

Proposition 3.3:

$$\text{(A)} \quad \bar{\Delta} P_{\underline{0}} = 0, \quad \bar{\Delta}_{a\underline{0}} = 0, \tag{3.29a}$$

$$\text{(B)} \quad P_{\underline{0}} \bar{\Delta} = P_{\underline{0}} (\mathscr{S} - I), \quad \bar{\Delta}_{\underline{0}a} = \bar{\delta}_{\underline{0},a}, \tag{3.29b}$$

$$\text{(C)} \quad \bar{\Delta} P_{\underline{1}} = (\mathscr{S} - I) P_{\underline{1}}, \quad \bar{\Delta}_{a\underline{1}} = \bar{\delta}_{a,\underline{1}}, \tag{3.29c}$$

$$\text{(D)} \quad P_{\underline{1}} \bar{\Delta} = 0, \quad \bar{\Delta}_{\underline{1}a} = 0. \tag{3.29d}$$

Results (3.29a) and (3.29d) preclude the existence of an inverse for $\bar{\Delta}$ on the whole lattice since it has a row and column of zeroes. (See the examples in the Appendix.) We can construct an inverse, however, if we remove this row and column. We define the restriction of $\bar{\Delta}$ to a map of $\mathscr{L}^{*} \rightarrow \mathscr{L}'$ by

$$\hat{\Delta} \equiv Q_{\underline{1}} \bar{\Delta} Q_{\underline{0}}. \tag{3.30a}$$

The restriction

$$\tilde{\Delta} = Q_{\underline{0}\underline{1}} \bar{\Delta} Q_{\underline{0}\underline{1}}, \tag{3.30b}$$

considered as a map from $\mathscr{L}'^{*}$ to $\mathscr{L}'^{*}$ is also useful. The inverses of these matrices exist as we see below by explicit construction. As a consequence we have the equations

$$\hat{\Delta} \hat{\Delta}^{-1} = Q_{\underline{1}}, \quad \sum_c{}^{*} \bar{\Delta}_{ac} \hat{\Delta}_{cb}^{-1} = \delta_{ab} \bar{\delta}_{a\underline{1}}, \tag{3.31a}$$

$$\hat{\Delta}^{-1} \hat{\Delta} = Q_{\underline{0}}, \quad \sum_c{}' \hat{\Delta}_{ac}^{-1} \bar{\Delta}_{cb} = \delta_{ab} \bar{\delta}_{a\underline{0}}, \tag{3.31b}$$

$$\tilde{\Delta} \tilde{\Delta}^{-1} = \tilde{\Delta}^{-1} \tilde{\Delta}$$

$$= Q_{\underline{0}\underline{1}}, \quad \sum_c{}'^{*} \bar{\Delta}_{ac} \tilde{\Delta}_{cb}^{-1} = \delta_{ab} \bar{\delta}_{a\underline{1}} \quad (a \neq \underline{0}), \tag{3.31c}$$

$$\sum_c{}'^{*} \tilde{\Delta}_{ac}^{-1} \bar{\Delta}_{cb} = \delta_{ab} \bar{\delta}_{a\underline{0}} \quad (a \neq \underline{1}). \tag{3.31d}$$

Note that the left-hand sum in Eq. (3.31c) does *not* vanish when $a = \underline{0}$, nor that of (3.31d) when $a = \underline{1}$. This is discussed in conjuction with Theorem 2 of Sec. IV

Theorem 3.3:

(A) The inverse $\hat{\Delta}^{-1}$ mapping $\mathscr{L}' \rightarrow \mathscr{L}^{*}$ is given by the restriction of the matrix

$$X = \Delta^{-1} (\mathscr{S} P_{\underline{0}} - Q_{\underline{1}}) = -Q_{\underline{0}} \Delta^{-1} Q_{\underline{1}} \tag{3.32a}$$

to the appropriate domain and range. This implies the matrix elements

$$(\hat{\Delta}^{-1})_{ab} = -\Delta_{ab}^{-1}, \quad a \neq \underline{0}, \ b \neq \underline{1}. \tag{3.32b}$$

(B) The inverse $\tilde{\Delta}^{-1}$ mapping $\mathscr{L}'^{*} \rightarrow \mathscr{L}'^{*}$ is given by the restriction of the matrix

$$Y = Q_{\underline{0}\underline{1}} \Delta^{-1} Q_{\underline{0}\underline{1}} (\mathscr{S} \bar{C} - I) \tag{3.33a}$$

to the space $\mathscr{L}'^{*}$, where the matrix $\bar{C}$ is given by $\bar{C}_{ab} = \delta_{ab} (-C_a / C_{\underline{0}})$. The matrix elements of $\tilde{\Delta}^{-1}$ are given by

$$(\tilde{\Delta}^{-1})_{ab} = -\Delta_{ab}^{-1} - D_a C_b / C_{\underline{0}}, \quad a, b \neq \underline{1}, \underline{0}. \tag{3.33b}$$

Proof: We begin by proving part (A). Let X be the imbedding of $\hat{\Delta}^{-1}$ in $\mathscr{L}$, i.e., the matrix elements of X are the same as those of $\hat{\Delta}^{-1}$ when the indices coincide and X has its last row and first column equal to zero. We require

$$XQ_{\underline{1}} = X, \tag{3.34a}$$

$$Q_{\underline{0}}X = X. \tag{3.34b}$$

Since

$$Q_{\underline{1}}\bar{\Delta} = \bar{\Delta}, \tag{3.35a}$$

$$\bar{\Delta}Q_{\underline{0}} = \bar{\Delta}, \tag{3.35b}$$

the matrix X satisfies

$$X\bar{\Delta} = Q_{\underline{0}}, \tag{3.36a}$$

$$\bar{\Delta}X = Q_{\underline{1}}. \tag{3.36b}$$

If we multiply (3.36a) on the right by $\Delta^{-1}Q_{\underline{1}}$ we then obtain, using (2.10),

$$XQ_{\underline{1}}(\mathcal{S}\Delta^{-1} - 1)Q_{\underline{1}} = Q_{\underline{0}}\Delta^{-1}Q_{\underline{1}}. \tag{3.37}$$

Since $\mathcal{S}P_{\underline{1}}\mathcal{S} = \mathcal{S}$ we see that (3.10b) implies the sum rule

$$\mathcal{S}\Delta^{-1} = \mathcal{S}P_{\underline{1}}, \tag{3.38}$$

which when combined with (3.37) yields (3.32). The same result follows from (3.36b) and the sum rule

$$\Delta^{-1}\mathcal{S} = P_{\underline{0}}\mathcal{S} \tag{3.39}$$

which is a consequence of (3.10a) and the identity $\mathcal{S}P_{\underline{0}}\mathcal{S} = \mathcal{S}$. We note from (3.38) and [3.39] that Δ^{-1} is left (right) $\mathcal{S}$-equivalent to $P_{\underline{1}}$ ($P_{\underline{0}}$).

We now prove the result (B). Equation (3.24a) implies that

$$\mathcal{S}CQ_{\underline{1}}\mathcal{S} = \mathcal{S}, \tag{3.40}$$

and therefore, using (3.16a) and (2.10a) that

$$\mathcal{S}CQ_{\underline{1}}\bar{\Delta}Q_{\underline{1}} = 0, \tag{3.41}$$

where we have introduced the diagonal matrix C by $C_{ab} = \delta_{\underline{ab}}C_a$. If we put $I = P_{\underline{0}} + Q_{\underline{0}}$ on either side of the matrix $\bar{\Delta}$ in (3.41), Proposition 3.3 yields the result

$$\mathcal{S}C\tilde{\Delta} = -\mathcal{S}CQ_{\underline{1}}P_{\underline{0}}(\mathcal{S} - I)Q_{\underline{1}} = -C_{\underline{0}}\mathcal{S}Q_{\underline{01}}.$$

We therefore have

$$\mathcal{S}\bar{C}\tilde{\Delta} = (\bar{\Delta} + \Delta)Q_{\underline{01}}$$

and

$$Q_{\underline{01}}(\mathcal{S}\bar{C} - I)\tilde{\Delta} = Q_{\underline{01}}\Delta Q_{\underline{01}}. \tag{3.42}$$

Since Δ^{-1} exists on $\mathcal{L}^{*\prime}$, we conclude that

$$Q_{\underline{01}}\Delta Q_{\underline{01}}\Delta^{-1}Q_{\underline{01}} = Q_{\underline{01}}\Delta^{-1}Q_{\underline{01}}\Delta Q_{\underline{01}} = Q_{\underline{01}}. \tag{3.43}$$

After multiplying (3.40) by Δ^{-1} we obtain

$$Y\bar{\Delta} = Q_{\underline{01}}.$$

Similar manipulations that begin with

$$\mathcal{S}DQ_{\underline{0}}\mathcal{S} = \mathcal{S} \tag{3.44}$$

and

$$Q_{\underline{0}}\bar{\Delta}DQ_{\underline{0}}\mathcal{S} = 0 \tag{3.45}$$

yield (B). (D is the diagonal matrix with elements D_a.)■

Sum rules may be obtained for $\hat{\Delta}^{-1}$ in a manner analogous to the ones obtained for Δ^{-1}, i.e., by beginning with the analogous parts of Proposition 3.3 and carrying out similar manipulations. They may also be obtained directly by observing from (3.32b) that the matrices are, in fact, nearly identical. We easily find

Lemma 3.2:

$$\text{(A) } \hat{\Delta}^{-1}\mathcal{S}P_{\underline{1}} = P_{\underline{1}}, \quad \text{or } \sum_{\underline{b}}{}'(\hat{\Delta}^{-1})_{ab} = \delta_{a\underline{1}}, \tag{3.46a}$$

$$\text{(B) } P_0 Q_{\underline{1}}\mathcal{S}\hat{\Delta}^{-1} = P_{\underline{0}}, \quad \text{or } \sum_{a}{}^{*}(\hat{\Delta}^{-1})_{ab} = \delta_{\underline{0}b}. \tag{3.46b}$$

Proof: To obtain (A) we multiply (3.29c) on the left by Q_1, on the right by $Q_{\underline{0}}$, and observe that P_0 and Q_1 commute. This transforms the $\bar{\Delta}$ into a $\hat{\Delta}$ giving

$$\hat{\Delta}P_{\underline{1}} = Q_{\underline{1}}\mathcal{S}P_{\underline{1}}.$$

If we now multiply by $\hat{\Delta}^{-1}$ on the left and use (3.31b) we obtain (3.46a). Part (B) is proved in an analogous manner beginning with (3.29b). Note that the results (3.46) are dual to (interchange $\underline{0}$ and $\underline{1}$) the sum rule for Δ^{-1}, Eqs. (3.13) although the summations in (3.46) are restricted.■

Observe that the relation of $\hat{\Delta}^{-1}$ to Δ^{-1} implies the following important results:

$$(\hat{\Delta}^{-1})_{a\underline{0}} = D_a, \quad a \neq \underline{0}, \tag{3.47a}$$

$$(\hat{\Delta}^{-1})_{\underline{1}b} = C_b, \quad b \neq \underline{1}, \tag{3.47b}$$

and therefore

$$\sum_b{}'^{*}\hat{\Delta}^{-1}_{ab} = -D_a + \delta_{a\underline{1}} \tag{3.48a}$$

and

$$\sum_a{}'^{*}\hat{\Delta}^{-1}_{ab} = -C_b + \delta_{\underline{0}b} \tag{3.48b}$$

follow immediately from the sum rules of Lemma 3.2. We also obtain

Theorem 3.4:

$$\sum_a{}' C_a\bar{\Delta}_{ab} = \delta_{\underline{1}b}, \quad b \neq \underline{0}, \tag{3.49a}$$

$$\sum_b{}' \bar{\Delta}_{ab}D_b = \delta_{\underline{0}a}, \quad a \neq \underline{1}. \tag{3.49b}$$

Proof: These results follow in a straightforward manner from Eqs. (3.40) and (3.41) and Eqs. (3.44) and (3.45), respectively.■

Sum rules for $\tilde{\Delta}^{-1}$ may also be obtained, though in a somewhat indirect fashion. Proposition 3.3 affords us little aid since in all of the results the $\bar{\Delta}$ has either a $P_{\underline{0}}$ or a $P_{\underline{1}}$ next to it. If we attempted to multiply these equations by $Q_{\underline{01}}$ on the left and right to change $\bar{\Delta}$ into $\tilde{\Delta}$ we could only obtain $0 = 0$. The explicit form given in (3.30b) may be used to obtain the sum rules, or the sum rules for Δ^{-1} may be transformed using (2.10). The results are

Lemma 3.3:

$$\sum_b{}'^{*}(\tilde{\Delta}^{-1})_{ab} = -D_a/C_{\underline{0}}, \quad a \neq \underline{0},\underline{1}, \tag{3.50a}$$

$$\sum_a{}'^{*}(\tilde{\Delta}^{-1})_{ab} = -C_b/C_{\underline{0}}, \quad b \neq \underline{0},\underline{1}. \tag{3.50b}$$

Proof: These are easily obtained from the explicit form of $\tilde{\Delta}^{-1}$ given by Eq. (3.33), the sum rules on Δ^{-1}, Eqs. (3.35), and the sum rules for C and D, Eqs. (3.19). An important intermediate step is the observation that if $a,b \in \mathcal{L}'^{*}$, then

$$\sum_a{}'^{*}\Delta^{-1}_{ab} = C_b, \quad b \neq \underline{0},\underline{1}, \tag{3.51a}$$

$$\sum_b{}'^{*} \Delta^{-1}_{ab} = D_a, \quad a \neq \underline{0},\underline{1}. \tag{3.51b}$$

These follow directly from (3.15).■

We conclude this section with a discussion of the product relations for Δ and $\bar{\Delta}$. These relations control the connectivity structure in products of operators of the form (2.11) and therefore are among the most important and powerful results in the lattice theory arsenal.

The basic results are that $a \supseteq b$ and $a \supseteq c$ iff $a \supseteq b \cup c$; and that $a \supseteq c$ and $b \supseteq c$ iff $a \cap b \supseteq c$. Expressed in terms of Δ and $\bar{\Delta}$ these take the form

Proposition 3.4:

$$\text{(A)} \quad \Delta_{ab}\Delta_{ac} = \Delta_{a,b\cup c}, \tag{3.52a}$$

$$\text{(B)} \quad \Delta_{ac}\Delta_{bc} = \Delta_{a\cap b,c}, \tag{3.52b}$$

$$\text{(C)} \quad \Delta_{ab}\bar{\Delta}_{ac} = \Delta_{ab}\bar{\Delta}_{a,b\cup c}, \tag{3.52c}$$

$$\text{(D)} \quad \Delta_{ac}\bar{\Delta}_{bc} = \Delta_{ac}\bar{\Delta}_{a\cap b,c}. \tag{3.52d}$$

These may be generalized to an arbitrary number of products as follows.

Corollary: If a, b, a_i and b_i, $i = 1 \cdots n$, are arbitrary partitions, then

$$\prod_{i=1}^{n} \Delta_{ab_i} = \Delta_{a,\bigcup_{i=1}^{n} b_i}, \tag{3.53a}$$

$$\prod_{i=1}^{n} \Delta_{a_i b} = \Delta_{\bigcap_{i=1}^{n} a_i,b}. \tag{3.53b}$$

Proof is by induction using Proposition 3.4.

Proposition 3.4 may be used to obtain sum rules similar to (3.19):

Theorem 3.5:

$$\text{(A)} \quad \sum_c{}' C_c \Delta_{a\cap c,b} = \Delta_{ab}\bar{\delta}_{\underline{1}b}, \tag{3.54a}$$

$$\text{(B)} \quad \sum_c{}^{*} \Delta_{a,b\cup c} D_c = \bar{\delta}_{a\underline{0}}\Delta_{ab}, \tag{3.54b}$$

$$\text{(C)} \quad \Delta_{ab}\sum_c{}' C_c \bar{\Delta}_{a\cap c,b} = \delta_{a\underline{1}}\delta_{b\underline{1}}, \tag{3.54c}$$

$$\text{(D)} \quad \Delta_{ab}\sum_c{}^{*} \bar{\Delta}_{a,b\cup c} D_c = \delta_{a\underline{0}}\delta_{b\underline{0}}. \tag{3.54d}$$

Proof: We obtain (A) by multiplying (3.52b) by Δ^{-1}_{da} summing over all $a \in \mathscr{L}$, and setting $d = \underline{1}$. Equation (3.14a) then allows us to express the result as

$$\sum_c C_c \Delta_{a\cap c,b} = -\Delta_{ab}\delta_{\underline{1}b}. \tag{3.55a}$$

The form given in the statement of the theorem is obtained by explicating the term $c = \underline{1}$. The result (B) is obtained similarly but with the use of (3.14b). The unrestricted form for this case is

$$\sum_c \Delta_{a,b\cup c} D_c = -\delta_{a\underline{0}}\Delta_{ab}. \tag{3.55b}$$

The last two results are proved in a similar fashion, *mutatis mutandis*. ■

Results for arbitrary numbers of unions or intersections may be obtained by repeated applications of parts (A) and (B) of the theorem.

Corollary:

$$\text{(A)} \quad \sum_{a_1 \cdots a_n}{}' C_{a_1} C_{a_2} \cdots C_{a_n} \Delta_{a_1\cap a_2 \cdots a_n,b} = \bar{\delta}_{\underline{1}b}, \tag{3.56a}$$

$$\text{(B)} \quad \sum_{a_2 \cdots a_n}{}' C_{a_2} \cdots C_{a_n} \Delta_{a_1\cap a_2 \cdots \cap a_n,b} = \Delta_{a_1,b}\bar{\delta}_{\underline{1}b}, \tag{3.56b}$$

$$\text{(C)} \quad \sum_{b_1 \cdots b_n}{}' \Delta_{a,b_1\cup b_2 \cdots \cup b_n} D_{b_1} D_{b_2} \cdots D_{b_n} = \bar{\delta}_{a\underline{0}}, \tag{3.56c}$$

$$\text{(D)} \quad \sum_{b_2 \cdots b_n}{}' \Delta_{a,b_1\cup b_2 \cdots \cup b_n} D_{b_2} \cdots D_{b_n} = \bar{\delta}_{a\underline{0}}\Delta_{ab_1}. \tag{3.56d}$$

Among the lattice theory results useful in scattering theory are the combinations of Δ and $\bar{\Delta}$ traced with a C or D matrix. These permit simple analysis of connectivity structures by means of the following (see Theorem 4.4)

Theorem 3.6:

$$\text{(A)} \quad \sum_c{}' C_c \bar{\Delta}_{ca}\Delta_{cb} = \delta_{\underline{1},a\cup b}\bar{\delta}_{\underline{1}b} \quad \text{or} \quad \bar{\Delta}^{t} C Q_{\underline{1}} \Delta = U_{\underline{1}} Q_{\underline{1}}, \tag{3.57a}$$

$$\text{(B)} \quad \sum_c{}^{*} \Delta_{ac}\bar{\Delta}_{bc} D_c = \bar{\delta}_{a\underline{0}}\delta_{a\cap b,\underline{0}} \quad \text{or} \quad \Delta D Q_{\underline{0}} \bar{\Delta}^{t} = Q_{\underline{0}} U_{\underline{0}}, \tag{3.57b}$$

where we have defined the matrices $(U_{\underline{1}})_{ab} = \delta_{\underline{1},a\cup b}$ and $(U_{\underline{0}})_{ab} = \delta_{\underline{0},a\cap b}$. The superscript "$t$" means matrix transpose. Note also the transposed forms of the equations, viz,

$$(\text{A}^T) \quad \sum_c{}' C_c \Delta_{ca}\bar{\Delta}_{cb} = \bar{\delta}_{\underline{1}a}\delta_{\underline{1},a\cup b} \quad \text{or} \quad \Delta^{t} Q_{\underline{1}} C \bar{\Delta} = Q_{\underline{1}} U_{\underline{1}}, \tag{3.58a}$$

$$(\text{B}^T) \quad \sum_c{}^{*} \bar{\Delta}_{ac}\Delta_{bc} D_c = \delta_{a\cap b,\underline{0}}\bar{\delta}_{b\underline{0}} \quad \text{or} \quad \bar{\Delta} Q_{\underline{0}} D \Delta^{t} = U_{\underline{0}} Q_{\underline{0}}. \tag{3.58b}$$

Proof: To prove (A) consider the expression on the left-hand side. Equation (2.10) allows us to transform it into

$$\sum_c{}' C_c \bar{\Delta}_{ca}\Delta_{cb} = \sum_c{}' C_c (1 - \Delta_{ca})\Delta_{cb}.$$

Equations (3.52b) and (3.19a) then reduce the right-hand side of the preceding equations to

$$\bar{\delta}_{\underline{1}b} - \bar{\delta}_{\underline{1},a\cup b} = \delta_{\underline{1},a\cup b} - \delta_{\underline{1}b}.$$

If $b = \underline{1}$ this vanishes. Otherwise, we get $\delta_{\underline{1},a\cup b}$. Part (B) is proved analogously.■

The restriction of the above theorems to the sets of partitions with $n_a = 2$ or $N-1$ is useful for applications to scattering theory. We observe that these subsets of the lattice are distinguished. We define the sets[54]

$$\mathscr{L}_n = \{a \in \mathscr{L} \mid n_a = n\} \tag{3.59}$$

and the associated matrix projectors

$$(P_n)_{ab} = \delta_{ab}\delta_{nn_a},$$

$$Q_n = I - P_n. \tag{3.60}$$

The sets $\mathscr{L}_{N-1}$ and $\mathscr{L}_2$ are distinguished since $\mathscr{L}$ satisfies the condition that every element of $\mathscr{L}$ can be generated as finite unions of elements of $\mathscr{L}_{N-1}$ or as finite intersections of elements of $\mathscr{L}_2$. We use the notation a_m to indicate $a \in \mathscr{L}_m$. The useful results are

Corollary:

(A) $(\Delta\,'Q_1 C\bar{\Delta})P_{N-1} = P_2\bar{\Delta}P_{N-1}$

$$\text{or } \sum_c{}' C_c \Delta_{ca}\bar{\Delta}_{cb_{N-1}} = \delta_{2,n_a}\bar{\Delta}_{ab_{N-1}}, \quad (3.61a)$$

(B) $P_2(\Delta\,'Q_1 C\bar{\Delta}) = P_2\bar{\Delta}$

$$\text{or } \sum_c{}' C_c \Delta_{ca_2}\bar{\Delta}_{cb} = \bar{\Delta}_{a_2,b}, \quad (3.61b)$$

(C) $(\bar{\Delta}Q_{\underline{0}}D\Delta\,')P_{N-1} = \bar{\Delta}P_{N-1}$

$$\text{or } \sum_c{}^{*}\bar{\Delta}_{ac}\Delta_{b_{N-1}c}D_c = \bar{\Delta}_{ab_{N-1}}, \quad (3.61c)$$

(D) $P_2(\bar{\Delta}Q_{\underline{0}}D\Delta\,') = P_2\bar{\Delta}P_{N-1}$

$$\text{or } \sum_c{}^{*}\bar{\Delta}_{a_2c}\Delta_{bc}D_c = \bar{\Delta}_{a_2b}\delta_{N-1,n_b}. \quad (3.61d)$$

These results follow immediately from the theorem and the following:

Lemma:

(A) $Q_1 U_{\underline{1}} P_{N-1} = P_2\bar{\Delta}P_{N-1}$.

(If $n_b = N-1$ and $a \neq \underline{1}$, then $a \cup b = \underline{1}$ iff $n_a = 2$ and $a \not\supseteq b$.) (3.62a)

(B) $P_2 U_{\underline{1}} = P_2\bar{\Delta}$.

(If $n_a = 2$, then $a \cup b = \underline{1}$ iff $a \not\supseteq b$.) (3.62b)

(C) $P_2 U_{\underline{0}} Q_{\underline{0}} = P_2\bar{\Delta}P_{N-1}$.

(If $n_a = 2$ and $b \neq \underline{0}$, then $a \cap b = \underline{0}$ iff $n_b = N-1$ and $a \not\supseteq b$.) (3.62c)

(D) $U_{\underline{0}} P_{N-1} = \bar{\Delta}P_{N-1}$.

(If $n_b = N-1$, then $a \cap b = \underline{0}$ iff $a \not\supseteq b$.) (3.62d)

IV. OPERATOR RELATIONS

In this section we consider the application of the lattice theoretic results of the previous section to the case of operators with cluster expansions. Some of the relevant terms have already beeen introduced in Sec. II. The duality structure discussed in the previous chapter suggests the introduction of some additional quantities.

We define five labelling maps of $\mathscr{L} \times \mathscr{C} \to \mathscr{C}$. The connectivity map C: $(a,A) \mapsto [A]_a$, the interior map $M_{\rm I}$: $(a,A) \mapsto A_a$, and the exterior map $M_{\rm E}: (a,A) \mapsto A^a$ have already been defined. We here introduce the containment and exclusion maps, $N_{\rm C}: (a,A) \mapsto {}_aA$ and $N_{\rm E}: (a,A) \mapsto {}^aA$. The *a-containing part of* A, ${}_aA$, is defined by

$${}_aA = \sum_b [A]_b \Delta_{ba}. \quad (4.1)$$

It is the sum of those parts of A with connectivities containing the partition a. The *a-excluding part of* A is the sum of those parts of A with connectivities which do not contain the partition a, viz.,

$${}^aA = \sum_b [A]_b \bar{\Delta}_{ba}. \quad (4.2)$$

From the definitions (4.1), (4.2), (2.7), and (2.8) it is clear that for any operator which possesses a cluster expansion

$$A = {}_aA + {}^aA = A_a + A^a. \quad (4.3)$$

Containment and exclusion labelled operators are not unknown in the scattering literature. For example, in a three-body problem modelled after deuteron-nucleus scattering Dodd and Greider introduced the operators[55] (we use the three-body partition labels given in the Appendix)

$${}^aH = H_0 + V_b + V_c, \quad {}_aH = V_a.$$

If particle 1 is taken to be the heavy nucleus, then ${}^aH$ repesents the Hamiltonian which describes the interaction of the two particles of the deuteron with the nuclear field, if we ignore the interaction between them.

Additional useful definitions concern the extraction of parts of operators having particular classes of connectivity. We define the *disconnected part* of an operator as the operator less its completely connected part:

$$[A]^D = A - [A]_{\underline{1}}. \quad (44)$$

The *partly-connected part* of an operator is the operator less its completely disconnected part:

$$[A]^C \equiv A - [A]_{\underline{0}}. \quad (45)$$

The *part of an operator of connectivity m*, where m is an integer, is the sum of all the parts of an operator having the connectivity of an m-cluster partition:

$$[A]_{(m)} \equiv \sum_b \delta_{n_b m}[A]_b. \quad (46)$$

We say that an operator has *homogeneous connectivity n* if it consists only of parts with connectivity n. In this case we have

$$A = [A]_{(n)}. \quad (47)$$

One of the most interesting sets of results which can be obtained from the lattice properties of Sec. III are the cluster/cumulant expansions and their generalizations. To each operator $A \in \mathscr{C}$ we have associated five vectors of operators labelled by $\mathscr{L}$: $[A]_a$, A_a, A^a, ${}_aA$, and ${}^aA$. As we shall see below, knowledge of any of these sets suffices to give the full operator. We refer to these as *decomposition sets*. If we could always calculate all the matrix elements of all our operators these decompositions would not be of interest. In many cases, however, the full operator is of such complexity that there is no hope of performing an exact calculation. The decompositions then become exceedingly useful for the purpose of bringing various properties to the fore to facilitat various approximation strategies.

Relations between the decomposition sets can easily be obtained using the inverses constructed in the previous section. The classical cluster/cumulant expansions are closey related to Eq. (2.7) and its inversion given in (4.8a) below. We describe the relations in the next theorem as *generalized cluster/cumulant inversions*. These results follow directly from the properties of the inverses of Δ and $\bar{\Delta}$, Theorems 3.1 and 3.3.

Theorem 4.1:

(A) $$[A]_a = \sum_b{}' \Delta_{ab}^{-1}{}' A_b, \qquad a \neq \underline{1}, \quad (4.4a)$$

(B) $$[A]_a = -\sum_b{}' \Delta_{ab}^{-1}{}' A^b, \quad a \neq \underline{0}, \quad (4.4b)$$

(C) $$[A]_b = \sum_a{}^{*}\, {}_aA\Delta_{ab}^{-1}{}', \qquad b \neq \underline{0}, \quad (4.4c)$$

(D) $[A]_b = -\sum_a{}^{*a} A\Delta^{-1}_{ab}, \quad b \neq \underline{1}.$ (4.8d)

Proof: If we apply Δ^{-1} to (2.7) we get

$$\sum_c \Delta^{-1}_{ac} A_c = [A]_a .$$

The segregation of the term $c = \underline{1}$ from the sum yields

$$[A]_a = \sum_c{}' \Delta^{-1}_{ac} A_c + \delta_{a\underline{1}} A. \tag{4.9}$$

Result (A) follows if we restrict a to be different from $\underline{1}$. To get (B) we must be a bit more careful because of the restriction of the range of $\hat{\Delta}^{-1}$. If we multiply (2.8) by $\hat{\Delta}^{-1}$, we find using Eq. (3.31b) that

$$\sum_c{}' \hat{\Delta}^{-1}_{ac} A^c = \sum_c{}' \sum_b \hat{\Delta}^{-1}_{ac} \bar{\Delta}_{cb} [A]_b = \bar{\delta}_{a\underline{0}} [A]_a .$$

By restricting a to be different from $\underline{0}$ and applying the specific form of $\hat{\Delta}^{-1}$ given in (3.32b) we obtain the result (B). The final two results follow from similar manipulations if we begin with (4.1) and (4.2) and operate on the right with the appropriate inverse matrices.■

These results, together with the cluster expansion (2.1′), allow us to obtain expressions for the full operator in terms of the decomposition sets. These require the sum rules of Theorems 3.2 and 3.4 and the associated lemmas and corollaries. In order to avoid trivialities, we exclude the terms

$$\begin{aligned}
&A_{\underline{1}} = A, && A^{\underline{1}} = 0,\\
&A^{\underline{0}} = A - [A]_{\underline{0}}, && A_{\underline{0}} = [A]_{\underline{0}},\\
&{}_{\underline{0}}A = A, && {}^{\underline{0}}A = 0,\\
&{}^{\underline{1}}A = A - [A]_{\underline{1}}, && {}_{\underline{1}}A = [A]_{\underline{1}},
\end{aligned} \tag{4.10}$$

from the sums. The completely connected and completely diconnected terms are distinguished in some of the relations as a consequence.

Theorem 4.2:

$$A = \sum_b{}' C_b A_b + [A]_{\underline{1}}, \tag{4.11a}$$

$$A = -\frac{1}{C_{\underline{0}}} \sum_b{}^{*} C_b A^b + [A]_{\underline{0}} + \frac{1}{C_0}[A]_{\underline{1}}, \tag{4.11b}$$

$$A = \sum_a{}^{*}{}_a A D_a + [A]_{\underline{0}}, \tag{4.11c}$$

$$A = -\frac{1}{C_{\underline{0}}} \sum_a{}'^{a} A D_a + \frac{1}{C_{\underline{0}}}[A]_{\underline{0}} + [A]_{\underline{1}}. \tag{4.11d}$$

Proof: To prove the first equation we sum (4.8a) on all $a \neq \underline{1}$. Equation (3.12a) leads immediately to (4.11a). Equation (4.11c) is obtained in an analogous way. Equations (4.11b) and (4.11d) may be obtained using Theorem 4.1 and (3.15) or by the following

Lemma:

$$\sum_c{}'^{*} \tilde{\Delta}^{-1}_{ac} \bar{\Delta}_{c\underline{1}} = -D_a / C_{\underline{0}}, \quad a \neq \underline{0}, \underline{1}, \tag{4.12a}$$

$$\sum_c{}'^{*} \bar{\Delta}_{\underline{0}c} \tilde{\Delta}^{-1}_{cb} = -C_b / C_{\underline{0}}, \quad b \neq \underline{0}, \underline{1}. \tag{4.12b}$$

These provide the extensions of (3.28c) and (3.28d) mentioned in the previous section. They are easily proved using $\bar{\Delta}_{c\underline{1}} = \bar{\delta}_{c\underline{1}}$ and $\bar{\Delta}_{\underline{0}c} = \bar{\delta}_{\underline{0}c}$ together with the sum rules (3.50). If we multiply (2.8) by $\tilde{\Delta}^{-1}$ and sum on $\mathscr{L}'^{*}$, the result [using (3.31d) and (4.12a)] is

$$\sum_c{}'^{*} \tilde{\Delta}^{-1}_{ac} A^c = [A]_a \bar{\delta}_{a\underline{0}} - \frac{D_a}{C_{\underline{0}}} [A]_{\underline{1}} .$$

We now sum a over $\mathscr{L}'^{*}$ and use (3.12c), (2.1′), and (3.21c) to obtain the desired result. Equation (4.11d) is obtained by the dual argument.■

For some applications involving connectivity it is the completely connected part of the operator which is of interest. The results of Theorem 4.2 may then be more useful in the following form:

Corollary:

$$[A]^D = \sum_b{}' C_b A_b, \tag{4.13a}$$

$$[A]_{\underline{1}} = \sum_b{}' C_b A^b, \tag{4.13b}$$

$$[A]^C = \sum_a{}^{*}{}_a A D_a, \tag{4.13c}$$

$$[A]_{\underline{0}} = \sum_a{}^{*a} A D_a . \tag{4.13d}$$

The proof is omitted.

We now turn to the class of results which follow from the product rules of Proposition 3.4 and Theorem 3.5. We refer to these as *union and intersection properties*. Proposition 3.4 immediately implies the operator properties

$$A_{ab} = A_{a\cap b}, \tag{4.14a}$$

$$A^{\,b}_{a} = A^{\,a\cap b}_{a}, \tag{4.14b}$$

$$_{ab}A = {}_{a\cup b}A, \tag{4.14c}$$

$$^{b}_{a}A = {}^{a\cup b}_{a}A. \tag{4.14d}$$

A double index (left or right) means that both conditions are applied [using the obvious generalization of Eq. (2.11)]. With these, we may immediately deduce a number of useful relations

Theorem 4.3:

(A) $\sum_b{}' C_b A_{a\cap b} = A_a, \quad a \neq \underline{1},$ (4.15a)

(B) $\sum_b{}' C_b A^{\,a\cap b}_{a} = 0, \quad a \neq \underline{1},$ (4.15b)

(C) $\sum_a{}^{*}{}_{a\cup b} A D_a = {}_b A, \quad b \neq \underline{0},$ (4.15c)

(D) $\sum_a{}^{*a\cup b}_{\;\;\,b} A D_a = 0, \quad b \neq \underline{0}.$ (4.15d)

Proof: These proofs may be obtained either by using Theorem 3.5 or directly from Theorem 4.2 and its corollary. To see how this works, we consider case (A). Apply Eq. (4.13a) to the operator A_a with $a \neq \underline{1}$. Since A_a has no completely connected part the result is

$$A_a = \sum_b{}' C_b A_{ab} .$$

Equation (4.14a) immediately leads to the desired result. Similar manipulations yield (B)–(D). Note that the corresponding results for $a = \underline{1}$ and $b = \underline{0}$ are simply the results of the corollary, Eqs. (4.13), since $\underline{1} \cap b = b$ and $\underline{0} \cup b = b$.■

The extension of these results to an arbitrary number of intersections and unions is easily obtained by repeated applications of the theorem or by using Eqs. (3.56), e.g.,

$$\sum_{a_1 \cdots a_n} {}' C_{a_1} \cdots C_{a_n} A_{a_1 \cap a_2 \cdots \cap a_n} = [A\,]^D, \tag{4.16a}$$

$$\sum_{a_2 \cdots a_n} {}' C_{a_2} \cdots C_{a_n} A_{a \cap a_2 \cdots \cap a_n} = A_a, \quad a \neq \underline{1}, \tag{4.16b}$$

$$\sum_{b_1 \cdots b_n} {}^{*}{}_{b_1 \cup b_2 \cdots \cup b_n} A D_{b_1} D_{b_2} \cdots D_{b_n} = [A\,]^C, \tag{4.16c}$$

$$\sum_{b_2 \cdots b_n} {}^{*}{}_{b \cup b_2 \cdots \cup b_n} A D_{b_2} \cdots D_{b_n} = {}_b A, \quad b \neq \underline{0}. \tag{4.16d}$$

An extremely valuable set of operator theorems follow from Theorem 3.6. These are the *connectivity theorems* which have been used to construct connected-kernel scattering theories.

Theorem 4.4:

$$\sum_c {}' C_c A_c B^c = [[A\,]^D B\,]_{\underline{1}}, \tag{4.17a}$$

$$\sum_c {}' C_c A^c B_c = [A\,[B\,]^D]_{\underline{1}}. \tag{4.17b}$$

This theorem allows one to combine operators describing the subsystem in order to obtain completely connected operators. These theorems are useful in scattering theory because of the upper–lower index structure that arises in the second resolvent relations.[1] The reason for the presence of the $[\;]^D$ associated with the lower indexed quantity is due to the restriction of the sum to $\mathscr{L}'$. The term $[A\,]_{\underline{1}}$ never appears in a sum or A_c over $c \in \mathscr{L}'$.

The dual results arising from Eq. (3.50) are of the form

$$\sum_c A^c B D_c = \sum_{ab} [A\,]_a [B\,]_b \bar{\delta}_{a\underline{0}} \delta_{a \cap b, \underline{0}}. \tag{4.18}$$

Since the operator product satisfies (2.2) rather than the corresponding statement with $a \cup b \to a \cap b$ this relation has no obvious utility.

A class of practical results is associated with Eq. (3.27), the sum Δ_{ab} over all a having a fixed number of clusters. Upon multiplying (3.27) by $[A\,]_b$ and summing over all b we get

Theorem 4.5:

$$\sum_a \delta_{n_a m} A_a = \sum_n S_n^m [A\,]_{(n)}. \tag{4.19}$$

This result is most useful if A is an operator of homogeneous connectivity we then have:

Corollary: If A is an operator of homogeneous connectivity n then

$$\sum_a \delta_{n_a m} A_a = S_n^m A. \tag{4.20}$$

This is useful in scattering theory since one frequently considers the case where the particles only interact via pair potentials. The potential V is then an operator of homogeneous connectivity $N-1$. Applying the corollary to the operators V and V^b (any b) we get the well-known results[1,2]

$$\sum_a \delta_{n_a m} V_a = S_{N-1}^m V, \tag{4.21a}$$

$$\sum_a \delta_{n_a m} V_a^b = S_{N-1}^m V^b. \tag{4.21b}$$

One may easily obtain the generalizations

$$\sum_a \delta_{n_a m} V_{a \cap b} = S_{N-1}^m V_b, \tag{4.21c}$$

$$\sum_a \delta_{n_a m} V_{a \cap b}^c = S_{N-1}^m V_b^c, \tag{4.21d}$$

and so on, by applying the corollary to the operators V_b and V_b^c and using (4.14a).

The dual form of the theorem is not useful since summing the right index of Δ over the set of partitions having a fixed number of clusters depends on the specific left partition index, not just on its number of clusters. Similar results can, however, be easily found for $\bar{\Delta}$. These rely on the

Lemma:

$$\sum_a \delta_{n_a m} \bar{\Delta}_{ab} = S_N^m - S_{n_b}^m. \tag{4.22}$$

This follows directly from (3.24) by writing $\bar{\Delta} = 1 - \Delta$. The associated theorem is

Theorem 4.6:

$$\sum_a \delta_{n_a m} A^a = \sum_n (S_N^m - S_n^m)[A\,]_{(n)}. \tag{4.23}$$

Corollary: If A is an operator of homogeneous connectivity n, then

$$\sum_a \delta_{n_a m} A^a = (S_N^m - S_n^m)[A\,]_{(n)}. \tag{4.24}$$

In the particular case of scattering theory with only pair potentials so V is of homogeneous connectivity $N-1$, one obtains

$$\sum_a \delta_{n_a m} V^a = (S_N^m - S_{N-1}^m)V, \tag{4.25a}$$

$$\sum_a \delta_{n_a m} V_b^a = (S_N^m - S_{N-1}^m)V_b, \tag{4.25b}$$

$$\sum_a \delta_{n_a m} V^{ab} = (S_N^m - S_{N-1}^m)V^b, \tag{4.25c}$$

by applying the corollary to V, V_b, and V^b, respectively.

Two set of interesting results can be obtained from the corollary to Theorem 3.6. The first set is generalization of a lemma given in Ref. 1.

Theorem 4.7:

$$\sum_c {}' \Delta_{ca} C_c [A^c]_{(N-1)} = \delta_{n_a 2} [A^a]_{(N-1)}, \tag{4.26a}$$

$$\sum_c {}^{*} \Delta_{ac} [{}^c A\,]_{(2)} D_c = \delta_{n_a N-1} [{}^a A\,]_{(2)}. \tag{4.26b}$$

Proof: The first result follows immediately if we multiply (3.61a) on the right by $[A\,]_b$ and sum over $b \in \mathscr{L}_{N-1}$. The second follows after multiplication of (3.61d) by $[A\,]_a$ and summation over $a \in \mathscr{L}_2$.■

For the case of an operator of homogeneous connectivity N-1, the first part of the theorem yields the result of Ref. 1.

$$\sum_c{}' \Delta_{ca} C_c V^c = \delta_{n_a,2} V^a. \tag{4.27}$$

If we had an operator K of homogeneous connectivity 2 (see, for example, the kernel operators discussed in Ref. 56), the second part of the theorem would yield

$$\sum_a{}^* \Delta_{ac}{}^c K D_c = \delta_{n_a, N-1}{}^a K. \tag{4.28}$$

Similar operations applied to the other free index in (3.61a) and (3.61d) lead to relations between the left and right labelled quantities. By performing manipulations analogous to those in the previous theorem we obtain

Theorem 4.8:

$$\sum_c{}' \bar{\Delta}_{cb} C_c A_c = [{}^{b_N}\ {}^1 A\ ]_{(2)}, \tag{4.29a}$$

$$\sum_c {}^* \bar{\Delta}_{a_2,c}\ {}_c A D_c = [A^{a_2}]_{(N-1)}. \tag{4.29b}$$

Equations (3.61b) and (3.61c) can be used in a similar manner but do not lead to interesting results.

V. SUMMARY AND CONCLUSIONS

In this paper we have considered the combinatoric operator relations which have recently been developed for N-particle scattering theory from an abstract point of view. The concept of connectivity is defined abstractly by the introduction of a connectivity structure as a labelling operation mapping each operator in an allowed set into a partition-labelled set of operators. This allows us to see that a number of distinct connectivity structures are possible, some of which are exhibited in Sec. II.

The properties of the partition lattice are then studied in terms of this abstract structure using the algebra of real-valued partition-labelled matrices (the incidence algebra). This is done, in particular, through the properties of the inclusion and exclusion matrices Δ and $\bar{\Delta}$. The inverse of Δ is the matrix of anticluster coefficients (the Möbius function of the lattice). Explicit expressions are given for Δ^{-1}. The exclusion matrix $\bar{\Delta}$ is singular, but inverses of restrictions may be found in two ways. Explicit expressions are found for both these inverses, $\hat{\Delta}^{-1}$ and $\tilde{\Delta}^{-1}$. The sums of the rows and columns of each of the three inverse matrices have specific values which we obtain for each case. The values in the boundary row and column of the anticluster matrix, C_a and D_a, are important numbers characterizing the partitions. Explicit forms are given for these numbers. They also satisfy summation conditions. Finally, we study product rules—the relations which specify the union and intersection rules in the lattice. These lead to a number of theorems which eventually lead to connectivity conditions.

These abstract lattice theory results are then applied to the construction of rules for partition-labelled operators. In addition to the standard partition-labelled interior and exterior operators, it appears natural to define sets of including and excluding operators. These operators have received only tentative use in scattering theory up till now. A wide variety of theorems are obtained including generalizations of the traditional cluster/cumulant expansions and rules for extracting the connected and disconnected parts of an operator from the various sets of partition-labelled operators. Sharp generalizations of previously known connectivity and distribution theorems are stated.

The structure of the partition lattice plays a deep and fundamental role in constraining the treatment of clustering in the N-body problem. Until recently the implications of the underlying presence of this structure have been almost totally ignored by mathematical physicists studying clustering. In this paper we have tried to demonstrate how this structure is manifested in the treatment of operators and the power of the lattice theory techiques.

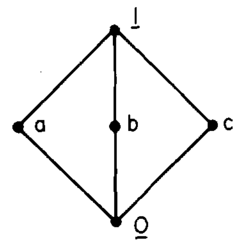


FIG. 2. The lattice for the partitions of three distinguishable objects.

ACKNOWLEDGMENTS

This work has been supported in part by the U. S. Department of Energy and in part by the National Science Foundation. We are grateful to the University of Maryland Computer Science Center for supporting the calculations displayed in the Appendix.

APPENDIX

In this appendix we display the partition lattices and the associated matrices for the case of three and four distinguishable objects.

For the case of three objects labelled 1, 2, 3, the possible partitions are

$$\underline{1} = (123), \quad \begin{array}{l} a = 1(23) \\ b = 2(13), \\ c = 3(12), \end{array} \quad \underline{0} = (1)(2)(3).$$

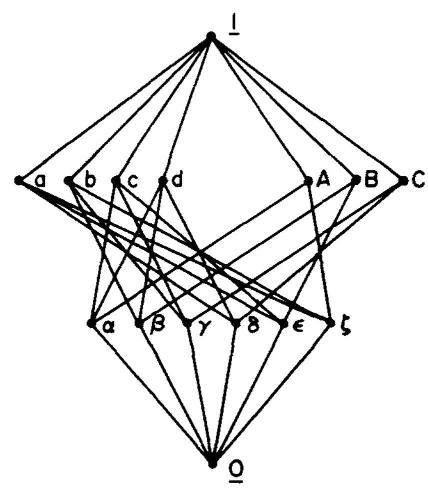


FIG. 3. The lattice for the partitions of four distinguishable objects.

The containment structure of the lattice is shown in Fig. 2. We take as the standard order $\underline{1} > a > b > c > \underline{0}$.

For the case of four objects labelled 1,2,3,4, the possible partitions are

$$
\begin{array}{llll}
\underline{1} = (123), & a = 1(234), & \alpha = (12)(3)(4), & \\
 & b = 2(134), & \beta = (13)(2)(4), & \\
 & c = 3(124), & \gamma = (14)(2)(3), & \\
 & d = 4(123), & \delta = (23)(1)(4), & \underline{0} = (1)(2)(3)(4) \\
 & A = (12)(34), & \epsilon = (24)(1)(3), & \\
 & B = (13)(24), & \zeta = (34)(1)(2), & \\
 & C = (14)(23), & &
\end{array}
$$

We take as standard order

$\underline{1} > a > b > c > d > A > B > C > \alpha > \beta > \gamma > \delta > \epsilon > \zeta > \underline{0}$.

The containment structure is shown in Fig. 3.

The interesting matrices for the case $N = 3$ are (labels in standard order)

$$
\Delta = \begin{pmatrix} 1 & 1 & 1 & 1 & 1 \\ 0 & 1 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 & 1 \\ 0 & 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 0 & 1 \end{pmatrix},
$$

$$
\Delta^{-1} = \begin{pmatrix} 1 & -1 & -1 & -1 & 2 \\ 0 & 1 & 0 & 0 & -1 \\ 0 & 0 & 1 & 0 & -1 \\ 0 & 0 & 0 & 1 & -1 \\ 0 & 0 & 0 & 0 & 1 \end{pmatrix},
$$

$$
\bar{\Delta} = \begin{pmatrix} 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 1 & 1 & 0 \\ 1 & 1 & 0 & 1 & 0 \\ 1 & 1 & 1 & 0 & 0 \\ 1 & 1 & 1 & 1 & 0 \end{pmatrix}, \quad \hat{\Delta} = \begin{pmatrix} 1 & 0 & 1 & 1 \\ 1 & 1 & 0 & 1 \\ 1 & 1 & 1 & 0 \\ 1 & 1 & 1 & 1 \end{pmatrix},
$$

$$
\hat{\Delta}^{-1} = \begin{pmatrix} 1 & 1 & 1 & -2 \\ -1 & 0 & 0 & 1 \\ 0 & -1 & 0 & 1 \\ 0 & 0 & -1 & 1 \end{pmatrix},
$$

$$
\tilde{\Delta} = \begin{pmatrix} 0 & 1 & 1 \\ 1 & 0 & 1 \\ 1 & 1 & 0 \end{pmatrix},
$$

$$
\tilde{\Delta}^{-1} = \tfrac{1}{2} \begin{pmatrix} -1 & 1 & 1 \\ 1 & -1 & 1 \\ 1 & 1 & -1 \end{pmatrix}.
$$

For the case $N = 4$, we block off the matrices by heavy lines between partitions with different numbers of clusters, and by dotted lines beween partitions with the same number of clusters but with clusters of different sizes. The relevant matrices are

$$
\Delta = \left(\begin{array}{c|cccc|ccc|cccccc|c}
1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ \hline
 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 1 & 1 \\
 & & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 0 & 0 & 1 & 1 \\
 & & & 1 & 0 & 0 & 0 & 0 & 1 & 0 & 1 & 0 & 1 & 0 & 1 \\
 & & & & 1 & 0 & 0 & 0 & 1 & 1 & 0 & 1 & 0 & 0 & 1 \\ \hline
 & & & & & 1 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 1 & 1 \\
 & & & & & & 1 & 0 & 0 & 1 & 0 & 0 & 1 & 0 & 1 \\
 & & & & & & & 1 & 0 & 0 & 1 & 1 & 0 & 0 & 1 \\ \hline
 & & & & & & & & 1 & 0 & 0 & 0 & 0 & 0 & 1 \\
 & & & & & & & & & 1 & 0 & 0 & 0 & 0 & 1 \\
 & & & & & & & & & & 1 & 0 & 0 & 0 & 1 \\
 & & & & & & & & & & & 1 & 0 & 0 & 1 \\
 & & & & & & & & & & & & 1 & 0 & 1 \\
 & & & & & & & & & & & & & 1 & 1 \\ \hline
 & & & & & & & & & & & & & & 1
\end{array}\right),
$$

(all elements below the diagonal vanish),

$$
\Delta^{-1} = \left(\begin{array}{c|cccc|ccc|cccccc|c}
1 & -1 & -1 & -1 & -1 & -1 & -1 & -1 & 2 & 2 & 2 & 2 & 2 & 2 & -6 \\
\hline
 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & -1 & -1 & -1 & 2 \\
 & & 1 & 0 & 0 & 0 & 0 & 0 & 0 & -1 & -1 & 0 & 0 & -1 & 2 \\
 & & & 1 & 0 & 0 & 0 & 0 & -1 & 0 & -1 & 0 & -1 & 0 & 2 \\
 & & & & 1 & 0 & 0 & 0 & -1 & -1 & 0 & -1 & 0 & 0 & 2 \\
\hline
 & & & & & 1 & 0 & 0 & -1 & 0 & 0 & 0 & 0 & -1 & 1 \\
 & & & & & & 1 & 0 & 0 & -1 & 0 & 0 & -1 & 0 & 1 \\
 & & & & & & & 1 & 0 & 0 & -1 & -1 & 0 & 0 & 1 \\
\hline
 & & & & & & & & 1 & 0 & 0 & 0 & 0 & 0 & -1 \\
 & & & & & & & & & 1 & 0 & 0 & 0 & 0 & -1 \\
 & & & & & & & & & & 1 & 0 & 0 & 0 & -1 \\
 & & & & & & & & & & & 1 & 0 & 0 & -1 \\
 & & & & & & & & & & & & 1 & 0 & -1 \\
 & & & & & & & & & & & & & 1 & -1 \\
\hline
 & & & & & & & & & & & & & & 1
\end{array}\right)
$$

(all elements below the diagonal vanish),

$$
\tilde{\Delta} = \left(\begin{array}{c|cccc|ccc|cccccc|c}
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
\hline
1 & 0 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 0 & 0 & 0 & 0 \\
1 & 1 & 0 & 1 & 1 & 1 & 1 & 1 & 1 & 0 & 0 & 1 & 1 & 0 & 0 \\
1 & 1 & 1 & 0 & 1 & 1 & 1 & 1 & 0 & 1 & 0 & 1 & 0 & 1 & 0 \\
1 & 1 & 1 & 1 & 0 & 1 & 1 & 1 & 0 & 0 & 1 & 0 & 1 & 1 & 0 \\
\hline
1 & 1 & 1 & 1 & 1 & 0 & 1 & 1 & 0 & 1 & 1 & 1 & 1 & 0 & 0 \\
1 & 1 & 1 & 1 & 1 & 1 & 0 & 1 & 1 & 0 & 1 & 1 & 0 & 1 & 0 \\
1 & 1 & 1 & 1 & 1 & 1 & 1 & 0 & 1 & 1 & 0 & 0 & 1 & 1 & 0 \\
\hline
1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 0 & 1 & 1 & 1 & 1 & 1 & 0 \\
1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 0 & 1 & 1 & 1 & 1 & 0 \\
1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 0 & 1 & 1 & 1 & 0 \\
1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 0 & 1 & 1 & 0 \\
1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 0 & 1 & 0 \\
1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 0 & 0 \\
\hline
1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 0
\end{array}\right),
$$

$$\hat{\Delta} = \left(\begin{array}{c|cccc|ccc|cccccc}
1 & 0 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 0 & 0 & 0 \\
1 & 1 & 0 & 1 & 1 & 1 & 1 & 1 & 1 & 0 & 0 & 1 & 1 & 0 \\
1 & 1 & 1 & 0 & 1 & 1 & 1 & 1 & 0 & 1 & 0 & 1 & 0 & 1 \\
1 & 1 & 1 & 1 & 0 & 1 & 1 & 1 & 0 & 0 & 1 & 0 & 1 & 1 \\
\hdashline
1 & 1 & 1 & 1 & 1 & 0 & 1 & 1 & 0 & 1 & 1 & 1 & 1 & 0 \\
1 & 1 & 1 & 1 & 1 & 1 & 0 & 1 & 1 & 0 & 1 & 1 & 0 & 1 \\
1 & 1 & 1 & 1 & 1 & 1 & 1 & 0 & 1 & 1 & 0 & 0 & 1 & 1 \\
\hline
1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 0 & 1 & 1 & 1 & 1 & 1 \\
1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 0 & 1 & 1 & 1 & 1 \\
1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 0 & 1 & 1 & 1 \\
1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 0 & 1 & 1 \\
1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 0 & 1 \\
1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 0 \\
\hline
1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1
\end{array}\right),$$

$$\tilde{\Delta} = \left(\begin{array}{cccc|ccc|cccccc}
0 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 0 & 0 & 0 \\
1 & 0 & 1 & 1 & 1 & 1 & 1 & 1 & 0 & 0 & 1 & 1 & 0 \\
1 & 1 & 0 & 1 & 1 & 1 & 1 & 0 & 1 & 0 & 1 & 0 & 1 \\
1 & 1 & 1 & 0 & 1 & 1 & 1 & 0 & 0 & 1 & 0 & 1 & 1 \\
\hdashline
1 & 1 & 1 & 1 & 0 & 1 & 1 & 0 & 1 & 1 & 1 & 1 & 0 \\
1 & 1 & 1 & 1 & 1 & 0 & 1 & 1 & 0 & 1 & 1 & 0 & 1 \\
1 & 1 & 1 & 1 & 1 & 1 & 0 & 1 & 1 & 0 & 0 & 1 & 1 \\
\hline
1 & 1 & 1 & 1 & 1 & 1 & 1 & 0 & 1 & 1 & 1 & 1 & 1 \\
1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 0 & 1 & 1 & 1 & 1 \\
1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 0 & 1 & 1 & 1 \\
1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 0 & 1 & 1 \\
1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 0 & 1 \\
1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 0
\end{array}\right),$$

$$
\hat{\Delta}^{-1} = \left(\begin{array}{cccc|ccc|cccccc|c}
1 & 1 & 1 & 1 & 1 & 1 & 1 & -2 & -2 & -2 & -2 & -2 & -2 & 6 \\
\hline
-1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 1 & -2 \\
 & -1 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 0 & 0 & 1 & -2 \\
 & & -1 & 0 & 0 & 0 & 0 & 1 & 0 & 1 & 0 & 1 & 0 & -2 \\
 & & & -1 & 0 & 0 & 0 & 1 & 1 & 0 & 1 & 0 & 0 & -2 \\
\hdashline
 & & & & -1 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 1 & -1 \\
 & & & & & -1 & 0 & 0 & 1 & 0 & 0 & 1 & 0 & -1 \\
 & & & & & & -1 & 0 & 0 & 1 & 1 & 0 & 0 & -1 \\
\hline
 & & & & & & & -1 & 0 & 0 & 0 & 0 & 0 & 1 \\
 & & & & & & & & -1 & 0 & 0 & 0 & 0 & 1 \\
 & & & & & & & & & -1 & 0 & 0 & 0 & 1 \\
 & & & & & & & & & & -1 & 0 & 0 & 1 \\
 & & & & & & & & & & & -1 & 0 & 1 \\
 & & & & & & & & & & & & -1 & 1
\end{array}\right),
$$

(all elements below the subdiagonal vanish)

$$
\tilde{\Delta}^{-1} = \left(\begin{array}{cccc|ccc|cccccc}
-2/3 & 1/3 & 1/3 & 1/3 & 1/3 & 1/3 & 1/3 & -2/3 & -2/3 & -2/3 & 1/3 & 1/3 & 1/3 \\
1/3 & -2/3 & 1/3 & 1/3 & 1/3 & 1/3 & 1/3 & -2/3 & 1/3 & 1/3 & -2/3 & -2/3 & 1/3 \\
1/3 & 1/3 & -2/3 & 1/3 & 1/3 & 1/3 & 1/3 & 1/3 & -2/3 & 1/3 & -2/3 & 1/3 & -2/3 \\
1/3 & 1/3 & 1/3 & -2/3 & 1/3 & 1/3 & 1/3 & 1/3 & 1/3 & -2/3 & 1/3 & -2/3 & -2/3 \\
\hdashline
1/6 & 1/6 & 1/6 & 1/6 & -5/6 & 1/6 & 1/6 & 2/3 & -1/3 & -1/3 & -1/3 & -1/3 & 2/3 \\
1/6 & 1/6 & 1/6 & 1/6 & 1/6 & -5/6 & 1/6 & -1/3 & 2/3 & -1/3 & -1/3 & 2/3 & -1/3 \\
1/6 & 1/6 & 1/6 & 1/6 & 1/6 & 1/6 & -5/6 & -1/3 & -1/3 & 2/3 & 2/3 & -1/3 & -1/3 \\
\hline
-1/6 & -1/6 & -1/6 & -1/6 & -1/6 & -1/6 & -1/6 & -2/3 & 1/3 & 1/3 & 1/3 & 1/3 & 1/3 \\
-1/6 & -1/6 & -1/6 & -1/6 & -1/6 & -1/6 & -1/6 & 1/3 & -2/3 & 1/3 & 1/3 & 1/3 & 1/3 \\
-1/6 & -1/6 & -1/6 & -1/6 & -1/6 & -1/6 & -1/6 & 1/3 & 1/3 & -2/3 & 1/3 & 1/3 & 1/3 \\
-1/6 & -1/6 & -1/6 & -1/6 & -1/6 & -1/6 & -1/6 & 1/3 & 1/3 & 1/3 & -2/3 & 1/3 & 1/3 \\
-1/6 & -1/6 & -1/6 & -1/6 & -1/6 & -1/6 & -1/6 & 1/3 & 1/3 & 1/3 & 1/3 & -2/3 & 1/3 \\
-1/6 & -1/6 & -1/6 & -1/6 & -1/6 & -1/6 & -1/6 & 1/3 & 1/3 & 1/3 & 1/3 & 1/3 & -2/3
\end{array}\right)
$$

[1]P. Benoist-Gueutal, M. L'Huillier, E. F. Redish, and P. C. Tandy, Phys. Rev. C **17**, 1924 (1978).
[2]K. L. Kowalski, Phys. Rev. C **16**, 7 (1977).
[3]V. Vanzani, in *Few-Body Physics*, edited by G. Pisent, V. Vanzani, and L. Fonda, Proceedings of Trieste Workshop, March, 1978 (IAEA, Austria, 1978), p. 57.
[4]W. Polyzou and E. F. Redish, Ann. Phys. (N.Y.) **119**, 1 (1979).
[5]K. L. Kowalski, in *Few Body Systems and Nuclear Forces II* (Springer-Verlag, Berlin, 1978), p. 393.
[6]K. L. Kowalski, Ann. Phys. (N.Y.) **120**, 328 (1979).
[7]L. Ludeking and J. Vary, Phys. Rev C **20**, 2188 (1979).
[8]W. Polyzou, Ph. D. Thesis, University of Maryland, 1979, unpublished.
[9]W. Polyzou, J. Math. Phys. **21**, 506 (1980).
[10]S. Weinberg, Phys. Rev. **133**, B 232 (1964).
[11]C. Van Winter, Mat. Fys. Skr. Danske Vid. Selsk. **2** no. 8 (1964).
[12]L. D. Faddeev, *Mathematical Aspects of the Three-Body Problem in the Quantum Scattering Theory*, Trudy Mat. Inst. imemi V. A. Steklova LXIX [English translation by Israel Program for Scientific Translations (Sivan, Jerusalem, 1965)].
[13]O. A. Yakubovskii, Yad. Fiz. **5**, 1312 (1967) [Sov. J. Nucl. Phys. **5**, 937 (1967)].
[14]K. L. Kowalski, Phys. Rev. Lett. **15**, 798, 908 E (1965).
[15]H. P. Noyes, Phys. Rev. Lett. **15**, 538 (1965).
[16]W. Hunziker, in *The Schrödinger Equation*, edited by W. Thirring and P. Urban (Springer-Verlag, Wien, 1977), p. 43, and references therein.
[17]J. Ginibre, in Ref. 16, p. 95.
[18]N. Nakanishi, *Graph Theory and Feynman Integrals* (Gordon and Breach, New York, 1971); *Graph Theory and Theoretical Physics*, edited by F. Harary (Academic, New York, 1967).
[19]N. M. Hugenholtz, Physica **23**, 481 (1957).
[20]G. Birkhoff, *Lattice Theory* (American Mathematical Society, Providence, R. I., 1967).
[21]G. Szasz, *Introduction to Lattice Theory* (Academic, New York, 1963).
[22]See also footnote 12 of Ref. 10.
[23]H. Cramer, *Mathematical Methods of Statistics* (Princeton U. P., Princeton, 1964), p. 185.
[24]R. Kubo, J. Phys. Soc. Japan **17**, 1100 (1962); J. Math Phys. **4**, 174 (1963).
[25]H. D. Ursell, Proc. Cambridge Philos. Soc. **23**, 685 (1927).
[26]J. E. Mayer and M. G. Mayer, *Statistical Mechanics* (Wiley, New York, 1941).
[27]S. A. Rice and P. Gray, *The Statistical Mechanics of Simple Liquids* (Interscience, New York, 1965), Chap. 2.
[28]P. Kristensen, Dan. Mat. Fys. Medd. **28**, 12 (1954).
[29]E. Freese, Nuovo Cimento **2**, 50 (1955).
[30]R. Haag, Phys. Rev. **112**, 669 (1958).
[31]W. Zimmerman, Nuovo Cimento **13**, 503 (1959).
[32]R. Jost, *The General Theory of Quantized Fields* (American Mathematical Society, Providence, R. I., 1965), p. 66.
[33]E. H. Wichmann and J. H. Crichton, Phys. Rev. **132**, 2788 (1963).
[34]J. K. Percus, Commun. Math. Phys. **40**, 283 (1975).
[35]D. J. Ernst, J. T. Londergan, G. A. Miller, and R. M. Thaler, Phys. Rev. C **16**, 537 (1977); E. R. Siciliano and R. M. Thaler, Phys. Rev C **16**, 1322 (1977).
[36]Gy. Bencze, Phys. Lett. B **61**, 139 (1976).
[37]I. H. Sloan, Phys. Rev. C **6**, 1945 (1972).
[38]Gy. Bencze, Nucl. Phys. A **210**, 568 (1973).
[39]E. F. Redish, Nucl. Phys. A **225**, 16 (1974).
[40]M. Baer and D. J. Kouri, J. Math. Phys. **14**, 1631 (1973).
[41]D. J. Diestler and H. Krüger, Chem. Phys. **4**, 151 (1974).
[42]D. J. Kouri and F. S. Levin, Phys. Lett. B **50**, 421 (1974); Nucl. Phys. A **253**, 395 (1975).
[43]W. Tobocman, Phys. Rev. C **9**, 2466 (1974).
[44]Gy. Bencze and P. C. Tandy, Phys. Rev. C **16**, 564 (1977).
[45]The combinatorics which we use here apply to an abstact set of N objects, although we are nearly always concerned solely with the particular realization of N distingusishable particles.
[46]The term "connectivity expansion" would perhaps be more appropriate, however the former expression is well established in the literature so we use it here.
[47]M. Goldberger and K. M. Watson, *Collision Theory* (Wiley, New York, 1964).
[48]In lattice theory (Ref. 51) the set of real valued matrices labelled by lattice indices is referred to as the *incidence algebra*.
[49]In addition, the authors have been unable to agree among themselves which representation is simplest.
[50]E. F. Redish, Nucl. Phys. A **235**, 82 (1974).
[51]G. C. Rota, Z. Wahrsch. Verw. Gebiete. **2**, 340 (1964).
[52]Such matrices are called channel coupling arrays (Ref. 2,42, and 43) in the literature.
[53]*Handbook of Mathematical Fuctions*, edited by M. Abramowitz and I. Stegun (NBS, 1965), pp. 821-825.
[54]Notice that in this notation the projector we have been calling $P_{\underline{0}}$ is equal to P_N and $P_{\underline{1}} = P_1$. While it would have been more natural to refer to the N-cluster partition as "N" we have avoided doing so since the subscript "N" is often used in scattering theory to describe the fully interacting system, i.e., the partition $\underline{1}$.
[55]L. R. Dodd and K. R. Greider, Phys. Rev. **139** B 1472 (1965).
[56]M. L'Huillier, E. F. Redish, and P. C. Tandy, J. Math. Phys. **19**, 1276 (1978).

Eigenvalues of an anharmonic oscillator

R. N. Kesarwani
Department of Mathematics, University of Ottawa, Ottawa, KIN 9B4, Canada
Y. P. Varshni
Department of Physics, University of Ottawa, Ottawa, KIN 9B4, Canada

(Received 7 November 1980; accepted for publication 6 March 1981)

The five-term WKBJ approximation is applied to calculate the eigenvalues for the potential $V(x) = \frac{1}{2}kx^2 + ax^4$, $k > 0$ and $a > 0$. Numerical results are compared with those of Hioe and Montroll. It is found that the accuracy of the calculated eigenvalues improves rapidly with increase in the quantum number n. At $n = 4$, a seven significant figure accuracy is achieved and at $n = 6$, a nine significant figure accuracy.

PACS numbers: 03.65.Sq, 03.65.Ge

I. INTRODUCTION

The potential

$$V(x) = \tfrac{1}{2}kx^2 + ax^4, \quad k > 0 \text{ and } a > 0 \tag{1}$$

is of considerable importance in physics. Field theorists are interested[1–3] in it because of the possibility of constructing a field theory based on it. Molecular physicists are interested[4,5] in it because there are certain ring compounds whose vibrations can be described by such a potential. Consequently, there has been a great deal of interest[1–18] in the analytical as well as numerical study of potential (1).

Eigenvalues of high accuracy for this potential have been obtained by Biswas *et al.*,[16] and by Hioe and Montroll[17] (HM) by different methods. Biswas *et al.*[16] postulated wavefunctions to be of the form

$$\psi = \exp(-\tfrac{1}{2}x^2)\sum_{n=0}^{\infty} c_n x^{2n}. \tag{2}$$

On substituting this expression in the Schrödinger equation for potential (1), they obtained a three-term difference equation for the $\{c_n\}$ and the energy E. In order to assure the existence of solutions, they had to set the infinite determinant of the coefficients equal to zero. The energy levels were then found numerically from the resulting Hill determinant. The numerical method used by Biswas *et al.*[16] was to truncate the determinants, calculate the eigenvalues at different levels of truncation, and search for the limits of successive estimates as the truncated determinants were increased in size. Determinants of orders as high as 100×100 were used when a was large as compared to k. Biswas *et al.*[16] give energy values to 15 significant figures for the ground state, and to 8 significant figures for excited states up to and including $n = 7$. HM have developed rapidly converging algorithms, using the Bargmann representation, and they have calculated energy levels accurate to seven or more significant figures for $n = 0$ to $n = 8$. They have tabulated results to nine significant figures.

Wentzel's quantization condition for the WKBJ approximation of any order has been justified and put on an irrefutable basis by Fröman and Fröman.[19] Recently, using Dunham's[20] method, we have derived[21] an expression for the five-term WKBJ approximation. The higher-order terms in the WKBJ approximation have also been considered by Bender *et al.*[22] It was of interest to apply the five-term WKBJ approximation to the potential (1) to ascertain what sort of accuracy can be obtained from it for the excited states. (It is well known that the WKBJ method gives low accuracy at very small quantum numbers, except when the potential is close to a harmonic oscillator potential.) Also, these calculations will help to throw light on the accuracy of such previous results as have been obtained by other methods (e.g., Refs. 4, 16, and 17). We shall compare our results with those of HM as these authors give the largest number of significant figures in their results for excited states.

The WKBJ method has been used previously to calculate the eigenvalues for the potential (1). Lu and his collaborators[9,14] have calculated the eigenvalues using the modified WKBJ method given by Miller and Good.[23] It was pointed out by Fröman *et al.*[18] that the quantization conditions derived by Lu *et al.*[14] in their investigation on the anharmonic oscillator is the same as the well-known WKBJ quantization condition, particularized to a three-term WKBJ approximation. Fröman *et al.*[18] have also pointed out a number of corrections in the expressions derived by Lu *et al.*[14] and have corrected the numerical results obtained therefrom. The starting point of the approach of Handelsman and Lew[10] is the WKBJ approximation. The integrals are evaluated, and expanded in series. A technique is then developed for reversion of the resulting series, by which the eigenvalues are given directly as an expansion in powers of two well-defined variables. A second-order approximation obtained from this double series is evaluated for the potential (1) (and three others). The tabulated results[10] have the same sort of accuracy as a two-term WKBJ calculation; however, in some regions, the results given by Handelsman and Lew[10] are less accurate than those obtained from a two-term WKBJ calculation. In addition to the work referred to above, HM have also derived simple formulas for the eigenvalues, using the one-term WKBJ approximation.

An anonymous referee has drawn our attention to a forthcoming paper by Hioe *et al.*[24] in which these authors have used a higher-order WKBJ method to obtain energy levels for the potential (1).

II. APPLICATION OF THE FIVE-TERM WKBJ APPROXIMATION

The quantization for the five-term WKBJ approximation is

$$n+\tfrac{1}{2}=I_1+I_2+I_3+I_4+I_5, \tag{3}$$

where[21]

$$I_1=\frac{(2m)^{1/2}/\hbar}{\pi}\int_{r_1}^{r_2}(\epsilon-V)^{1/2}\,dx$$

$$I_2=-\frac{\hbar/(2m)^{1/2}}{\pi\,4!}\frac{d}{d\epsilon}\int_{r_1}^{r_2}V''(\epsilon-V)^{-1/2}\,dx,$$

$$I_3=\frac{[\hbar/(2m)^{1/2}]^3}{4\pi\,6!}\times\frac{d^3}{d\epsilon^3}\int_{r_1}^{r_2}(7V''^2-5V'V''')(\epsilon-V)^{-1/2}\,dx,$$

$$I_4=-\frac{[\hbar/(2m)^{1/2}]^5}{2\pi\,9!}\left[\frac{d^5}{d\epsilon^5}\int_{r_1}^{r_2}(93V''^3-224V'V''V'''+35V^{(4)}V'^2)(\epsilon-V)^{-1/2}\,dx+216\frac{d^4}{d\epsilon^4}\int_{r_1}^{r_2}V'''^2(\epsilon-V)^{-1/2}\,dx\right],$$

and

$$I_5=\frac{[\hbar/(2m)^{1/2}]^7}{48\pi\,10!}\left[\frac{d^7}{d\epsilon^7}\int_{r_1}^{r_2}(1143V''^4+2065V'^2V''V^{(4)}-175V'^3V^{(5)})(\epsilon-V)^{-1/2}\,dx-\frac{d^6}{d\epsilon^6}\int_{r_1}^{r_2}(352V''V'''^2+6511V''^2V^{(4)})(\epsilon-V)^{-1/2}\,dx-20\frac{d^5}{d\epsilon^5}\int_{r_1}^{r_2}(29V^{(4)2}+173V'''V^{(5)})(\epsilon-V)^{-1/2}\,dx\right].$$

Here r_1 and r_2 are the real roots of $\epsilon-V(x)=0$, and $V^{(n)}$ represents the nth derivative of V. To apply (3) to the potential (1), we set

$$\epsilon-V=\epsilon-\tfrac{1}{2}kx^2-ax^4=y^2, \tag{4}$$

and

$$\beta=\{[(k^2+16a\epsilon)^{1/2}-k]/4a\}^{1/2}. \tag{5}$$

Then

$$I_1=\frac{(2m)^{1/2}/\hbar}{\pi}\int_{-\beta}^{\beta}y\,dx,$$

$$I_2=-\frac{\hbar/(2m)^{1/2}}{\pi\,4!}\frac{d}{d\epsilon}\int_{-\beta}^{\beta}(k+12ax^2)y^{-1}\,dx,$$

$$I_3=\frac{[\hbar/(2m)^{1/2}]^3}{4\pi\,6!}\frac{d^3}{d\epsilon^3}\int_{-\beta}^{\beta}(7k^2+48akx^2+528a^2x)y^{-1}\,dx,$$

$$I_4=-\frac{[\hbar/(2m)^{1/2}]^5}{2\pi\,9!}\left[\frac{d^5}{d\epsilon^5}\int_{-\beta}^{\beta}(93k^3-1188ak^2x^2-39\,120a^2kx^4-83\,904a^3x^6)y^{-1}\,dx+124\,416a^2\frac{d^4}{d\epsilon^4}\int_{-\beta}^{\beta}x^2y^{-1}\,dx\right],$$

and

$$I_5=\frac{[\hbar/(2m)^{1/2}]^7}{48\pi\,10!}\left[\frac{d^7}{d\epsilon^7}\int_{-\beta}^{\beta}[1143k^4+104\,424ak^3x^2+1978\,752a^2k^2x^4+13\,451\,136a^3kx^6+33\,216\,768a^4x^8]y^{-1}\,dx-24a\frac{d^6}{d\epsilon^6}\int_{-\beta}^{\beta}[6511k^2+164\,712akx^2+1\,038\,960a^2x^4]y^{-1}dx-334\,080a^2\frac{d^5}{d\epsilon^5}\int_{-\beta}^{\beta}y^{-1}\,dx\right]. \tag{6}$$

Note that y^2 is a quartic polynomial in x and the expressions for I_1,I_2,I_3,I_4, and I_5 all involve elliptic integrals.[25] It is therefore, possible to evaluate them explicitly in terms of $K(\omega)$ and $E(\omega)$, the complete elliptic integrals of the first and second kind. We shall find it convenient to express our results in terms of λ, ω, and z, which are defined as follows:

$$\lambda=a\hbar/(mk^3)^{1/2}, \tag{7}$$

$$\omega=[\tfrac{1}{2}-\tfrac{1}{2}(1+16a\epsilon/k^2)^{-1/2}]^{1/2}, \tag{8}$$

and

$$z=\omega^2(1-\omega^2)\quad\text{or}\quad 1-2\omega^2=(1-4z)^{1/2}. \tag{9}$$

Our λ has the same significance as the λ of HM.

The final expressions for I_1,I_2,I_3,I_4, and I_5 are gien below. The essential steps in the evaluation of I_5 are gen in the Appendix. The evaluation of the other four is sinlar.

$$I_1=[(1-4z)^{-3/4}/3\pi\lambda]\{2z\,K(\omega)-(1-4z)^{1/2}[E(\omega)-(1-\omega^2)\,K(\omega)]\},$$

$$I_2=[\lambda(1-4z)^{3/4}/6\pi]\{-4(1-4z)^{1/2}K(\omega)-(1/z+8)[E(\omega)-(1-\omega^2)K(\omega)]\},$$

$$I_3=\frac{\lambda^3(1-4z)^{9/4}}{360\pi}\left\{\left(-\frac{28}{z^2}+\frac{66}{z}-5664+28\,672z\right)K(\omega)+(1-4z)^{1/2}\left(\frac{56}{z^3}-\frac{69}{z^2}+\frac{144}{z}-14\,336\right)[E(\omega)-(1-\omega^2)\,K(\omega)]\right\},$$

$$I_4=\frac{\lambda^5(1-4z)^{15/4}}{315\pi}\left\{(1-4z)^{1/2}\left(\frac{124}{z^4}-\frac{441}{4z^3}+\frac{587}{4z^2}+\frac{1108}{z}-551\,424+4063\,232z\right)K(\omega)\right.$$

$$-\left(\frac{248}{z^5}-\frac{1867}{2z^4}+\frac{10\,189}{16z^3}+\frac{2135}{2z^2}-\frac{11\,624}{z}+1\,610\,752-8\,126\,464z\right)[E(\omega)-(1-\omega^2)K(\omega)]\bigg\},$$

$$I_5=\frac{\lambda^7(1-4z)^{21/4}}{315\pi}\bigg\{\bigg(-\frac{1524}{z^6}+\frac{5883}{z^5}-\frac{55\,079}{16z^4}-\frac{7253}{64z^3}+\frac{32\,591}{2z^2}+\frac{83\,596}{z}$$
$$-149\,719\,040+2\,128\,740\,352z-6\,392\,119\,296z^2\bigg)K(\omega)+(1-4z)^{1/2}\bigg(\frac{3048}{z^7}-\frac{8337}{z^6}+\frac{15\,169}{4z^5}+\frac{316\,085}{128z^4}$$
$$-\frac{87\,275}{4z^3}-\frac{158\,246}{z^2}+\frac{1\,696\,768}{z}-465\,108\,992+3\,196\,059\,648z\bigg)[E(\omega)-(1-\omega^2)K(\omega)]\bigg\}. \qquad (10)$$

III. RESULTS AND DISCUSSION

The energy eigenvalues were calculated by solving (3). The elliptic integrals are expressed in terms of hypergeometric functions[26]:

$$K(\omega)=\tfrac{1}{2}\pi\,{}_2F_1(\tfrac{1}{2},\tfrac{1}{2};1;\omega^2),\quad |\omega|<1, \qquad (11)$$
$$E(\omega)=\tfrac{1}{2}\pi\,{}_2F_1(-\tfrac{1}{2},\tfrac{1}{2};1;\omega^2),\quad |\omega|<1,$$

and thus can be expanded as series in powers of ω^2. The number of terms which need be retained depends on the accuracy desired and on the eigenvalue. The root is determined by the Newton–Raphson method. In the same units as those used by HM, the eigenvalue for the state with quantum number n is given by

$$E_n=[1/(1-2\omega^2)^2-1]/16\lambda, \qquad (12)$$

where ω^2 is the root of (3).

Eigenvalues were calculated for $n=0$ to $n=8$ for the same values of λ for which results are tabulated by HM. Our calculated results to nine significant figures are compared with those of HM in Table I. One point may be noted here. The series expansion from which (3) is obtained is, in general, semiconvergent.[27,28] Consequently, if one finds a situation in which $|I_{j+1}/I_j|$ is greater than one, it would be appropriate to take terms only up to and including I_j on the right-hand side of (3).

In actual practice, calculations were carried out in stages for one-, two-, three-, four-, and five-term WKBJ approximations. A few cases were encountered for $n=0$ and $n=1$ for which $|I_4/I_3|$ was greater than one. Such cases are identified by a superscript a in Table I, and the results shown are the three-term WKBJ ones. All other eigenvalues are from the five-term WKBJ approximation. At each stage

TABLE I. Calculated eigenvalues for the anharmonic oscillator for $n=0$ to $n=8$. For each value of λ, the first line shows the values calculated in this paper, and the second, the values given by Hioe and Montroll. All eigenvalues except those indicated by a superscript a are from the five-term WKBJ approximation.

λ	E_0	E_1	E_2	E_3	E_4	E_5	E_6	E_7	E_8
0.002	0.501489663	1.50741939	2.51920212	3.53674413	4.55995556	5.58875005	6.62304460	7.66275933	8.70781730
	0.501489660	1.50741939	2.51920212	3.53674413	4.55995556	5.58875005	6.62304460	7.66275933	8.70781730
0.006	0.504409708	1.52180565	2.55597230	3.60618633	4.67180037	5.75223087	6.84694847	7.95547029	9.07735366
	0.504409710	1.52180565	2.55597230	3.60618633	4.67180037	5.75223087	6.84694847	7.95547029	9.07735366
0.010	0.507256204	1.53564828	2.59084580	3.67109494	4.77491312	5.90102667	7.04832688	8.21583781	9.40269231
	0.507256200	1.53564828	2.59084580	3.67109494	4.77491312	5.90102667	7.04832688	8.21583781	9.40269231
0.050	0.532634621	1.65343643	2.87397965	4.17633891	5.54929781	6.98496310	8.47739734	10.0219318	11.6147761
	0.532642750	1.65343601	2.87397963	4.17633891	5.54929781	6.98496310	8.47739734	10.0219318	11.6147761
0.100	0.558760543	1.76951479	3.13862403	4.62888281	6.22030090	7.89976723	9.65783999	11.4873156	13.3824748
	0.559146330	1.76950264	3.13862431	4.62888281	6.22030090	7.89976723	9.65783999	11.4873156	13.3789698
0.300	0.631866408[a]	2.09510885[a]	3.84477025	5.79657396	7.91175270	10.1664889	12.5442587	15.0327712	17.6224482
	0.637991780	2.09464199	3.84478265	5.79657363	7.91175273	10.1664889	12.5442587	15.0327713	17.6224482
0.500	0.682803225[a]	2.32541709[a]	4.32749159	6.57840305	9.02877865	11.6487207	14.4176692	17.3204242	20.3451930
	0.696175820	2.32440635	4.32752498	6.57840195	9.02877872	11.6487207	14.4176692	17.3204242	20.3451931
0.700	0.723724040[a]	2.51044542	4.71027250	7.19326729	9.90261059	12.8039297	15.8736836	19.0945183	22.4529987
	0.743903500	2.50922810	4.71032810	7.19326528	9.90261070	12.8039297	15.8736836	19.0945183	22.4529996
1.000	0.774649833[a]	2.73974461	5.17920454	7.94240736	10.9635829	14.2031391	17.6340491	21.2364355	24.9949364
	0.803770650	2.73789227	5.17929169	7.94240399	10.9635831	14.2031394	17.6340492	21.2364362	24.9949457
2	0.900418402[a]	3.29637221	6.30370845	9.72733046	13.4812755	17.5141324	21.7909564	26.2861252	30.9798828
	0.951568470	3.29286782	6.30388057	9.72732319	13.4812759	17.5141324	21.7909564	26.2861250	30.9798830
50	2.26141836[a]	8.93357114	17.4360160	27.1926924	37.9384997	49.5164187	61.8203488	74.7728287	88.3143280
	2.49970877	8.91509636	17.4369921	27.1926458	37.9385022	49.5164187	61.8203488	74.7728290	88.3143280
200	3.53933065[a]	14.0899288	27.5498021	43.0053495	60.0339891	78.3856234	97.8913315	118.427830	139.900396
	3.93093134	14.0592268	27.5514347	43.0052709	60.0339933	78.3856232	97.8913315	118.427830	139.900400
1000	6.01457421[a]	24.0257544	47.0144834	73.4192519	102.516150	133.876891	167.212258	202.311200	239.011578
	6.69422085	23.9722061	47.0173387	73.4191140	102.516157	133.876891	167.212258	202.311200	239.011580
8000	11.9997577[a]	47.9986931	93.9548434	146.745790	204.922696	267.628499	334.284478	404.468346	477.855698
	13.3669076	47.8907687	93.9606046	146.745512	204.922711	267.628498	334.284478	404.468350	477.855700
20000	16.2800935[a]	65.1333236	127.501009	199.145503	278.100218	363.201844	453.664875	548.916141	648.515329
	18.1372291	64.9866757	127.508839	199.145124	278.100238	363.201843	453.664875	548.916140	648.515330

of calculation, such checks were employed as were possible. For instance, our three-term WKBJ results tally exactly with those obtained by Fröman *et al.*[18]

To compare our results with those of HM in a convenient fashion, in Table II, we show the ratio R_n, where

$$R_n = [E_n(\text{ours}) - E_n(\text{HM})]/E_n(\text{HM}). \quad (13)$$

This type of representation has the advantage that the number of zeros that follow the decimal point in the value of R_n also give, in most cases, the number of significant figures to which the two values tally. For example, $R_0 = 0.60 \times 10^{-8}$ indicates that our results agree with those of HM to eight significant figures. In some cases there is a difference of one in the number of zeros and the number of significant figures which agree. In addition, this representation offers the great convenience that one can readily see the broad pattern of the results. In column two of Table II, we also show α, which is sometimes used to represent the degree of anharmonicity and is related to λ by

$$\alpha = \lambda^{2/3}/(1 + \lambda^{2/3}). \quad (14)$$

It will be noticed from Table II that the WKBJ results are the poorest for high λ and low n. As n increases there is a rapid improvement in results, in as much as at $n = 4$, the WKBJ results agree with the HM results to seven significant figures, which is the claimed accuracy of HM results. The preponderance of zeroes in the right-hand top corner of Table II may be noted. The fact that in many cases our results and HM results agree to nine significant figures implies that these values are correct to nine significant figures, as the two sets of values were calculated by different methods. There is excellent agreement between our values and those of HM for $n \geqslant 4$, except for one case, namely $\lambda = 0.1$, $n = 8$; we suspect that in this case the HM value is inaccurate. For $n = 6$, our results and those of HM agree to nine significant figures in all cases but one. However, for $n = 7$ and $n = 8$ there are a number of cases where the agreement between our results and those of HM is to less than nine significant figures. Naturally, the question arises, which set of results are more accurate? We shall adduce evidence to indicate that our results for $n = 7$ and $n = 8$ are accurate to nine significant figures. We shall represent the eigenvalue obtained from a j-term WKBJ approximation by $E^{(j)}$. In Table III, for $n = 8$, we show in the first line against each λ, the eigenvalues $E^{(1)}$, $E^{(2)}$, $E^{(3)}$, $E^{(4)}$, and $E^{(5)}$. The second line shows the differences $E^{(2)} - E^{(1)}$, $E^{(3)} - E^{(2)}$, $E^{(4)} - E^{(3)}$, and $E^{(5)} - E^{(4)}$ in columns three to six respectively. While there is no known method of obtaining exact error bounds for energy eigenvalues obtained by the WKBJ method, an examination of $|E^{(j+1)} - E^{(j)}|$ is helpful for this purpose. It will be noticed (Table III) that $|E^{(j+1)} - E^{(j)}|$ decreases rapidly as j increases. Column six shows that in most cases $E^{(5)} - E^{(4)}$ is zero, only in three cases there is a difference of one in the ninth significant figure between the two eigenvalues. The trend of $|E^{(j+1)} - E^{(j)}|$ values indicates that the effect of including the sixth term in the WKBJ approximation will show up in the tenth significant figure or beyond. It would be reasonable to infer from this that the eigenvalues calculated from the five-term WKBJ approximation are accurate to nine significant figures for $n = 8$. The results for $n = 7$ are very similar. Beyond $n = 8$, as n becomes larger, the trend of R_n values in Table II indicates that one can expect an even better accuracy.

We note that in the WKBJ method, the eigenvalue of each quantum state is calculated independently of others so

TABLE II. Values of the ratio R_n, defined by Eq. (13). The number in parenthesis is the power of 10 by which the preceding figure is to multiplied. Thus 0.60(− 8) stands for 0.60×10^{-8}

λ	α	R_0	R_1	R_2	R_3	R_4	R_5	R_6	R_7	R_8
0.002	0.01563	0.60(−08)	0	0	0	0	0	0	0	0
0.006	0.03196	−0.40(−08)	0	0	0	0	0	0	0	0
0.010	0.04436	0.79(−08)	0	0	0	0	0	0	0	0
0.050	0.11950	−0.15(−04)	0.25(−06)	0.70(−08)	0	0	0	0	0	0
0.100	0.17726	−0.69(−03)	0.69(−05)	−0.89(−07)	0	0	0	0	0	0.26(−03)
0.300	0.30946	−0.96(−02)	0.22(−03)	−0.32(−05)	0.57(−07)	−0.38(−08)	0	0	−0.67(−08)	0
0.500	0.38649	−0.19(−01)	0.43(−03)	−0.77(−05)	0.17(−06)	−0.78(−08)	0	0	0	−0.49(−08)
0.700	0.44063	−0.27(−01)	0.49(−03)	−0.12(−04)	0.28(−06)	−0.11(−07)	0	0	0	−0.40(−07)
1.000	0.50000	−0.36(−01)	0.68(−03)	−0.17(−04)	0.42(−06)	−0.18(−07)	−0.21(−07)	−0.57(−08)	−0.33(−07)	−0.37(−06)
2	0.61351	−0.54(−01)	0.11(−02)	−0.27(−04)	0.75(−06)	−0.30(−07)	0	0	0.76(−08)	−0.65(−08)
50	0.93136	−0.95(−01)	0.21(−02)	−0.56(−04)	0.17(−05)	−0.66(−07)	0	0	−0.40(−08)	0
200	0.97159	−0.10	0.22(−02)	−0.59(−04)	0.18(−05)	−0.70(−07)	0.26(−08)	0	0	−0.29(−07)
1000	0.99010	−0.10	0.22(−02)	−0.61(−04)	0.19(−05)	−0.68(−07)	0	0	0	−0.84(−08)
8000	0.99751	−0.10	0.23(−02)	−0.61(−04)	0.19(−05)	−0.73(−07)	0.37(−08)	0	−0.99(−08)	−0.42(−08)
20000	0.99864	−0.10	0.23(−02)	−0.61(−04)	0.19(−05)	−0.72(−07)	0.28(−08)	0	0.18(−08)	−0.15(−08)

that small errors in the ground state calculations do not propagate with amplification into excited state energy calculations.

The methods used in Refs. 16 and 17 tend to become cumbersome and probably less accurate as n is increased. On the other hand, the accuracy of the results from the WKBJ method improves with n. In that sense the WKBJ method complements the methods of Refs. 16 and 17.

ACKNOWLEDGMENT

This work was supported in part by a research grant from the Natural Sciences and Engineering Research Council of Canada to one of the authors (Y.P.V.).

APPENDIX

In this appendix we show the details of evaluation of the term I_5. Based on the procedure given in Ref. 25, we need only evaluate integrals of the form

$$U_n = \int_{-\beta}^{\beta} x^n y^{-1}\, dx, \quad n = 0,2,4,6,8, \tag{A1}$$

where y and β are given by (4) and (5).

Set

$$u_n(x) = \int x^n y^{-1}\, dx, \quad n = 0,2,4,6,8. \tag{A2}$$

It is easily seen that

$$\begin{aligned} u_4(x) &= (1/3a)[\epsilon u_0(x) - k u_2(x)] - xy/3a, \\ u_6(x) &= (1/15a^2)[(9a\epsilon + 2k^2)\, u_2(x) - 2k\epsilon u_0(x)] + (1/15a^2)(2k - 3ax^2)xy, \\ u_8(x) &= (1/105a^3)[(25a\epsilon + 6k^2)\epsilon u_0(x) - (52a\epsilon + 6k^2)\, k\, u_2(x)] + (1/15a^2)(25a\epsilon + 6k^2 - 9akx^2 + 15a^2x^4)xy. \end{aligned} \tag{A3}$$

Since U_n are definite integrals over an interval $(-\beta,\beta)$, it follows that

$$\begin{aligned} U_4 &= (1/3a)(\epsilon U_0 - kU_2), \\ U_6 &= (1/15a^2)[(9a\epsilon + 2k^2)U_2 - 2k\epsilon U_0], \\ U_8 &= (1/105a^3)[(25a\epsilon + 6k^2)\epsilon U_0 - (52a\epsilon + 6k^2)kU_2]. \end{aligned} \tag{A4}$$

For our purpose, it remains to determine U_0 and U_2.

Writing

$$y^2 = \epsilon - \tfrac{1}{2}kx^2 - ax^4 = a(\alpha^2 + x^2)(\beta^2 - x^2),$$

TABLE III. Calculated eigenvalues for $n = 8$, from one-, two-, three-, four- and five-term WKBJ approximations. The second line against each λ shows the differences between the eigenvalue in that column and that in the preceding column.

λ	$E^{(1)}$	$E^{(2)}$ $E^{(2)} - E^{(1)}$	$E^{(3)}$ $E^{(3)} - E^{(2)}$	$E^{(4)}$ $E^{(4)} - E^{(3)}$	$E^{(5)}$ $E^{(5)} - E^{(4)}$
0.002	8.707184414	8.707817271 0.000632857	8.707817301 0.000000030	8.707817301 0.	8.707817301 0.
0.006	9.075884279	9.077353401 0.001469122	9.077353657 0.000000256	9.077353657 0.	9.077353657 0.
0.01	9.400668581	9.402691829 0.002023249	9.402692306 0.000000477	9.402692306 0.	9.402692306 0.
0.05	11.61052500	11.61477617 0.00425117	11.61477611 -0.00000005	11.61477609 -0.00000002	11.61477609 0.
0.10	13.37702345	13.38247636 0.00545291	13.38247485 -0.00000151	13.38247481 -0.00000004	13.38247481 0.
0.30	17.61455962	17.62245330 0.00789368	17.62244829 -0.00000501	17.62244821 -0.00000009	17.62244821 0.
0.50	20.33585334	20.34520017 0.00934683	20.34519314 -0.00000703	20.34519304 -0.00000011	20.34519304 0.
0.70	22.44256188	22.45300730 0.01044542	22.45299881 -0.00000849	22.45299869 -0.00000012	22.45299869 0.
1.00	24.98319516	24.99494671 0.01175155	24.99493654 -0.00001017	24.99493641 -0.00000013	24.99493641 0.
2	30.96511726	30.97989691 0.01477965	30.97988300 -0.00001390	30.97988283 -0.00000017	30.97988284 0.
50	88.27130308	88.31437400 0.04307093	88.31432845 -0.00004555	88.31432797 -0.00000049	88.31432798 0.00000001
200	139.8321193	139.9004693 0.0683501	139.9003964 -0.0000729	139.9003956 -0.0000008	139.9003956 0.
1000	238.8948426	239.0117040 0.1168614	239.0115788 -0.0001252	239.0115775 -0.0000013	239.0115775 0.
6000	477.6222404	477.8559510 0.2337106	477.8557002 -0.0002507	477.8556976 -0.0000026	477.8556977 0.0000001
20000	648.1984809	648.5156724 0.3171915	648.5153320 -0.0003404	648.5153284 -0.0000036	648.5153285 0.0000001

where

$$\alpha^2 = \frac{(16a\epsilon + k^2)^{1/2} + k}{4a} \quad \text{and} \quad \beta^2 = \frac{(16a\epsilon + k^2)^{1/2} - k}{4a},$$

we get

$$U_0 = \int_{-\beta}^{\beta} y^{-1}\,dx = \frac{2}{a^{1/2}}\int_0^{\beta} [(\alpha^2 + x^2)(\beta^2 - x^2)]^{-1/2}\,dx$$
$$= \frac{2}{[a(\alpha^2 + \beta^2)]^{1/2}} \int_0^{\pi/2} \frac{d\theta}{(1 - \omega^2 \sin^2\theta)^{1/2}} = (8/k)^{1/2}(1 - 4z)^{1/4} K(\omega), \tag{A5}$$

and

$$U_2 = \int_{-\beta}^{\beta} x^2 y^{-1}\,dx = \frac{2}{a^{1/2}}\int_0^{\beta} x^2[(\alpha^2 + x^2)(\beta^2 - x^2)]^{-1/2}\,dx = [(2k)^{1/2}/a](1 - 4z)^{-1/4}[E(\omega) - (1 - \omega^2)K(\omega)]. \tag{A6}$$

The expression I_5 in (6) now becomes

$$I_5 = \{[\hbar/(2m)^{1/2}]^7/48\,\pi\,10!\}[(d^7/d\epsilon^7)(1143\,k^4\,U_0 + 104\,424\,ak^3\,U_2 + 1\,978\,752\,a^2k^2U_4$$
$$+ 13\,451\,136\,a^3kU_6 + 33\,216\,768\,a^4\,U_8)$$
$$- 24a(d^6/d\epsilon^6)(6511k^2U_0 + 164\,712\,ak\,U_2 + 1\,038\,960\,a^2\,U_4) - 334\,080\,a^2(d^5/d\epsilon^5)U_0].$$

On using the reduction formulas (A4), and the relations (A5) and (A6), we obtain

$$I_5 = \frac{[\hbar/(2m)^{1/2}]^7}{\pi 10!}\left\{\frac{d^7}{d\epsilon^7}\left[\left(\frac{381}{16}k^4 + \frac{1\,671\,688}{105}a\epsilon k^2 + \frac{1\,153\,360}{7}a^2\epsilon^2\right)U_0 - ak\left(\frac{2\,886\,521}{210}k^2 + \frac{6\,110\,072}{35}a\epsilon\right)U_2\right]\right.$$
$$\left. - a\frac{d^6}{d\epsilon^6}\left[\left(\frac{6511}{2}k^2 + 173\,160\,a\epsilon\right)U_0 - 90\,804\,ak\,U_2\right] - 6960a^2\frac{d^5}{d\epsilon^5}U_0\right\}$$
$$= \frac{[\frac{1}{2}\hbar(k/m)^{1/2}]^7}{\pi 10!}\left[\frac{2}{105}\frac{d^7}{d\epsilon^7}\{(1 - 4z)^{-7/4}(40\,005 + 6\,366\,712z - 8\,806\,528z^2)\,K(\omega)\right.$$
$$- (1 - 4z)^{-5/4}(11\,546\,084 - 9\,523\,904z)\,[E(\omega) - (1 - \omega^2)K(\omega)]\}$$
$$- (16a/k^2)(d^6/d\epsilon^6)\{(1 - 4z)^{-3/4}(6511 + 60\,536z)\,K(\omega) - 90\,804\,(1 - 4z)^{-1/4}[E(\omega) - (1 - \omega^2)K(\omega)]\}$$
$$\left. - 222\,720\,(a^2/k^4)(d^5/d\epsilon^5)\{(1 - 4z)^{1/4}K(\omega)\}\right].$$

To carry out the differentiation repeatedly in the above expression, we apply chain rule keeping in mind that

$$\frac{dz}{d\epsilon} = \frac{4a}{k^2}(1 - 4z)^2,$$
$$\frac{d\omega}{d\epsilon} = \frac{2a}{\omega k^2}(1 - 4z)^{3/2},$$

and the formulas

$$\frac{dK(\omega)}{d\omega} = \frac{E(\omega) - (1 - \omega^2)K(\omega)}{\omega(1 - \omega^2)},$$
$$\frac{d}{d\omega}[E(\omega) - (1 - \omega^2)K(\omega)] = \omega K(\omega).$$

We obtain

$$I_5 = [ak^{3/2}(\hbar/m^{1/2})^7/32\,\pi\,10!]\,[(d^6/d\epsilon^6)\{(1 - 4z)^{-3/4}(-9402 + 316\,736z - 195\,072z^2)\,K(\omega)\}$$
$$+ (1 - 4z)^{-1/4}(381/z - 494\,368 + 97\,536z)[E(\omega) - (1 - \omega^2)K(\omega)]\} - 55\,680\,(a/k^2)(d^5/d\epsilon^5)\{(1 - 4z)^{1/4}K(\omega)\}]$$
$$= [3a^2k^{-1/2}(\hbar/m^{1/2})^7/4\pi 10!](d^5/d\epsilon^5)\{(1 - 4z)^{1/4}(127/4z + 4444 + 55\,104z + 130\,048z^2)K(\omega)$$
$$- (1 - 4z)^{3/4}(127/2z^2 + 466/z + 39\,744 + 65\,024z)[E(\omega) - (1 - \omega^2)K(\omega)]\},$$

which finally gives

$$I_5 = \frac{\lambda^7}{315\pi}\left\{(1 - 4z)^{21/4}\left(-\frac{1524}{z^6} + \frac{5883}{z^5} - \frac{55\,079}{16z^4} - \frac{7253}{64z^3} + \frac{32\,591}{2z^2} + \frac{83\,596}{z}\right.\right.$$
$$\left. - 149\,719\,040 + 2\,128\,740\,352z - 6\,392\,119\,296z^2\right)K(\omega)$$
$$+ (1 - 4z)^{23/4}\left(\frac{3048}{z^7} - \frac{8337}{z^6} + \frac{15\,169}{4z^5} + \frac{316\,085}{128z^4} - \frac{87\,275}{4z^3}\right.$$
$$\left.\left. - \frac{158\,246}{z^2} + \frac{1\,696\,768}{z} - 465\,108\,992 + 3\,196\,059\,648z\right)[E(\omega) - (1 - \omega^2)\,K(\omega)]\right\}.$$

[1]C. M. Bender and T. T. Wu, Phys. Rev. Lett. **21**, 406 (1968).
[2]C. M. Bender and T. T. Wu, Phys. Rev. **184**, 1231 (1969).
[3]C. M. Bender and T. T. Wu, Phys. Rev. D **7**, 1620 (1973).
[4]S. I. Chan, D. Stelman, and L. E. Thompson, J. Chem. Phys. **41**, 2828 (1964).
[5]References given in R. N. Kesarwani and Y. P. Varshni, Z. Naturforsch. **33a**, 1581 (1978).
[6]R. McWeeny and C. A. Coulson, Proc. Cambridge Philos. Soc. **44**, 413 (1948).
[7]N. Bazley and D. Fox, Phys. Rev. **124**, 483 (1961).
[8]C. Reid, J. Chem. Phys. **43**, S186 (1965).
[9]P. Lu, J. Chem. Phys. **47**, 815 (1967); **53**, 845 (1970).
[10]R. A. Handelsman and J. S. Lew, J. Chem. Phys. **50**, 1182 (1969).
[11]J. Loeffel, A. Martin, B. Simon, and A. Wightman, Phys. Lett. B **30**, 656 (1969).
[12]B. Simon, Ann. Phys. **58**, 76 (1970).
[13]S. Graffi, V. Grecchi, and B. Simon, Phys. Lett. B **32**, 631 (1970).
[14]P. Lu, S. S. Wald, and B.-L. Young, Phys. Rev. D **7**, 1701 (1972).
[15]P. M. Mathews and K. Eswaran, Lett. Nuovo Cimento **5**, 15 (1972).
[16]S. N. Biswas, K. Dutta, R. P. Saxena, P. K. Srivastava, and V. S. Varma, J. Math. Phys. **14**, 1190 (1973).
[17]F. T. Hioe and E. W. Montroll, J. Math. Phys. **16**, 1945 (1975).
[18]P. O. Fröman, F. Karlsson, and M. Lakshmanan, Phys. Rev. D **20**, 3435 (1979).
[19]N. Fröman and P. O. Fröman, J. Math. Phys. **18**, 96 (1977). See also N. Fröman, Ark. Fys. **32**, 541 (1966).
[20]J. L. Dunham, Phys. Rev. **41**, 713, 721 (1932).
[21]R. N. Kesarwani and Y. P. Varshni, J. Math. Phys. **21**, 90 (1980); **21**, 2852 (1980).
[22]C. M. Bender, K. Olaussen, and P. S. Wang, Phys. Rev. D **16**, 1740 (1977).
[23]S. C. Miller and R. H. Good Jr., Phys. Rev. **91**, 174 (1953).
[24]F. T. Hioe, E. W. Montroll, and M. Yamawaki, in *Perspectives in Statistical Physics*, edited by H. J. Raveche (North-Holland, Amsterdam) (in press).
[25]M. Abramowitz and I. A. Stegun, *Handbook of Mathematical Functions* (Dover, New York, 1965).
[26]A. Erdelyi, W. Magnus, F. Oberhettinger, and F. G. Tricomi, *Higher Transcendental Functions*, Vol. 2 (McGraw-Hill, New York, 1953).
[27]G. D. Birkhoff, Bull. Am. Math. Soc. **39**, 696 (1933).
[28]E. C. Kemble, *The Fundamental Principles of Quantum Mechanics* (Dover, New York, 1958); this is a reprint of the 1937 edition published by McGraw-Hill.

A nonmanifold theory of space–time

J. S. Mashford
Department of Mathematics, University of Melbourne, Parkville, Victoria 3052, Australia

(Received 12 June 1980; accepted for publication 23 January 1981)

A Zeeman topology is defined in the general framework of any set W of events which has been equipped with an acyclic signal relation $\sim\to$. The assumption that the $\sim\to$ structure of w is locally that of Minkowski space and that the "piecing together" maps are smooth in an appropriate sense, allows a tangent bundle $p:E\to W$ to be defined. This bundle has, as structure group, the group G of linear causal automorphisms of Minkowski space.

PACS numbers: 04.20.Cv, 02.40. – k

1. INTRODUCTION

In general relativity, space–time is assumed to have the structure of a Lorentzian manifold, and therefore, locally, the Euclidean topology $\mathscr{E}^4$. It has been argued[1,2] that this topology has little physical significance.

Let $\langle\ ,\ \rangle:\mathbb{R}^4\times\mathbb{R}^4\to\mathbb{R}^4$ be the Minkowski inner product for $\mathbb{R}^4$ defined by $\langle x, y\rangle = x^{\dagger}\eta y$, where $\eta = \mathrm{diag}(1, -1, -1, -1)$. Using a theorem of Zeeman[1] (proved earlier by Alexandrov and Ovchinnikova[3]), the Lorentz group of linear $\langle\ ,\ \rangle$ preserving maps of $\mathbb{R}^4$ onto itself can, up to translations and positive dilatations, be characterized as the group of bijections $f:\mathbb{R}^4\to\mathbb{R}^4$ such that f, f^{-1} take light beams onto light beams. It is reasonable, therefore, that if space–time has a natural topology then this topology should be related to its light beam structure.

Zeeman[4] defined a number of topologies for Minkowski space $(\mathbb{R}^4, \langle\ ,\ \rangle)$ by requiring that the induced topology on timelike lines and or spacelike planes be Euclidean. Göbel[2,5], Hawking *et al.*[6], and others have investigated similar topologies for Lorentzian manifolds.

In the present paper a "Zeeman topology"[2,7] is defined in which two events are considered to be "close" only if they are interacting. This topology can be defined in the general context, independent of Lorentzian manifolds, of any nonempty set W with a relation $\sim\to\subset W\times W$ (specifying the interactions in W) for which $\sim\to$ satisfies a certain causality axiom.

Application of the equivalence principle that, locally, space–time has the structure of Minkowski space, leads to a manifold like model for space–time. An analog of the Gâteaux derivative is defined and it is shown that the derivative of any allowable coordinate transformation takes values in the group G, generated by the orthochronous Lorentz transformations and positive dilatations of Minkowski space. G therefore forms the structure group of the tangent bundle for space–time.

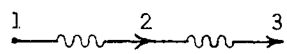
2. WEBS

Given a set S and a relation $\rho\subset S\times S$, consider the collection of all relations $\rho'\subset S\times S$ which are transitive extensions of ρ. When ordered by inclusion this collection has a least element $\bar{\rho}$ given by $x\bar{\rho}y$ if and only if there is a finite chain $\{x_0,\ldots,x_n\}\subset S$ (where $n\in\{1, 2, \cdots\}$) such that $x_0 = x$, $x_n = y$, and $x_{i-1}\rho x_i$ for $i = 1,\ldots,n$.

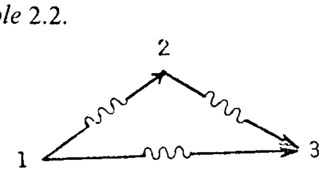
For any pair of events x,y in physical space–time M, let $x\sim\to y$ mean that a light signal, or photon, can travel from event x to event y. Suppose that $x, y, z\in M$ are such that $x\sim\to y$ and $y\sim\to z$. Since the velocity of light is finite, x must be distinct from z. By the same argument, if $x\,\overline{\sim\to}\,y$, then $x\neq y$. This requirement that there are no "causal loops" expresses the directedness of the flow of time and motivates the following definition:

A *web* is a nonempty set W together with a relation $\sim\to\subset W\times W$ such that $x\,\overline{\sim\to}\,y\Rightarrow x\neq y$.

Example 2.1.



Example 2.2.



Example 2.3. The *Minkowski web* L defined by $L = \mathbb{R}^4$, $x\sim\to y\Leftrightarrow Q(y-x) = 0,\ x^0 < y^0$, where $Q(x) = \langle x,x\rangle$, is the characteristic quadratic form for Minkowski space.

Example 2.4. Let $W = (W, \sim\to)$ be a web. Define $\sim\to_{\mathrm{rev}}\subset W\times W$ by $x\sim\to_{\mathrm{rev}}y\Leftrightarrow y\sim\to x$. Then $W_{\mathrm{rev}} = (W, \sim\to_{\mathrm{rev}})$ is a web.

Elements of W will be called *events* or points. $\overline{\sim\to}$ is a partial ordering of W and will be denoted by $<$. If S is a nonempty subset of a web $W = (W, \sim\to)$ then $(S, \sim\to\cap S\times S)$ is a web.

Define $\leftarrow\sim\to\ \subset W\times W$ by

$$x\leftarrow\sim\to y\Leftrightarrow x\sim\to y,\quad x = y \text{ or } y\sim\to x.$$

If x and y are distinct events such that $x\leftarrow\sim\to y$, then x and y will be said to be *interacting*. $\overline{\leftarrow\sim\to}$ is an equivalence relation on W and $\overline{\leftarrow\sim\to}$ equivalence classes will be called *connected components* of W. W will be said to be *connected* if it has only one connected component.

Let $(W_1, \sim\to_1)$, $(W_2, \sim\to_2)$ be webs. A map $f: W_1\to W_2$ will be said to be a *web morphism* (or $\sim\to$ morphism or simply morphism) if

 0022-2488/81/091990-04$01.00

$$(\forall x, y\in W_1) \quad x\sim\rightarrow_1 y \Rightarrow f(x)\sim\rightarrow_2 f(y).$$

f will be said to be an *isomorphism*, written $f: W_1 \approx W_2$, if f is a bijection such that f, f^{-1} are morphisms. Let $\mathrm{Aut}(W, \sim\rightarrow)$ [or simply $\mathrm{Aut}(W)$] denote the group of isomorphisms of W onto itself. Morphisms and isomorphisms for $\leftarrow\sim\rightarrow$ and $<$ are defined similarly.

Let Θ be the orthochronous Lorentz group, G as defined before, and G' the group generated by the orthochronous Poincaré transformations and positive dilatations of Minkowski space. By Zeeman's theorem,[1] $\mathrm{Aut}(L) = G'$.

Any $\sim\rightarrow$ preserving map preserves $<$. Therefore, every $\sim\rightarrow$ isomorphism is a $<$ isomorphism, that is, an isotony. However, not all isotonies are web isomorphisms. For example, let $(W_1, \sim\rightarrow_1)$ be the web of Example 2.1, $(W_2, \sim\rightarrow_2)$ the web of Example 2.2, and define $f: W_1 \rightarrow W_2$ by $f(i) = i$. Therefore, the web structure contains more information than the causal structure. If $x,y\in L$ then $x<y \Leftrightarrow Q(y-x)\geq 0, x^0<y^0$. For any web W, $\mathrm{Aut}(W,\sim\rightarrow) \subset \mathrm{Aut}(W,<)$. The inclusion may be strict, as is shown, for example, by the web $W = W_1 \cup W_2$ with W_1, W_2 as defined above. However, Nanda[8] proved a theorem which shows that $\mathrm{Aut}(L,<) = \mathrm{Aut}(L,\sim\rightarrow) = G'$.

3. THE $\mathscr{H}$ TOPOLOGY

A subset of W will be said to be a *light beam* if it is a maximal set of interacting events. Thus $\beta\subset W$ is a light beam if and only if

$$(\forall x,y\in\beta) \qquad x\leftarrow\sim\rightarrow y,$$

and

$$(\forall x\in W\sim\beta)(\exists y\in\beta) \quad x\leftarrow\sim\rightarrow y.$$

By the axiom of choice, if $S\subset W$ is such that $x\leftarrow\sim\rightarrow y$ for all $x,y\in S$, Then there exists a light beam containing S. Taking $S=\{x\}$ shows that every point is contained in some light beam. Light beams in L are the sets $a+\mathbb{R}e = \{a+te: t\in\mathbb{R}\}$, where $a\in\mathbb{R}^4$ and $e\in\{x\in\mathbb{R}^4: 0\sim\rightarrow x\}$. In L a light beam is determined by any two points in it; however, this is not true in general. If $f: W_1\rightarrow W_2$ then f is a $\leftarrow\sim\rightarrow$ morphism if and only if f takes light beams into light beams. Using a corollary to Zeeman's theorem[9] and the fact that for any $x,y\in L$, $Q(y-x)\geq 0$ if and only if every light beam through y meets some light beam through x, one can show that, as stated in the introduction, the group of all bijections $f: L\rightarrow L$ such that f, f^{-1} take light beams onto light beams is the group generated by the Poincaré transformations and the positive dilatations. Therefore, $\mathrm{Aut}(L,\leftarrow\sim\rightarrow)\cong G'\times\mathbb{Z}_2$.

Let $\mathscr{L} = \mathscr{L}_W$ be the collection of light beams in W. If $\beta\in\mathscr{L}$, then $<\cap\,\beta\times\beta$ is a total ordering of β. It is natural, therefore, to give β the interval topology[10] induced by its ordering. Light beams in L are isotonic to $(\mathbb{R},<)$ and therefore homeomorphic to the Euclidean line $\mathscr{E}^1$. Let $\mathscr{H} = \mathscr{H}_W$ be the topological union[7] of $\mathscr{L}$; that is, if $U\subset W = \cup\mathscr{L}$ then $U\in\mathscr{H}$ if and only if $\beta\cap U$ is open in β for all $\beta\in\mathscr{L}$. It is easy to show that $\mathscr{H}$ is a topology for W. If W is finite then $\mathscr{H}$ is discrete. Since $\mathscr{H}_L$ is an exact topological union[7] it can be characterized as the finest topology for L which induces the Euclidean topology on light beams. $\mathscr{H}_L$ is strictly finer than $\mathscr{E}^4$ and so Hausdorff. It induces the discrete topology on timelike curves and spacelike hypersurfaces. $(L,\mathscr{H}_L)$ is a topologically connected space.

If $C\subset W$ is a connected component of W, then $(\forall\beta\in\mathscr{L})$ $\beta\cap C\neq\emptyset\Rightarrow\beta\subset C$ and so $C\in\mathscr{H}$. Therefore, if a web is topologically connected it must be connected. However, not all connected webs are topologically connected, as is shown by the web $W = (-1,0)\cup(0,1)\subset\mathbb{R}$, where $x\sim\rightarrow y\Leftrightarrow x<y$.

If $x\in W$ let $\mathscr{L}_x$ be the set of all light beams in W containing x where two light beams are identified if they coincide near x. That is

$$\mathscr{L}_x = \{\underline{\beta}: x\in\beta\in\mathscr{L}\}, \text{ where } \underline{\beta} = \{\gamma\in\mathscr{L} :(\exists U\in\mathscr{H})\ x\in\gamma\cap U,\ \gamma\cap U = \beta\cap U\}$$

is the $\mathscr{L}$ germ of β at x. In L, elements of $\mathscr{L}_x$ are singletons, $\underline{\beta} = \{\beta\}$, and there is a natural one-to-one projection of $\mathscr{L}_x$ onto $S^2\subset\mathbb{R}^3$ given by $\underline{\beta}\mapsto$ the unique $\omega\in S^2$ such that $\beta = \{x+t(1,\omega): t\in\mathbb{R}\}$.

Any continuous path $\gamma:[a,b]\rightarrow\mathscr{E}^4$ can be uniformly approximated arbitrarily well by a polygonal path whose sides are light beams, and any such polygonal path is $\mathscr{H}$ continuous. Therefore the set $C([a,b],\mathscr{H})$ is dense in the space $C([a,b],\mathscr{E}^4)$ where $C([a,b],\mathscr{E}^4)$ has its usual supremum norm.

4. LOCALLY MINKOWSKIAN WEBS

If W_0, W are webs, then W will be said to be *locally* W_0 if every point in W has an open neighborhood isomorphic to an open subweb of W_0. The equivalence principle provides some justification for modelling space–time by locally L webs. If M is a Lorentzian manifold and $\sim\rightarrow$ is interpreted in the usual way in terms of null geodesics, then in general, M need not be a locally L web. However, if M is causal and conformally flat then it is a locally L web. Since $L\approx L_{\mathrm{rev}}$, a web W is locally L if and only if W_{rev} is locally L. If W is locally L, then light beams in W are causally unbounded and locally isotonic to $(\mathbb{R},<)$, and are therefore one manifolds.

For W a web and $x\in W$ denote $\{y\in W: x\sim\rightarrow y\}$ by $\{x\sim\rightarrow\}$. If X, Y are (real) vector spaces let $\mathrm{Hom}(X,Y)$ denote the space of linear maps $T: X\rightarrow Y$. Suppose that X is a vector space which has been equipped with a Hausdorff topology. Let $U\subset L$ be open and $f: U\rightarrow X$. f will be said to be *differentiable* at $a\in U$ if there exists a linear map $T\in\mathrm{Hom}(\mathbb{R}^4,X)$ such that

$$(\forall e\in\{0\sim\rightarrow\}) \quad h^{-1}(f(a+he)-f(a))\rightarrow Te \quad \text{as } h\rightarrow 0.$$

If the derivative exists it is unique and will be denoted by $T = Df(a)$. If X has a norm $|\cdot|: X\rightarrow[0,\infty)$ such that the norm topology is coarse in X, then a necessary condition for f to be differentiable at a with derivative $T\in\mathrm{Hom}(\mathbb{R}^4,X)$ is that

$$(\forall e\{0\sim\rightarrow\}) \quad |h|^{-1}|f(a+he)-f(a)-The|\rightarrow 0 \quad \text{as } h\rightarrow 0.$$

When X is normable this condition is sufficient. The condition is also sufficient when X is L and f is a $\leftarrow\sim\rightarrow$ morphism. If $f: U\rightarrow X$ is differentiable at $a\in U$, define $\partial_\alpha(x)(f) = (Df(x))e_\alpha$ where $\{e_\alpha\}$ is the standard basis for $\mathbb{R}^4$. The derivative at $x\in L$ of a function depends only on the behavior of the function at events interacting with x.

Let U, V be open subwebs of L, $f: U\rightarrow V$ a $\leftarrow\sim\rightarrow$ morphism differentiable at $a\in U$, $g:V\to L$ a $\leftarrow \sim \rightarrow$ morphism differentiable at $f(a)$, $T=Df(a)$, and $S=Dg(f(a))$. Then $g\circ f$ is a $\leftarrow \sim \rightarrow$ morphism and for any $e\in\{0\sim\rightarrow\}$, $h\neq 0$,

$$\begin{aligned}|h|^{-1}|g\circ f(a+he)-g\circ f(a)-SThe|\\ \leq |h|^{-1}|g(f(a+he)-g(f(a)))-S(f(a+he)-f(a))|\\ +|h|^{-1}|S(f(a+he)-f(a)-The)|\to 0 \quad \text{as } h\to 0,\end{aligned}$$

because $f(a+he)-f(a)\to 0$ along a fixed light beam as $h\to 0$. Therefore, $g\circ f$ is differentiable at a and the chain rule, $(D(g\circ f))(a)=(Dgf(a))Df(a)$, holds. If $U\subset L$ is Euclidean open and $f:U\to L=\mathbb{R}^4$ is a $\leftarrow \sim \rightarrow$ morphism differentiable at $a\in U$ in the usual Gâteaux sense, then f is differentiable and the two derivatives coincide.

Suppose that $f:U\approx V$ and f is a diffeomorphism, that is, f, f^{-1} are differentiable over their respective domains. Let $a\in U$ and $T=Df(a)$. By the chain rule, T is nonsingular. Let $e\in\{0\sim\rightarrow\}$. Choose $e'\in\{f(a+he)-f(a):h>0\}$. f is a $\sim\rightarrow$ morphism and so $e'\in\{0\sim\rightarrow\}$. For all $h>0$, $h^{-1}(f(a+he)-f(a))\in[0,\infty)e'$. Thus, since $h^{-1}(f(a+he)-f(a))\to Te$ as $h\to 0^+$, $[0,\infty)e'$ is closed and $Te\neq 0$, we have that $Te\in\{0\sim\rightarrow\}$. Hence $T\{0\sim\rightarrow\}\subset\{0\sim\rightarrow\}$ and so, by a well-known theorem,[11] $T\in G$. Therefore $Df(x)\in G$ for all $x\in U$.

By Zeeman's theorem, if $f:L\approx L$ then $f\in G'$ and so f is everywhere differentiable with derivative $f-f(0)$. There are, however, web isomorphisms $f:U\approx V$; U, $V\in\mathscr{H}_L$, other than elements of G'.

Example 4.1. (the conformal inversion)

Let $U=\{x\in\mathbb{R}^4:Q(x)>0,\ x^0>0\}$, $V=-U$. U and V are Euclidean open, and so open. Define $f:U\to V$ by

$$f(x)=-[Q(x)]^{-1}x.$$

Then f is a web isomorphism and is differentiable over U with derivative

$$Df(x)=[Q(x)]^{-2}[2x\langle x|-Q(x)I],$$

where $\langle x|\in\mathrm{Hom}(\mathbb{R}^4,\mathbb{R})$ is the map $\langle x|(v)=\langle x,v\rangle$. The standard matrix for $Df(x)$ is

$$[Df(x)]^\alpha{}_\beta=[Q(x)]^{-2}[2x^\alpha x_\beta-Q(x)\delta^\alpha{}_\beta],$$

where $x_\beta=\eta_{\beta\gamma}x^\gamma$. If $x\in U$, $T=Df(x)$ then $\langle Tu,Tv\rangle=[Q(x)]^{-2}\langle u,v\rangle$ for all $u, v\in\mathbb{R}^4$, so the "dilatation factor" at x is $[Q(x)]^{-1}$. $f^{-1}:V\approx U$ is given by $f^{-1}(\xi)=-[Q(\xi)]^{-1}\xi$ and so f^{-1} is differentiable over V.

If W is locally L let $\mathscr{C}=\mathscr{C}_W$ be the family of all coordinate systems, or charts, for W, that is $\phi\in\mathscr{C}\Leftrightarrow\phi:U\approx U'$ for some $U\in\mathscr{H}$, $U'\in\mathscr{H}_L$. A subcovering $\mathscr{D}\subset\mathscr{C}$ will be called a *differentiable structure* for W if it is a maximal family of differentiably compatible charts, where two charts ϕ_1, ϕ_2 are differentiably compatible if and only if $\mathrm{Domain}(\phi_1)\cap\mathrm{Domain}(\phi_2)=\varnothing$ or $\phi_2\circ\phi_1^{-1}$ is a diffeomorphism.

A theorem of Hawking[5,6] states that if M, M' are (time oriented strongly causal) C^∞ Lorentzian manifolds and $h:M\to M'$ is a manifold homeomorphism which takes null geodesics to null geodesics, then h is a C^∞ diffeomorphism. Let U, $V\subset L$ be Euclidean open and $f:U\approx V$. It is known[12] that f must be a Euclidean homeomorphism. Hawking's theorem then implies that f must be C^∞. Thus, Hawking's theorem implies that if $f:U\approx V$ where $U, V\in\xi^4$, then f is differentiable. If this result is true for arbitrary U, $V\in\mathscr{H}_L$, then every locally L web W has a unique differentiable structure $\mathscr{D}=\mathscr{C}$ consisting of all charts for W.

If $\mathscr{D}=\{\phi_i\}$ is a differentiable structure for W then the assignment $(\phi_i,\phi_j)\mapsto g_{ij}=(D\phi_j\circ\phi_i^{-1})\circ\phi_i$ satisfies the algebraic condition (the chain rule) necessary to be a $\mathscr{D}$ cocycle with values in G.[13] Therefore, there is a unique vector bundle $p:E\to W$ associated with $(W,\mathscr{D})$ for which the g_{ij} are the transition functions. The topology of the space $E=TW=\cup_{x\in W}T_xW$ will depend on the topology given to G, subject to the continuity of the g_{ij}. Denote $\mathscr{D}_x=\{\phi\in\mathscr{D}:x\in\mathrm{domain}(\phi)\}$. Then a tangent vector $A_x\in T_xW$ can be thought of as an assignment $A_x:\mathscr{D}_x\to\mathbb{R}^4$ such that

$$A_x(\phi_j)=g_{ij}(x)A_x(\phi_i),\quad \forall\phi_i,\ \phi_j\in\mathscr{D}_x$$

[this corresponds to the usual formula, $\bar{A}^\alpha=(\partial\bar{x}^\alpha/\partial x^\beta)A^\beta$]. In the construction of TW, G has its usual linear representation on $\mathbb{R}^4$. In general, any representation $\pi:G\to\mathrm{Aut}(X)$ of G, induces a bundle over W with fibers isomorphic to X and structure group $G/\mathrm{Ker}(\pi)$.

Let $F:G\times S^2\to S^2$ be defined by

$$F(T,\omega)=F_T(\omega)=([T(1,\omega)]^0)^{-1}([T(1,\omega)]^1,[T(1,\omega)]^2,[T(1,\omega)]^3).$$

F is a C^ω action of G on S^2. F is an extension to G of the usual action of $O(3)$ on S^2, with the usual imbedding of $O(3)$ in G and therefore, F is transitive. Suppose that $T\in\mathrm{Ker}(F)$. Then, for all $\omega\in S^2$, $([T(1,\omega)]^1,[T(1,\omega)]^2,[T(1,\omega)]^3)=[T(1,\omega)]^0\omega$; thus, $T(1,\omega)=[T(1,\omega)]^0(1,\omega)$. Therefore, $\omega\mapsto[T(1,\omega)]^0$ is a continuous function from S^2 into the set $\sigma(T)$ of eigenvalues of T. Since $\sigma(T)$ is discrete and T is orthochronous, $[T(1,\omega)]^0$ must equal a positive constant, so $T=\lambda I$, $\lambda>0$. Therefore, $\mathrm{Ker}(F)\cong\mathbb{R}^+$ is the dilatation subgroup of G and F is an effective action of Θ on S^2.

Let $\phi\in\mathscr{D}_x$, $\phi:U\approx U'$, and let $\underline{\beta}\in\mathscr{L}_x$. Choose $\beta\in\underline{\beta}$. $\phi(\beta\cap U)$ is an open subset of a light beam in L through $\phi(x)$ and is therefore taken by the projection map defined previously to some $\omega\in S^2$. If $\beta_1,\beta_2\in\underline{\beta}$, then $\phi(\beta_1\cap U)=\phi(\beta_2\cap U)$. Therefore the map $\underline{\beta}\mapsto\omega$ is a well-defined bijection of $\mathscr{L}_x$ onto S^2. Every $\phi\in\mathscr{D}_x$ provides an identification of $\mathscr{L}_x$ with S^2, and so every chart pair ϕ_i, $\phi_j\in\mathscr{D}_x$ induces a permutation of S^2. It is easy to show that this permutation is F_T, where $T=g_{ij}(x)$. Thus, $\mathscr{L}_x$ has the topology and C^ω differentiable structure of S^2, and the bundle $\mathscr{L}W=\cup_{x\in W}\mathscr{L}_x$ has structure group Θ.

ACKNOWLEDGMENT

I would like to thank G. H. Williams and P. F. Price for helpful discussions and advice.

[1]E. C. Zeeman, J. Math. Phys. **5**, 490 (1964).

[2]R. Göbel, Commun. Math Phys. **16**, 289 (1976).

[3]A. D. Alexandrov and V. V. Ovchinnikova, Vest. Leningr. Univ. **11**, 96 (1953).

[4]E. C. Zeeman, Topology **6**, 161 (1967).

[5]R. Göbel, J. Math. Phys. **17**, 845 (1976).

[6]S. W. Hawking, A. R. King, and P. J. McCarthy, J. Math. Phys. **17**, 174 (1976).

[7]F. A. Cel'nik, Soviet Math. Dokl. **9**, 1151 (1968).
[8]S. Nanda, J. Math. Phys. **12**, 394 (1970).
[9]L. Michel, in *Axiomatic Field Theory*, Brandeis University Summer Institute in Theoretical Physics 1965, Vol. 1, edited by M. Chretien and S. Deser (Gordon and Breach, New York, 1966), p. 296.
[10]I. L. Lynn, Proc. Amer. Math. Soc. **13**, 494 (1962).
[11]W. Rindler, *Special Relativity* (Oliver and Boyd, Edinburgh, 1960), p. 21.
[12]S. Nanda, Math. Proc. Cambridge Philos. Soc. **79**, 533 (1976).
[13]N. R. Wallach, *Harmonic Analysis on Homogeneous Spaces* (Marcel Dekker, New York, 1973), p. 2.

Local transverse-traceless tensor operators in general relativity[a)]

David K. Monroe[b)]
Department of Physics and Astronomy, University of North Carolina, Chapel Hill, North Carolina 27514

(Received 8 December 1980; accepted for publication 20 March 1981)

Two flat-space transverse-traceless tensor operators can be used to construct initial data for numerical solutions of the gravitational field equations. One of these operators is related to the conformal curvature 3-tensor and is shown to exist in a large class of nonflat 3-spaces. The second operator enjoys no such liberty. Important applications to gravitational wave scattering are suggested. It is argued that the number of operators available on a particular 3-space is related to the number of gravitational field modes that are excited in the space.

PACS numbers: 04.20.Cv, 04.30. + x

I. INTRODUCTION

Symmetric transverse-traceless (TT) tensors play an important role in general relativity. In the initial-value problem[1] it is found that one part of the freely specifiable data is the TT field momentum density $\pi^{ab}_{\rm TT}$. In the linearized theory, so important for understanding weak gravitational waves, one may always choose the TT gauge, in which the tensor $h_{ij} = g_{ij} - \delta_{ij}$ is transverse-traceless.

There is a method due to York[2] for constructing TT tensors in curved spaces. Given an arbitrary symmetric 3-tensor one can always decompose it into three mutually orthogonal terms, one of which is the TT part:

$$h_{ab} = h^{\rm TT}_{ab} + (LX)_{ab} + \tfrac{1}{3}g_{ab}h,$$
$$(LX)_{ab} \equiv \nabla_a X_b + \nabla_b X_a - \tfrac{2}{3}g_{ab}(\nabla_c X^c).$$

However, an elliptic differential equation

$$\nabla_a (LX)^a{}_b = \nabla_a h^a{}_b - \tfrac{1}{3}\nabla_b h$$

must be solved in carrying out this decomposition and in general it cannot be solved analytically. The solution of this equation by numerical methods is also troublesome because of the nature of elliptic equations.[3] Clearly it would be of great convenience if there were completely local prescriptions for generating TT tensors on curved spaces.

The purpose of this paper is to present a new technique for generating TT tensors by local methods in a large class of three-dimensional Riemannian spaces. As a first step I shall consider local TT operators in flat spaces. Reviewing this material will not only show how a more general TT operator can be constructed, it will also lead to some useful techniques for generating initial data for gravitational Cauchy problems. Then the subject of TT tensor operators on curved spaces will be considered. I will argue that the existence of one or more TT operators on a space depends upon how many gravitational field modes are already excited in the space. Applications of these operators will be suggested throughout this paper.

II. FLAT-SPACE TRANSVERSE-TRACELESS OPERATORS

Flat-space TT operators have been known for some time.[4] They are usually presented as direct generalizations of the transverse vector operators so frequently used in electromagnetism for the construction of multipole moments[5] and the use of Debye potentials[6] (see Appendix A). A straight forward method of deriving flat-space TT operators is to use Fourier analysis to reduce the differential condition $\partial_i h_{ij}(\mathbf{x}) = 0$ to the simpler algebraic one, $k_i h_{ij}(\mathbf{k}) = 0$. Combining this with the trace condition $h_{ii}(\mathbf{x}) = 0 = h_{ij}(\mathbf{k})$, one can solve for the most general $h_{ii}(\mathbf{k})$ that has the form $u_{ij}(\mathbf{k})\Phi(\mathbf{k})$. This particular form of $h_{ij}(\mathbf{k})$is chosen because it makes $h_{ij}(\mathbf{x})$ a TT tensor operator premultiplying a scalar $\Phi(\mathbf{x})$. The latter is called a Debye potential in analogy with the transverse-vector operator formalism.

When the equations for $u_{ij}(\mathbf{k})$ are solved, it is found that derivatives in **k**-space must be introduced, and that there are only two inequivalent solutions. Transforming back to **x**-space, TT operators L_{ij} and M_{ij} are obtained. After some manipulation these take the simple forms (see Appendix B)

$$L_{ij} = L_{(i}M_{j)} + L_{(i}\partial_{j)}, \quad M_{ij} = [ikl]\partial_k L_{lj},$$
$$L_i = [ikl]x_l\partial_K, \quad M_i = [ikl]\partial_k L_l.$$

An important property of L_{ij} and M_{ij} is that they are orthogonal; that is, for arbitrary Φ and Ψ

$$\int d^3x (L_{ij}\Phi)(M_{ij}\Psi) \equiv 0.$$

This means that, given an arbitrary TT Cartesian tensor $h^{\rm TT}_{ij}$, one can write

$$h^{\rm TT}_{ij} = L_{ij}\Phi + M_{ij}\Psi$$

and solve, without any physical ambiguity, for each of the Debye potentials Φ and Ψ. The solutions are unique up to arbitrary monopole and dipole terms (see Appendix B). However, L_{ij} and M_{ij} will annihilate these terms anyway, so they are of no significance.

III. APPLICATIONS OF THE FLAT-SPACE OPERATORS

An important use for the flat-space TT operators just discussed is in the construction of initial data to be evolved by numerical methods.[3] Solutions of the linearized field equations are valid in the wave zone. The computer program

a)Supported in part by NSF Grant PHY 74-14191.
b)Present address: Department of Physics and Astronomy, Appalachian State University, Boone, N.C. 28608.

 0022-2488/81/091994-07$01.00

can then advance the wave into stronger-field regions where the effects of self-interaction can be analyzed. This procedure is now outlined and examples will follow.

A wavelike effect can be described by the TT part of the tensor $h_{ij} = g_{ij} - \delta_{ij}$, which measures departure from flatness. In the wavezone and in the absence of sources, this TT part satisfies the wave equation:

$$\Box h_{ij}^{\mathrm{TT}} = 0.$$

In terms of Debye potentials these six equations become the two equations (see Appendix B)

$$\Box \Phi = 0, \quad \Box \Psi = 0.$$

Of particular interest are situations in which Φ and Ψ have simple mutipole structure. As has been mentioned already, monopole and dipole potentials are annihilated by the operators L_{ij} and M_{ij}. The first nontrivial result occurs when Φ or Ψ has the form

$$f(t,r) Y_{2m}(\theta,\phi),$$

where Y_{2m} is a spherical harmonic. An h_{ij}^{TT} so obtained is a pure-quadrupole TT tensor. Similarly, potentials of the form

$$f(t,r) Y_{3m}(\theta,\phi)$$

will generate pure-octupole TT tensors.

In order to calculate the components of h_{ij}^{TT} in these two cases, it is far more convenient to work in spherical coordinates than Cartesian. The spherical counterparts of L_{ij} and M_{ij} are easily calculated from the covariant definitions:

$$L_{ab} = L_{(a} M_{b)} + L_{(a} \nabla_{b)}, \qquad M_{ab} = \epsilon_{akl} \nabla^k L^l{}_b,$$
$$L^a = \epsilon^{abc} \nabla_b r_c, \qquad M^a = \epsilon^{abc} \nabla_b L_c,$$
$$r_a \equiv \tfrac{1}{2} \nabla_a (x^2 + y^2 + z^2), \qquad \epsilon_{abc} = (\sqrt{g})[abc].$$

In the case of azimuthal symmetry, the operators have the nontrival components[7]

$$L_{r\phi} = (s/2)\partial_\theta (L^2 + 2), \qquad L_{\theta\phi} = -(r/2)(r\partial_r + 2)s^2 \partial_\theta (1/s)\partial_\theta,$$
$$M_{rr} = (1/2r^2) L^2 (L^2 + 2), \qquad M_{r\theta} = -(1/2r)\partial_r r \partial_\theta (L^2 + 2),$$
$$\mathrm{M}_{\theta\theta} = -\tfrac{1}{2}(c/s)\partial_\theta (L^2 + 2) + (r/2)\partial_r (r\partial_r + 2) s \partial_\theta (1/s) \partial_\theta,$$

with the following definitions having been made:

$$s \equiv \sin\theta, \quad c \equiv \cos\theta,$$

$$L^2 \equiv L^a L_a = \frac{1}{s}\partial_\theta s \partial_\theta + \frac{1}{s^2}\partial_\phi^2 \to \frac{1}{\sin\theta}\partial_\theta \sin\theta \partial_\theta.$$

Thus $h_{ab}^{\mathrm{TT}} = L_{ab}\Phi + M_{ab}\Psi$ appears as

$$h_{ab}^{\mathrm{TT}} = \begin{bmatrix} M_{rr}\Psi & M_{r\theta}\Psi & L_{r\phi}\Phi \\ M_{r\theta}\Psi & M_{\theta\theta}\Psi & L_{\theta\phi}\Phi \\ L_{r\phi}\Phi & L_{\theta\phi}\Phi & M_{\phi\phi}\Psi \end{bmatrix}.$$

It is immediately apparent that $\Psi(r,\theta)$ generates those parts of h_{ab}^{TT} which are invariant under the reflection $\phi \to \pi - \phi$, while Φ generates those parts of h_{ab}^{TT} which change only in sign under this reflection. Therefore Ψ produces even-parity solutions, while Φ produces odd-parity ones. The reason for this can be traced back to the definitions of the operators L_{ab} and M_{ab}. The first is actually a pseudotensor operator, while the second is a tensor operator in the fullest sense.

A pure-quadrupole even-parity azimuthally symmetric TT tensor can be explicitly displayed. Choosing the Debye potential Ψ as

$$\Psi = f(t,r)(\cos^2\theta - \tfrac{1}{3}),$$

it is found that

$$M_{rr}\Psi = (12/r^2)\Psi = (12 f/r^2)(\cos^2\theta - \tfrac{1}{3}),$$
$$M_{r\theta}\Psi = -4[f_{,r} + (1/r)f]\sin\theta\cos\theta,$$
$$M_{\theta\theta}\Psi = -4f + [f_{,rr} + (3/r)f_{,r} - (7/r^2)f] r^2 \sin^2\theta,$$
$$M_{\phi\phi}\Psi = -4f \sin^2\theta - [f_{,rr} + (3/r)f_{,r} - (8/r^2)f] r^2 \sin^4\theta.$$

The wavelike TT tensor $h_{ab}^{\mathrm{TT}} = M_{ab}\Psi$ will be completely specified when one has chosen a solution of the following equation for the function $f(t,r)$:

$$0 = -f_{,tt} + f_{,rr} + (2/r)f_{,r} - (6/r^2)f.$$

Teukolsky[8] has produced just such a TT wavelike solution of the linearized Einstein equations by other methods. The solution is referred to as a linearized Brill[9] wave. To demonstrate that Teukolsky's solution is actually of the type shown above, the process used will be inverted and the Debye potentials found for Teukolsky's wave. The line-element in question is

$$\begin{aligned} ds^2 = & -dt^2 + [1 + 3E\sin^2\theta\cos^2\theta - A]d\rho^2 \\ & + 6\sin\theta\cos\theta\,[E\cos^2\theta + B - C]d\rho\, dz \\ & + [1 + 3E\cos^4\theta + 6(B - C)\cos^2\theta + 3C - A]dz^2 \\ & + \rho^2[1 + 3(A - C)\sin^2\theta - A]d\phi^2, \end{aligned}$$

where the function A, B, C, and E are

$$A = 6\left[\frac{1}{r^3}F''(r) - 3\frac{1}{r^4}F'(r) + 3\frac{1}{r^5}F(r)\right],$$

$$B = 2\left[\frac{1}{r^2}F'''(r) - 3\frac{1}{r^3}F''(r) + 6\frac{1}{r^4}F'(r) - 6\frac{1}{r^5}F(r)\right],$$

$$C = \tfrac{1}{2}\left[\frac{1}{r}F''''(r) - \frac{2}{r^2}F'''(r) + \frac{9}{r^3}F''(r) - \frac{21}{r^4}F'(r) + \frac{21}{r^5}F(r)\right],$$

$$E = A + C - 2B.$$

The tensor h_{ab} so defined can be verified to be transverse-traceless. In spherical coordinates it takes the form

$$h_{rr} = 3A(\cos^2\theta - \tfrac{1}{3}),$$
$$h_{r\theta} = -3Br\sin\theta\cos\theta,$$
$$h_{\theta\theta} = -r^2 A + 3r^2 C\sin^2\theta,$$
$$h_{\phi\phi} = -r^2 A\sin^2\theta + 3r^2(A - C)\sin^4\theta.$$

The Debye potential Φ can be found by solving the equation (see Appendix B)

$$-\tfrac{1}{2}L^2(L^2 + 2)\Phi = r^a L^b h_{ab} = r^3 L^3 h_{33} = 0.$$

Consequently $\Phi = 0$ up to irrelevant monopole and dipole terms. The other Debye potential Ψ is found by solving the equation

$$\tfrac{1}{2}L^2(L^2 + 2)\Psi = r^a r^b h_{ab} = 3Ar^2(\cos^2\theta - \tfrac{1}{3}).$$

The solution is easily found to be

$$\Psi = \tfrac{1}{4}r^2 A(\cos^2\theta - \tfrac{1}{3}),$$

again up to inconsequential monopole and dipole terms. Camparison of the Debye potentials for Teukolsky's solution with those for the more general quadrupole solution shows that they are equivalent if and only if

$$A = (4/r^2) f(t,r).$$

As a further illustration of these techniques, the pure-octapole even-parity azimuthally symmetric solutions of $\Box h^{\mathrm{TT}}_{ab} = 0$ will now be displayed. The Debye potentials are

$$\Phi = 0, \quad \Psi = f(t,r)\,(\cos^3\theta - \tfrac{3}{5}\cos\theta).$$

In spherical coordinates h^{TT}_{ab} has the components

$$\begin{aligned} h_{rr} &= 60(f/r^2)\,(\cos^3\theta - \tfrac{3}{5}\cos\theta),\\ h_{r\theta} &= -15(f_{,r} + 1/r)f)\,(\cos^2\theta - \tfrac{1}{5})\sin\theta,\\ h_{\theta\theta} &= -3r^2(f_{,rr} + (3/r)f_{,r} + (5/r^2)f)\cos^3\theta\\ &\quad + 3r^2(f_{,rr} + (3/r)f_{,r} + (1/r^2)f)\cos\theta,\\ h_{\theta\theta} &= -3r^2(f_{,rr} + (3/r)f_{,r} - (15/r^2)f)\cos^5\theta\\ &\quad + 6r^2(f_{,rr} + (3/r)f_{,r} - (13/r^2)f)\cos^3\theta\\ &\quad - 3r^2(f_{,rr} + (3/r)f_{,r} - (11/r^2)f)\cos\theta. \end{aligned}$$

This TT tensor is completely specified when one chooses an $f(r,t)$ which satisfies the equation

$$0 = -f_{,tt} + f_{,rr} + (2/r)f_{,r} - (12/r^2)f.$$

IV. TRANSVERSE-TRACELESS OPERATORS ON CURVED THREE-SPACES

First, conformally flat spaces will be analyzed. As is known[10] conformal scaling enables one to construct a TT tensor in a given space if one is already known in a conformally related space. For conformally flat spaces, described by $\tilde{g}_{ab} = \phi^4 f_{ab}$, this means that the flat-space TT operators L_{ab} and M_{ab} can be readily generalized to

$$\tilde{L}_{ab} = \phi^{-2} L_{ab}, \quad \tilde{M}_{ab} = \phi^{-2} M_{ab}.$$

The orthogonality of L_{ab} and M_{ab} translates into the modified orthogonality relation

$$\int d^3x \tilde{g}^{1/2} \phi^3 (\tilde{L}^{ab}\Phi)(\tilde{M}_{ab}\Psi) = 0.$$

Finally, the inversion formulas become

$$\begin{aligned} -\tfrac{1}{2}L^2(L^2+2)\Phi &= r^a L^b\{\phi^2 h^{\mathrm{TT}}_{ab}\},\\ \tfrac{1}{2}L^2(L^2+2)\Psi &= r^a r^b\{\phi^2 h^{\mathrm{TT}}_{ab}\}, \end{aligned}$$

with r^a and L^a borrowed from flat space.

The existence of these TT operators in conformally flat spaces is very fortunate since it means that wavelike perturbations of the Schwarzschild metric, whose standard spacelike slices $dt = 0$ are conformally flat, can be studied using Debye potentials. The perturbation equations can be written in terms of two scalars rather than a tensor and so it will be easier to determine how gravitational waves are scattered and absorbed by black holes. This particular problem has, of course, already been solved by others, notably Teukolsky,[11] but with a different formalism (Newman–Penrose). It would still be of great interest to compare the results of the two fundamentally different methods, and so the Debye potential approach to wavelike perturbations is being investigated by the author. An obvious advantage of the Debye-potential approach is that it better lends itself to physical interpretation, since the potentials are simply related to the 3-metric.

Now we pass on to the more general and more difficult case of 3-spaces which are neither flat nor conformally flat. It will be found that one of the operators, L_{ab}, can be generalized to a quite large class of curved spaces. As for the second operator M_{ab}, the situation is far more of a problem and it appears that this operator does not exist on any but conformally flat (or flat) spaces. However, this is only a conjecture.

The key to the problem of generalizing L_{ab} is to recognize its relationship to the conformal curvature 3-tensor[12] defined as

$$\beta^a{}_b \equiv [amn]\nabla_m S_{nb}, \quad S_{ab} \equiv R_{ab} - \tfrac{1}{4} g_{ab} R,$$

where R_{ab} is the Ricci curvature 3-tensor. Because of its structure and the Bianchi identity, this tensor density is symmetric and transverse-traceless. Furthermore, it vanishes if and only if the 3-metric from which it is constructed is conformally flat. To show the relationship between L_{ab} and $\beta^a{}_b$, one need only write

$$g_{ab} = \delta_{ab} + h_{ab}$$

and examine that part of $\beta^a{}_b$ which is of first order in h_{ab}. It is found that

$$\beta^a{}_b = \tfrac{1}{4}[akl]\partial_k\{h_{lc,cb} - h_{lb,cc}\} + \text{S.T.},$$

where "S.T." stands for an identical expression with the free indices ("a" and "b") interchanged. If h_{ab} is chosen to have the form

$$h_{ab} = 2x_a x_b \Phi$$

with Φ an arbitrary scalar, then $\beta^a{}_b$ becomes

$$\begin{aligned} \beta^a{}_b &= \tfrac{1}{2}[akl]\partial_k\{\partial_b\partial_c X_l X_c - \partial_c\partial_c X_l X_b\}\Phi + \text{S.T.}\\ &= \tfrac{1}{2}\{(\partial_b\partial_c X_c - \partial_c\partial_c X_b) + \partial_b\}L_a\Phi + \text{S.T.}\\ &= \tfrac{1}{2}\{M_b + \partial_b\}L_a\Phi + \text{S.T.}\\ &= \{M_{(a}L_{b)} + \partial_{(a}L_{b)}\}\Phi = L_{ab}\Phi. \end{aligned}$$

It has been proven, then, that the flat-space operator L_{ab} can be extracted from the conformal curvature tensor. Specifically, $L_{ab}\Phi$ is the first-order part of $\beta^a{}_b$ with a flat background metric and a perturbation of the special form $h_{ab} = 2x_a x_b \Phi$.

The path to generalizing L_{ab} is now clear. We must express the 3-metric in the form

$$g_{ab} = \bar{g}_{ab} + h_{ab},$$

with $\bar{g}_{ab}$ some given "base" metric, and calculate the first-order part of $\beta^a{}_b$ in the perturbation h_{ab}. If we agree to let bars denote quantities of order 0 in h_{ab} and asterisks denote those of order 1, then we find that

$$\begin{aligned} \Gamma^c_{ab} &= \bar{\Gamma}^c_{ab} + \Gamma^{*c}{}_{ab} + O(h^2),\\ \bar{\Gamma}^c_{ab} &= \tfrac{1}{2}\bar{g}^{cd}(\partial_a\bar{g}_{bd} + \partial_b\bar{g}_{ad} - \partial_d\bar{g}_{ab}),\\ \Gamma^{*c}_{ab} &= \tfrac{1}{2}\bar{g}^{cd}(\bar{\nabla}_a h_{bd} + \bar{\nabla}_b h_{ad} - \bar{\nabla}_d h_{ab}),\\ R_{ab} &= \bar{R}_{ab} + R^*{}_{ab} + O(h^2),\\ \bar{R}_{ab} &= \bar{\Gamma}^c_{ab,c} - \bar{\Gamma}^c_{ac,b} + \bar{\Gamma}^d_{cd}\bar{\Gamma}^c_{ab} - \bar{\Gamma}^c_{ad}\bar{\Gamma}^d_{bc},\\ R^*{}_{ab} &= \bar{\nabla}_c\Gamma^{*c}{}_{ab} - \bar{\nabla}_b\Gamma^{*c}{}_{ac},\\ S_{ab} &= \bar{S}_{ab} + S^*{}_{ab} + O(h^2),\\ \bar{S}_{ab} &= \bar{R}_{ab} - \tfrac{1}{4}\bar{g}_{ab}\bar{R},\\ S^*{}_{ab} &= R^*{}_{ab} - \tfrac{1}{4}h_{ab}\bar{R} - \tfrac{1}{4}\bar{g}_{ab}(\bar{g}^{cd}R^*{}_{cd} - h^{cd}\bar{R}_{cd}), \end{aligned}$$

$$\beta^a{}_b = \overline{\beta}^a{}_b + \beta^{*a}{}_b + O(h^2),$$
$$\overline{\beta}^a{}_b = [akl]\nabla_k S_{lb},$$
$$\beta^{*a}{}_b = [akl]\{\overline{\nabla}_k S^*_{lb} - \Gamma^{*c}_{kb}\overline{S}_{lc}\}.$$

The properties of $\beta^{*a}{}_b$ could be discovered by inspecting the expression just determined for it. However, it is far easier to infer them from the known properties of the full conformal curvature tensor $\beta^a{}_b$. From the fact that $\beta^a{}_b$ is tracefree, it is found that

$$0 = \beta^a{}_a = \overline{\beta}^a{}_a + \beta^{*a}{}_a + O(h^2)$$

and so

$$\overline{\beta}^a{}_a = 0, \quad \beta^{*a}{}_a = 0.$$

Thus $\beta^{*a}{}_b$ is tracefree for arbitrary tensors h_{ab}.

The symmetry of $\beta^{ab} = \beta^a{}_c g^{cb}$ implies that

$$0 = [cab]\beta^a{}_d g^{db}$$
$$= [cab]\overline{\beta}^{ab} + [cab](\beta^{*a}{}_d \overline{g}^{db} - \overline{\beta}^a{}_d h^{db}) + O(h^2),$$

and so

$$\overline{\beta}^{ab} = \overline{\beta}^{ba}, \quad \beta^{*ab} - \beta^{*ba} = \overline{\beta}^a{}_d h^{db} - \overline{\beta}^b{}_d h^{da}.$$

This is the first instance in which the properties of $\beta^{*a}{}_b$ have been linked to those of h_{ab}. In order that $\beta^*{}_{ab}$ be symmetric, it is necessary and sufficient that h_{ab} and $\overline{\beta}^a{}_b$ commute:

$$\overline{\beta}^a{}_c h^{cb} - \overline{\beta}^a{}_c h^{ca} = 0. \tag{1}$$

Finally, from the transverse nature of $\beta^a{}_b$, it is found that

$$0 = \nabla_a \beta^a{}_b = \overline{\nabla}_a \overline{\beta}^a{}_b + \overline{\nabla}_a \beta^{*a}{}_b - \Gamma^{*c}_{ab}\overline{\beta}^a{}_c + O(h^2),$$

which in turn implies that

$$\overline{\nabla}_a \overline{\beta}^a{}_b = 0, \quad \overline{\nabla}_a \beta^{*a}{}_b = \tfrac{1}{2}\overline{\beta}^{ac}\overline{\nabla}_b h_{ac}.$$

Here is a second linkage between $\beta^{*a}{}_b$ and h^{ab}. If $\beta^{*a}{}_b$ is to be transverse, then it is necessary and sufficient that

$$\overline{\beta}^{ac}\overline{\nabla}_b h_{ac} = 0. \tag{2}$$

In the flat-space case h_{ab} was chosen to be $h_{ab} = 2x_a x_b \Phi$ in order to convert $\beta^{*a}{}_b$ into $L_{ab}\Phi$. It is now clear that this freedom to choose h_{ab} at will existed because the background metric $\overline{g}_{ab} = \delta_{ab}$ was flat and so $\overline{\beta}^a{}_b = 0$. In the more general case, when $\overline{g}_{ab}$ is not flat, this freedom no longer exists because $\overline{\beta}^a{}_b$ need not vanish. Instead, the tensor $\overline{\beta}^a{}_b$ itself must govern the sturcture of h_{ab}.

If $\beta^{*a}{}_b$ is to be a local linear TT operator premultiplying a scalar function and is to have the same differential structure as the flat-space operator L_{ab}, then we must choose

$$h_{ab} = u_{ab}\Phi.$$

The function Φ is the Debye potential and so must be completely arbitrary. The tensor u_{ab}, on the other hand, will be determined so as to make $\overline{\beta}^a{}_b$ symmetric and TT. It will be expressed in terms of the background conformal curvature $\overline{\beta}^a{}_b$.

The two conditions, Eqs. (1) and (2), become, when expressed in terms of u_{ab} and Φ,

$$0 = [cab]h_{ad}\overline{\beta}^d{}_b = [cab]u_{ab}\overline{\beta}^d{}_b\Phi, \tag{3}$$
$$= [cab]\overline{\beta}^{ab} + [cab](\beta^{*a}{}_d\overline{g}^{db} - \overline{\beta}^a{}_d h^{db}) + O(h^2), \tag{4}$$

If Eq. (3) is to be satisfied for arbitrary scalars Φ, then u_{ab} must commute with $\overline{\beta}^a{}_b$. This will generally occur when u_{ab} has the same eigenvectors with respect to $\overline{g}_{ab}$ as does $\overline{\beta}^a{}_b$. Hence

$$\overline{\beta}^a{}_b = \sum_{A=1}^{3} \overline{\beta}_A \underset{A}{\xi}{}^a \underset{A}{\xi}{}_b, \quad \underset{A}{\xi}{}^a \underset{B}{\xi}{}_a = \delta_{AB}, \tag{5}$$

$$u_{ab} = \sum_{A=1}^{3} u_A \underset{A}{\xi}{}_a \underset{A}{\xi}{}_b, \tag{6}$$

with the three eigenvalues $\{u_A\}$ unspecified. Combining this last result with Eq. (4), it is found that

$$0 = \Phi \sum_{A=1}^{3} \overline{\beta}_A \overline{\nabla}_c u_A + 2\Phi \sum_{A=1}^{3} \overline{\beta}_A u_A \underset{A}{\xi}{}^b \overline{\nabla}_c \underset{A}{\xi}{}_b + (\overline{\nabla}_c\Phi)\sum_{A=1}^{3}\overline{\beta}_A u_A. \tag{7}$$

Because the vectors $\{\underset{A}{\xi}{}^a\}$ are of unit length, the second term in Eq. (7) vanishes. Then, since Φ is wholly arbitrary, the first and third terms of Eq. (7) must vanish separately:

$$0 = \sum_{A=1}^{3} \overline{\beta}_A \overline{\nabla}_c u_A, \tag{8}$$

$$0 = \sum_{A=1}^{3} \overline{\beta}_A u_A. \tag{9}$$

Each of these constraints upon the three scalars $\{u_A\}$ can be simplified by making use of the tracelessness of $\overline{\beta}^a{}_b$. Choosing to eliminate the eigenvalue $\overline{\beta}_3$, we find that since $\overline{\beta}_3 = -\overline{\beta}_1 - \overline{\beta}_2$, then

$$\overline{\beta}_1\partial_c\alpha_1 + \overline{\beta}_2\partial_c\alpha_2 = 0, \tag{10}$$
$$\overline{\beta}_1\alpha_1 + \overline{\beta}_2\alpha_2 = 0, \tag{11}$$

where α_1 and α_2 are defined by

$$\alpha_k \equiv u_k - u_3 \quad (k = 1,2).$$

This system consist of four equations for two functions and is therefore overdetermined. No general solution exists and so $\beta^{*a}{}_b$ will not always lead to a generalization of the flat-space TT operator L_{ab}. The problem now is to find those situations in which a solution, and so an operator, exists.

Before doing this, it is worthwhile to identify a particular choice of h_{ab} which yields a trivial (vanishing) $\beta^{*a}{}_b$. Such an h_{ab}, or u_{ab}, must be discarded because it does not lead to a useful local TT operator. Furthermore, any nontrivial solution is unique only up to terms involving this or any other trivial h_{ab}. Suppose we calculate $\beta^{*a}{}_b$ when

$$h_{ab} = 2g_{ab}H,$$

that is, h_{ab} is proportional to the base metric. The quantities $\Gamma^{*c}{}_{ab}$, $R^*{}_{ab}$, and $S^*{}_{ab}$ and $\beta^{*a}{}_d$ reduce to

$$\Gamma^{*c}{}_{ab} = (\overline{\nabla}_a H)\delta^c_b + (\overline{\nabla}_b H)\delta^c_a - \overline{g}_{ab}(\overline{\nabla}^c H),$$
$$R^*{}_{ab} = -\overline{\nabla}_a\overline{\nabla}_b H - \overline{g}_{ab}\underline{\Delta} H,$$
$$S^*{}_{ab} = -\overline{\nabla}_a\overline{\nabla}_b H,$$
$$\beta^{*a}{}_b = [akl]\{\overline{\nabla}_l\overline{\nabla}_k\overline{\nabla}_b H - \overline{S}_{lb}\overline{\nabla}_k H + \overline{g}_{kb}\overline{S}_{lc}\overline{\nabla}^c H\}.$$

To show that $\beta^{*a}{}_b$ vanishes, we first make use of the rule for interchanging covariant derivatives,

$$[akl]\overline{\nabla}_l\overline{\nabla}_k\overline{\nabla}_b H = +\tfrac{1}{2}[akl]\overline{R}^s{}_{bKl}\overline{\nabla}_s H.$$

Next we use the fact that in 3-spaces R_{abkl} reduces to

$$\bar{S}_{ak}\bar{g}_{bl} + \bar{S}_{bl}\bar{g}_{ak} - \bar{S}_{al}\bar{g}_{bk} - \bar{S}_{bk}\bar{g}_{al},$$

where $\bar{S}_{ab} = \bar{R}_{ab} - \frac{1}{4}\bar{g}_{ab}\bar{R}$. Therefore

$$[akl]\bar{\nabla}_l\bar{\nabla}_k\bar{\nabla}_b H = [akl]\{\bar{S}_{lb}\bar{\nabla}_k H - \bar{g}_{kb}\bar{S}_{lc}\bar{\nabla}^c H\}.$$

Clearly, then $\beta^{*a}{}_b$ vanishes, and $h_{ab} = 2H\bar{g}_{ab}$ is a trivial choice. It should be remarked that this fact could have been inferred from the knowledge that under infinitesimal conformal transformations $\delta g_{ab} = \lambda g_{ab}$, the mixed-index conformal curvature tensor is invariant:

$$\delta\beta^a{}_b = 0.$$

Now the equations of constraint on u_{ab} will be solved in order to determine when $\beta^{*a}{}_b$ is a TT operator. Because these equations relate the eigenvalues of u_{ab} to those of $\bar{\beta}^a{}_b$, the analysis will be broken down into the following three cases:

(A) No eigenvalue ($\bar{\beta}_A$) vanishes.
(B) One eigenvalue (say $\bar{\beta}_3$) vanishes.
(C) Two (and so all) eigenvalues vanish.

In case (A) there are only two solutions to Eqs. (10) and (11);

$$\alpha_1 = 0,$$

$$\bar{\beta}_1 = c\bar{\beta}_2 \quad (c = \text{const}).$$

The first one, $\alpha_1 = 0$, implies also that $\alpha_2 = 0$, and so u_{ab} is of the form

$$u_{ab} = u_3 \sum_{A=1}^{3} \underset{A}{\xi}_a \underset{A}{\xi}_b = u_3\bar{g}_{ab}.$$

This solution must be rejected because it leads to $\beta^{*a}{}_b = 0$, as just shown. The second solution, $\bar{\beta}_1 = c\bar{\beta}_2$, yields $\alpha_2 = -c\alpha_1$ and so

$$u_1 = \alpha_1 + u_3, \quad u_2 = -c\alpha_1 + u_3.$$

Thus u_{ab} becomes

$$u_{ab} = \alpha_1\left[\underset{1}{\xi}_a\underset{1}{\xi}_b - c\underset{2}{\xi}_a\underset{2}{\xi}_b\right] + u_3\bar{g}_{ab},$$

with α_1 arbitrary. The second term, being proportional to the base metric, can be deleted with no loss of generality. In conclusion, then, when none of the eigenvalues of $\bar{\beta}^a{}_b$ vanish, a nontrivial local TT operator exists if only if these eigenvalues are in constant ratio to one another. An appropriate choice for h_{ab} is

$$h_{ab} = (\bar{\beta}_2\underset{1}{\xi}_a\underset{1}{\xi}_b - \bar{\beta}_1\underset{2}{\xi}_a\underset{2}{\xi}_b)\Phi,$$

and the generalization of the flat-space TT operator L_{ab} is

$$\hat{L}^a{}_b\Phi = (1/\sqrt{\bar{g}})\beta^{*a}{}_b[h\cdot\cdot].$$

In case (B) one of the eigenvalues of the conformal curvature tensor vanishes. We will take this one to be $\bar{\beta}_3$, and so $\bar{\beta}_1 + \bar{\beta}_2 = 0$ because of the tracelessness of $\bar{\beta}^a{}_b$. The equations which constrain u_{ab}, Eqs. (10) and (11), then become

$$\bar{\beta}_1\partial_c(\alpha_1 - \alpha_2) = 0, \quad \bar{\beta}_1(\alpha_1 - \alpha_2) = 0.$$

The only solution to these equations is

$$\alpha_1 = \alpha_2,$$

and this implies that $u_1 = u_2$. Therefore, u_{ab} is

$$u_{ab} = u_1(\underset{1}{\xi}_a\underset{1}{\xi}_b + \underset{2}{\xi}_a\underset{2}{\xi}_b) + u_3\underset{3}{\xi}_a\underset{3}{\xi}_b,$$

with u_1 and u_3 arbitrary. Because terms proportional to $\bar{g}_{ab}$ are irrelevant, u_{ab} can also be written

$$u_{ab} = \underset{3}{\xi}_a\underset{3}{\xi}_b.$$

Therefore, when one of the eigenvalues of $\bar{\beta}^a{}_b$, vanishes, say $\bar{\beta}_3$, a generalized L_{ab} exists:

$$\hat{L}^a{}_b\Phi = (1/\sqrt{\bar{g}})\beta^{*a}{}_b[h\cdot\cdot],$$

with an appropriate choice for h_{ab} being

$$h_{ab} = \underset{3}{\xi}_a\underset{3}{\xi}_b\Phi.$$

Finally, in case (C), all of the eigenvalues of $\bar{\beta}^a{}_b$ vanish and of course $\bar{\beta}^a{}_b$ itself vanishes. The constraints upon u_{ab} are now automatically satisfied and so a generalized L_{ab} always exists:

$$\hat{L}^a{}_b\Phi = (1/\sqrt{\bar{g}})\beta^{*a}{}_b[h\cdot\cdot].$$

Here $h_{ab} = u_{ab}\Phi$ and u_{ab} is totally arbitrary. A natural choice for the tensor u_{ab} is

$$u_{ab} = F(\phi)r_a r_b,$$

where r_a is the (covariant) position vector in the associated flat space whose metric is

$$f_{ab} = \phi^{-4}\bar{g}_{ab}.$$

A tedious calculation shows that when $F(\phi) = 2\phi^4$, the operator $\hat{L}_{ab}$ extracted from $\beta^{*a}{}_b$ is none other than the one that can be obtained by conformally scaling the corresponding flat-space operator L_{ab}.[13] Of course in this case not one but two TT operators exist, as pointed out earlier in this section.

If the results of cases (A), (B), and (C) are collected, a simple conclusion can be drawn. In the first two cases the operator $\hat{L}_{ab}$ exists whenever $\bar{\beta}^a{}_b$ has not more than one independent eigenvalue. In the third case when there are no independent eigenvalues at all, not one but two operators exist.

V. WHEN DOES THE SECOND OPERATOR M_{ab} EXIST?

A natural question to ask at this stage is whether a generalization of the flat-space operator M_{ab} exists in spaces more general than flat or conformally flat ones. Though no definite answer is known, some facts about the problem are clear. First of all, the curl of $(1/\sqrt{\bar{g}})\beta^{*a}{}_b$ is not in general a TT operator as *is* the case in flat spaces. In fact, even in conformally flat spaces the curl of $\tilde{L}_{ab} = \phi^{-2}L_{ab}$ is not TT and is not the same quantity as $\tilde{M}_{ab} = \phi^{-2}M_{ab}$. Furthermore, efforts to add compensating terms to the curl of $(1/\sqrt{\bar{g}})\beta^{*a}{}_b$ in order to make it TT only lead to *nonlocal* conditions upon the terms added. This situation is clearly unacceptable if the program outlined in this paper is to succeed, and so it is conjectured that $\tilde{M}_{ab}$ is the only generalization of the flat-space operator M_{ab}.

It is interesting that the conjectured nonexistence of $\tilde{M}_{ab}$ in spaces not conformally flat (or flat) closely parallels York's conjecture[12] that the conformal curvature tensor singles out the field excitations of the gravitational field. As has been illustrated in Ref. 7, there is a correlation between the number of field excitations exhibited by a 3-metric and the number of independent eigenvalues of the associated conformal curvature tensor. When all three eigenvalues are zero (i.e., no independent eigenvalues), there are no field modes excited. In this case the space is flat or conformally flat and so two orthogonal local TT operators exists for the construction of wavelike tensors. If either one eigenvalue is zero or all eigenvalues are nonzero but are in constant ratio to one another, then there is but one independent eigenvalue. In these two cases the flat-space operator L_{ab} can be generalized to $L_{ab} = (1/\sqrt{g})\beta^{*a}{}_b$ [recall cases (A) and (B) in Sec. 4]. Finally, when no eigenvalues are zero and any two of these are linearly independent, not even L_{ab} can be generalized by the methods considered in this paper. Thus, as the number of excited field modes increases from 0 to 1 to 2, it appears that the number of local TT operators available decreases from 2 to 1 to 0, respectively.

This conjectured relation between the number of excited field modes and the number of available TT operators can be illustrated with examples. In the case of the Schwarzschild solution natural slicing[14] yields a conformally flat 3-space. The conformal curvature vanishes identically and so all of its eigenvalues are zero. Consistent with this, the number of excited field modes is zero, there being only a nondynamical Newtonian-like potential appearing as a conformal factor. Two orthogonal local TT operators are available for constructing wavelike perturbations of the metric, or for constructing candidates for the TT part of the field momentum.

As an example of the case of one field mode being excited, consider the exact plane-wave solution with one polarization state.[15] The natural slicing $dt = 0$ yields a 3-metric whose conformal curvature has one independent eigenvalue corresponding to the single wavelike mode. One local TT operator is available for constructing TT tensors. A second example of this case is the Kerr–Newman geometry in Boyer–Lindquist coordinates with slicing $dt = 0$.[16] Again the conformal curvature has one independent eigenvalue. There is a field mode excited in this situation, but it is not wavelike. As can easily be shown,[7] the nonvanishing eigenvalues of $\beta^{a}{}_b$ are not time-dependent and furthermore they fall off much too fast to describe a wave carrying off mass–energy. This excited mode is analogous to the stationary magnetic field mode in the vicinity of a rotating charged sphere. At any rate, one local TT operator is available in the case of the Kerr–Newman geometry.

Lastly, the case of two excited field modes is nicely illustrated by the exact plane-wave solution with two polarization states.[17] Here the conformal curvature has two independent eigenvalues. None of the eigenvalues is zero and no two of them are linearly dependent. In this case not even $\beta^{*a}{}_b$ will yield a local TT operator and apparently none is available.

APPENDIX A: DEBYE POTENTIALS FOR VECTORS

In flat spaces an arbitrary vector can be decomposed into three terms as follows[18]:

$$V_a = L_a \Phi + M_a \Psi + \partial_a \lambda.$$

The operator L_a is essentially the angular momentum operator of quantum mechanics,

$$L_a = -[abc]x_b \partial_c,$$

and M_a is its curl,

$$M_a = [abc]\partial_b L_c.$$

The three parts of V_a are mutually orthogonal, provided that the three potentials fall off at least as fast as $1/r$.

The scalar λ is the usual "longitudinal" potential of V_a, while Φ and Ψ are the Debye potentials from which its transverse part is constructed. Given V_a one can quickly find the three potentials by solving elliptic differential equations:

$$\Delta\lambda = \partial_a V_a, \quad L^2\Phi = L_a V_a,$$
$$L^2\Psi = -x_a V_a + x_a \partial_a \lambda.$$

As is well known, Debye potentials are of great practical use in electromagnetism, especially in the construction of multipole moments. This subject is adequately covered by Jackson.[5] In constructing source-free solutions of Maxwell's equations it is easily shown that if the vector potential is written

$$A_a = L_a \Phi + M_a \Psi + \partial_a \lambda,$$

then the equations of motion and constraint become

$$\Box\Phi = 0, \quad \Box\Psi = 0, \quad \Delta(\dot{\lambda} + \Omega) = 0,$$

where Ω is the scalar potential of electromagnetism.

APPENDIX B: THE FLAT-SPACE OPERATORS L_{ab} and M_{ab}

Of the multitude of expressions for the TT operators L_{ab} and M_{ab}, the most convenient appear to be (in Cartesian coordinates)[4]

$$L_{ab} = \partial_{(a}L_{b)} + M_{(a}L_{b)} = L_{(a}\partial_{b)} + L_{(a}M_{b)},$$
$$M_{ab} = [akl]\partial_k L_{lb}.$$

These easily generalized to curvilinear coordinates.

If a Cartesian TT tensor $h^{\rm TT}_{ab}$ is written in the form

$$h^{\rm TT}_{ab} = L_{ab}\Phi + M_{ab}\Psi,$$

then the Debye potentials Φ and Ψ can be recovered by solving the bielliptic differential equations

$$\tfrac{1}{2}L^2(L^2+2)\Phi = \hat{\Phi} = -x_b L_b h^{\rm TT}_{ab},$$
$$\tfrac{1}{2}L^2(L^2+2)\Psi = \hat{\Psi} = x_a x_b h^{\rm TT}_{ab}.$$

The operaotr L^2 is the square of the operator L and is the "angular part" of the ordinary Laplacian, i.e.,

$$L^2 = \frac{1}{\sin\theta}\partial_\theta \sin\theta\,\partial_\theta + \frac{1}{\sin^2\theta}\partial^2\phi.$$

Because L^2 annihilates terms of the type $f(r)Y_{00}(\theta,\phi)$ and (L^2+2) annihilates those like $f(r)Y_{lm}(\theta,\phi)$, the bielliptic operator $\tfrac{1}{2}L^2(L^2+2)$ reduces both pure-monopole and pure-dipole terms to zero. It can also be shown that L_{ab} and M_{ab} annihilate these same kinds of terms.[7] Another important property common to both of the TT operators L_{ab} and M_{ab} is that the Laplacian commutes with each of them in flat space.

APPENDIX C: SIX-FOLD DECOMPOSITION OF CARTESIAN TENSORS

The York decomposition,[2] when applied to a Cartesian tensor h_{ab}, breaks this tensor into three mutually orthogonal terms:

$$h_{ab} = h^{TT}_{ab} + (LX)_{ab} + \tfrac{1}{3}\delta_{ab} h,$$

$$(LX)_{ab} = \partial_a X_b + \partial_b X_a - \tfrac{2}{3}\delta_{ab}\partial_c X_c.$$

The TT part can be written in terms of the two Debye potentials Φ and Ψ:

$$h^{TT}_{ab} = L_{ab}\Phi + M_{ab}\Psi.$$

The trace of h_{ab} represents of third potential.

I will now show how it is possible to extract three more potentials and their corresponding operators from the longitudinal term $(LX)_{ab}$. First, write the vector X_a in terms of the vector operators discussed in Appendix A;

$$X_a = L_a\alpha + M_a\beta + \partial_a\gamma.$$

Then calculate $(LX)_{ab}$ to obtain

$$(LX)_{ab} = 2\partial_{(a}L_{b)}\alpha + 2(\partial_{(a}M_{b)}\beta + 2(\partial_{(a}\partial_{b)} - \tfrac{1}{3}\delta_{ab}\Delta)\gamma.$$

We now have h_{ab} decomposed into six terms, each term an operator premultiplying a scalar potential[7]:

$$h_{ab} = L_{ab}\Phi + M_{ab}\Psi + N_{ab}\alpha + O_{ab}\beta + P_{ab}\gamma + Q_{ab}h,$$

$$N_{ab} = 2\partial_{(a}L_{b)}, \quad O_{ab} = 2\partial_{(a}M_{b)},$$

$$P_{ab} \equiv 2(\partial_{(a}\partial_{b)} - \tfrac{1}{3}\delta_{ab}\Delta), \quad Q_{ab} \equiv \tfrac{1}{3}\delta_{ab}.$$

A little investigation reveals that the six terms are mutually orthogonal provided that the potentials fall off at least as fast as $1/r$. This condition can be met whenever the gravitational field has a localized source and is asymptotically flat. (The $1/r$–fall-off condition can be loosened somewhat for at least some of the potentials.[7])

ACKNOWLEDGMENTS

I wish to express my deepest gratitude to J. W. York, Jr., who made the writing of this paper possible through his patience and hospitality. Many thanks also to Comer Duncan, with whom I had many helpful discussions.

[1] J. W. York, Jr., "Conformally Invariant Orthogonal Decomposition of Symmetric Tensors on Reimannian Manifolds and the Initial-Value Problem in General Realtivity," J. Math. Phys. **14**, 456 (1973).

[2] J. W. York, Jr., "Covariant Decomposition of Symmetric Tensors in the Theory of Graviation," Ann. Inst. Henri Poincare **XXI**, 319 (1974).

[3] Special thanks to Comer Duncan for his helpful discussion of this subject.

[4] W. Campbell and T. Morgan, "Debye Potentials for the Gravitational Field," Physica **53**, 264 (1971).

[5] J. Jackson, *Classical Electrodynamics* (Wiley, New York, 1962).

[6] A. Nisbet, "Hertzian Electromagnetic Potentials and Associated Gauge Transformations," Proc. Soc. London Ser. A **231**, 250 (1955).

[7] See D. Monroe, "Conformal Curvature of Spacelike Cauchy Data and Gravitational Radiation," Ph. D. thesis, University of North Carolina, Chapel Hill (1976).

[8] S. Teukolsky, unpublished note (1976).

[9] D. Brill, "On the positive definite mass of the Bondi-Wèber-Wheeler time-symmetric gravitational waves," Ann. Phys. (NY) 7, 466–483 (1959).

[10] The conformal scaling technique referred to is presented in its most general form in Ref. 1.

[11] S. Teukolsky, "Rotating Black Holes: Separable Wave Equations for Gravitational and Electromagnetic Perturbations," Phys. Rev. Lett. **29**, 1114 (1973).

[12] J. W. York, Jr., "Gravitational Degrees of Freedom and the Initial-Value Problem," Phys. Rev. Lett. **26**, 1656 (1971).

[13] The details of this calculation can be found in Ref. 7. It is not presented here because the result is not crucial to the content of the paper.

[14] See C. Misner, K. Thorne, and J. Wheeler, *Gravitation* (Freeman, San Francisco, 1973), p.840.

[15] Reference 14, p. 957.

[16] Reference 14, p. 877.

[17] H. Bondi, F. Pirani, and I. Robinson, "Gravitational Waves in General Relativity. III," Proc. R. Soc. London Ser. A **251**, 519 (1959).

[18] L. Eyges, *The Classical Electromagnetic Field* (Addison-Wesley, Reading, MA, 1972).

Weyl tensors for asymmetric complex curvatures

Colber G. Oliveira
Departmento de Física, Universidade de Brasília, Brazil

(Received 8 October 1980; accepted for publication 18 March 1981)

Considering a second-rank Hermitian field tensor and a general Hermitian connection we construct the associated complex curvature tensor. The Weyl tensor that corresponds to this complex curvature is determined. The formalism is applied to the Weyl unitary field theory and to the Moffat gravitational theory.

PACS numbers: 04.20.Cv, 04.50. + h

I. INTRODUCTION

The study of complex geometry in relativity is currently a subject of increasing interest.[1] In the definition of algebraically special spaces the null tetrad formalism of Newman–Penrose is generalized to a complex null tetrad.[2] The curvature and Weyl tensors become complex objects and their algebraic properties have been studied.[3] Presently we consider the problem of determination of the Weyl tensors associated with a complex curvature derivable from a general Hermitian connection. Along with the complex affinity we consider a Hermitian field tensor $g_{\mu\nu}$ which generalizes the usual symmetric metric tensor of general relativity. No relation is imposed, *a priori*, between the connection and the field tensor. Each particular relationship between these quantities is specific to each particular theory as for instance general relativity, the Weyl theory, or complex formalisms such as the Moffat gravitational theory. It is shown that the general curvature is a fourth-rank tensor skew symmetric in the last pair of indices, which contains several complex Ricci tensors.

The method used for the determination of the complex Weyl tensor is a decomposition process which separates the curvature into components with a well determined symmetry. This method has the advantage of providing a direct method for the determination of the spinors associated to the components of the complex curvature and of the complex Weyl tensors.

It is the purpose of this paper to apply these results to some unitary field theories of interest, such as the Weyl theory and the Moffat theory. However, the general geometric results may also be applied as an extension of the use of complex geometries in "complex relativity," since presently the *vierbeins* are completely general complex quantities that generate the Hermitian field tensor. The formula giving the field tensor in terms of the *vierbeins* is the same formula used in the Moffat theory.

Although the complex *vierbeins* are implicitly contained in the definition of the Hermitian metric field, they are not used in the present paper since all results apply directly to the connection and curvature as functions of the Hermitian field tensor. As is well known, the use of *vierbeins* is necessary for the determination of the spinors associated to the curvature and Weyl tensors, and as we have said before, the present formalism is easily translated in terms of complex *vierbeins* and two-component spinors.

In Sec. II we define the affine curvature tensor constructed with a Hermitian connection, and decompose this complex tensor into components with determined symmetry. The Ricci tensors and scalars of curvature are obtained by contractions with the Hermitian field tensor. In Sec. III we derive the complex Weyl tensors associated with each part of the previous decomposition. Finally, in Sec. IV we apply the formalism to some unitary field theories of interest.

II. THE CONNECTION AND THE CURVATURE

A general asymmetric connection may give rise to different forms of defining the parallel transport of vectors. For a contravariant vector field we may use the two possible definitions:

$$\delta A^{\mu} = -\Gamma^{\mu}_{\alpha\beta} A^{\alpha} dx^{\beta}, \quad \delta A^{\mu} = -\Gamma^{\mu}_{\beta\alpha} A^{\alpha} dx^{\beta}. \tag{2.1}$$

In order to distinguish these two forms of definition of the variation in the coordinates of the vector A under infinitesimal parallel transport, we use the notation suggested by Einstein[4]: we denote the first of the Eqs. (2.1) by $\delta A^{\overset{\mu}{+}}$ and the second one by $\delta A^{\overset{\mu}{-}}$. Accordingly, two possible definitions of covariant derivatives follow as

$$dx^{\nu} A^{\overset{\mu}{+}}_{;\nu} = dA^{\mu} - \delta A^{\overset{\mu}{+}}, dx^{\nu} A^{\overset{\mu}{-}}_{;\nu} = dA^{\mu} - \delta A^{\overset{\mu}{-}};$$

explicitly one has

$$A^{\overset{\mu}{+}}_{;\nu} = A^{\mu}_{,\nu} + \Gamma^{\mu}_{\alpha\nu} A^{\alpha}, \tag{2.2}$$

$$A^{\overset{\mu}{-}}_{;\nu} = A^{\mu}_{,\nu} + \Gamma^{\mu}_{\nu\alpha} A^{\alpha}. \tag{2.3}$$

The expressions for $\delta A_{\underset{+}{\mu}}$, $\delta A_{\underset{-}{\mu}}$ are directly derived from the definitions of Eq. (2.1) by imposing the condition that the length of the vector A is unchanged under parallel transport. Consequently, covariant derivatives of the classes $+$ and $-$ may be written for an arbitrary tensor field.

The asymmetric connection may be decomposed into symmetric and antisymmetric parts. In this paper we take the antisymmetric part of the connection as a purely imaginary third-rank tensor. With this choice the connection $\Gamma^{\mu}_{\alpha\beta}$ becomes Hermitian with respect to the covariant indices $\Gamma^{*\mu}_{\alpha\beta} = \Gamma^{\mu}_{\beta\alpha}$. Complex tensor fields were used by Einstein[4] in his complex nonsymmetric unitary theory. Recently this theory has been reviewed and generalized by Moffat.[5] In these theories the metric tensor is a second-order Hermitian tensor. This complex property of the metric allows for a simple expression of $g_{\mu\nu}$ in terms of a field of complex *vierbeins*.[6] Thus, one of the advantages of the complex formulation is the direct possibility of translating the theory in terms of

 0022-2488/81/092001-05$01.00

two-component spinors; if we work with a real asymmetric metric the process of determination of the corresponding spinors is more complicated.

The affine curvature tensor may be introduced by the commutator

$$A^{\sigma}_{\ ;\mu;\nu}-A^{\sigma}_{\ ;\nu;\mu}=R^{\ \sigma}_{\lambda\mu\nu}+2\Gamma^{\lambda}_{[\mu\nu]}A^{\sigma}_{\ ;\lambda}$$

and has the value

$$R^{\ \sigma}_{\lambda\mu\nu}(\Gamma)=\partial_\nu\Gamma^{\sigma}_{\lambda\mu}-\partial_\mu\Gamma^{\sigma}_{\lambda\nu}+\Gamma^{\rho}_{\lambda\mu}\Gamma^{\sigma}_{\rho\nu}-\Gamma^{\rho}_{\lambda\nu}\Gamma^{\sigma}_{\rho\mu}. \tag{2.4}$$

This same curvature tensor may be derived by calculating the variation in the components of a vector field around an infinitesimal closed loop, and is also called, for this reason, the curvature of rotation. The general curvature tensor $R^{\ \sigma}_{\lambda\mu\nu}$ is presently a complex fourth-rank tensor antisymmetric in the indices μ, ν. Accordingly, we may write

$$R^{\ \sigma}_{\lambda\mu\nu}(\Gamma)=T^{\ \sigma}_{\lambda\mu\nu}+iS^{\ \sigma}_{\lambda\mu\nu}, \tag{2.5}$$

$$T^{\ \sigma}_{\lambda\mu\nu}=-T^{\ \sigma}_{\lambda\nu\mu},\quad S^{\ \sigma}_{\lambda\mu\nu}=-S^{\ \sigma}_{\lambda\nu\mu}.$$

The explicit expression of these two tensors is

$$T^{\ \sigma}_{\lambda\mu\nu}=G^{\ \sigma}_{\lambda\mu\nu}+\Gamma^{\sigma}_{[\alpha\nu]}\Gamma^{\alpha}_{[\lambda\mu]}-\Gamma^{\sigma}_{[\alpha\mu]}\Gamma^{\alpha}_{[\lambda\nu]}+2\Gamma^{\alpha}_{[\nu\mu]}\Gamma^{\sigma}_{[\lambda\alpha]}, \tag{2.6}$$

$$S^{\ \sigma}_{\lambda\mu\nu}=\Gamma^{\sigma}_{[\lambda\mu];\nu}-\Gamma^{\sigma}_{[\lambda\nu];\mu}, \tag{2.7}$$

where $G^{\ \sigma}_{\lambda\mu\nu}$ is an affine curvature tensor constructed with the symmetric part of the affinity $\Gamma^{\alpha}_{\mu\nu}$. We mention that here the quantity $\Gamma^{\alpha}_{(\mu\nu)}$ is not necessarily equal to the Christoffel symbols. The formulas (2.6) and (2.7) show explicitly the covariance property of the decomposition (2.5). There is *a priori* no relationship between the connection $\Gamma^{\alpha}_{\mu\nu}$ and the Hermitian tensor field $g_{\mu\nu}$. This relationship will be characteristic of each particular theory considered, as for instance, general relativity, a semi-metric theory such as the unitary Weyl theory, or the asymmetric unitary theory suggested by Moffat. Thus, we keep the formalism in a general form, but we need the Hermitian tensor field for lowering the contravariant index of curvature, since a discussion of the symmetry properties of the curvature will be necessary.

The conventions for lowering and raising indices with a Hermitian metric are well known;

$$A_\alpha=g_{\alpha\beta}A^{\beta},\quad A^{\alpha}=g^{\alpha\beta}A_\beta,$$

$$g^{\lambda\alpha}g_{\alpha\beta}=g^{\alpha\lambda}g_{\beta\alpha}=\delta^{\lambda}_{\beta}.$$

Thus,

$$R_{\alpha\lambda\mu\nu}(\Gamma)=g_{\alpha\sigma}R^{\ \sigma}_{\lambda\mu\nu}=F_{\alpha\lambda\mu\nu}+iH_{\alpha\lambda\mu\nu}. \tag{2.8}$$

In Sec. III we will determine the Weyl tensor associated with the general curvature tensor $R_{\alpha\lambda\mu\nu}$. For the determination of this tensor we will use the method of decomposition of the curvature into a sum of factors with a determined symmetry. The following notation will be used:

$$S_{\mu\nu}=S_{\underline{\mu\nu}}+S_{\underset{\vee}{\mu\nu}},$$

$$S^{*}_{\underline{\mu\nu}}=S_{\underline{\nu\mu}},\quad S^{*}_{\underset{\vee}{\mu\nu}}=-S_{\underset{\vee}{\nu\mu}}.$$

$S_{\mu\nu}$ being any arbitrary second rank complex tensor, and $S_{\underline{\mu\nu}}$, $S_{\underset{\vee}{\mu\nu}}$ denote its Hermitian and Anti-Hermitian parts. This decomposition may be extended for a fourth rank complex tensor according to

$$R_{\mu\nu\rho\sigma}=R_{\underline{\mu\nu}\,\underline{\rho\sigma}}+R_{\underline{\mu\nu}\,\underset{\vee}{\rho\sigma}}+R_{\underset{\vee}{\mu\nu}\,\underline{\rho\sigma}}+R_{\underset{\vee}{\mu\nu}\,\underset{\vee}{\rho\sigma}}, \tag{2.9}$$

where, for instance,

$$R_{\underline{\mu\nu}\,\underline{\rho\sigma}}=\tfrac{1}{4}(R_{\mu\nu\rho\sigma}+R^{*}_{\mu\nu\sigma\rho}+R^{*}_{\nu\mu\rho\sigma}+R_{\nu\mu\sigma\rho}),$$

with similar expressions for the remaining components in Eq. (2.9). From Eq. (2.8) we have for the several terms in Eq. (2.9) (Ref. 7)

$$R_{\underline{\mu\nu}\,\underline{\rho\sigma}}=iH_{[\mu\nu][\rho\sigma]}, \tag{2.10}$$

$$R_{\underset{\vee}{\mu\nu}\,\underline{\rho\sigma}}=iH_{(\mu\nu)[\rho\sigma]}, \tag{2.11}$$

$$R_{\underline{\mu\nu}\,\underset{\vee}{\rho\sigma}}=F_{(\mu\nu)[\rho\sigma]}, \tag{2.12}$$

$$R_{\underset{\vee}{\mu\nu}\,\underset{\vee}{\rho\sigma}}=F_{[\mu\nu][\rho\sigma]}. \tag{2.13}$$

It is of interest to introduce the quantities

$$B_{\mu\nu\rho\sigma}=\tfrac{1}{2}(F_{[\mu\nu][\rho\sigma]}+F_{[\rho\sigma][\mu\nu]}), \tag{2.14}$$

$$E_{\mu\nu\rho\sigma}=\tfrac{1}{2}(F_{[\mu\nu][\rho\sigma]}-F_{[\rho\sigma][\mu\nu]}), \tag{2.15}$$

$$J_{\mu\nu\rho\sigma}=\tfrac{1}{2}(F_{(\mu\nu)[\rho\sigma]}+F_{(\rho\sigma)[\mu\nu]}), \tag{2.16}$$

$$I_{\mu\nu\rho\sigma}=\tfrac{1}{2}(F_{(\mu\nu)[\rho\sigma]}-F_{(\rho\sigma)[\mu\nu]}), \tag{2.17}$$

$$L_{\mu\nu\rho\sigma}=(i/2)(H_{[\mu\nu][\rho\sigma]}+H_{[\rho\sigma][\mu\nu]}), \tag{2.18}$$

$$M_{\mu\nu\rho\sigma}=(i/2)(H_{[\mu\nu][\rho\sigma]}-H_{[\rho\sigma][\mu\nu]}), \tag{2.19}$$

$$K_{\mu\nu\rho\sigma}=(i/2)(H_{(\mu\nu)[\rho\sigma]}+H_{(\rho\sigma)[\mu\nu]}), \tag{2.20}$$

$$U_{\mu\nu\rho\sigma}=(i/2)(H_{(\mu\nu)[\rho\sigma]}-H_{(\rho\sigma)[\mu\nu]}). \tag{2.21}$$

Accordingly, one gets

$$R_{\underline{\mu\nu}\,\underline{\rho\sigma}}=L_{\mu\nu\rho\sigma}+M_{\mu\nu\rho\sigma}, \tag{2.22}$$

$$R_{\underset{\vee}{\mu\nu}\,\underline{\rho\sigma}}=K_{\mu\nu\rho\sigma}+U_{\mu\nu\rho\sigma}, \tag{2.23}$$

$$R_{\underline{\mu\nu}\,\underset{\vee}{\rho\sigma}}=J_{\mu\nu\rho\sigma}+I_{\mu\nu\rho\sigma}, \tag{2.24}$$

$$R_{\underset{\vee}{\mu\nu}\,\underset{\vee}{\rho\sigma}}=B_{\mu\nu\rho\sigma}+E_{\mu\nu\rho\sigma}. \tag{2.25}$$

The tensors B, E, J, and I satisfy the symmetry properties

$$B_{\mu\nu\rho\sigma}=-B_{\nu\mu\rho\sigma}=-B_{\mu\nu\sigma\rho}=B_{\rho\sigma\mu\nu}, \tag{2.26}$$

$$E_{\mu\nu\rho\sigma}=-E_{\nu\mu\rho\sigma}=-E_{\mu\nu\sigma\rho}=-E_{\rho\sigma\mu\nu}, \tag{2.27}$$

$$J_{\mu\nu\rho\sigma}=J_{\rho\sigma\mu\nu},\quad I_{\mu\nu\rho\sigma}=-I_{\rho\sigma\mu\nu}. \tag{2.28}$$

The tensors L, M, K, and U satisfy the same sequence of symmetries. The curvature $R_{\mu\nu\rho\sigma}$ has 192 independent components. According to our decomposition this total number of components is separated into the tensor B with 21 components, E with 15, $J+I$ with 60 and the imaginary parts L, M, and $K+U$ with the same number of independent elements. We write down the Ricci tensors that correspond to these eight elements which compose the curvature in sequence. All contractions are carried out using the complex metric $g_{\mu\nu}$. For the components $B_{\mu\nu\rho\sigma}$ we have from Eq. (2.26)

$$B_{\nu\sigma}=g^{\rho\nu}B_{\mu\nu\rho\sigma}=B^{*}_{\sigma\nu}. \tag{2.29}$$

Thus, the Ricci tensor of the curvature $B_{\mu\nu\rho\sigma}$ is a Hermitian second-rank tensor. In what follows all contractions are taken as in Eq. (2.29). For the remaining components we find

$$E_{\nu\sigma}=-E^{*}_{\sigma\nu},\ J_{\nu\sigma}=J^{*}_{\sigma\nu},\ I_{\nu\sigma}=-I^{*}_{\sigma\nu},$$

$$L_{\nu\sigma} = -L^*_{\sigma\nu}, M_{\nu\sigma} = M^*_{\sigma\nu},$$
$$LK_{\nu\sigma} = -K^*_{\sigma\nu},\ U_{\nu\sigma} = U^*_{\sigma\nu}.$$

A further contraction generates the scalars of curvature. They satisfy the conditions

$$B = B^*, E = -E^*, J = J^*, I = -I^*, L = -L^*,$$
$$M = M^*, K = -K^*, U = U^*.$$

With these results we can determine the several elements that compose the Weyl tensor associated with the general complex curvature.

III. THE WEYL TENSOR OF THE CURVATURE $R_{\mu\nu\rho\sigma}$

In the determination of the Weyl tensor of the curvature $R_{\mu\nu\rho\sigma}$ it is of interest to introduce the four-index quantity $g_{\mu\nu\rho\sigma}$ given by

$$g_{\mu\nu\rho\sigma} = g_{\mu\rho} g_{\nu\sigma} - g_{\mu\sigma} g_{\nu\rho}; \tag{3.1}$$

since $g_{\mu\nu} = \underline{g_{\mu\nu}}$ we have

$$g_{\mu\nu\rho\sigma} = -g_{\mu\nu\sigma\rho} = -g_{\nu\mu\rho\sigma} = g^*_{\rho\sigma\mu\nu}. \tag{3.2}$$

The Weyl tensor will be composed of four elements ${}^{(i)}W_{\mu\nu\rho\sigma} (i = 1\cdots4)$; each of these elements corresponds to some combination of the several terms in Eqs. (2.22)–(2.25). First we select the components B and M and define the tensors

$$A_{\mu\nu\rho\sigma} = B_{\mu\nu\rho\sigma} + M_{\mu\nu\rho\sigma} = A^*_{\rho\sigma\mu\nu},$$
$$P_{\mu\nu\rho\sigma} = \tfrac{1}{2}(g_{\mu\nu\lambda\sigma} a^{\lambda}_{\cdot\rho} - g_{\mu\nu\lambda\rho} a^{\lambda}_{\cdot\sigma}) = P^*_{\rho\sigma\mu\nu},$$

where

$$a^{\lambda}_{\cdot\rho} = A^{\ \lambda}_{\cdot\rho} - (A/4)\delta^{\lambda}_{\rho},\ A = B + M = A^*.$$

The Weyl tensor "of the class (1)" is given by

$${}^{(1)}W_{\mu\nu\rho\sigma} = A_{\mu\nu\rho\sigma} - P_{\mu\nu\rho\sigma} - (A/12) g_{\mu\nu\rho\sigma}, \tag{3.3}$$

and has the same symmetries as the tensor $A_{\mu\nu\rho\sigma}$. In addition, it satisfies

$${}^{(1)}W_{\nu\sigma} = {}^{(1)}W^{\mu}_{\ \nu\mu\sigma} = 0.$$

Continuing with this process we select the components E and L and introduce the tensors

$$Q_{\mu\nu\rho\sigma} = E_{\mu\nu\rho\sigma} + L_{\mu\nu\rho\sigma} = -Q^*_{\rho\sigma\mu\nu},$$
$$T_{\mu\nu\rho\sigma} = \tfrac{1}{2}(g_{\mu\nu\lambda\sigma} q^{\lambda}_{\cdot\rho} - g_{\mu\nu\lambda\rho} q^{\lambda}_{\cdot\sigma}) = -T^*_{\rho\sigma\mu\nu},$$

with

$$q^{\lambda}_{\cdot\rho} = Q^{\ \lambda}_{\cdot\rho} - (Q/4)\delta^{\lambda}_{\rho},\ Q = E + L = -Q^*.$$

The Weyl tensor ${}^{(2)}W$ is given by the expression

$${}^{(2)}W_{\mu\nu\rho\sigma} = Q_{\mu\nu\rho\sigma} - T_{\mu\nu\rho\sigma} - (Q/12) g_{\mu\nu\rho\sigma}; \tag{3.4}$$

all symmetries presented by $Q_{\mu\nu\rho\sigma}$ are satisfied as well by ${}^{(2)}W_{\mu\nu\rho\sigma}$ and we have the conditions

$${}^{(2)}W_{\nu\sigma} = 0.$$

By a similar choice we finally take the tensors (J,U) and (I,K) and form the quantities:

$$\Omega_{\mu\nu\rho\sigma} = J_{\mu\nu\rho\sigma} + U_{\mu\nu\rho\sigma} = \Omega^*_{\rho\sigma\mu\nu},$$
$$\Lambda_{\mu\nu\rho\sigma} = I_{\mu\nu\rho\sigma} + K_{\mu\nu\rho\sigma} = -\Lambda^*_{\rho\sigma\mu\nu},$$
$$\Delta_{\mu\nu\rho\sigma} = \tfrac{1}{2}(g_{\mu\nu\lambda\sigma} \omega^{\lambda}_{\cdot\rho} - g_{\mu\nu\lambda\rho} \omega^{\lambda}_{\cdot\sigma}) = \Delta^*_{\rho\sigma\mu\nu},$$
$$\Phi_{\mu\nu\rho\sigma} = \tfrac{1}{2}(g_{\mu\nu\lambda\sigma} \phi^{\ \lambda}_{\cdot\rho} - g_{\mu\nu\lambda\rho} \phi^{\ \lambda}_{\cdot\sigma}) = -\phi^*_{\rho\sigma\mu\nu},$$

where

$$\omega^{\lambda}_{\cdot\rho} = \Omega^{\ \lambda}_{\rho} - (\Omega/4)\delta^{\lambda}_{\rho},\ \phi^{\ \lambda}_{\rho} = \Lambda^{\ \lambda}_{\cdot\rho} - (\Lambda/4)\delta^{\lambda}_{\rho}.$$

These elements generate the Weyl tensors "of the classes (3) and (4)" according to

$${}^{(3)}W_{\mu\nu\rho\sigma} = \Omega_{\mu\nu\rho\sigma} - \Delta_{\mu\nu\rho\sigma} - (\Omega/12) g_{\mu\nu\rho\sigma}, \tag{3.5}$$
$${}^{(4)}W_{\mu\nu\rho\sigma} = \Lambda_{\mu\nu\rho\sigma} - \Phi_{\mu\nu\rho\sigma} - (\Lambda/12) g_{\mu\nu\rho\sigma}. \tag{3.6}$$

${}^{(3)}W$ and ${}^{(4)}W$ have the same symmetry as the tensors Ω and Λ, and satisfy

$${}^{(3)}W_{\nu\sigma} = {}^{(4)}W_{\nu\sigma} = 0.$$

Thus, the Weyl tensor associated to the general, complex, curvature is given by the sum of the four elements ${}^{(i)}W$,

$$W_{\mu\nu\rho\sigma} = \sum_{i=1}^{4} {}^{(i)}W_{\mu\nu\rho\sigma},$$

and has the form

$$W_{\mu\nu\rho\sigma} = R_{\mu\nu\rho\sigma} - \tfrac{1}{2}[g_{\mu\nu\lambda\sigma}(R^{\ \lambda}_{\cdot\rho} - (R/4)\delta^{\lambda}_{\rho})$$
$$- g_{\mu\nu\lambda\sigma}(R^{\ \lambda}_{\cdot\sigma} - (R/4)\delta^{\lambda}_{\ \sigma})] - \tfrac{1}{12} R g_{\mu\nu\rho\sigma}, \tag{3.7}$$

where

$$R^{\lambda}_{\cdot\rho} = g^{\lambda\mu} g^{\beta\alpha} R_{\alpha\mu\beta\rho}.$$

The complex Weyl tensor $W_{\mu\nu\rho\sigma}$ has the symmetry property $W_{\mu\nu\rho\sigma} = -W_{\mu\nu\sigma\rho}$, which is the only symmetry property presented by the curvature tensor $R_{\mu\nu\rho\sigma}$. A formula like Eq. (3.7) could have been written without the necessity of going through the process of decomposition used in Sec. II and Sec. III. However, we have used this process of decomposition of the curvature $R_{\mu\nu\rho\sigma}$ since it allows directly the determination of the two-component spinors that correspond to the several parts of the curvature tensor and of the Weyl tensors ${}^{(i)}W$. Equation (3.7) along with the symmetry property of the Weyl tensor W do not contain sufficient information for the determination of these curvature spinors.

IV. APPLICATIONS OF THE FORMALISM

In the previous sections we have considered a general, complex, curvature associated with a general Hermitian connection $\Gamma^{\mu}_{\nu\lambda}$, and have used a Hermitian field tensor $g_{\mu\nu}$. No relation between the connection and the field tensor was used. Now we particularize this general formalism for two cases of interest: the asymmetric, complex, field theory of Moffat and the semi-metric, real, unitary theory proposed by Weyl.[8]

A. Moffat's theory

In this theory the metric and the connection are Hermitian objects which satisfy the field equations (in absence of fermion sources)

$$g_{\underset{+-}{\mu\nu},\lambda} - g_{\alpha\nu} \Gamma^{\alpha}_{\cdot\mu\lambda} - g_{\mu\alpha} \Gamma^{\alpha}_{\cdot\lambda\nu} = 0, \tag{4.1}$$
$$g^{[\mu\nu]}_{\ ,\nu} = 0, \tag{4.2}$$
$${}^*R_{(\mu\nu)}(\Gamma) = 0, \tag{4.3}$$
$${}^*R_{[\mu\nu]}(\Gamma) = -\tfrac{2}{3} w_{[\mu,\nu]}. \tag{4.4}$$

The quantity w_{μ} is a vector gauge field constructed as the vector of torsion of an affine connection $W^{\alpha}_{\cdot\mu\lambda}$ that is related to the $\Gamma^{\alpha}_{\cdot\mu\lambda}$ by

$$W^{\alpha}_{\mu\lambda} = \Gamma^{\alpha}_{\mu\lambda} - \tfrac{2}{3}\delta^{\alpha}_{\mu} W_{\lambda},$$

with

$$W_{\lambda} = W^{\alpha}_{[\lambda\alpha]} = i w_{\lambda}.$$

Consequently the vector of torsion of the Hermitian connection vansihes: $\Gamma_{\mu} = \Gamma^{\alpha}_{[\lambda\alpha]} = 0$. The remaining quantities in Eqs. (4.2)–(4.4) are given by

$$\underset{-}{g}^{[\mu\nu]} = (-g)^{1/2} g^{[\mu\nu]}, g = |g_{\mu\nu}|,$$

$${}^{*}R_{\mu\nu}(\Gamma) = R_{\mu\nu}(\Gamma) + (4\pi G/k^2c^4)\tau_{\mu\nu},$$

$$\tau_{\mu\nu} = -i(g_{[\mu\nu]} + g^{[\lambda\beta]}g_{\beta\nu}g_{\mu\lambda} + \tfrac{1}{2}g^{[\beta\lambda]}g_{\beta\lambda}g_{\mu\nu}) = \tau^{*}_{\nu\mu}.$$

We recall that the Hermitian field tensor may be written as $g_{\mu\nu} = g_{(\mu\nu)} + i g_{[\mu\nu]}$. G is the gravitational constant, and $k = iK$, where K has the dimension $L^{1/2}M^{-1/2}T$. The Ricci tensor $R_{\mu\nu}(\Gamma)$ is Hermitian, and consequently has the same form as the $g_{\mu\nu}$ written above. The same conclusion holds for the tensor ${}^{*}R_{\mu\nu}(\Gamma)$. Finally, the following identification is made:

$$A_{\mu} = (Kc^4/12\pi G) w_{\mu}.$$

The vector A_{μ} in the Einstein–Maxwell limit of the theory (which is obtained for $K\to 0$) generates the field strength $F_{\mu\nu}$ [this is obtained from Eq. (4.4)]; the Maxwell equations follow from Eq. (4.2) in this limit.

Going back to our general curvature tensor given by Eqs. (2.5)–(2.7), we can write

$$R^{\sigma}_{\cdot\lambda\mu\nu}(\Gamma) = Y^{\sigma}_{\cdot\lambda\mu\nu} + iV^{\sigma}_{\cdot\lambda\mu\nu},$$

with

$$Y^{\sigma}_{\cdot\lambda\mu\nu} = G^{\sigma}_{\cdot\lambda\mu\nu} - \Gamma^{\rho}{}_{[\lambda\mu]}\Gamma^{\sigma}{}_{[\rho\nu]} + \Gamma^{\rho}{}_{[\lambda\nu]}\Gamma^{\sigma}{}_{[\rho\mu]},$$

$$V^{\sigma}_{\cdot\lambda\mu\nu} = \partial_{\nu}\Gamma^{\sigma}{}_{[\lambda\nu]} - \partial_{\mu}\Gamma^{\sigma}{}_{[\lambda\nu]} + \Gamma^{\rho}{}_{[\lambda\mu]}\Gamma^{\sigma}{}_{[\rho\nu]}$$
$$+ \Gamma^{\alpha}{}_{(\rho\nu)}\Gamma^{\rho}{}_{[\lambda\mu]} - \Gamma^{\rho}{}_{(\lambda\nu)}\Gamma^{\sigma}{}_{[\rho\mu]} - \Gamma^{\sigma}{}_{(\rho\mu)}\Gamma^{\rho}{}_{[\lambda\nu]}.$$

Accordingly, the complex Ricci tensor may be written as $R_{\lambda\nu}(\Gamma) = Y_{\lambda\nu}(\Gamma) + iV_{\lambda\nu}(\Gamma)$; imposing the condition that the vector of torsion of the connection Γ vanishes, one obtains

$$Y_{\lambda\nu}(\Gamma) = G_{\lambda\nu} - \Gamma^{\rho}{}_{[\lambda\sigma]}\Gamma^{\sigma}{}_{[\rho\nu]} = Y_{\nu\lambda}(\Gamma),^{9}$$

$$V_{\lambda\nu}(\Gamma) = -\partial_{\sigma}\Gamma^{\sigma}_{[\lambda\nu]} + \Gamma^{\rho}{}_{(\lambda\sigma)}\Gamma^{\sigma}{}_{[\rho\nu]}$$
$$+ \Gamma^{\sigma}{}_{(\rho\nu)}\Gamma^{\rho}{}_{[\lambda\sigma]} - \Gamma^{\sigma}{}_{(\rho\sigma)}\Gamma^{\rho}_{[\lambda\nu]} = -V_{\nu\lambda}(\Gamma).$$

Thus, the conditions $\Gamma_{\mu} = 0$ imply that $R_{\nu\lambda}(\Gamma)$ is Hermitian. Accordingly, in the application of our formalism to the Moffat theory we have to impose the conditions

$$E_{\lambda\nu} = I_{\lambda\nu} = L_{\lambda\nu} = K_{\lambda\nu} = 0.$$

In this case the components of the Weyl tensor assume the form

$${}^{(1)}W_{\mu\nu\rho\sigma} = A_{\mu\nu\rho\sigma} - P_{\mu\nu\rho\sigma} - (A/12)g_{\mu\nu\rho\sigma},$$

$${}^{(2)}W_{\mu\nu\rho\sigma} = Q_{\mu\nu\rho\sigma},$$

$${}^{(3)}W_{\mu\nu\rho\sigma} = Q_{\mu\nu\rho\sigma} - \Delta_{\mu\nu\rho\sigma} - (\Omega/12)g_{\mu\nu\rho\sigma},$$

$${}^{(4)}W_{\mu\nu\rho\sigma} = \Lambda_{\mu\nu\rho\sigma}.$$

The explicit value for the Weyl tensor is obtained from the field Eqs. (4.3) and (4.4) and from the above equations;

$$W_{\mu\nu\rho\sigma} = R_{\mu\nu\rho\sigma} - (4\pi G/Kc^4)\big[i(F_{\mu\rho}g_{\nu\sigma} - F_{\nu\rho}g_{\mu\sigma} - F_{\mu\sigma}g_{\nu\rho}$$
$$+ F_{\nu\sigma}g_{\mu\rho}) + (1/2K)(\tau_{\mu\rho}g_{\nu\sigma} - \tau_{\nu\rho}g_{\mu\sigma} - \tau_{\mu\sigma}g_{\nu\rho} + \tau_{\nu\sigma}g_{\mu\rho})\big]$$
$$- \tfrac{1}{12}\Big[(4\pi G/K^2c^4)g^{(\rho\sigma)}\alpha_{(\rho\sigma)}$$
$$- g^{[\rho\sigma]}(4\pi G/Kc^4)(2F_{\sigma\rho} + \frac{1}{K}\beta_{[\sigma\rho]})\Big]g_{\mu\nu\rho\sigma}, \tag{4.5}$$

where

$$\tau_{\mu\nu} = \alpha_{(\mu\nu)} + i\beta_{[\mu\nu]},$$

$$F_{\mu\nu} = 2A_{[\mu,\nu]}.$$

In the Einstein–Maxwell limit of the theory, we have

$$(1/K)\beta_{[\mu\nu]} \to -2F_{\mu\nu}, \quad -(1/K^2)\alpha_{(\mu\nu)} \to 2T_{\mu\nu},$$

where $T_{\mu\nu}$ is the Maxwell energy-momentum tensor. It is easy to verify that in this limit $W_{\mu\nu\rho\sigma} \to R_{\mu\nu\rho\sigma}$ in empty space[10] (no charges and currents). For the full theory, which is unitary field theory, the expression for the complex Weyl tensor is given by Eq. (4.5), since in this situation the source terms cannot be distinguished from the other dynamical factors.

B. Weyl's theory

For the unitary field theory proposed by Weyl we have

$$\Gamma^{\rho}_{\mu\nu} = \{{}_{\mu}{}^{\rho}{}_{\nu}\} + \tfrac{1}{2}(\delta^{\rho}_{\mu}\phi_{\nu} + \delta^{\rho}_{\nu}\phi_{\mu} - g_{\mu\nu}\phi^{\rho}),$$
$$\Gamma^{\rho}{}_{[\mu\nu]} = 0, \tag{4.6}$$

where the metric $g_{(\mu\nu)}$ and the gauge vector field ϕ_{μ} are subjected to the gauge transformations

$$g'_{\mu\nu} = \lambda g_{\mu\nu}, \ \phi'_{\mu} = \phi_{\mu} - \partial_{\mu}\ln\lambda.$$

The affinity given by Eq. (4.6) is invariant under these transformations. Accordingly, the curvature tensor associated to the Weyl connection is also gauge invariant. In this case we have

$$L_{\mu\nu\rho\sigma} = M_{\mu\nu\rho\sigma} = K_{\mu\nu\rho\sigma} = U_{\mu\nu\rho\sigma} = 0.$$

The only remaining components of the curvature are the quantities B, E, J, and T. Thus, the components of the Weyl tensor have the form

$${}^{(1)}W_{\mu\nu\rho\sigma} = B_{\mu\nu\rho\sigma} - P_{\mu\nu\rho\sigma} - (B/12)g_{\mu\nu\rho\sigma}, \tag{4.7}$$

$${}^{(2)}W_{\mu\nu\rho\sigma} = E_{\mu\nu\rho\sigma} - T_{\mu\nu\rho\sigma} - (E/12)g_{\mu\nu\rho\sigma}, \tag{4.8}$$

$${}^{(3)}W_{\mu\nu\rho\sigma} = J_{\mu\nu\rho\sigma} - \Delta_{\mu\nu\rho\sigma} - (J/12)g_{\mu\nu\rho\sigma}, \tag{4.9}$$

$${}^{(4)}W_{\mu\nu\rho\sigma} = I_{\mu\nu\rho\sigma} - \Phi_{\mu\nu\rho\sigma} - (I/12)g_{\mu\nu\rho\sigma}. \tag{4.10}$$

A long but straightforward calculation gives

$${}^{(1)}W_{\rho\mu\nu\sigma} = C_{\rho\mu\nu\sigma} + \tfrac{1}{2}g_{\rho[\mu}\phi_{\nu]}\phi_{\mu} + \tfrac{1}{2}g_{\mu[\nu}\phi_{\sigma]}\phi_{\rho}$$
$$+ g_{\sigma[\mu}\phi_{\rho]}\phi_{\nu} + g_{\nu[\rho}\phi_{\mu]}\phi_{\sigma},$$

$${}^{(2)}W_{\rho\mu\nu\sigma} = 0,$$

$${}^{(3)}W_{\rho\mu\nu\sigma} = \tfrac{1}{2}(g_{\rho\mu}\phi_{\nu\sigma} + g_{\nu\sigma}\phi_{\rho\mu}),$$

$${}^{(4)}W_{\rho\mu\nu\sigma} = \tfrac{1}{2}(g_{\rho[\mu}\phi_{\nu]\sigma} + g_{\sigma[\rho}\phi_{\mu]\nu} + g_{\nu[\sigma}\phi_{\rho]\mu}),$$

where $C_{\rho\mu\nu\sigma}$ is the Weyl tensor for the Riemann-Christoffel curvature. $\phi_{\mu\nu}$ is the field tensor associated with the potentials ϕ_{μ} by the definition $\phi_{\mu\nu} = 2\phi_{[\mu,\nu]}$. In any application of the theory only gauge-invariant quantities have physical significance. As was mentioned before, the Weyl curvature tensor is a quantity of this type; $R'^{\mu}{}_{\nu\rho\sigma} = R^{\mu}{}_{\nu\rho\sigma}$. An inspection of Eqs. (4.7)–(4.10) shows that the several terms involved in these equations change under a gauge transformation by a multiplicative factor λ. Thus, we have for all components of the Weyl tensor, ${}^{(i)}\underset{\sim}{W}'_{\mu\nu\rho\sigma} = \lambda\, {}^{(i)}W_{\mu\nu\rho\sigma}$. This implies that the physical components of the Weyl tensor are given by

$$ {}^{(i)}\underset{\sim}{W}_{\mu\nu\rho\sigma} = {}^{(i)}W_{\mu\nu\rho\sigma}/(-g)^{1/4}. $$

The gauge invariant Weyl tensor associated with the curvature tensor $R^{\mu}{}_{\nu\rho\sigma}(\Gamma)$ is of the form

$$ \underset{\sim}{W}_{\mu\nu\rho\sigma} = \sum_{i=1}^{4} {}^{(i)}\underset{\sim}{W}_{\mu\nu\rho\sigma}. $$

The expression for $\underset{\sim}{W}_{\mu\nu\rho\sigma}$ has a general form, since in the previous calculations we have not used any set of possible field equations. For each choice of field equations we can particularize the expression of the Ricci-Christoffel tensor in Eq. (4.7).

V. CONCLUSION

We have seen how to determine the four complex components of the Weyl tensor that correspond to a general complex non-Riemannian curvature. The method used here for the calculation of the complex Weyl tensors may be applied directly for the determination of the four Weyl spinors that correspond to these tensors. In the determination of these spinors we have to work with a general complex set of *vierbeins*. Thus, the process of projection of the two-dimensional complex spin space has to be properly redefined. After the determination of the components of the Weyl spinor we may obtain its Petrov classification, and determine the necessary conditions for the existence of radiative fields for general complex geometries. The knowledge of such conditions is clearly important for any application of this formalism; in particular, it is of direct interest for the case of the two unitary field theories considered in this paper.

ACKNOWLEDGMENTS

I want to express my appreciation to Dr. P. Letelier for some interesting discussions on the subject. I also want to thank Mr. Edmo D. Campos for his help in part of the calculations.

[1] E. T. Newman and J. Winicour, J. Math. Phys. **15**, 426 (1974); J. Math. Phys. **15**, 113 (1974).

[2] C. W. Fette, A. Janis, and E. T. Newman, J. Math. Phys. **17**, 660 (1975).

[3] J. Plebański, J. Math. Phys. **16**, 2395 (1975).

[4] A. Einstein, Ann. Math. **46**, 578 (1945); Rev. Mod. Phys. **20**, 35 (1948); A. Einstein and G. Strauss, Ann. Math. **47**, 731 (1946).

[5] J. W. Moffat, Phys. Rev. D **15**, 3520 (1977).

[6] D. W. Sciama, J. Math. Phys. **2**, 472 (1961).

[7] Here $H_{[\mu\nu]\,[\rho\sigma]} = \frac{1}{2}(H_{\mu\nu\rho\sigma} - H_{\nu\mu\rho\sigma})$. Recall that $H_{\mu\nu\rho\sigma}$ is antisymmetric on (ρ,σ) as a consequence of Eq. (2.5).

[8] H. Weyl, Z. Physik **56**, 330 (1929).

[9] We recall that for any metrical theory, such as the Einstein–Cartan theory, we have $\Gamma^{\rho}{}_{\mu\nu} = \{{}_{\mu}{}^{\rho}{}_{\nu}\} + g^{\rho\sigma}(\Gamma_{[\mu\sigma]\nu} + \Gamma_{[\nu\sigma]\mu} + \Gamma_{[\mu\nu]\sigma})$. Then, in our case $\Gamma^{\rho}{}_{\rho\nu} = \{{}_{\rho}{}^{\rho}{}_{\nu}\} = \partial_{\nu}(-g)^{1/2} = \Gamma^{\rho}{}_{(\rho\nu)}$. This implies that $G_{\lambda\nu} = G_{\nu\lambda}$.

[10] In this case $R_{\mu\nu\rho\sigma}$ is interpreted as the real Riemann–Christoffel tensor.

On the analyticity of stationary gravitational fields at spatial infinity [a)]

P. Kundu

The Enrico Fermi Institute, The University of Chicago, Chicago, Illinois 60637

(Received 4 November 1980; accepted for publication 23 January 1981)

It is proved that all stationary, vacuum solutions of Einstein's equations which satisfy certain weak differentiability conditions characterizing asymptotic flatness, possess an analytic structure near spatial infinity. This analyticity theorem implies the existence of a multipole expansion whose coefficients can be expressed in terms of the Geroch–Hansen multipole moments defined at the point at infinity on the conformal manifold. This proves a longstanding conjecture that these moments uniquely determine the local structure of a stationary, asymptotically flat, vacuum metric.

PACS numbers: 04.20.Jb

I. INTRODUCTION

In general relativity, the problem of obtaining exact physically realistic solutions has long been recognized as an extremely difficult one. Even for the few exact solutions available in closed form, the physical interpretation is often obscured by the particular coordinate system employed. A familiar example is provided by the class of static axisymmetric vacuum solutions discovered by Weyl as early as 1917. While all the Weyl metrics are "known" when expressed in the canonical cylindrical coordinates, there does not seem to be a simple way of singling out a particular member (such as the Schwarzschild metric) tailored to fit a specific situation.[1] Within the framework of Newtonian gravitation, a given source distribution determines the multipole moments which in turn completely determine the gravitational potential through the familiar multipole expansion. Motivated by this elementary result, considerable attention has been directed in recent years toward the study of multipole moments within the framework of general relativity as a means of extracting physical information from a given metric.[2] Several (generally inequivalent) definitions are given in the literature from different points of view. In view of the large gauge (i.e., coordinate) freedom available (or, in other words, lack of a preferred coordinate system) in general relativity—a crucial feature that is absent in the Newtonian theory—coordinate-free definitions seem to be preferrable over others. Utilizing conformal techniques, originally due to Penrose,[3] Geroch[4] has formulated an elegant coordinate-free definition of the multipole moments of a static vacuum metric. This definition has been extended by Hansen[5] to the more general case of stationary vacuum fields. One starts by considering the quotient manifold $\mathscr{V} = \mathscr{M}/\mathscr{G}$ obtained from a stationary spacetime $(\mathscr{M}, g)$ (signature $+\,-\,-\,-$) by means of the canonical projection

$$\pi: x \to [x] = \{x' \in \mathscr{M} \,|\, x \sim x'\}, \quad x \in \mathscr{M},$$

where $\sim$ denotes the equivalence relation induced by the action of the one parameter timelike group of motions $\mathscr{G}$ on $(\mathscr{M}, g)$. Thus $\mathscr{V}$ represents the manifold of trajectories of the associated Killing vector field $X = \xi^\mu \partial/\partial x^\mu$ on $(\mathscr{M}, g)$ satisfying the Killing equation $\nabla_{(\mu}\xi_{\nu)} = 0$ with respect to the metric $g_{\mu\nu}$. Topologically $\mathscr{V}$ is assumed to be $\mathbb{R}^3 - \mathscr{B}$ where $\mathscr{B}$ is a closed 3-ball containing the source. The norm $\lambda = \xi_\mu \xi^\mu$ and the twist ω (satisfying $\nabla_\mu \omega = e_{\mu\nu\rho\sigma}\xi^\nu \nabla^\rho \xi^\sigma$ on account of the Einstein equations $R_{\mu\nu} = 0$) define a complex potential

$$\Phi = \Phi_M + i\Phi_J = \tfrac{1}{4}\lambda^{-1}(\mathscr{E} - 1)(\mathscr{E}^* + 1) \tag{1.1}$$

and an induced positive definite metric

$$h = -\lambda \pi_* g \tag{1.2}$$

on $\mathscr{V}$, where $\mathscr{E} = \lambda + i\omega$ is the familiar Ernst potential.[6] The next step is to introduce a notion of asymptotic flatness at spatial infinity. The main idea is to attach a point Λ to the manifold $\mathscr{V}$ to represent the "spatial infinity" so that the spatial metric h_{ij} and the complex potential Φ give rise to a set of smooth (i.e., C^∞) fields

$$\tilde{\Phi} = \Omega^{-1/2}\Phi, \quad \tilde{h}_{ij} = \Omega^2 h_{ij} \tag{1.3}$$

on the conformal completion $\tilde{\mathscr{V}} = \mathscr{V} \cup \Lambda$. The conformal factor Ω is a (non-negative) scalar field (which is at least C^2) on $\tilde{\mathscr{V}}$ satisfying the conditions

$$[\Omega]_\Lambda = 0, \quad [\Omega_{;i}]_\Lambda = 0, \quad [\Omega_{;ij} - 2\tilde{h}_{ij}]_\Lambda = 0, \tag{1.4}$$

where the semicolon represents the action of the covariant derivative $\tilde{\nabla}$ on the conformal manifold $(\tilde{\mathscr{V}}, \tilde{h})$. These conditions express in a covariant manner the requirement that Ω has to fall off "as r^{-2} at spatial infinity" (i.e., as $r \to \infty$). The multipole moments $Q_{i_1 \cdots i_s}$ are then defined as a sequence of traceless symmetric tensors at Λ by the recurrence relations

$$\begin{aligned} Q_{i_1\cdots i_s} &= [P_{i_1\cdots i_s}]_\Lambda, \\ P &= \tilde{\Phi}, \quad P_{i_1} = \tilde{\Phi}_{;i_1}, \\ P_{i_1\cdots i_{s+1}} &= \mathscr{C}\,[P_{(i_1\cdots i_s;i_{s+1})} - \tfrac{1}{2}s(2s-1)P_{(i_1\cdots i_{s-1}}\tilde{R}_{i_s i_{s+1})}], \end{aligned} \tag{1.5}$$

where $\mathscr{C}[\cdots]$ represents the trace-free part of a tensor and (square) round brackets denote (anti-)symmetrization of the indices. (The real and imaginary parts of $Q_{i_1\cdots i_s}$ are respectively the 2^s-mass and angular momentum multipole moments.) It has been conjectured that the geometry of a stationary vacuum metric is uniquely specified by its multipole moments. Given a stationary vacuum metric, one has to find a suitable conformal factor Ω in order to calculate its moments. For the simplest examples such as the Schwarzschild, Weyl, and Kerr solutions, a suitable Ω can be found by in-

[a)] Work supported by a grant from the National Science Foundation.

spection. The converse problem of obtaining asymptotic solution of the vacuum field equations by expressing the rescaled field variables in terms of the Geroch–Hansen moments has been addressed in a recent paper.[7] The procedure was to select $\Omega = \sigma$, the square of the geodesic distance from Λ on $(\tilde{\mathscr{V}}, \tilde{h})$, as the conformal factor and expand all the rescaled field variables $\tilde{\Phi}$ and $\tilde{h}_{ij}$ in normal coordinates defined around Λ. However, the difficulty was that the rescaled field equations and the standard asymptotic flatness conditions themselves did not as such guarantee the analyticity property of $\tilde{\Phi}$ and $\tilde{h}_{ij}$ in a neighborhood of Λ, although they were certainly consistent with it. One had to impose the extra condition that $(\tilde{\mathscr{V}}, \tilde{h})$ be an analytic manifold from which the analyticity of $\tilde{\Phi}$ could then be established. This is unlike the Newtonian theory where the physical boundary conditions on Laplace's equation automatically guarantee the analyticity of the rescaled scalar potential in a neighborhood of Λ.

The purpose of this paper is to establish rigorously an analyticity theorem (Theorem 1 of Sec. II) for all the field variables in the case of a stationary vacuum metric starting from weak differentiability requirements. Again a special conformal factor is selected intrinsically on $(\tilde{\mathscr{V}}, \tilde{h})$ which is, however, different from the previous one. This time the standard existence theorems for elliptic systems come to rescue the situation. The main obstacle is to prove the ellipticity of the system of equations at hand in a neighborhood of Λ. Although the ellipticity and the consequent analyticity property in a finite region are well known,[8] the situation becomes much more delicate at spatial infinity, i.e., at the point Λ, due to the appearance of certain terms in the rescaled field equations that are formally singular at Λ. In order to resolve this difficulty, the extra singular terms are treated as auxiliary field variables for which additional regular field equations are derived so that the entire extended set of field equations becomes a second order (nonlinear) elliptic system in a neighborhood of Λ with analytic coefficients. The analyticity of the solutions then follows from some very powerful theorems due to Morrey.[9,10]

The above mentioned idea of proving analyticity of the rescaled field variables by considering an extended regular coupled elliptic system originates in the work of Hansen[5] where the smoothness of $\tilde{\Phi}_M$ and $\tilde{\Phi}_J$ was established in this manner. Recently Beig and Simon[11] have independently utilized this idea to derive an analyticity theorem for the static vacuum metrics for a specially chosen conformal factor and have also proved the multipole expansion theorem. In this special case fortunately the field equations can be written in a much simpler form which considerably simplifies the arguments. Prior to these efforts, some partial results on the Geroch–Hansen asymptotic multipole theory have been obtained by Xanthopoulos.[12] He proved that if the angular momentum multipole moments of a stationary asymptotically flat spacetime vanish (i.e., $\mathrm{Im}\, Q_{i_1\cdots i_s} = 0$), then the spacetime is static; moreover, if $Q_{i_1\cdots i_s} = 0$, then the spacetime is flat. His proof made use of a "unique continuation" theorem of Aronszajn and did not require analyticity of the potentials.

II. THE ANALYTICITY THEOREM

In this section we shall give a precise formulation of the analyticity theorem mentioned in Sec. I. Following Refs. 5 and 13, it can be shown that the Einstein equations $R_{\mu\nu} = 0$ for a general stationary vacuum metric imply the following set of equations on $\mathscr{V}$:

$$(h^{ij}\nabla_i\nabla_j - 2R)\Phi = 0, \tag{2.1}$$

$$R_{ij} = 2[\nabla_i\Phi\nabla_j\Phi^* - (1 + 4|\Phi|^2)^{-1}\nabla_i|\Phi|^2\nabla_j|\Phi|^2], \tag{2.2}$$

where R_{ij} is the Ricci tensor of $(\mathscr{V}, h)$ and $R = h^{ij}R_{ij}$. We choose the conformal factor Ω by requiring the scalar potential $\tilde{\Phi}$ to satisfy

$$|\tilde{\Phi}|^2 = \mu^2 = \text{const} \tag{2.3}$$

everywhere on the conformal manifold $(\tilde{\mathscr{V}}, \tilde{h})$. Ω is thus a positive definite function on $\tilde{\mathscr{V}}$ which is equal to $\mu^{-2}|\Phi|^2$ on $\mathscr{V}$ and which vanishes at Λ. The constant μ in (2.3) is to be identified with $(m^2 + l^2)^{1/2}$, where $m = [\tilde{\Phi}_M]_\Lambda$ is the mass of the system and $l = [\tilde{\Phi}_J]_\Lambda$ is the angular momentum monopole moment (NUT parameter). However, considering the field equations in the presence of a stationary distribution of matter, Hansen[5] has shown that for an asymptotically flat stationary solution l must vanish in the case when the entire orbit manifold including the region occupied by the sources is topologically $\mathbb{R}^3$. We therefore assume that the system has a nonzero mass so that $\mu \neq 0$ and introduce a new field α on $(\tilde{\mathscr{V}}, \tilde{h})$ by the relation

$$\tilde{\Phi} = \mu\exp(i\alpha). \tag{2.4}$$

Upon performing the conformal transformation (1.3) the rescaled field equations on $(\tilde{\mathscr{V}}, \tilde{h})$ become (with $\tilde{\Delta}$ denoting the covariant Laplacian operator corresponding to the metric $\tilde{h}_{ij}$),

$$(\tilde{\Delta} - 2\tilde{R})\tilde{\Phi} = \tfrac{15}{2}\Omega^{-1}[\tilde{\Delta}\Omega - \tfrac{3}{2}w]\tilde{\Phi} \tag{2.5}$$

or, equivalently (separating the real and imaginary parts),

$$(\tilde{\Delta}\Omega - \tfrac{3}{2}w) = -\tfrac{2}{15}(\alpha_m\alpha^m + 2\tilde{R})\Omega, \tag{2.6a}$$

$$\tilde{\Delta}\alpha = 0 \tag{2.6b}$$

and

$$\begin{aligned}\tilde{R}_{ij} = &-\Omega^{-1}[\Omega_{;ij} + \tilde{h}_{ij}(\tilde{\Delta}\Omega - 2w) - \tfrac{1}{2}\mu^2\Omega_{;i}\Omega_{;j}] \\ &+ 2\mu^2[\Omega\alpha_i\alpha_j - \mu^2(1 + 4\mu^2\Omega)^{-1}\Omega_{;i}\Omega_{;j}],\end{aligned} \tag{2.7}$$

where we have set $\alpha_m := \alpha_{;m}$ and

$$w := \Omega^{-1}\Omega_{;k}\Omega^{;k}. \tag{2.8}$$

$\tilde{R}_{ij}$ is the Ricci tensor of $(\tilde{\mathscr{V}}, \tilde{h})$, $\tilde{R} = \tilde{h}^{ij}\tilde{R}_{ij}$ and the indices $i, j, k, \cdots$ are raised and lowered by the conformal metric $\tilde{h}_{ij}$. The main result of this paper is given by the following theorem.

Theorem 1: If

(1) Ω, α, and $\tilde{h}_{ij}$ satisfy the vacuum Einstein equations, namely, Eqs. (2.6a), (2.6b), and (2.7) on an open domain $\mathscr{D}_0 \subset \tilde{\mathscr{V}}$ containing Λ,

(2) Ω and α are of class C^2 and $\tilde{h}_{ij}$ is of class C^4 on $\mathscr{D}_0$,

(3) $\Omega > 0$ on $\mathscr{D}_0 - \{\Lambda\}$ and $\Omega = 0$ at Λ, and

(4) $\mu \neq 0$,

then there exists a chart $(\mathscr{D}, \phi)$ consisting of an open set $\mathscr{D}$ containing Λ with $\bar{\mathscr{D}} \subset \mathscr{D}_0$ and a diffeomorphism ϕ:

$\mathscr{D}\to\mathbb{R}^3$ with respect to which the basic fields Ω, α, and $\tilde{h}_{ij}$ as well as the two auxiliary fields w and K_{ij} defined respectively by Eq. (2.8) and the relation

$$K_{ij} := \Omega^{-1}(-\Omega_{;ij} + \tfrac{1}{2}\mu^2\Omega_{;i}\Omega_{;j} + \tfrac{1}{2}\tilde{h}_{ij}w), \tag{2.9}$$

are analytic functions (i.e., of class C^ω) on $\mathscr{D}$.

The basic tools for our proof are the following two theorems on elliptic systems of equations due to Morrey specialized to second order systems. A system

$$F^A(x,u^B,u^B{}_{,i},u^B{}_{,ij}) = 0 \quad (A,B = 1,2,\dots,N) \tag{2.10}$$

of N coupled second order nonlinear differential equations for N functions $u^B(x)$ is said to be elliptic on an open neighborhood $\mathscr{D}_0$ if, at every point $x\in\mathscr{D}_0$

$$\det|L^A{}_B(x,\zeta)|\neq 0 \tag{2.11}$$

for each nonzero $\zeta = (\zeta_i)\in\mathscr{T}^*_x(\mathscr{D}_0)$, the cotangent space at x, where

$$L^A{}_B(x,\zeta) := \zeta_k\zeta_l\left[\frac{\partial F^A}{\partial u^B{}_{,kl}}\right](x) \tag{2.12}$$

and the comma denotes the ordinary derivative at x.

Theorem 2 (Morrey[9]): If the functions $u^B(x)$ are of class C^2 and satisfy Eqs. (2.10) and (2.11) on a bounded domain $\mathscr{D}_0$ and F^A are of class C^1 on the set $\bar{\mathscr{R}}$, the closure of a set $\mathscr{R}$, which contains the compact set $\{x,u^B(x),u^B{}_{,k}(x),u^B{}_{,kl}(x)|x\in\bar{\mathscr{D}}_1\}$, with $\bar{\mathscr{D}}_1\subset\mathscr{D}_0$, then $u^B(x)$ are of class $C^{2+\nu}$, i.e., $u^B{}_{,kl}(x)$ are uniformly Hölder continuous with exponent ν $(0<\nu<1)$[14] on any domain $\mathscr{D}_2$ such that $\bar{\mathscr{D}}_2\subset\mathscr{D}_1$.

If, moreover, F^A are of class $C^{n+\nu}$, i.e., the nth derivatives of F^A are uniformly Hölder-continuous with exponent ν on $\bar{\mathscr{R}}$ (n = integer$\geqslant 1$, $0<\nu<1$), then $u^B(x)$ are of class $C^{n+2+\nu}$ on $\mathscr{D}_1$.

Theorem 3 (Morrey[10]): If the solutions $u^B(x)$ of Eqs. (2.10), (2.11) are of class C^2 on $\mathscr{D}_0$ and F^A are of class C^ω (i.e., analytic) with respect to all its arguments on $\bar{\mathscr{R}}$, then $u^B(x)$ are also of class C^ω on $\mathscr{D}_1$ with $\bar{\mathscr{D}}_1\subset\mathscr{D}_0$.[15]

Proof of Theorem 1: As has been already noted, the field equations (2.6a) and (2.7) themselves are formally singular at Λ. The main idea of the proof is to obtain a set of second order equations which can be shown to be elliptic in a neighborhood of Λ.

We begin by noting that since α satisfies a Laplace equation [Eq. (2.6b)] and $\tilde{h}_{ij}\in C^4\subset C^{3+\nu}$ $(0<\nu<1)$ on an open set $\mathscr{D}_0$ containing Λ, Theorem 2 implies that $\alpha\in C^{4+\nu}$ and consequently $\alpha_m\in C^{3+\nu}$ on an open set $\mathscr{D}_1$ containing Λ with $\bar{\mathscr{D}}_1\subset\mathscr{D}_0$.

Next, contracting Eq. (2.7), we obtain

$$\tilde{\Delta}\Omega - \tfrac{3}{2}w = \tfrac{1}{4}\Omega\{-\tilde{R} + \tfrac{1}{2}\mu^2 w + 2\mu^2\Omega \times[\alpha_m\alpha^m - \mu^2(1+4\mu^2\Omega)^{-1}w]\}. \tag{2.13}$$

Upon equating the right-hand sides of Eqs. (2.6a) and (2.13), we get, after cancelling a factor Ω, the relation

$$\tilde{R} + 8\alpha_m\alpha^m(1+\tfrac{15}{4}\mu^2\Omega) + \tfrac{15}{2}\mu^2 w(1+4\mu^2\Omega)^{-1} = 0, \tag{2.14}$$

which must hold everywhere on $\mathscr{D}_0 - \{\Lambda\}$ where $\Omega>0$. However, we observe that $\tilde{h}_{ij}\in C^4$ on $\mathscr{D}_0$ and, consequently, $\tilde{R}\in C^2$ on $\mathscr{D}_0$ including at Λ. Further, $\Omega\in C^2$ on $\mathscr{D}_0$ and $\alpha_m\in C^{3+\nu}$ on $\mathscr{D}_1$. Therefore, by virtue of Eq. (2.14), the function w defined by Eq. (2.8) can be extended to a function of class C^2 on $\mathscr{D}_1$, including at the point Λ where it is formally undefined. Introducing the tensor

$$L_{ij} := \tilde{R}_{ij} - \tfrac{1}{4}\tilde{h}_{ij}\tilde{R}, \tag{2.15}$$

we can express Eq. (2.7) in the form

$$L_{ij} = K_{ij} - \mu^2\Theta_{ij}, \tag{2.16}$$

where K_{ij} is defined by Eq. (2.9) and

$$\Theta_{ij} := \tfrac{1}{2}\tilde{h}_{ij}[\Omega\alpha_m\alpha^m + \tfrac{1}{4}w(1+4\mu^2\Omega)^{-1}] - 2[\Omega\alpha_i\alpha_j - \mu^2(1+4\mu^2\Omega)^{-1}\Omega_{;i}\Omega_{;j}]. \tag{2.17}$$

Rewriting Eq. (2.9) in the form

$$\Omega_{;ij} = -\Omega K_{ij} + \tfrac{1}{2}\mu^2\Omega_{;i}\Omega_{;j} + \tfrac{1}{2}\tilde{h}_{ij}w =: \mathscr{F}^{(1)}_{ij}(\Omega,\Omega_{;m},w,\tilde{h}_{mn},K_{mn}), \tag{2.18}$$

we obtain a second order equation for Ω:

$$\tilde{\Delta}\Omega = \mathscr{F}^{(1)}(\Omega,w,\tilde{h}_{mn},K_{mn}), \tag{2.19}$$

where, explicitly,

$$\mathscr{F}^{(1)} := \tilde{h}^{ij}\mathscr{F}^{(1)}_{ij} = \tfrac{3}{2}w + \Omega(\tfrac{1}{2}\mu^2 w - K_m{}^m) \tag{2.20a}$$

$$= \tfrac{3}{2}w + 2\Omega[\alpha_m\alpha^m(1+4\mu^2\Omega) + \mu^2 w(1+4\mu^2\Omega)^{-1}]. \tag{2.20b}$$

The second alternative form of $\mathscr{F}^{(1)}$ results on eliminating $\tilde{R}$ from Eq. (2.6a) with the help of Eq. (2.14). Since $\Omega\in C^2$, $w\in C^2\subset CC^{1+\nu}$ and $\alpha_m\in C^{3+\nu}$ on $\mathscr{D}_1$, it follows from Eqs. (2.19) and (2.20b) by an application of Theorem 2, that $\Omega\in CC^{3+\nu}$ on an open set $\mathscr{D}_2$ containing Λ such that $\bar{\mathscr{D}}_2\subset\mathscr{D}_1$. Hence, by virtue of Eqs. (2.16) and (2.17), K_{ij} can be extended to functions of class C^2 on $\mathscr{D}_2$ including at Λ where it is formally singular. We can also rewrite Eq. (2.7) in the form

$$\tilde{R}_{ij} = \mathscr{F}^{(2)}_{ij}(\Omega,\Omega_{;m},w,\tilde{h}_{mn},\alpha_m,K_{mn}), \tag{2.21}$$

where

$$\mathscr{F}^{(2)}_{ij} := K_{ij} + \tilde{h}_{ij}(K_m{}^m - \tfrac{1}{2}\mu^2 w) + 2\mu^2[\Omega\alpha_i\alpha_j - \mu^2(1+4\mu^2\Omega)^{-1}\Omega_{;i}\Omega_{;j}]. \tag{2.22}$$

Next we note that

$$w_{;i} = \mu^2 w\Omega_{;i} - 2K_i{}^k\Omega_{;k}. \tag{2.23}$$

From Eqs. (2.18) and (2.23) it follows that $w_{;ij}$ can be expressed as

$$w_{;ij} = \mathscr{F}^{(3)}_{ij}(\Omega,\Omega_{;m},w,\tilde{h}_{mn},K_{mn},K_{mn;p}). \tag{2.24}$$

Thus w satisfies an elliptic equation

$$\tilde{\Delta}w = \mathscr{F}^{(3)}, \tag{2.25}$$

where we have set

$$\mathscr{F}^{(3)} := \tilde{h}^{ij}\mathscr{F}^{(3)}_{ij}. \tag{2.26}$$

Obtaining an equation for K_{ij} requires somewhat lengthy but straightforward calculations. Taking the curl of ΩK_{ij}, one finds

$$\Omega K_{i[j;k]} = \tfrac{1}{2}\tilde{h}_{i[j}w_{;k]} - \tfrac{1}{2}\mu^2\Omega_{;i[j}\Omega_{;k]} + \tilde{h}_{i[j}L_{k]l}\Omega^{;l} - \mu^2\Theta_{i[j}\Omega_{;k]},$$

where use has been made of the Ricci identity for $\Omega_{;i[jk]}$ and the formula $\tilde{R}_{iljk} = 4\tilde{h}_{[i][j}L_{k]|l]}$ for the Riemann tensor in three dimensions. Next, using the expressions for $\Omega_{;ij}$ and $w_{;i}$ given by Eqs. (2.18) and (2.23), one obtains the equation

$$\Omega K_{i[j;k]} = \mu^2[\tfrac{1}{2}\Omega K_{i[j}\Omega_{;k]} + \tfrac{1}{4}\tilde{h}_{i[j}\Omega_{;k]}w - \Theta_{i[j}\Omega_{;k]} - \tilde{h}_{i[j}\Theta_{k]l}\Omega^{;l}]. \tag{2.27}$$

Inserting the expression (2.17) for Θ_{ij} into Eq. (2.27), one finally arrives at an equation of the form

$$\Omega\,[K_{i[j;k]} - \mu^2 W_{ijk}(\Omega,\Omega_{;m},w,\tilde{h}_{mn},\alpha_m,K_{mn})] = 0. \tag{2.28}$$

The quantity within the square brackets itself therefore vanishes on $\mathscr{D}_0 - \{\Lambda\}$, where $\Omega > 0$. At this point it is crucial to observe that this quantity is C^1 on the open set $\mathscr{D}_2$, with $\bar{\mathscr{D}}_2 \subset \mathscr{D}_0$, including at Λ and, consequently, must vanish also at Λ. Hence the relation

$$K_{i[j;k]} = \mu^2 W_{ijk} \tag{2.29}$$

holds everywhere on $\mathscr{D}_0$. Operating on Eq. (2.29) with $\tilde{\nabla}^k$ and commuting the derivatives with the help of the Ricci identity, we obtain

$$\tilde{\Delta}K_{ij} = (K_i{}^k{}_{;k})_{;j} - 4\tilde{h}_{[i][j}L_{k]|m]}K^{km} + K_i{}^m L_{jm} + \tfrac{1}{4}K_{ij}\tilde{R} + \mu^2 W_{ij}{}^k{}_{;k}. \tag{2.30}$$

The first term can be further reduced with the help of Eq. (2.16) and the contracted Bianchi identity

$$L_i{}^k{}_{;k} = \tfrac{1}{4}\tilde{R}_{;i}. \tag{2.31}$$

Finally, using Eqs. (2.14), (2.16)–(2.18), and (2.24), after some tedious but straightforward calculations, the desired second order equation for the auxiliary variables K_{ij} can be brought to the form

$$\tilde{\Delta}K_{ij} = \mathscr{F}^{(4)}_{ij}(\Omega,\Omega_{;m},w,\tilde{h}_{mn},\alpha_m,\alpha_{(m;n)},\alpha_{(m;n)p},K_{mn},K_{mn;p}). \tag{2.32}$$

It is important to notice that the right-hand side of Eq. (2.32) depends on derivatives of α_m to second order! Hence, to complete the system, we need an equation for α_m. Commuting the derivatives, it is easy to show that

$$\tilde{\Delta}\alpha_i - (\tilde{\Delta}\alpha)_{;i} = L_{ij}\alpha^j + \tfrac{1}{4}\tilde{R}\alpha_i.$$

Making use of Eqs. (2.6b). (2.14), (2.16), and (2.17), we get an equation of the form

$$\tilde{\Delta}\alpha_i = \mathscr{F}^{(5)}_i(\Omega,\Omega_{;m},w,\tilde{h}_{mn},\alpha_m,K_{mn}). \tag{2.33}$$

We now introduce a harmonic coordinate system $\{x'^i\}$ on an open neighborhood of $\mathscr{V}$ with the point Λ as the origin $\{x'^i = 0\}$ by the conditions

$$\tilde{\Delta}x'^i = 0, \quad (\partial x'^i/\partial x^j)_\Lambda = \delta^i{}_j. \tag{2.34}$$

Since $\tilde{h}_{mn} \in C^4 \subset C^{3+\nu}$ on $\mathscr{D}_0$, the coefficients of the linear elliptic differential operator $\tilde{\Delta}$ are of class $C^{2+\nu}$ on $\mathscr{D}_0$. Hence, by the second part of Theorem 2, it follows that the solutions $x'^i(x)$ are of class $C^{4+\nu}$ on $\mathscr{D}_1$. Moreover, there exists an open neighborhood $\mathscr{N} \subset \mathscr{D}_1$ of Λ such that the Jacobian $\det|\partial x'^i/\partial x^j| \neq 0$ on $\mathscr{N}$. In this new coordinate system Eqs. (2.19), (2.21), (2.25), (2.32), and (2.33) constitute the desired system of second order coupled equations of the form (2.10) for the 17 functions

$$u^B(x') = \{\Omega,w,\tilde{h}'_{mn},\alpha'_m,K'_{mn}\},$$

which are C^2 on $\mathscr{N}$. Writing $\zeta^2 = \zeta_i\zeta^i$, the 17-dimensional square matrix $\|L^A{}_B(x',\zeta)\|$ can be expressed in the form

$$\|L^A{}_B(x',\zeta)\| = \begin{bmatrix} \zeta^2 & & & & \\ 0 & \zeta^2 & & 0 & \\ 0 & 0 & \zeta^2 I^{(6)} & & \\ 0 & 0 & M & \zeta^2 I^{(3)} & \\ 0 & 0 & M' & M'' & \zeta^2 I^{(6)} \end{bmatrix},$$

where $I^{(N)}$ stands for the $N \times N$ unit matrix and M, M', M'' are nonzero matrices of appropriate dimensions. Since obviously $\det|L^A{}_B| = (\zeta^2)^{17} \neq 0\ \forall \zeta \neq 0$, we have established the fact that the extended system is indeed elliptic on $\mathscr{N}$. Hence from Theorem 3 it follows that the functions $u^B(x')$ are C^ω on an open neighborhood $\mathscr{D}$ containing Λ such that $\bar{\mathscr{D}} \subset \mathscr{N}$. Of course, analyticity of α'_m implies the analyticity of α on $\mathscr{D}$. This completes the proof.

One can define a collection $\mathscr{S}$ of C^ω-related charts based on harmonic coordinate neighborhoods, in each of which the field variables are analytic. This collection $\mathscr{S}$ defines an analytic structure on $(\tilde{\mathscr{V}}, \tilde{h})$, which then becomes an analytic Riemannian manifold.

A similar procedure can be employed to prove an analyticity theorem for static electrovac spacetimes using the potentials defined by Hoenselaers.[16] The problem of finding an appropriate set of potentials suitable for describing a general stationary electrovac spacetime still remains open.

III. THE MULTIPOLE EXPANSION

The analyticity theorem proved in the last section enables one to construct a multipole expansion for stationary vacuum fields by a method analogous to that of the previous paper.[7] The procedure is to expand the field variables in a normal coordinate system centered around Λ. Since $(\tilde{\mathscr{V}}, \tilde{h})$ is an analytic manifold, the exponential map $\exp_\Lambda$ which maps a star-shaped open neighborhood $\mathscr{N}_0 \subset \mathscr{T}_\Lambda$ onto a normal neighborhood $\mathscr{N}_\Lambda \subset \mathscr{N}$ of Λ, is an analytic diffeomorphism. The pair $(\mathscr{N}_\Lambda, \exp_\Lambda)$ is therefore an analytic chart in $\tilde{\mathscr{V}}$. Thus one can expand Ω, α, and $\tilde{h}_{ij}$ around Λ in normal coordinates which are in many respects the most natural generalization of the Cartesian coordinates in a general curved space. The coefficients of the Taylor expansion of any analytic tensor field **T** in normal coordinates[17] are actually covariant tensorial objects involving only the various covariant derivatives of **T** and the Riemann tensor evaluated at the origin and consequently have a coordinate-free significance. Moreover, the effect of curvature is shown explicitly through the terms involving the Riemann tensor and its derivatives. The coefficients of the normal coordinate expansion of Ω, α, and $\tilde{h}_{ij}$ can be expressed completely in terms of the multipole moments. In order to prove this, one needs only to show that the covariant derivatives of Ω, α, and $\tilde{R}_{ij}$ at Λ are uniquely determined by the moments themselves. This can be demonstrated by a simple inductive argument. Since the procedure is quite straightforward though rather tedious, we confine ourselves only to a brief outline of the proof. It is analogous to the static case treated by Beig and

Simon, but somewhat more complicated in its details. To the lowest few orders we have

$$Q = m + il, \quad Q_i = i(m + il)[\alpha_i]_\Lambda \equiv (m + il)[A_i]_\Lambda,$$
$$Q_{i_1 i_2} = (m + il)\mathscr{C}[A_{i_1 i_2} - \tfrac{1}{2}\widetilde{R}_{i_1 i_2}]_\Lambda, \tag{3.1}$$

where we have set

$$(e^{i\alpha})_{;(i_1 \cdots i_n)} = e^{i\alpha} A_{i_1 \cdots i_n} \tag{3.2}$$

and $[\alpha]_\Lambda = \tan^{-1}(l/m)$. We also note the restriction $Q^* Q_i = i\mu^2[\alpha_i]_\Lambda$, which serves to fix the "origin" of the multipole expansion. When $l = 0$, $\mathrm{Re}Q_i = 0$, showing that our choice of the conformal factor corresponds to an expansion around the center of mass in the physical space. We further note that $[\alpha_i]_\Lambda = \mathrm{Im}Q_i/m$. Also since $[\Omega_{;ij} - 2\widetilde{h}_{ij}]_\Lambda = 0$, it follows from Eq. (2.13) that $[w]_\Lambda = 4$. From Eq. (2.14) one then obtains the value of $[\widetilde{R}]_\Lambda$ in terms of the moments while Eqs. (3.1) and (3.2) yield the values of $\mathscr{C}[\alpha_{;i_1 i_2}] = \alpha_{;i_1 i_2}$ and $\mathscr{C}[\widetilde{R}_{i_1 i_2}]$ at the point Λ. One therefore also obtains the value of the full Ricci tensor $\widetilde{R}_{i_1 i_2}$ at Λ. Using Eq. (2.16), we can rewrite Eqs. (2.18) and (2.23) as

$$\Omega_{;ij} = \Omega L_{ij} + M_{ij}(\Omega, \Omega_{;m}, w) \tag{3.3}$$

and

$$w_{;i} = \mu^2 w \Omega_{;i} - 2(L_i{}^m - \mu^2 \Theta_i{}^m)\Omega_{;m}. \tag{3.4}$$

Equation (3.4) then immediately gives $[w_{;i}]_\Lambda = 0$.

Now, in order to set up the induction, we start with the assumption that the multipole moments uniquely determine the covariant derivatives of Ω, α, w, and $\widetilde{R}_{ij}$ at Λ up to certain orders; let us suppose that the derivatives of Ω and α are known up to rth order, that of w up to $(r-1)$th order, and that of $\widetilde{R}_{ij}$ up to $(r-2)$th order. Then, taking $(r-1)$ derivatives of Eq. (3.3), we get the $(r+1)$th derivatives of Ω at Λ; the right-hand side involves derivatives of $\widetilde{R}_{ij}$ only up to order $(r-3)$ since $[\Omega]_\Lambda = 0$ and $[\Omega_{;i}]_\Lambda = 0$. Similarly, taking $(r-1)$ derivatives of Eq. (3.4), we get the rth derivative of w at Λ; again, the right-hand side involves derivatives of $\widetilde{R}_{ij}$ only up to order $(r-2)$. Next, taking $(r-1)$ derivatives of Eq. (2.14), one can calculate $[\widetilde{R}_{;i_1 \cdots i_{r-1}}]_\Lambda$. Finally, the 2^{r+1} moments combined with all the previous information determine $\mathscr{C}[\alpha_{;(i_1 \cdots i_{r+1})}]$ and $\mathscr{C}[\widetilde{R}_{(i_1 i_2; i_3 \cdots i_{r+1})}]$ at Λ. From the latter the value of $\widetilde{R}_{i_1 i_2; i_3 \cdots i_{r+1}}$ at Λ can be calculated reducing its various trace and antisymmetric parts to lower order terms by commuting the derivatives and using the Ricci and Bianchi identities. The trace and the antisymmetric parts of $\alpha_{;i_1 \cdots i_{r+1}}$ are evaluated in the same way by noting that $\widetilde{\Delta}\alpha = 0$ and $\alpha_{;ij} - \alpha_{;ji} = 0$.

Thus with the knowledge of the covariant derivatives of the quantities Ω, α, w, and $\widetilde{R}_{ij}$ at Λ up to a certain order in terms of the moments, we have succeeded in evaluating their next order derivatives at Λ in terms of the multipole moments alone. This completes the proof by induction.

IV. CONCLUDING REMARKS

The main result of this paper is an analyticity theorem for stationary vacuum metrics satisfying the Einstein equations. We have shown that if the manifold of the Killing trajectories admits a conformal completion satisfying the conditions of Theorem 1 for a certain intrinsically selected conformal factor Ω, then the spacetime is analytic in a neighborhood of spatial infinity. The requirements on the conformal completion can also be expressed in terms of an arbitrary conformal factor $\overline{\Omega}$ which is assumed to be C^2 on an open neighborhood $\mathscr{D}_0$ of Λ with $\overline{\Omega} > 0$ on $\mathscr{D}_0 - \{\Lambda\}$ and $\overline{\Omega} = 0$ at Λ. Setting $\overline{\Phi} = \overline{\Omega}^{-1/2}\Phi$, $[|\overline{\Phi}|^2]_\Lambda = \mu^2 (\neq 0)$ and requiring the conformal metric $\overline{h}_{ij} = \overline{\Omega}^2 h_{ij}$ as well as $\chi = \mu^{-2}|\overline{\Phi}|^2$ to be C^4 on $\mathscr{D}_0$, one easily recovers the differentiability conditions imposed on the preferred conformal factor $\Omega = \chi\overline{\Omega}$ and the corresponding conformal metric $\widetilde{h}_{ij} = \chi^2 \overline{h}_{ij}$.[18] (The condition on $\alpha = \arg\overline{\Phi}$ is obviously independent of the choice of the conformal factor.) Moreover the asymptotic conditions

$$[\overline{\Omega}]_\Lambda = 0, \quad [\overline{\nabla}_i \overline{\Omega}]_\Lambda = 0, \quad [\overline{\nabla}_i \overline{\nabla}_j \overline{\Omega} - 2\overline{h}_{ij}]_\Lambda = 0$$

($\overline{\nabla}_i$ being the covariant derivative with respect to $\overline{h}_{ij}$), imply the asymptotic conditions (1.4) for Ω. The metric is therefore asymptotically flat at spatial infinity in the sense of Geroch. The arguments outlined in Sec. III then show that the structure of all such spacetimes are completely determined near spatial infinity by the Geroch–Hansen multipole moments. The asymptotic conditions (1.4) have, in fact, been explicitly used there to relate the multipole moments to the basic field variables describing the stationary vacuum metric. These considerations strongly suggest that the conditions of Theorem 1 (with $l = 0$) indeed encompass all the stationary vacuum metrics which one might want to regard as asymptotically flat at spatial infinity from a physical standpoint. The condition that the rescaled metric $\widetilde{h}_{ij}$ be C^4 near Λ is admittedly a little awkward which one might wish to replace by the minimal requirement that $\widetilde{h}_{ij}$ be C^2 near Λ. The present method of proof of the analyticity theorem is then no longer valid. This, however, only seems to be a minor technical point since if $\widetilde{h}_{ij}$ is C^2 near Λ but not, say C^3 (so that $\mathscr{C}[\widetilde{R}_{ij}]_\Lambda$ is only C^0 there[19]), then the multipole moments themselves are not expected to be well defined for orders higher than the quadrupole.

In conclusion it may be remarked that, although we have shown that the multipole moments completely determine the local geometry of an asymptotically flat, stationary, vacuum metric, the program of obtaining explicit formulae for the coefficients of the multipole expansions in terms of the moments appears to present a formidable task in view of the complexity of the rescaled field equations. However, one might expect considerable simplification in the special case of the static solutions (e.g., the Weyl metrics) where only the Ricci terms contribute to the multipole moments.

ACKNOWLEDGMENT

The author is pleased to thank Professor R. Geroch for many stimulating discussions.

[1]D. Zipoy, J. Math. Phys. 7, 1137 (1966); B. H. Voorhees, Phys. Rev. D 2, 2119 (1970).

[2]For a recent extensive survey, see K. Thorne, Rev. Mod. Phys. 52, 2, Part I, (1980). This article contains a detailed bibliography of earlier works in this field.

[3]R. Penrose, in *Battelle Rencontres*, edited by B. S. DeWitt and J. A. Wheeler (Benjamin, New York, 1968).

[4]R. Geroch, J. Math. Phys. **11**, 2580 (1970).
[5]R. O. Hansen, J. Math. Phys. **15**, 46 (1974).
[6]F. J. Ernst, Phys. Rev. **167**, 1175 (1968); J. Math. Phys. **12**, 2395 (1971).
[7]P. Kundu, University of Rochester Report No. UR-754 (to be published in J. Math. Phys.).
[8]H. Müller zum Hagen, Proc. Camb. Phil. Soc. **67**, 415 (1970); *ibid.* **68**, 199 (1970).
[9]C. B. Morrey, in *Contributions to Theory of Partial Differential Equations*, Annals of Mathematics Studies No. 33 (Princeton U. P., Princeton, N.J., 1954), pp. 101–56.
[10]C. B. Morrey, Am. J. Math. **80**, 198–237 (1958).
[11]R. Beig and W. Simon, Universität Wien Preprint No. UWThPh-80-14.
[12]B. C. Xanthopoulos, J. Phys. A **12**, 1025 (1979).
[13]R. Geroch, J. Math. Phys. **12**, 918 (1971).
[14]A function $f(x)$ is said to be uniformly Hölder-continuous with exponent ν $(0<\nu<1)$ on an open set $\mathscr{D}\subset\mathbb{R}^m$ if there exists a constant K such that

$$|f(x') - f(x'')| < K\,\|x' - x''\|^{\nu} \quad \forall x',x''\in\mathscr{D},$$

where $\|x' - x''\|$ is the Euclidean distance between x' and x'' in $\mathscr{D}$.

[15]In this theorem it is enough to assume $u^B(x)$ to be C^2 since the first part of Theorem 2 then automatically implies that $u^B(x)$ will be $C^{2+\nu}$.

[16]C. Hoenselaers, Prog. Theor. Phys. **55**, 466 (1976).
[17]A. Z. Petrov, *Einstein Spaces* (Pergamon, Oxford, 1969).
[18]These conditions seem to be quite reasonable. In fact it is easy to show that if in the spirit of Refs. 5 and 7 one assumes $\bar{\Omega},\bar{\Phi}$ and the auxiliary quantity $\bar{w} = \bar{\Omega}^{-1}\bar{h}^{mn}\bar{\nabla}_m\Omega\bar{\nabla}_n\Omega$ to be C^2 and $\bar{h}_{mn}$ to be $C^{4+\nu}$ $(0<\nu<1)$ near Λ, then $\bar{\Phi}$ is automatically $C^{4+\nu}$ near Λ. *Outline of proof*: By the Lemma 3 of Ref. 7, the quantity

$$\bar{\kappa}^4 = \tfrac{1}{2}\lambda^{-2}\bar{h}^{mn}(\bar{\nabla}_m\lambda\bar{\nabla}_n\lambda + \bar{\nabla}_m\omega\bar{\nabla}_n\omega)$$

is C^2 near Λ. The considerations of Lemma 1 and 2 then show (remembering that $\bar{h}_{mn}$ is now $C^{4+\nu}$ and Hansen's elliptic system involves third derivatives of $\bar{h}_{mn}$) that $\bar{\kappa}$ is of class $C^{3+\nu}$ near Λ. The coefficients of the elliptic equation

$$(\bar{\Delta} - \tfrac{1}{8}\bar{R})\bar{\Phi} = \tfrac{15}{8}\bar{\kappa}^4\bar{\Phi}$$

are then at least $C^{2+\nu}$ near Λ from which our assertion follows.

[19]This can be seen by the following argument: Since $\tilde{h}_{ij}$ is supposed to be only C^2 near Λ but not C^3, $\tilde{R}_{ij}$ is only C^0 which via Eq. (2.14) imply that w is of class C^0 near Λ. Equations (2.16) and (2.17) then show that K_{ij} is C^0 near Λ. From Eq. (2.23) it follows that $w_{;j}$ is C^0, i.e., w is C^1 near Λ. A second application of Eq. (2.14) then further shows that $\tilde{R}$ is C^1 near Λ. Thus $\mathscr{C}[\tilde{R}_{ij}]$ can only be C^0 near Λ.

Thermal fluctuations in quantum gravity: A semiclassical description

Anne Magnon-Ashtekar

Département de Mathématiques, Université de Clermont-Fd., 63170 Aubière, France[a]
and Raman Research Institute, Hebbal P. O., Bangalore 560006, India

(Received 28 May 1980; accepted for publication 24 April 1981)

By Wick rotating the Levi-Civita metric in an appropriate coordinate range, a positive definite, asymptotically Euclidean metric is obtained on R^4. This metric has the following properties: (a) it is smooth everywhere except on a 3-sphere Σ, where it suffers (finite) discontinuities; (b) it is Ricci-flat everywhere except on Σ; and (c) it is periodic in "imaginary time" both inside and outside Σ but with two *distinct* periods (reflecting the fact that the corresponding Lorentzian section covers the region bounded by two distinct horizons). In the Euclidean approach to quantum gravity, each region with a fixed period may be regarded as being at a fixed temperature. Therefore, in the semiclassical approximation, the metric represents an interesting extension of the familiar states of thermal equilibrium of the gravitational field.

PACS numbers: 04.60. + n

1. INTRODUCTION

In the quantum theory of interacting fields most approaches, presently available, are the perturbative ones. Over the past few years, however, the importance of the possible nonperturbative effects has gradually been appreciated. In the absence of an exact theory, one is forced to resort to semiclassical methods in order to describe these effects. Several such methods have been proposed. In particular, mathematical aspects of Feynman path integrals have led to the idea of Euclideanization and a lot of effort has been devoted to the investigation of properties of solutions to the Yang–Mills equations in the Euclidean space. The viewpoint here is that Euclidean solutions with certain properties signal the occurrence of physically interesting quantum processes, which escape the perturbative expansions. Particularly interesting examples of such solutions are instantons and merons. There does not yet exist a comprehensive set of rules to decide which properties in the Euclidean domain are to play important roles in quantum physics; Euclidean physics is still very young. Any solution with mathematically interesting properties is therefore a potential candidate for being useful; the only criterion is whether or not a natural physical interpretation can be associated with it in, e.g., the path integral scheme.

In the case of quantum gravity, the issue is even less settled. Due to the absence of a preferred time, what is meant by Euclideanization is not *a priori* clear. For instance, one can decide to complexify Lorentzian solutions of Einstein's equation and investigate their Euclidean sections. The difficulty is of course that relatively few solutions admit such Euclidean sections. Another viewpoint[1] is to look for Euclidean solutions in their own right and then search for a suitable physical interpretation. This had led to the notion of gravitational instantons.[2] However, the role played by gravitational instantons has turned out to be much more diverse than that of their Yang–Mills analogs. At least three types of gravitational instantons, with distinct physical meanings, have emerged so far: The black-hole instantons describe states in thermal equilibrium (for which, incidentally, the label instanton is a misnomer, since they are time *independent*; the term instanton was originally coined for objects "localized in Euclidean time"!); the compact instantons contribute to the space–time foam picture[3]; while the locally asymptotically flat instantons, as in the Yang–Mills case, signal quantum tunnelling processes. Thus, in the gravitational case, the rules underlying Euclidean physics are, at the moment, even more flexible: any Euclidean object which is likely to have a simple physical interpretation should be considered.

The purpose of this note is to present such an object. As we shall see, its existence seems to describe, in the full quantum domain, a thermal fluctuation of the gravitational field involving the presence of two regions at different temperatures.

2. THE MODEL

Recall that a state of thermal equilibrium of the gravitational field is described by a black-hole instanton, periodic in imaginary time. The frequency of this "time rotation" is identified with the temperature of the equilibrium state. We now wish to find an Euclidean solution of Einstein's equation, representing a state with two regions at different temperatures. Although, *a priori*, such a solution need not arise from complexification of a Lorentzian space–time, for simplicity let us restrict ourselves to the case when it does. Then, intuitively, one would expect this space–time to exhibit two horizons and a Killing vector field $\partial/\partial t$, which remains timelike in the region connecting these horizons. Such a situation is of course uncommon. Consider, for instance, the Kerr space–time. It does exhibit two horizons but the Killing field $\partial/\partial t$ fails to be timelike in the region in between. The Kerr instanton, based on the asymptotic region (where $\partial/\partial t$ is timelike) describes in fact a state of thermal equilibrium (its temperature being determined by the area of the outer horizon). Therefore, most of the Euclidean solutions describing thermal fluctuations are likely to exist in their own right, without any reference to a Lorentzian space–time. Nevertheless, as we shall show now, there is an appropriate setup in which such a Euclidean solution arises in connection with

[a]Permanent address.

 0022-2488/81/092012-05$01.00

a Lorentzian section. This setup is provided by the vacuum Lorentzian C-metric. That such a situation can occur may be traced back to the fact that the C-metric, in its maximal extension, appears to represent the gravitational field of a pair of black holes. (Recall that solutions describing an isolated black hole give rise, in the Euclidean domain, only to equilibrium states.)

A. Properties of the C-metric

The (hyperbolic) vacuum C-metric[4–6] is given by[7]

$$ds^2 = r^2\{ -F(y)dt^2 + G(x)dz^2 + F^{-1}(y)\,dy^2 + G^{-1}(x)\,dx^2\}, \tag{1}$$

with $r = A^{-1}(x+y)^{-1}, G(x) = 1 - x^2 - 2mAx^3$, $F(y) = -G(-y)$. *A priori*, no restriction is made on the range of coordinates x, y, z, t, or the sign of the arbitrary parameters m and A. This metric is considered as the gravitational analog of the Born solution in electrodynamics and is interpreted as the gravitational field of a pair of uniformly accelerated black holes.[6] Since three coefficients of the metric have the same sign, and since $\partial/\partial t, \partial/\partial z$ are Killing vector fields, Euclidean Einstein solutions will be obtained along the section $\tau = it$, provided $F(y)$ and $G(x)$ have the same sign. Consequently, we must exclude the values $|mA| \geqslant 27^{-1/2}$ (corresponding to double root or single root for the cubics F and G) since, in this case, regions where F and G have the appropriate sign are connected to the curvature singularity at $r = 0$. Let us assume that $|mA| < 27^{-1/2}$ and denote by $x_3 < x_2 < x_1$ the three distinct roots of $G(x)$. It is easy to find, in the manifold of orbits of the two Killing fields, the region which is located in between the Killing horizons and therefore hidden from the curvature singularity at $r = 0$. In the chart x, y this region is defined by $x_3 < x < x_1$ and $-x_1 < y < -x_3$. A simple inspection of the graph of the cubics F and G [related by $F(y) = -G(-y)$] shows the existence of two and only two subregions in which $F(y)\cdot G(x) > 0$. These are region I: $\{x_2 < x < x_1\} \times \{-x_2 < y < -x_3\}$ and region II: $\{x_3 < x < x_2\} \times \{-x_1 < y < -x_2\}$. These regions are shown in Figs. 1 and 2 which correspond, respectively, to $27^{-1/2} > mA > 0$ and $-27^{-1/2} < mA < 0$. (Recall that the Minkowskian limit is obtained for $m = 0$.) The "boundary at infinity" gets projected, on the manifold of orbits, along the line $y + x = 0$, i.e., $|r| = \infty$. The pullbacks to space–time of regions I and II are therefore asymptotically flat, since the Riemann tensor goes to zero[6] as r^{-3}. (They are connected to infinity at the point Λ where $x = x_2$ and $y = -x_2$.) Notice that the product $F(y)\cdot G(x)$ is not positive in regions III and IV; these two regions are asymptotically flat, the Killing field $\partial/\partial t$ being spacelike in both regions.

It will be convenient to introduce new coordinates covering regions I and II. These are

$$x^*(x) = \int_{x_2}^{x} |G(u)|^{-1/2}\,du, \tag{2a}$$

taking values in $[0, x^*(x_1)]$ for region I and $[x^*(x_3), 0]$ for region II; and

$$y^*(y) = \int_{-x_2}^{y} |F(v)|^{-1/2}\,dv, \tag{2b}$$

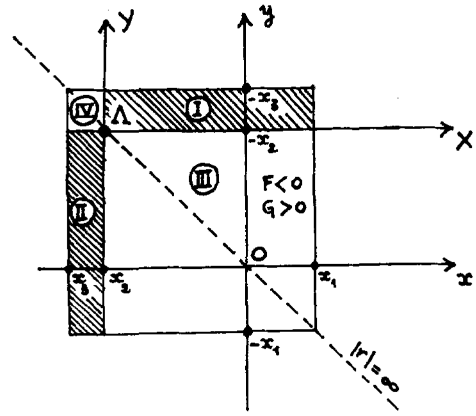


FIG. 1. $0 < mA < (27)^{-1/2}$.

taking values in $[0, y^*(-x_3)]$ for region I, and $[y^*(-x_1), 0]$ for region II. In these coordinates, the metric on the manifold of orbits takes the simple form

$$\pm\, d\sigma^2 = \pm\, r^2(dx^{*2} + dy^{*2})$$

in regions I and II, respectively. [The integrals converge since $G(u) \sim (u - x_2)$ and $F(v) \sim (v + x_2)$ in the neighborhood of $u = x_2$ and $v = -x_2$, respectively.] It is also useful to notice the existence of a value of the parameter $(mA)^2$ for which regions I and II appear as squares in the chart x, y. This comes about as follows. The graph of G (respectively, F) is symmetric with respect to the point $x = x_2$ (respectively, $y = -x_2$) if and only if $(mA)^2 = (2\times 27)^{-1}$: for this value, one has $G_{mA}(x) = (1/6mA)X - 2mAX^3 = -G_{-mA}(x) = -F_{mA}(-x)$, where we have set $X = x - x_2$ (with $x_2 = -1/6mA$). Consequently, there exists a discrete isometry between region I (respectively, II) of Fig. 1 and region II (respectively, I) of Fig. 2: one has the equality

$$d\sigma^2_{mA}(X, Y) = d\sigma^2_{-mA}(-X, -Y), \tag{3}$$

where we have set $Y = y + x_2$.

From now onwards, we shall fix $mA = (2\times 27)^{-1/2}$ and

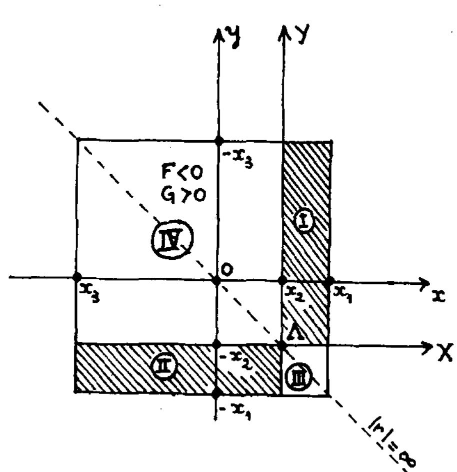


FIG. 2. $-(27)^{-1/2} < mA < 0$.

focus on region I; as we shall see, region II leads to an identical structure.

B. The Euclidean section

We now wish to study the Euclidean solution of Einstein's equation, based on region I of Fig. 1. This solution is of course obtained by restricting oneself to the section $\tau = it$ of the complexified space–time, and is represented by the metric

$$dS^2 = r^2(F(y)\,d\tau^2 + G(x)\,dz^2 + G^{-1}(x)\,dx^2 + F^{-1}(y)\,dy^2). \tag{4}$$

In order to study the topology and differential structure of the underlying manifold, one might consider, as in Ref. 6, the 2-dimensional sections obtained by keeping y and τ fixed ($y \neq -x_2$). These sections can be viewed as topological 2-spheres with north pole at $x = x_2$ and south pole at $x = x_1$. The south pole, however, exhibits a nodal singularity.[6] Similarly, the sections obtained by keeping x and z fixed ($x \neq x_2$), appear as 2-spheres with a nodal singularity at the south pole $y = -x_3$. To reveal the presence of these nodal singularities, let us cover the "manifold" of orbits (i.e., region I) by the following two charts: (x, y) in $[x_2,x_1[\times[-x_2,-x_3[$, originating at $(x_2,-x_2)$ and (x,y) in $]x_2,x_1]\times]-x_2,-x_3]$ originating at $(x_1,-x_3)$. These charts are convenient to compute the periods of the Killing rotations. Let us consider $\partial/\partial\tau$. In the first chart, its period is given by $4\pi/F'(-x_2) = -2\pi[mA(x_1-x_2)(x_3-x_2)]^{-1}$, while in the second chart [originating at $(x_1,-x_3)$], by[8] $4\pi/F'(-x_3) = -2\pi[mA(x_1-x_3)(x_2-x_3)]^{-1}$. The two periods are distinct. A possible interpretation is to consider that the two charts are incompatible. As a result one is faced with a 2-dimensional "sheet of nodal singularities". If we choose to work in the first chart, these singularities appear at the south poles of the $y-\tau$ 2-spheres, i.e., the points not covered by the chart.

We shall take an alternative viewpoint and show that these 2-dimensional "sheets" of nodal singularities can be avoided by choosing an appropriate differential structure. Let us define $\bar r = (x^* + y^*)$ and $\bar\rho = (x^* - y^*)$, for which the metric of Eq. (4) becomes

$$dS^2 = r^2(F(y)\,d\tau^2 + G(x)\,dz^2 + \tfrac{1}{2}d\bar r^2 + \tfrac{1}{2}\,d\bar\rho^2). \tag{5}$$

Recall that we have restricted ourselves to the case $mA = (2\times 27)^{-1/2}$, for which region I is a "square". For this value of mA, $x_1 - x_2 = x_2 - x_3$ and x^*, y^* define the same monotonic (increasing) function on $[x_2,x_1]$ and $[-x_2,-x_3]$, respectively, with values in $[0,x^*(x_1) = M]$. Hence $\bar r$ takes its values in $[0,2M]$ and $\bar\rho$ in $[-M,M]$. (Note that x^* and y^* are not smoothly related to x and y at the zeros of F and G). Let us further introduce, in the region $\bar r < M$ (respectively, $\bar r > M$) the following coordinates: $\phi = z\tilde k_a^{-1}$, $\psi = \tau k_a^{-1}$ (respectively, $\phi = z\tilde k_i^{-1}, \psi = \tau k_i^{-1}$) both in $[0,2\pi[$. Here $k_a,\tilde k_a$ (respectively, $k_i,\tilde k_i$) are the periods computed in Ref. 8. Our manifold structure will be defined by the chart $(\bar r, \bar\rho, \phi, \psi)$. We shall now show that the 3-surfaces $\Sigma(\bar r)$, defined by $\bar r =$ constant, are diffeomorphic to 3-spheres. The 3-surface $\bar r = M$—which corresponds in Fig. 3 to the diagonal $\bar r_d$—will

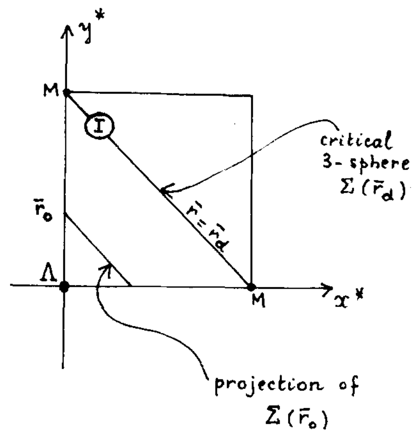


FIG. 3. $mA = (54)^{-1/2}$.

require further study, since it exhibits a metric discontinuity. As we shall see, this is due to the fact that $\Sigma(\bar r_d)$ connects two roots of $F(y)$ and $G(x)$. In the two regions $\bar r < \bar r_d$ (called "asymptotic region") and $\bar r > \bar r_d$ (called "inner region") the metric will have no pathology whatsoever.

Fix a 3-surface $\Sigma(\bar r_0), \bar r_0 \neq \bar r_d$, say, $\bar r_0 < \bar r_d$. Then we have

Proposition: $\Sigma(\bar r_0)$ is diffeomorphic to S^3, the induced metric on it being regular everywhere.

Proof: Consider a 2-plane R^2_a, with polar coordinates $(\rho_a,\theta_a) = (x^*,\phi)$, $x^* \geqslant 0$ and $0 \leqslant \phi < 2\pi$, equipped with a metric $dS^2_a = G(x)\tilde k^2_a d\phi^2 + (dx^*)^2$ and a 2-plane R^2_b, with coordinates $(\rho_b,\theta_b) = (y^*,\psi)$, $y^* \geqslant 0$, $0 \leqslant \psi < 2\pi$, and metric $dS^2_b = F(y)k^2_a d\psi^2 + (dy^*)^2$. Let $V_4 = R^2_a \times R^2_b$. It is easy to verify that the metric on V_4 is well defined everywhere, including points with $x^* = 0$ or $y^* = 0$. Consider the 3-dimensional surface $\mathscr{S}(\bar r_0)$ of V_4, defined by $x^* + y^* = \bar r_0$. $\mathscr{S}(\bar r_0)$ is a smooth, compact, metric submanifold of V_4. Furthermore, it is easy to show, e.g., using the Cartesian coordinates on V_4, that every line through the origin of V_4 intersects $\mathscr{S}(\bar r_0)$ in exactly two points. Hence $\mathscr{S}(\bar r_0)$ is diffeomorphic to S^3. Finally, by inspection, $\mathscr{S}(\bar r_0)$ is conformally isometric to $\Sigma(\bar r_0)$, the conformal factor r^2 being regular everywhere. Hence the result.[9]□

Consider, now, the critical value $\bar r = \bar r_d$. Since the manifold structure is defined by the chart $(\phi,\psi,\bar r,\bar\rho)$, it is clear that $\Sigma(\bar r_d)$ is diffeomorphic to a 3-sphere. On this 3-sphere, however, the space–time metric (5) exhibits a discontinuity. This is due to the fact that $\tau = k_a\psi$ outside $\Sigma(\bar r_d)$ while $\tau = k_i\psi$ inside (respectively, $z = \tilde k_a\phi$ outside and $z = \tilde k_i\phi$ inside). Such a discontinuity could be expected from the jump in the periods of the imaginary time rotation $\partial/\partial\tau$, or the fact that the two horizons of the Killing field $\partial/\partial t$ have distinct areas.

The above results enable us to conclude that the Euclidean "solution" based on region I appears as a family of nested 3-spheres originating at one point ($x = x_1, y = -x_3$) (where the orbits of $\partial/\partial\tau$ and $\partial/\partial z$ degenerate to a single point) and expanding up to infinity (i.e., the point Λ of Fig. 3) after crossing the critical sphere at $\bar r = \bar r_d$, where the metric is discontinuous. The underlying topology is R^4. The region connected to Λ (i.e., $\bar r < \bar r_d$) is asymptotically Euclidean.[10]

3. DISCUSSION

We have presented an Euclidean solution attached to a particular value of the parameter mA: $mA = (54)^{-1/2}$. What is the situation for the other values satisfying $0 < |mA| < (27)^{-1/2}$? It turns out that the above results are easy to extend. Let us focus, for instance, on the situation depicted in Fig. 1, region I. (The results are identical for any other permissible region.) Recall that region I can be covered by the chart (x^*, y^*) [as defined via Eqs. (2a) and (2b)]. Here, $x^* \in [0,M]$ and $y^* \in [0,N]$. Set $\bar{r} = y^* + (N/M)x^*$, $\bar{r} \in [0,2N]$ and $\bar{\rho} = y^* - (N/M)x^*$, $\bar{\rho} \in [-N,N]$. Consider, next, the 3-surfaces $\Sigma(\bar{r}_0)$ obtained for a fixed value $\bar{r}_0$ of $\bar{r}$. We claim that $\Sigma(\bar{r}_0)$ is again diffeomorphic to S^3, the induced metric being regular on this S^3 if $\bar{r}_0 \neq N$; if $\bar{r}_0 = N$, the metric has a discontinuity on the corresponding sphere $\Sigma(N)$. (The proof is completely analogous to that of the Proposition in Sec. 2. B.) Thus, the Euclidean "solution" appears again as a family of nested 3-spheres, the underlying topology being R^4. The metric is regular outside the critical 3-sphere, and is asymptotically Euclidean. Finally, there exists a 2-parameter family of such Euclidean "solutions," parameters m and A being restricted by the condition $0 < |mA| < (27)^{-1/2}$.

What is the physical interpretation associated with those solutions? Recall, to begin with, the situation in the case of the Schwarzschild instanton. In the Lorentzian section, one considers the region $r \geqslant 2m_0$, in which the Killing field $\partial/\partial t$ is everywhere nonspacelike. The instanton is obtained by taking the Euclidean section of this region. The Killing field $\partial/\partial\tau$ now appears as a rotation with a period $8\pi m_0$, m_0 being the Schwarzschild mass. The instanton is interpreted as representing, *in the semiclassical approximation*, a state of thermal equilibrium (with temperature $1/8\pi m_0$) of the quantized gravitational field. In the case of the C-metric, we have considered the space–time region bounded by *two* horizons, in which the Killing field $\partial/\partial t$ is everywhere nonspacelike. In the Euclidean section, $\partial/\partial\tau$ appears, as expected, as a rotation. However, it exhibits two *distinct* periods[11]—$2\pi m[(x_1 - x_3)(x_2 - x_3)]^{-1}$ for the inner region and $-2\pi m[(x_1 - x_2)(x_3 - x_2)]^{-1}$ for the asymptotic one—reflecting the existence of two horizons in the Lorentzian section. Therefore, it appears natural to interpret the Euclidean C-metric as representing, again in the semiclassical approximation, a state of the quantized gravitational field involving two regions at different temperatures: $T_i = (x_1 - x_3)\times(x_2 - x_3)/2\pi m$ (associated with the inner region) and temperature $T_a = (x_2 - x_1)(x_3 - x_2)/2\pi m$ (associated with the asymptotic region).[11] As one might expect on intuitive grounds, $T_i > T_a$ for all permissible values of the parameters, i.e., for all m and A subject to $0 < |mA| < (27)^{-1/2}$. Thus, in the semiclassical approximation, the model describes a thermal fluctuation of the quantum gravitational field.

Let us now examine the action of these Euclidean fields. Recall[1] that this action is given by $\int\int\int\int_M R(\sqrt{-g})d^4x + \int\int\int_{\partial M}\pi(\sqrt{h})\,d^3x$, where ∂M is the boundary of the manifold; h_{ab}, the 3-metric induced on ∂M by dS^2, and π, the extrinsic curvature of ∂M. In practice, for a solution whose boundary is at infinity, ∂M is first chosen at finite distance and a limit is then taken as "∂M goes to infinity", i.e., becomes the actual boundary of the solution. Unfortunately, this procedure leads to a divergent integral even in the case of flat space. Consequently, a renormalization procedure has to be introduced: one replaces the integrand $\pi\sqrt{h}$ by $(\pi - \pi_0)\sqrt{h}$, where π_0 is so chosen that the integral vanishes in the limit to flat space. What is the situation for Euclidean C-metrics? Since g_{ab} exhibits a discontinuity on the 3-sphere $\Sigma(\bar{r}_d)$, it is convenient to view the solution as the union of two disconnected regions: the inner region with $\bar{r} > \bar{r}_d$ and the asymptotic region with $0 < \bar{r} < \bar{r}_d$. The inner region has one boundary, namely, the sphere $\Sigma(\bar{r}_d)$, while the asymptotic region has two boundaries, $\Sigma(\bar{r}_d)$ and Λ, the point at infinity. The action of the inner region is obtained as the boundary term $\lim_{\bar{r}\to\bar{r}_d}\int\int\int_{\Sigma(\bar{r})}\pi(\bar{r},\bar{\rho})(\sqrt{h})\,d^3x$, where $\bar{r} > \bar{r}_d$. Similarly, the action of the asymptotic region is the sum

$$\lim_{\bar{r}\to\bar{r}_d}\int\int\int_{\Sigma(\bar{r})}\pi(\bar{r},\bar{\rho})(\sqrt{h})\,d^3x$$
$$+\lim_{\bar{r}\to 0}\int\int\int_{\Sigma(\bar{r})}[\pi(\bar{r},\bar{\rho}) - \pi_0(\bar{r},\bar{\rho})](\sqrt{h})\,d^3x,$$

with $\bar{r} < \bar{r}_d$. The extrinsic curvature π of any 3-sphere $\Sigma(\bar{r})$, $\bar{r} \neq \bar{r}_d$, can be easily computed. In the chart (ϕ,ψ,x,y), we have $\pi(x,y) = -3A/(\sqrt{2})(\sqrt{kF} + \sqrt{\tilde{k}G}) + [2(\sqrt{2})r]^{-1}[F'(\sqrt{k/F}) + G'(\sqrt{\tilde{k}/G})]$, where k and $\tilde{k}$ are, respectively, equal to $k_i, \tilde{k}_i$ and $k_a, \tilde{k}_a$ in the inner and asymptotic regions. Therefore, the integral $\int\int\int_{\Sigma(\bar{r})}\pi(\sqrt{h})\,d^3x$ has finite but distinct limits as $\bar{r}\to\bar{r}_d$ with $\bar{r} < \bar{r}_d$ and $\bar{r} > \bar{r}_d$. Thus, the discontinuity on the critical 3-sphere makes a finite contribution to the action. The boundary term at infinity, however, is not easy to compute: since the cubics F and G degenerate into quadrics in the flat space limit, an obvious choice of imbedding of the boundary into flat space, and hence, that of π_0, is simply not available. However, one can argue that, since the Riemann tensor has the same asymptotic falloff as that of the Schwarzschild instanton, the contribution of the boundary term at infinity will also be finite.

By definition, gravitational instantons are positive definite, complete, regular metrics satisfying Einstein's equation. Therefore, the above solutions are *not* instantons; they possess metric discontinuities. In the path integral approach, however, there appears to be no *a priori* reason to exclude such solutions, provided the discontinuities make only finite contributions to the total action. Indeed, in evaluating path integrals for systems with a finite number of degrees of freedom, one cannot restrict oneself to paths which are C^∞; one must include paths with a finite number of discontinuities. Furthermore, our analysis indicates that such discontinuities may well turn out to be an essential feature of all Euclidean solutions representing fluctuations due to regions of different temperature. The discontinuities, it would appear, serve as sinks or sources of heat; their occurrence is the "mildest" pathology required for the Killing field $\partial/\partial\tau$ to change its period. If one were to allow matter fields, one could arrange the situation so that there is an exchange of heat between the gravitational field and the matter fields without the occurrence of pathologies. In the absence of matter, however, external sinks of heat are essential and the discontinuities are well suited to serve this purpose. Perhaps one might gain new insight into the thermodynamics of the gravitational field by exploring this issue further.

There exist, in the literature, two generalizations of the Lorentzian C-metric: one due to Plebanski and Demianski,[12] which introduces a rotation parameter "a" by allowing the two Killing fields to possess twist, and the other, due to Ernst,[13] which introduces a new parameter "λ" which, if suitably chosen, enables one to get rid of the nodal singularities.[14] We believe that our analysis can be generalized in a straightforward way to incorporate the rotating case. This generalization will be very analogous to the one which leads one from the Schwarzschild to the Kerr instanton. Ernst's extension, on the other hand, *cannot* be Euclideanized: a systematic examination shows that, if $\lambda \neq 0$, *every* Euclidean section has a curvature singularity![15] Thus, direct extensions of the present ideas will lead only to a 3-parameter family of solutions, labeled by m, A, and a, the parameters being restricted to suitable ranges to obtain positive definite metrics free of curvature singularities.

ACKNOWLEDGMENTS

I am grateful to Tevian Dray and especially Abhay Ashtekar for numerous discussions. Thanks are also due to Don Page and an anonymous referee for suggestions concerning the physical interpretation of the solution. The final version of the paper was prepared during a visit to Syracuse University; I thank the physics department for its hospitality.

[1]S. W. Hawking, *In General Relativity: An Einstein Centenary Survey*, edited by S. W. Hawking and W. Israel (Cambridge University, London, 1979).

[2]G. W. Gibbons, "Gravitational Instantons: A Survey," talk presented at the International Conference on Mathematical Physics, Lausanne (1979).

[3]S. W. Hawking, Nucl. Phys. B **144**, 349 (1978).

[4]J. Levi-Civita, Atti. Accad. Nazl. Lincei. Rend. **27**, 343 (1918).

[5]E. T. Newman and L. Tamburino, J. Math. Phys. **2**, 667 (1961).

[6]W. Kinnersley and M. Walker, Phys. Rev. D **2**, 1359 (1970); A. Ashtekar and T. Dray, "On the Existence of Solutions to Einstein's Equation with Non-Zero Bondi News," Commun. Math. Phys. (in press).

[7]In the Lorentzian section, our signature is $(-\ +\ +\ +)$.

[8]This period is computed as follows. Note first that $r^2[-F(y)\,dt^2 + F(y)^{-1}\,dy^2] = -r^2F(y)[dt^2 - d\tilde{y}^2] = -r^2F(y)k_a^2(v'w')^{-1}\,dw'dv'$, where we have successively introduced

$$\tilde{y} = \int_{-x_2}^{y} F(y)^{-1}dy, w' = \exp[(t+\tilde{y})/k_a]$$

and $v' = -\exp[(\tilde{y} - t)/k_a]$. Choosing $k_a^{-1} = -mA(x_1 - x_2)(x_3 - x_2) = \frac{1}{2}F'(-x_2)$ reduces the above metric to

$$(2r^2/mA)\frac{(y+x_1)(y+x_3)}{(x_1-x_2)^2(x_3-x_2)^2}\exp\left\{\frac{(x_3-x_2)}{(x_3-x_1)}\ln|y+x_1| + \frac{(x_1-x_2)}{(x_1-x_3)}\ln|y+x_3|dw'\,dv'\right\},$$

which is regular at $y = -x_2$. Finally, setting $2x' = v' + w'$ and $2t' = v' - w'$ the metric becomes "manifestly conformally flat" and $\partial/\partial\tau$ is now given by $\partial/\partial\tau = k_a^{-1}(x'\partial/\partial\tau' - \tau'\partial/\partial x')$. Thus, $\partial/\partial\tau$ appears as a rotational Killing field with a period $2\pi k_a = 4\pi/F'(-x_2)$. An analogous calculation in the neighborhood of $y = -x_3$ gives us the period $2\pi k_i = 4\pi/F'(-x_3)$. The same results are valid for the Killing rotation $\partial/\partial z$. In the neighborhood of $x = x_2$ its period is given by $4\pi/G'(x_2) = 2\pi\tilde{k}_a$ and in the neighborhood of $x = x_1$ by $4\pi/G'(x_1) = 2\pi\tilde{k}_i$.

[9]Note that on each 3-sphere $\Sigma(\bar{r}_0)$, the orbit of $\partial/\partial t$ or $\partial/\partial z$ degenerates to a point at the roots of F or G. For each Killing field, the locus of these points is a circle, the orbit of the nonvanishing Killing field.

[10]In the neighborhood of Λ, $F(Y) \sim Y\cdot F'(-x_2)$ and $G(X) \sim X\cdot G'(x_2)$. Thus,

$$y^* = \int_0^Y \{Y|F'(-x_2)|\}^{-1/2}dY \sim 2Y^{1/2}|F'(-x_2)|^{\ \ 1/2}$$

and therefore $Y \sim 1/4|F'(-x_2)|(y^*)^2 = a(y^*)^2$. Similarly, $X \sim a(x^*)^2$. It follows that $dS^2 = \{A^{-2}/a^2[(x^*)^2 + (y^*)^2]^2\}[a(y^*)^2dt^2 + a(x^*)^2dz^2 + dy^{*2} + dx^{*2}]$. Set $\rho = r^{1/2}$, $x^* = (1/\rho)\cos\theta$, $y^* = (1/\rho)\sin\theta$, then $dS^2 = A^{-2}[d\rho^2 + \rho^2(a\sin^2\theta\,d\tau^2 + a\cos^2\theta dz^2 + d\theta^2)]$. Since $|G'(x_2)| = |F'(-x_2)|$, $\partial/\partial\tau$ and $\partial/\partial z$ have the same period in the neighborhood of infinity. Consequently, the metric is asymptotically Euclidean, the expansion of the 3-spheres $\rho =$ constant being the same in all principal directions. (This remark is due to D. Page).

[11]The periods computed above are dimensionless, due to the fact that the parameter t in the C-metric has no dimensions. To obtain the temperature, therefore, a rescaling of the affine parameter τ by $m(mA)$ is necessary.

[12]J. F. Plebanski and M. Demianski, Ann. Phys. (NY) **98**, 98 (1976).

[13]F. J. Ernst, J. Math. Phys. **19**, 1986 (1978).

[14]T. Dray and M. Walker, Lett. Math. Phys. **4**, 15 (1980).

[15]In this case, one has again to select the region in which $F\cdot G > 0$. All such regions are connected either to the curvature singularity which persists at $r = 0$ or to the point at infinity where the Riemann tensor now diverges as $\exp[\lambda^2r^4FG]$.

Validity of the Bethe–Yang hypothesis in the delta-function interaction problem

M. K. Fung
Institute for Theoretical Physics, State University of New York at Stony Brook, Stony Brook, New York 11794

(Received 17 March 1981; accepted for publication 1 May 1981)

We give an explicit verification of the Bethe–Yang hypothesis in the one-dimensional delta-function interaction problem.

PACS numbers: 05.30.Fk

I. INTRODUCTION

The one-dimensional fermion problem with repulsive delta function interaction,

$$H = -\sum_{1}^{N} \frac{\partial^2}{\partial x_i^2} + 2c\sum_{i<j}\delta(x_i - x_j), \quad c>0, \tag{1}$$

has been solved by Yang.[1] He applied Bethe's hypothesis twice, the second time in a generalized form. The results of this second application were stated in his paper without the details of the steps. The purpose of this paper is to supply these details. In other words, we shall show that (Y18), which we shall call the Bethe–Yang hypothesis, solves the eigenvalue problem (Y14).

Recently the Bethe–Yang hypothesis has been used to obtain exact solutions of a number of models, notably the Gross–Neveu model[2] and the Kondo model.[3] Because of these developments there has been renewed interest in the Bethe–Yang hypothesis. Since some crucial steps were omitted in Yang's paper, we supply them here.

II. THE MATRIX EIGENVALUE PROBLEM

$M = 1$ case

We first consider the one down spin case. In this case the Bethe–Yang hypothesis states that

$$\Phi = F(\Lambda, y), \tag{2}$$

Where F is defined in (Y19). We shall write Φ as a $N\times 1$ coulumn matrix whose explicit form is

$$\Phi = \begin{pmatrix} F(\Lambda,1) \\ F(\Lambda,2) \\ \vdots \\ F(\Lambda,N)\end{pmatrix}. \tag{3}$$

Here Λ is defined by (Y20) which, in this case, reduces to

$$\prod_{j=1}^{N} \frac{ip_j - i\Lambda - c'}{ip_j - i\Lambda + c'} = 1. \tag{4}$$

There is one Λ in this case and Eq. (4) is the equation which determines it. It is easy to check that (4) guarantees the cyclic boundary condition

$$F(\Lambda, N+1) = F(\Lambda, 1) \quad (\text{with } p_{N+1} = p_1). \tag{5}$$

We consider now (Y14) for $j = 1$. The eigenvalue problem (Y14) is

$$\mu_1 \Phi = \Omega_1 \Phi, \tag{6}$$

where

$$\Omega_1 = X'_{21} X'_{31} \cdots X'_{N1}. \tag{7}$$

In order to discuss how Ω_1 acts on Φ, it is convenient to introduce the concept of a unit component vector $|1\rangle_y$ which is a column matrix with entry 1 at the y^{th} position and 0 elsewhere. This unit component vector represents a down spin at position y. P_{ij} operates by interchanging the spins at the i^{th} and j^{th} positions. It is easy to see that Ω_1 operating on $|1\rangle_y$ gives

$$\begin{aligned}\Omega_1|1\rangle_y &= a_{y1}|1\rangle_y + b_{y1}b_{(y-1)1}|1\rangle_{y-1} + b_{y1}a_{(y-1)1}b_{(y-2)1}|1\rangle_{y-2} \\ &\quad + \cdots + b_{y1}a_{(y-1)1}\cdots a_{31}b_{21}|1\rangle_2 + b_{y1}a_{(y-1)1}\cdots a_{31}a_{21}|1\rangle_1, \\ &\qquad \text{for } y \neq 1,\end{aligned} \tag{8}$$

and

$$\begin{aligned}\Omega_1|1\rangle_1 &= b_{N1}|1\rangle_N + a_{N1}b_{(N-1)1}|1\rangle_{N-1} + \cdots + a_{N1}a_{(N-1)1} \\ &\quad \times \cdots a_{31}b_{21}|1\rangle_2 + a_{N1}a_{(N-1)1}\cdots a_{31}a_{21}|1\rangle_1,\end{aligned} \tag{9}$$

where we have used the notation

$$X'_{ij} = a_{ij} + b_{ij}P_{ij} \tag{10}$$

for X'_{ij} defined in (Y15). Notice that there are at most two interchanges. With the aid of (8) and (9) we shall check (6) successively for $y = N, N-1, \cdots, 1$; i.e., we want to check

$$\mu_1\langle N|\Phi\rangle = \langle N|\Omega_1\Phi\rangle, \tag{11}$$

$$\mu_1\langle N-1|\Phi\rangle = \langle N-1|\Omega_1\Phi\rangle, \tag{12}$$

$$\mu_1\langle y|\Phi\rangle = \langle y|\Omega_1\Phi\rangle, \qquad y\neq N \text{ or } 1 \tag{13}$$

and

$$\mu_1\langle 1|\Phi\rangle = \langle 1|\Omega_1\Phi\rangle, \tag{14}$$

where μ_1 is given, in this case, by (Y21) as

$$\mu_1 = \frac{ip_1 - i\Lambda - c'}{ip_1 - i\Lambda + c'}. \tag{15}$$

For $y = N$, the right-hand side of (6) is

$$b_{N1}F(\Lambda,1) + a_{N1}F(\Lambda,N) = \left(\frac{ip_1 - i\Lambda - c'}{ip_1 - i\Lambda + c'}\right)F(\Lambda,N), \tag{16}$$

where we have used the cyclic condition (5). Thus (11) is verified.

For $y = N-1$, the expression $\Omega_1\Phi$ becomes

$$\begin{aligned}&a_{N1}b_{(N-1)1}F(\Lambda,1) + b_{N1}b_{(N-1)1}F(\Lambda,N) \\ &\quad + a_{(N-1)1}F(\Lambda,N-1) = \left(\frac{ip_{N-1} - i\Lambda - c'}{ip_1 - i\Lambda + c'}\right)b_{(N-1)1} \\ &\quad \times F(\Lambda,N-1) + a_{(N-1)1}F(\Lambda,N-1).\end{aligned} \tag{17}$$

 0022-2488/81/092017-03$01.00

We have summed the first two terms. The sum here is a special consequence ($y = N-1$) of the following general relation:

$$a_{N1}a_{(N-1)1}\cdots a_{(y+1)1}F(\Lambda,N+1)+b_{N1}a_{(N-1)1}\cdots a_{(y+1)1}\times F(\Lambda,N)+b_{(N-1)1}a_{(N-2)1}\cdots a_{(y+1)1}F(\Lambda,N-2)+\cdots+b_{(y+1)1}F(\Lambda,y+1)=\left(\frac{ip_y-i\Lambda-c'}{ip_1-i\Lambda+c'}\right)F(\Lambda,y). \quad (18)$$

Relation (18) can be easily proved by induction, and we shall have many occasions to use it. Now

$$\left(\frac{ip_{N-1}-i\Lambda-c'}{ip_1-i\Lambda+c'}\right)b_{(N-1)1}F(\Lambda,N-1)+a_{(N-1)1}F(\Lambda,N-1)=\left(\frac{ip_1-i\Lambda-c'}{ip_1-i\Lambda+c'}\right)F(\Lambda,N-1). \quad (19)$$

Hence (12) is satisfied.

For general $y\neq 1$ or N, again we have

$$\begin{aligned}&a_{N1}\cdots a_{(y+1)1}b_{y1}F(\Lambda,1)+b_{N1}\cdots a_{(y+1)1}b_{y1}F(\Lambda,N)+\cdots\\&\quad+b_{(y+1)1}b_{y1}F(\Lambda,y+1)+a_{y1}F(\Lambda,y)\\&\quad=\left(\frac{ip_y-i\Lambda-c'}{ip_1-i\Lambda+c'}\right)b_{y1}F(\Lambda,y)+a_{y1}F(\Lambda,y) \quad (20)\\&\quad=\left(\frac{ip_1-i\Lambda-c'}{ip_1-i\Lambda+c'}\right)F(\Lambda,y).\end{aligned}$$

We have thus verified (13).

Finally, to establish Eq. (14), i.e., for $y = 1$, we need to verify

$$a_{N1}\cdots a_{21}F(\Lambda,1)+b_{N1}\cdots a_{21}F(\Lambda,N)+\cdots+b_{21}F(\Lambda,2)=\left(\frac{ip_1-i\Lambda-c'}{ip_1-i\Lambda+c'}\right)F(\Lambda,1), \quad (21)$$

which follows from (18)

Now we have proved that the Bethe–Yang hypothesis works for (6), where μ_1 given by (15) is the eigenvalue. However, there is no peculiarity about the role of Ω_1 because of the cyclic boundary condition. We can thus draw the conclusion that

$$\mu_j=\frac{ip_j-i\Lambda-c'}{ip_j-i\Lambda+c'} \quad (22)$$

and the Bethe-Yang hypothesis works for every j. We have now finished proving that the Bethe-Yang hypothesis is valid for $M = 1$.

$M = 2$ case

For two down spins, (Y18) says that Φ can be written as

$$\Phi=\alpha F(\Lambda_\alpha,x)F(\Lambda_\beta,y)+\beta F(\Lambda_\beta,x)F(\Lambda_\alpha,y), \quad (23)$$

where α and β are constant coefficients. Again, we write Φ as an $N(N-1)/2\times 1$ matrix

$$\Phi=\begin{pmatrix}\alpha F(\Lambda_\alpha,1)F(\Lambda_\beta,2)+\beta F(\Lambda_\beta,1)F(\Lambda_\alpha,2)\\ \alpha F(\Lambda_\alpha,1)F(\Lambda_\beta,3)+\beta F(\Lambda_\beta,1)F(\Lambda_\alpha,3)\\ \vdots\\ \alpha F(\Lambda_\alpha,N-1)F(\Lambda_\beta,N)+\beta F(\Lambda_\beta,N-1)F(\Lambda_\alpha,N)\end{pmatrix}. \quad (24)$$

There are two Λ 's and the cyclic boundary condition

$$\alpha F(\Lambda_\alpha,1)F(\Lambda_\beta,y)+\beta F(\Lambda_\beta,1)F(\Lambda_\alpha,y)=\alpha F(\Lambda_\alpha,y)F(\Lambda_\beta,N+1)+\beta F(\Lambda_\beta,y)F(\Lambda_\alpha,N+1) \quad (25)$$

defines the following equations for Λ_α and Λ_β:

$$\prod_{j=1}^{N}\frac{ip_j-i\Lambda_\alpha-c'}{ip_j-i\Lambda_\alpha+c'}=\frac{\alpha}{\beta} \quad (26a)$$

and

$$\prod_{j=1}^{N}\frac{ip_j-i\Lambda_\beta-c'}{ip_j-i\Lambda_\beta+c'}=\frac{\beta}{\alpha}. \quad (26b)$$

To solve (6), we follow the same procedure as before by introducing the concept of a unit component vector $|1\rangle_{xy}$, which is a column matrix with entry 1 at the $(x, y)^{\text{th}}$ position and 0 elsewhere. The action of Ω_1 on one of these component vectors is

$$\begin{aligned}\Omega_1|1\rangle_{2x}&=a_{y1}\,[a_{x1}|1\rangle_{xy}+b_{x1}b_{(x-1)1}|1\rangle_{(x-1)y}+\cdots+b_{x1}\cdots a_{21}|1\rangle_{1y}]\\&\quad+b_{y1}b_{(y-1)1}\,[a_{x1}|1\rangle_{x(y-1)}+b_{x1}b_{(x-1)1}|1\rangle_{(x-1)(y-1)}+\cdots+b_{x1}\cdots a_{21}|1\rangle_{1(y-1)}]+\cdots\\&\quad+b_{y1}\cdots b_{(x+1)1}\,[a_{x1}|1\rangle_{x(x+1)}+b_{x1}b_{(x-1)1}|1\rangle_{(x-1)(x+1)}+\cdots+b_{x1}\cdots a_{21}|1\rangle_{1(x+1)}]\\&\quad+b_{y1}\cdots a_{(x+1)1}\,[b_{(x-1)1}|1\rangle_{(x-1)x}+a_{(x-1)1}b_{(x-2)1}|1\rangle_{(x-2)x}+\cdots\\&\quad+a_{(x-1)1}\cdots b_{21}|1\rangle_{2x}+a_{(x-1)1}\cdots a_{21}|1\rangle_{1x}]\end{aligned} \quad (27)$$

for $x\neq 1$

and

$$\begin{aligned}&\Omega_1|1\rangle_{1y}=\\&b_{N1}\,[a_{y1}|1\rangle_{yN}+b_{y1}b_{(y-1)1}|1\rangle_{(y-1)N}+\cdots+b_{y1}\cdots a_{21}|1\rangle_{1N}]+\cdots+a_{N1}\cdots b_{(y+1)1}\,[a_{y1}|1\rangle_{y(y+1)}+b_{y1}b_{(y-1)1}|1\rangle_{(y-1)(y+1)}\\&+\cdots+b_{y1}\cdots a_{21}|1\rangle_{1(y+1)}]+a_{N1}\cdots a_{(y+1)1}\,[b_{(y-1)1}|1\rangle_{(y-1)y}+\cdots+a_{(y-1)1}\cdots a_{21}|1\rangle_{1y}].\end{aligned} \quad (28)$$

(27) and (28) are the analogs of (8) and (9), except that now there are at most four interchanges of spins.

To fix α/β, we check the following equation:

$$\mu_1\langle N-1,N|\Phi\rangle=\langle N-1,N|\Omega_1\Phi\rangle, \quad (29)$$

requiring μ_1 to be of the form

$$\mu_1=\left(\frac{ip_1-i\Lambda_\alpha-c'}{ip_1-i\Lambda_\alpha+c'}\right)\left(\frac{ip_1-i\Lambda_\beta-c'}{ip_1-i\Lambda_\beta+c'}\right). \quad (30)$$

Indeed,

$$\langle N-1,N|\Omega_1\Phi\rangle = b_{N1}a_{(N-1)1}\alpha F(\Lambda_\alpha,1)F(\Lambda_\beta,N-1) + a_{N1}a_{(N-1)1}\alpha F(\Lambda_\alpha,N-1)F(\Lambda_\beta,N)$$
$$+ b_{(N-1)}\alpha F(\Lambda_\alpha,1)F(\Lambda_\beta,N) + \{\alpha\leftrightarrow\beta\}$$
$$= \left(\frac{ip_1 - i\Lambda_\beta - c'}{ip_1 - i\Lambda_\beta + c'}\right)\left(\frac{ip_1 - i\Lambda_\alpha - c'}{ip_1 - i\Lambda_\alpha + c'}\right)\alpha F(\Lambda_\alpha,N-1)F(\Lambda_\beta,N)$$
$$-\left(\frac{ip_1 - i\Lambda_\beta - c'}{ip_1 - i\Lambda_\beta + c'}\right)b_{(N-1)}\left(\frac{ip_{N-1} - i\Lambda_\alpha - c'}{ip_1 - i\Lambda_\alpha + c'}\right)\alpha F(\Lambda_\alpha,N-1)F(\Lambda_\beta,N)$$
$$+ b_{(N-1)1}\alpha F(\Lambda_\alpha,N)F(\Lambda_\beta,N+1) + \{\alpha\leftrightarrow\beta\}, \tag{31}$$

where we have used cyclic condition (26). For μ_1 to be given by (30), we obtain

$$\frac{\alpha}{\beta} = \frac{i\Lambda_\alpha - i\Lambda_\beta + c}{i\Lambda_\alpha - i\Lambda_\beta - c}. \tag{32}$$

Equations (26a), (26b), and (32) give (Y20), as equations defining Λ_α and Λ_β. Now we check the equation

$$\mu_1\langle x,y|\Phi\rangle = \langle x,y|\Omega_1\Phi\rangle, \quad \text{for } x\neq 1 \text{ and } y\neq N. \tag{33}$$

Using (27) and (28), we can obtain

$$\langle x,y|\Omega_1\Phi\rangle = \left(\frac{ip_1 - i\Lambda_\beta - c'}{ip_1 - i\Lambda_\beta + c'}\right)[a_{x1}\ \alpha F(\Lambda_\alpha,x)F(\Lambda_\beta,y) + b_{(x+1)1}b_{x1}\alpha F(\Lambda_\alpha,x+1)F(\Lambda_\beta,y) + \cdots$$
$$+ b_{(y-1)1}\cdots b_{x1}\alpha F(\Lambda_\alpha,y-1)F(\Lambda_\beta,y)] + a_{N1}\cdots a_{(y+1)1}a_{(y-1)1}\cdots b_{x1}\alpha F(\Lambda_\alpha,1)F(\Lambda_\beta,y)$$
$$+ b_{N1}\cdots a_{(y+1)1}a_{(y-1)1}\cdots b_{x1}\alpha F(\Lambda_\alpha,y)F(\Lambda_\beta,N) + \cdots + b_{(y+1)1}a_{(y-1)1}\cdots b_{x1}\alpha F(\Lambda_\alpha,y)F(\Lambda_\beta,y+1) + \{\alpha\leftrightarrow\beta\}, \tag{34}$$

where each term inside the square bracket is originally a sum like (20). Using (18), we can perform the summation further, giving

$$\langle x,y|\Omega_1\Phi\rangle$$
$$= \alpha\left(\frac{ip_1 - i\Lambda_\beta - c'}{ip_1 - i\Lambda_\beta + c'}\right)\left[\left(\frac{ip_1 - i\Lambda_\alpha - c'}{ip_1 - i\Lambda_\alpha + c'}\right)F(\Lambda_\alpha,x)F(\Lambda_\beta,y) - \left(\frac{ip_{y-1} - i\Lambda_\alpha - c'}{ip_1 - i\Lambda_\alpha + c'}\right)a_{(y-1)1}\cdots b_{x1}F(\Lambda_\alpha,y-1)F(\Lambda_\beta,y)\right]$$
$$+ \alpha\left(\frac{ip_y - i\Lambda_\beta - c'}{ip_1 - i\Lambda_\beta + c'}\right)a_{(y-1)1}\cdots b_{x1}F(\Lambda_\alpha,y)F(\Lambda_\beta,y) + \{\alpha\leftrightarrow\beta\}. \tag{35}$$

Again the eigenvalue μ_1 is given by (30) if the second and third terms and their $\{\alpha\leftrightarrow\beta\}$ in (35) sum to zero leading to (32). So (29) is satisfied.

Our remaining job is to check (6) for other values of x and y. The algebra is straightforward. We omit these steps. We reach the conclusion that the Bethe–Yang hypothesis is valid for (6). By arguments of cyclicity we observe that the Bethe–Yang hypothesis works for every j with

$$\mu_j = \left(\frac{ip_j - i\Lambda_\alpha - c'}{ip_j - i\Lambda_\alpha + c'}\right)\left(\frac{ip_j - i\Lambda_\beta - c'}{ip_j - i\Lambda_\beta + c'}\right). \tag{36}$$

We have thus finished the proof that the Bethe-Yang hypothesis is valid for $M = 2$.

For $M > 2$, similar algebraic manipulation can be used to verify the hypothesis. The algebra is messy but no new features of difficulty appear. The general formulae are given by (Y20) and (Y21).

ACKNOWLEDGMENTS

The author is grateful to Dr. C. N. Yang for his guidance throughout the writing of this paper. This work is supported in part by NSF Contract No. 790637A01.

[1]C. N. Yang, Phys. Rev. Lett. **19**, 1312 (1967). We follow the same notation here. Formula (m) in Yang's paper is referred to as (Ym). Also, see M. Gaudin, Phys. Lett. A **24**, 55 (1967).

[2]N. Andrei and J. H. Lowenstein, Phys. Rev. Lett. **43**, 1698 (1979); A. A. Belavin, Phys. Lett. B **87**, 117 (1979).

[3]N. Andrei, Phys. Rev. Lett. **45**, 379 (1980); N. Andrei and J. H. Lowenstein, Phys. Rev. Lett. **46**, 356 (1981); P. B. Wiegmann, Pis'ma Zh. Eksp. Teor. Fiz. **31**, 392 (1980) [JETP Lett. **31**, 364 (1980)].

Renormalization group convergence for small perturbations of Gaussian random fields with slowly decaying correlations

V. A. Malyshev
Mathematical and Mechanical Department, University of Moscow, Moscow, USSR
B. Tirozzi
Istituto Matematico "G. Castelnuovo," University of Rome, Rome, Italy

(Received 30 May 1979; accepted for publication 6 March 1980)

We find that certain random fields obtained by perturbing Gaussian fields with a self-interaction potential have the same limit properties as do the random fields of statistical mechanics.

PACS numbers: 05.50. + q

INTRODUCTION

The limit theorems of the theory of probability have been applied many times to Gibbs random fields of various kinds.

The integral central limit theorem has been proved by many authors[1,2] and also the local one.[3-5] These theorems are very important because they describe the properties of Gibbs random fields outside the critical region. The critical behavior can be described by other kinds of limit theorems, that is by the existence of fixed points of the renormalization group.[6-10] In this paper we find both types of limit theorems. The first one (the integral theorem) is obtained in the case of a Gaussian random field with slowly decreasing correlations but integrable, perturbed with a small self-interaction potential; the second one is obtained in the case of a Gaussian random field with nonintegrable correlation perturbed by a potential of the same kind as before. In this case we obtain a Gaussian isotropical automodel random field with a correlation asymptotically equal to the correlation of the unperturbed Gaussian random field. As we know this is the first example of renormgroup convergence for non-Gaussian Gibbs translation invariant field to a nontrivial Gaussian field. We proceed now to exact statements.

1. FORMULATION OF RESULTS

Let σ_t be a translation invariant Gaussian random field on a lattice $\mathbb{Z}^\nu$, with

$$\langle \sigma_t \rangle_0 = 0\,, \quad \langle \sigma_t\, \sigma_{t'} \rangle_0 = \varphi(t - t')\,.$$

We denote μ_0 the corresponding probability measure on the space $\Omega = \mathbb{R}^{\mathbb{Z}^\nu}$ of configurations with Borel σ-albegra Σ; $\langle\cdot\rangle_0$ is the expectation w.r.t. μ_0. Let $u(x)$ be a real function on $\mathbb{R}$ bounded from below and such that

$$\langle u^2(\sigma_t)\rangle_0 < \infty\,.$$

Let us consider a new measure μ_Λ on Ω with density

$$d\mu_\Lambda / d\mu_0 = Z_\Lambda^{-1} \exp\Big(- \sum_{t\in\Lambda} \epsilon u(\sigma_t)\Big)$$

w.r.t. μ_0; Λ is a finite subset of $\mathbb{Z}^\nu$, $Z_\Lambda = \Big\langle e^{-\sum_{t\in\Lambda}\epsilon u(\sigma_t)}\Big\rangle_0$, $\langle\cdot\rangle_\Lambda$ is the expectation w.r.t. μ_Λ. We denote Σ_Λ the minimal σ-subalgebra of Σ such that any σ_t, $t\in\Lambda$, is measurable w.r.t. Σ_Λ. We shall write $F = F_\Lambda$ if F is Σ_Λ-measurable and $|\Lambda| < \infty$. We shall use the following:

Theorem 1.1: If

$$d = \sum_{t:0\neq t\in\mathbb{Z}^\nu} |\varphi(t)| < \varphi(0) < \infty\,, \tag{1.1}$$

then there exists $\epsilon_0 > 0$ such that for any $0 < \epsilon < \epsilon_0$ and any bounded F_Λ the following limits

$$\langle F_\Lambda \rangle \equiv \lim_{\Lambda\to\mathbb{Z}^\nu} \langle F_\Lambda \rangle_\Lambda$$

exist and have convergent cluster expansion defined in Sec. 2. The moments $\langle F_\Lambda\rangle$ uniquely define a translation invariant measure μ on Ω.

This theorem was proved in Ref. 11. We give independent proof for the sake of completeness. We use the methods of this proof also in the proof of the following theorems.

Theorem 1.2: Under the conditions of theorem 1.1 the central limit theorem holds for the random field σ_t with probability measure μ. We shall prove this theorem in Sec. 4. Let us put $t = (t^1,\ldots,t^\nu)\in\mathbb{Z}^\nu$ and

$$S_t^{(k)} = \sum_{t^i k \leqslant \tilde t^i < (t^i+1)k,\, i=1,\ldots,\nu} \sigma_{\tilde t}\,.$$

$\sigma_t^{(k)} = S_t^{(k)}/\sqrt{DS_t^{(k)}}$. (w.r.t. measure μ); $DS_t^{(k)} = \langle S_t^{(k)2}\rangle$
This transformation is the renormalization group transformation.[6-9] The easy generalization of Theorem 1.2 is the following.

Theorem 1.3: The random field $\sigma_t^{(k)}$ converges weakly (i.e., finite dimensional distributions converge) to the trivial Gaussian field ξ_t with $\langle \xi_t\rangle_0 = 0$, $\langle \xi_t \xi_{t'}\rangle_0 = \delta_{tt'}$. We want now to find conditions when the convergence is to a nontrivial Gaussian field.

Let us consider some Gaussian field with measure μ_0 such that

$$\langle \sigma_t \rangle_{0'} = 0, \quad \langle \sigma_t \sigma_{t'}\rangle_0 = \psi(t - t')\,.$$

We shall not suppose (1.1), but assume that

$$|t|^{\nu-2}\psi(t) \to C, \quad |t| \to \infty\,. \tag{1.2}$$

This field exists when $\nu \geqslant 3$. Let us define new random variables

$$\hat\sigma_t = \hat\sigma_t^{\,i,j,k} = \big[(\sigma_t - \sigma_{t+e_i}) - (\sigma_{t+e_j} - \sigma_{t+e_i+e_j})\big] - \big[(\sigma_{t+e_k} - \sigma_{t+e_k+e_i}) - (\sigma_{t+e_k+e_j} - \sigma_{t+e_i+e_j+e_k})\big]\,,$$

where $e_1,\ldots,e_\nu$ are the unit coordinate vectors and e_i, e_j, e_k are three different elements of this basis. Let us consider the new measure

$$\frac{d\mu_\Lambda}{d\mu_0} = \widehat{Z}_\Lambda^{-1}\exp\Big(-\sum_{t\in\Lambda}\epsilon u(\widehat{\sigma}_t)\Big)$$

$$\widehat{Z}_\Lambda = \Big\langle\exp\Big(-\sum_{t\in\Lambda}\epsilon u(\widehat{\sigma}_t)\Big)\Big\rangle_0 .$$

Assumption: We assume that μ_0 is such that $\varphi(t) = \langle\widehat{\sigma}_0,\widehat{\sigma}_t\rangle$ satisfies (1.1).

It is an easy matter to construct such examples. We can consider arbitrary Gaussian field σ_t satisfying (1.2) and make the transformation

$$\sigma_t \rightarrow \sigma'_{t'} = \sigma_{mt} . \tag{1.3}$$

We note that for all m

$$\sum_t |\langle\widehat{\sigma}'_0,\widehat{\sigma}'_t\rangle| < \infty , \tag{1.4}$$

and for sufficiently large m (1.1) is valid with

$$\varphi(t) = \langle\widehat{\sigma}'_0,\widehat{\sigma}'_t\rangle .$$

Theorem 1.4: For $\epsilon > 0$ sufficiently small μ_Λ tends weakly to some non-Gaussian measure μ. There is convergent cluster expansion in this case. The proof is quite similar to the proof of Theorem 1.1. Let σ_t be the non-Gaussian random field with measure μ constructed in Theorem 1.4. Now let $\nu = 3$.

Theorem 1.5: The sequence of random fields

$$\sigma_t^{(k)} = S_t^{(k)}/(DS_t^{(k)})^{1/2}$$

(w.r.t. measure μ) weakly converges when $k\rightarrow\infty$ to a Gaussian isotropic automodel vector field with correlation function $K(t,s)$ given by

$$K(t,s) = \text{const}\int_{\Lambda_t}\int_{\Lambda_s} du\, dv \frac{1}{|u-v|^{\nu-2}} , \tag{1.5}$$

where

$$\Lambda_t = \{u\in\mathbb{R}^\nu : u = t + u^{(0)}, u^{(0)} = (u_1^{(0)},...,u_\nu^{(0)}), 0\leqslant u_k^{(0)}\leqslant 1\} . \tag{1.6}$$

2. VACUUM CLUSTER EXPANSION

Let us denote

$$f(x) = \exp[(-\epsilon u(x))] , \quad f_T = \prod_{t\in T} f(\sigma_t) .$$

Then

$$Z_\Lambda = \Big\langle\prod_{t\in\Lambda} f(\sigma_t)\Big\rangle_0 = \langle f_\Lambda\rangle_0 .$$

and

$$Z_\Lambda\langle F_A\rangle_\Lambda = \langle F_A f_\Lambda\rangle_0 . \tag{2.1}$$

The right-hand side of (2.1) is the moment of $N+1$ random variables

$$F_A, f(\sigma_{t_1}),...,f(\sigma_{t_N}), \text{ where } \Lambda = \{t_1,...,t_N\} .$$

So one can expand this moment in semi-invariants and after resummation we get

$$\langle F_A f_\Lambda\rangle = \sum_{T_1,T_2,...,T_p}\langle F_A f_{T_1}\rangle_0^C\langle f_{T_2}\rangle_0^C\cdots\langle f_{T_p}\rangle_0^C$$
$$= \sum_{T_1\subset\Lambda}\langle F_A f_{T_1}\rangle_0^C Z_{\Lambda - T_1} , \tag{2.2}$$

where T_1 can be empty and where, e.g., $\langle F_A f_{T_1}\rangle_0^C$ is the semi-invariant of $T_1 + 1$ random variables $F_A, f_t, t\in T_1$, w.r.t. measure μ_0. This gives the desirable expansion in finite volume Λ:

$$\langle F_A\rangle_\Lambda = \sum_{T\subset\Lambda}\langle F_A f_T\rangle_0^C g_T^{(\Lambda)} ,$$
$$g_T^{(\Lambda)} = Z_{\Lambda - T}/Z_\Lambda . \tag{2.3}$$

In order to prove convergence we use the following systems of equations for g_T. For any finite nonempty $A\subset\mathbf{Z}^\nu$ we fix some point $t_A\in A$ and expand $Z_{\Lambda-(A-t_A)}$ in semi-invariants

$$Z_{\Lambda-(A-t_A)} = \sum_{T_1,...,T_k}\langle f_{T_1}\rangle_0^C\cdots\langle f_{T_k}\rangle_0^C ,$$

where the summation is through all partitions $T_1\cup\cdots\cup T_k$ of $\Lambda - (A - t_A)$. We can assume that $t_A\in T_1$. After resummation over $T_2,...,T_k$ we get

$$Z_{\Lambda-(A-t_A)} = \sum_T\langle f_T\rangle_0^C Z_{\Lambda-(A\cup T)} , \tag{2.4}$$

where the summation is over all T such that $t_A\in T\subset\Lambda - (A - t_A)$.

Let us denote

$$K_T = \langle f_T\rangle_0^C .$$

Then $K_{\{t_A\}}\equiv K_0$ does not depend on t_A and from (2.4) we get

$$g_{A-\{t_A\}}^{(\Lambda)} = K_0 g_A^{(\Lambda)} + \sum_T{}' K_T g_{A\cup T}^{(\Lambda)} ,$$

where in Σ' the summation is over all T such that $|T| > 1$, $t_A\in T\subset\Lambda - (-A - t_A)$. Alternatively we have a system of equations

$$g_A^{(\Lambda)} - K_0^{-1}g_{A-t_A}^{(\Lambda)} + K_0^{-1}\sum_T{}' K_T g_{A\cup T}^{(\Lambda)} = 0 , \quad |A| > 1 ,$$
$$g_A^{(\Lambda)} + K_0^{-1}\sum_T{}' K_T g_{A\cup T}^{(\Lambda)} = K_0^{-1} , \quad |A| = 1 . \tag{2.5}$$

We shall consider also the limiting system of equations

$$g_A - K_0^{-1}g_{A-t_A} + K_0^{-1}\sum_T{}' K_T g_{A\cup T} = 0 , \quad |A| > 1 ,$$
$$g_A + K_0^{-1}\sum_T{}' K_T g_{A\cup T} = K_0^{-1} , \quad |A| = 1 , \tag{2.6}$$

where in Σ' the summation is over all T such that $|T| > 1$ and $t_A\in T\subset\mathbf{Z}^\nu - (A - \{t_A\})$.

We shall prove that (2.5) and (2.6) have unique solutions in the Banach space $\mathscr{B}_\Lambda$ (respectively $\mathscr{B}$) of functions $\psi = (\psi_A)$ on the set of all nonempty finite subsets $A\subset\Lambda$ (respectively $A\subset\mathbf{Z}^\nu$) with the norm

$$\|\psi\| = \sup_A[(K_0/2)^{|A|}|\psi_A|].$$

Let us consider the linear operator $L = E - R + K$ in, e.g., $\mathscr{B}$, where E is the identity, $R\psi = \psi'$, where $\psi' = (\psi'_A)$, and

$$\psi'_A = \begin{cases} K_0^{-1}\psi_{A-t_A} , & |A|\geqslant 2 \\ 0 & |A| = 1 . \end{cases}$$

It is easy to see that $\|R\|\leqslant 1/2$. The operator K transforms the vector (ψ_A) into $(\psi_A^{''})$, where

$$\psi_A^{''}=K_0^{-1}\sum_{T:t_i\in T\subset Z^\nu-(A-t_i),\ |T|>1}K_T\psi_{A\cup T}$$

Let us note that

$$0<C_1\leqslant K_0=\langle e^{-\epsilon u(\sigma_i)}\rangle_0\leqslant C_2<\infty\ , \tag{2.7}$$

uniformly on $0\leqslant\epsilon\leqslant\epsilon_0$. It follows that

$$\|K\|\leqslant K_0^{-1}\sum_{T:0\in T\subset Z^\nu|T|>1}|K_T\|2/K_0|^{|T|-1}\ . \tag{2.8}$$

Lemma 2.1: Under the conditions of Theorem 1.1 for any $\delta>0$ there exists $\epsilon_0>0$ such that for $0<\epsilon\leqslant\epsilon_0$ one has $\|K\|<\delta$. This lemma will be proved below.

It follows that $L=E-R+K$ is invertible in $\mathscr{B}$ (and in $\mathscr{B}_A$). Thus there exists a unique solution of (2.6) [and (2.5)]. The solution $g_A^{(A)}$ of (2.5) tends to the solution g_A of (2.6) when $A\to\mathbb{Z}^\nu$. We shall not prove this fact here (see Ref. 11), as the proof is standard (see Ref. 12).

If $\|K\|\leqslant 1/4$ then $\|(E-R+K)^{-1}\|\leqslant 4$ and so

$$|g_A|\leqslant 2(K_0/2)^{-|A|}. \tag{2.9}$$

We shall prove below also

Lemma 2.2:

$$\sum_{T:|T|=n,\ T\subset Z^\nu}|\langle F_A f_T\rangle_0^C|\leqslant C(F_A)(C\epsilon)^n, \tag{2.10}$$

where C does not depend on ϵ and $F_A, C(F_A)$ depends only on F_A.

It follows that the cluster expansion

$$\langle F_A\rangle=\sum_{T\subset A}\langle F_A f_T\rangle_0^C g_T \tag{2.11}$$

is absolutely convergent. Thus Theorem 1.1 is proved.

3. BOUNDS ON SEMI-INVARIANTS OF FUNCTIONS OF A GAUSSIAN RANDOM FIELD

Let us put

$$\hat f(x)=f(x)-1=\int_0^\epsilon[-u(x)]e^{-\epsilon' u(x)}\,d\epsilon' \tag{3.1}$$

and note that

$$\langle f(\sigma_{t_1}),\dots,f(\sigma_{t_m})\rangle_0^C\equiv\langle\hat f(\sigma_{t_1}),\dots,\hat f(\sigma_{t_n})\rangle_0^C\ . \tag{3.2}$$

We shall estimate the semi-invariants (3.2) under the conditions of Theorem 1.1. Let us put $t_1=0$ and we shall always assume that $t_1,t_2,\dots,t_n$ are pairwise different. Let $\Sigma^{(n)}$ be the sum over all lexicografically ordered sequences $(t_2,\dots,t_n)$, the components of which are pairwise different and $0\neq t_i\in\mathbb{Z}^\nu$, i.e., over all subsets $T=\{t_2,\dots,t_n\}\subset\mathbb{Z}^\nu-\{0\}$. Lemma 2.1 evidently follows from Lemma 3.1.

Lemma 3.1:

$$\Sigma^{(n)}|\langle f(\sigma_0),f(\sigma_{t_2}),\dots,f(\sigma_{t_n})\rangle_0^C|\leqslant(C\epsilon)^n. \tag{3.3}$$

We shall prove this lemma now. Let $h_n(\sigma_t)$ be the normed Hermite polynomials w.r.t. μ_0, i.e., $\langle h_n^2(\sigma_t)\rangle_0=1$; we can assume that $\varphi(0)=\langle\sigma_t^2\rangle_0=1$ without loss of generality. Then $d<1$.

Let us denote

$$:\sigma_t^{(n)}:=\sqrt{n!}\,h_n(\sigma_t)$$

the corresponding Wick polynomials.

We expand $\hat f(\sigma_t)$ in Hermite polynomials in $L_2(d\mu_0)$.

$$\hat f(\sigma_t)=\sum_{m=0}^\infty a_m h_m(\sigma_t)=\sum_{m=0}^\infty\frac{a_m}{\sqrt{m!}}:\sigma_t{}^m:\ .$$

Then

$$\sum_{m=0}^\infty(a_m)^2=\int|\hat f|^2\,d\mu_0\leqslant(C\epsilon)^2,$$

and so

$$|a_m|\leqslant C\epsilon\ . \tag{3.4}$$

Then

$$\langle\hat f(\sigma_0),\hat f(\sigma_{t_2}),\dots,\hat f(\sigma_{t_n})\rangle_0^C$$
$$=\sum_{m_1=1}^\infty\cdots\sum_{m_n=1}^\infty\frac{a_{m_1}\cdots a_{m_n}}{\sqrt{m_1!\cdots m_n!}}\langle:\sigma_0^{m_1}:,\dots,:\sigma_{t_n}^{m_n}:\rangle_0^C\ , \tag{3.5}$$

if the series in the right-hand side of (3.5) is absolutely convergent. Let us fix the ordered sequence $(m_1,m_2,\dots,m_n)$.

Lemma 3.2:

$$\sum_{(t_2,\dots,t_n)}\left|\langle:\sigma_0{}^{m_1}:,:\sigma_{t_2}{}^{m_2}:,\dots,:\sigma_{t_n}{}^{m_n}:\rangle_0^C\right|$$
$$\leqslant(n-1)!d^{N/2}\prod_{i=1}^n m_i(m_i-2)(m_i-4)\cdots\ , \tag{3.6}$$

where $N=m_1+\cdots+m_n$ and the sum $\Sigma_{(t_2,\dots,t_n)}$ is over all the *ordered* sequences $(t_2,\dots,t_n)$ of points $t_i\in\mathbb{Z}^\nu$ such that $t_i\neq 0, t_i\neq t_j$, if $i\neq j$. We shall prove this lemma below and now using (3.6) we shall prove Lemma 3.1.

Proof of Lemma 3.1: We have

$$\Sigma^{(n)}|\langle f(\sigma_0),f(\sigma_{t_2}),\dots,f(\sigma_{t_n})\rangle_0^C|$$
$$\leqslant\sum_{m_1=1}^\infty\cdots\sum_{m_n=1}^\infty\frac{|a_{m_1}|\cdots|a_{m_n}|}{\sqrt{m_1!\cdots m_n!}\,(n-1)!}(n-1)!d^{N/2}$$
$$\times\prod_{i=1}^n m_i(m_i-2)(m_i-4)$$
$$\leqslant\sum_{m_1=1}^\infty\cdots\sum_{m_n=1}^\infty|a_{m_1}|\cdots|a_{m_n}|2m_1\cdots 2m_n d^{(m_1+\cdots+m_n)/2}$$
$$\leqslant(C_1\epsilon)^n\ , \tag{3.7}$$

where C_1 depends on $(1-d)^{-1}$. Thus Lemma 3.1 is proved.

Proof of Lemma 3.2:

We shall use the diagrammatic representation

$$\sum_{(t_2,\dots,t_n)}\langle:\sigma_0^{m_1}:,:\sigma_{t_2}^{m_2}:,\dots,:\sigma_{t_n}^{m_n}:\rangle_0^C$$
$$=\sum I_G\ , \tag{3.8}$$

which we shall now define exactly.

A diagram G consists of the following objects: (1) An ordered set of vertices $\{1,\dots,n\}$ and of one-to-one mapping τ of $\{1,\dots,n\}$ in $\mathbb{Z}^\nu$ such that $\tau(1)=0,\tau(i)=t_i$. (2) Each vertex has an ordered set of m_i legs $(i,1),(i,2),\dots,(i,m_i)$. (3) There is a partition of the set of all $N = m_1 + \cdots + m_n$ legs onto pairs such that each pair (line of the diagram G) has legs from different vertices and the resulting graph is connected. (4). The Gaussian random variable $\sigma_{ij} = \sigma_{t_i}$ corresponds to the leg (i,j). The contribution of the diagram G is then

$$I_G = \prod \langle \sigma_{ij}, \sigma_{kl} \rangle_0 ,$$

where the product is taken over all pairs (ij,kl) of contracted legs (i.e., over lines of G). Formula (3.8) is then Wick's theorem for semi-invariants.

Let $\mathfrak{A}_N$ be the set of all sequences $\alpha = (x_1,\ldots,x_{N/2})$ with $0 \neq x_i \in \mathbb{Z}^\nu, N = m_1 + \cdots + m_n$. Let us fix a permutation

$$\Pi = \begin{pmatrix} 2 & \cdots & n \\ \pi(2) & \cdots & \pi(n) \end{pmatrix}$$

and $\hat{m}_1 = m_1, \hat{m}_i = m_{\Pi(i)}, i \geqslant 2$. For each $\alpha \in \mathfrak{A}_N$ and each Π we shall construct a set $\mathcal{G}_\pi(\alpha)$ of connected diagrams with fixed $(m_1,\ldots,m_n)$. A set $\mathcal{G}_\pi(\alpha)$ can be empty or otherwise it contains no more than $\Pi_{i=1}^n m_i(m_i - 2)(m_i - 4)\cdots$ diagrams and moreover α, π uniquely define vertices $t_2,\ldots,t_n$ which are the same for all $G \in \mathcal{G}_\pi(\alpha)$. Moreover the contribution of each $G \in \mathcal{G}_\pi(\alpha)$ *is equal to* $\Pi_{i=1}^{N/2} \varphi(x_i)$. We describe now the algorithm of construction of $\mathcal{G}_\pi(\alpha)$. This algorithm constructs vertices and contracts legs step by step.

First step: We begin with the vertex $t_1 = 0$ and contract leg (1,1) with arbitrary of $\hat{m}_2$ legs of vertex $\hat{t}_2 = x_1$. Thus the first step is finished and we proceed by induction.

Let $T = (t_1, \hat{t}_2,\ldots,\hat{t}_l), l \leqslant k$, be the vertices with $\hat{m}_1,\ldots,\hat{m}_l$ legs which have been constructed after k steps. Suppose that $r_1,\ldots,r_l$ legs of the correspondingly $\hat{m}_1,\ldots,\hat{m}_l$ are yet uncontracted. So on each step one line and 0 or 1 new vertex is constructed.

$(k+1)$st *step*: Let us consider the vertex $\hat{t}_i, \leqslant i \leqslant l$, such that $\hat{t}_i = x_1 + x_2 + \cdots x_k$. It exists by inductive construction. There can be two cases: $r_i > 0$ or $r_i = 0$.

(1) If $r_i > 0$ then consider the point $\hat{t}_i + x_{k+1}$ is not equal to any $\hat{t}_1,\ldots,\hat{t}_l$ then we construct new a vertex $\hat{t}_{l+1} = \hat{t}_i + x_{k+1}$ and we contract the *first* of r_i uncontracted legs of $\hat{t}_i$ with *arbitrary* of $\hat{m}_{l+1}$ legs of $\hat{t}_{l+1}$. If $\hat{t}_i + x_{k+1}$ is equal to some $\hat{t}_j, j \leqslant l$, then we contract the *first* of r_i uncontracted legs of $\hat{t}_i$ with *any* of the r_j uncontracted legs of $\hat{t}_j$. If $r_j = 0$ then we define $\mathcal{G}_\pi(\alpha)$ to be empty. Otherwise we proceed to the next step.

(2) If $r_i = 0$ then we take the *first* $\hat{t}_j, 1 \leqslant j \leqslant l$, such that $r_j \neq 0$ and we take the *first* of its uncontracted legs and we contract it with *any* of the uncontracted legs of the vertex $\hat{t}_p = \hat{t}_j + x_{k+1}$ (as in case 1). If all $r_j, 1 \leqslant j \leqslant l$, are equal to 0, then we define $\mathcal{G}_\pi(\alpha)$ to be empty.

We get a combinatorial factor

$$\hat{m}_i(\hat{m}_i - 2)\cdots$$

for each $\hat{t}_i$ since after each *arbitrary* choice of a leg we immediately take the *first* leg of $\hat{t}_i$. Evidently each diagram G with given $(m_1,\ldots,m_n)$ is in at least one $\mathcal{G}_\pi(\alpha)$. Lemma 3.2 is proved.

Proof of Lemma 2.2: It is sufficient to consider the case when

$$F_A = \prod_{i=1}^k f_i(x_i) ,$$

where $|f_i| \leqslant 1$ and $A = \{x_1,\ldots,x_k\}$. We use the following formula

$$\langle f_1(\sigma_{x_1}) f_k(\sigma_{x_k}), f(\sigma_{t_1}),\ldots,f(\sigma_{t_n}) \rangle_0^C = \sum \prod_{j=1}^p \langle f_i(\sigma_{x_i}), x_i \in T_j, f(\sigma_t), t \in T_j \rangle_0^C , \tag{3.9}$$

where the sum is over all the partitions $(T_1,\ldots,T_p)$ of the set $\{x_1,\ldots,x_k,t_1,\ldots,t_n\}$ such that each T_j has non empty intersection with $\{x_1,\ldots,x_k\}$. If all the t_i are different from all x_j then as in the proof of Lemma 3.1 we expand f_i and f in Hermite polynomials and we use diagrams. The diagram G will give contribution to the rhs of (3.9) iff each of its vertices t_i is connected by some path with some of the vertices $x_1,\ldots,x_k$. The proof then repeats the proof of Lemma 3.1.

If in

$$\langle f_i(x_i), x_i \in T_j, f(\sigma_t), t \in T_j \rangle_0^C , \tag{3.10}$$

some $x_i \in T_j$ is equal to some $t \in T_j$ then we use the formula

$$\begin{aligned} \langle \psi(y_1),\ldots,\psi(y_q) \rangle_0^C &= \langle \psi(y_1)\psi(y_2), \psi(y_3),\ldots,\psi(y_q) \rangle_0^C \\ &\quad - \sum_{T' \subset \{y_3,\ldots,y_q\}} \langle \psi(y_1), \psi(y_j), j \in T' \rangle_0^C \\ &\quad \times \langle \psi(y_2), \psi(y_j), y_j \in \{3,\ldots,q\} - T' \rangle_0^C \end{aligned}$$

to exclude these cases.

We use (3.11) subsequently for each pair $x_i = t_i$ and for all semi-invariants (3.10).

After this procedure we shall also have the sum of diagrams such that each of its points is connected with some of the points $x_1,\ldots,x_k$. We get also factors depending on k. Our construction is again applicable to this sum of "connected" diagrams.

4. THE CENTRAL LIMIT THEOREM

We shall prove Theorem 2 here. Let $\langle S_A,\ldots,S_A \rangle^C$ be the nth semi-invariant of the random variable $S_A = \Sigma_{t \in A} \sigma_t$. We want to show that

$$\lim_{|A| \to \infty} \frac{1}{(DS_A)^{n/2}} \langle S_A,\ldots,S_A \rangle^C = 0, \quad n \geqslant 3 . \tag{4.1}$$

We have

$$\langle S_A,\ldots,S_A \rangle^C = \sum_{t_1,\ldots,t_n \in A} \langle \sigma_{t_1},\ldots,\sigma_{t_n} \rangle^C .$$

We shall define new random variables

$$\widetilde{\sigma}_t = \sum_{\alpha=1}^n \sigma_t^{(\alpha)} \omega^\alpha ,$$

where $\omega = e^{2\pi i/n}$ and $\sigma_t^{(\alpha)} = 1,\ldots n$, are new random fields which are independent copies of the random field σ_t with measure μ. Then we have (see Ref. 13)

$$\langle \widetilde{\sigma}_{t_1} \cdots \widetilde{\sigma}_{t_k} \rangle = 0, \quad k < n ,$$

and

$$\langle \sigma_{t_1},\ldots,\sigma_{t_n} \rangle^C = (1/n) \langle \widetilde{\sigma}_{t_1},\ldots,\widetilde{\sigma}_{t_n} \rangle^C = (1/n) \langle \widetilde{\sigma}_{t_1} \cdots \widetilde{\sigma}_{t_n} \rangle . \tag{4.2}$$

Using the cluster expansion for our n independent copies of μ [it is similar to (2.11)]

$$\langle\tilde\sigma_{t_1}\cdots\tilde\sigma_{t_n}\rangle = \sum_{\{x_1,\dots,x_k\}\subset Z^\nu}\langle\tilde\sigma_{t_1}\cdots\tilde\sigma_{t_n},F(x_1),\dots,F(x_k)\rangle_0^C, \times(g_{\{x_1,\dots,x_k\}})^n, \tag{4.3}$$

where

$$F(x_i)=\prod_{\alpha=1}^n f(\sigma_{x_i}^{(\alpha)}),$$

we have

$$\langle\tilde\sigma_{t_1}\cdots\tilde\sigma_{t_n},F(x_1),\dots,F(x_k)\rangle_0^C, = \langle\tilde\sigma_{t_1},\dots,\tilde\sigma_{t_n},F(x_1),\dots,F(x_k)\rangle_0^C, \tag{4.4}$$

and

$$\langle\tilde\sigma_{t_1}\cdots\tilde\sigma_{t_p},F(x_1),\dots,F(x_k)\rangle_0^C=0,\quad p<n.$$

The proof of (4.4) is similar to that of (4.2) due to the symmetry of $F(x)$ w.r.t. permutations of $\{1,\dots,n\}$. Then

$$|\langle\tilde\sigma_{t_1},\dots,\tilde\sigma_{t_n},F(x_1),\dots,F(x_k)\rangle_0^C| \leqslant n^n\sup_{\alpha_1',\dots,\alpha_n'}|\langle\sigma_{t_1}^{(\alpha_1')},\dots,\sigma_{t_n}^{(\alpha_n')},F(x_1),\dots,F(x_k)\rangle_0^C|. \tag{4.5}$$

We expand each $f(\sigma_t^{(\alpha)})$ in the series of Wick polynomials and get as in (3.5)

$$\begin{aligned}\langle\sigma_{t_1}^{(\alpha_1')},\dots,\sigma_{t_n}^{(\alpha_n')},F(x_1),\dots,F(x_k)\rangle_0^C &= \sum_{m_1^1,\dots,m_n^1>0}\frac{a_{m_1^1}\cdots a_{m_n^1}}{\sqrt{m_1^1!\cdots m_n^1!}}\\ &\cdots\sum_{m_1^k,\dots,m_n^k>0}\frac{a_{m_1^k}\cdots a_{m_n^k}}{\sqrt{m_1^k!\cdots m_n^k!}}\\ &\times\langle\sigma_{t_1}^{(\alpha_1')},\dots,\sigma_{t_n}^{(\alpha_n')},:(\sigma_{x_1}^{(1)})^{m_1^1}:\cdots\\ &:(\sigma_{x_1}^{(n)})^{m_n^1}:,\dots,:(\sigma_{x_k}^{(1)})^{m_1^k}:\cdots:(\sigma_{x_k}^{(n)})^{m_n^k}:\rangle_0^C.\end{aligned} \tag{4.6}$$

From (4.2), (4.3), (2.9) we get

$$|\langle\sigma_{t_1},\dots,\sigma_{t_n}\rangle^C| \leqslant \sum_{\{x_1,\dots,x_k\}\subset Z^\nu}2^nB^{nk}|\langle\tilde\sigma_{t_1},\dots,\tilde\sigma_{t_n},F(x_1),\dots,F(x_k)\rangle_0^C|, \tag{4.7}$$

$B=2/K_0$.

We want to represent the semi-invariants in the rhs of (4.6) as the sum of connected diagrams G with $n+k$ vertices $t_1,\dots,t_n,\dots,x_k$. We fix $t_1=0$ and we want to prove that

$$\sum_{t_2,\dots,t_n\in Z^\nu}|\langle\sigma_0,\dots,\sigma_{t_n}\rangle^C|\leqslant\text{const.} \tag{4.8}$$

(4.8) follows from:

Lemma 4.1:

$$\sum_{t_1,\dots,t_n\in\Lambda}|\langle\sigma_{t_1},\dots,\sigma_{t_n}\rangle^C|\leqslant\text{const}|\Lambda|. \tag{4.9}$$

Proof of (4.8): Let us denote by χ the semi-invariants in the rhs of (4.6). It depends on m_i^j,α_i',x_i,t_j. Then using (4.6), (4.7) we can bound (4.8) by

$$\begin{aligned}&\sum_{t_2,\dots,t_n}\sum_{\{x_1,\dots,x_k\}\subset Z^\nu}2^nB^{nk}\sum_{m_i^j>0}\prod_{i,j}\frac{|a_{m_i^j}|}{\sqrt{m_i^j!}}|\chi|\\ &\leqslant\sum_{t_2,\dots,t_n}2^n\sum_{\{x_1,\dots,x_k\}}(BC\epsilon)^{nk}\sum_{m_i^j>0}\prod_{i,j}\cdot\frac{1}{\sqrt{m_i^j!}}|\chi|\\ &\leqslant\frac{2^n}{k!}\sum_{\substack{t_2,\dots,t_n,x_1,\dots,x_k\\ x_i\neq x_j}}(BC\epsilon)^{nk}\sum_{m_i^j>0}\prod_{i,j}\frac{1}{\sqrt{m_i^j!}}|\chi|.\end{aligned}$$

We note that for all i

$$\sum_j m_j^i\geqslant1. \tag{4.10}$$

We fix, n,k,m_i^j, and shall prove that

$$\sum_{\substack{t_2,\dots,t_n,x_1,\dots,x_k\\ x_i\neq x_j}}|\chi| \leqslant k!k^n(1+d)^n d^{\Sigma m_i^j-n/2} \times\left[\prod_{i,j}m_i^j(m_i^j-2)(m_i^j-4)\cdots\right]\prod_{i,j}m_i^j. \tag{4.11}$$

From (4.10) and (4.11) we get (4.8), (4.9).

In order to complete the proof we must prove (4.11). We show that this proof is quite similar to the proof of Lemma 3.2. In fact there are no lines between legs of the same vertex x_i since $\sigma_{x_i}^{(\alpha)},\sigma_{x_i}^{(\alpha')}$ are independent for $\alpha\neq\alpha'$. To construct the diagrams using our method we define $\mathfrak{A}_{N,n}$ to be the set of sequences

$$(y_1,\dots,y_n,z_1,\dots,z_{N-n/2}),$$

$$N=\sum_{i,j}m_i^j,\ y_i,z_j\in\mathbb{Z}^\nu,\quad z_j\neq0.$$

Let us note that if we delete vertices $0,t_2,\dots,t_n$ and delete lines connecting them to the other vertices then the remaining part of the diagram remains connected. Thus we begin with the vertex 0 and construct the second vertex. $x_{\pi(1)}=y_1$ and the line between 0 and $x_{\pi(1)}$. This gives a factor $(1+d)$.

As earlier, we fix a permutation π of $\{1,\dots,k\}$. This gives a factor $k!$. Then for each $i\in\{2,\dots,n\}$ we choose one of the k vertices $x_1,\dots,x_k$ and choose one leg from the chosen vertex to be contracted with the vertex t_i. This gives a factor $k^{n-1}\Pi_{i,j}m_i^j$. Each such a construction generates a factor $(1+d)$. The remaining connected part of the diagram is constructed in the closest similarity to the construction in the proof of Lemma 3.2. The proof of (4.8) and of Lemma 4.1 is thus completed.

Lemma 4.2:

$$\langle S_\Lambda^2\rangle\geqslant Q|\Lambda|,\quad Q>0.$$

The proof is quite similar to the preceding one. We have

$$\langle S_\Lambda^2\rangle=\sum_{t,t'\in\Lambda}\langle\sigma_t,\sigma_{t'}\rangle^C.$$

In order to bound $\langle\sigma_t,\sigma_{t'}\rangle_0^C$ we use cluster expansion. It contains the "main term" $\langle\sigma_t,\sigma_{t'}\rangle_0^C$. The sum of the remaining terms being of order $O(\epsilon)$.

We have for the "main terms"

$$\frac{1}{|\Lambda|}\sum_{t,t'\in\Lambda}\langle\sigma_t,\sigma_{t'}\rangle_0^C=\frac{1}{|\Lambda|}\langle S_\Lambda^2\rangle_0\underset{|\Lambda|\to\infty}{\to}Q_1.$$

Then Lemma 4.2 follows with $Q\geqslant Q_1-\epsilon$. It is not difficult to prove that

$$\frac{1}{|\Lambda|}\langle S_\Lambda^2\rangle\underset{|\Lambda|\to\infty}{\to}Q.$$

5. RENORMGROUP CONVERGENCE TO NONTRIVIAL GAUSSIAN AUTOMODEL FIELD

We have to show that $\langle \sigma_{t_1}^{(k)},\ldots,\sigma_{t_n}^{(k)}\rangle^C \underset{k\to\infty}{\to} 0\,, \qquad (5.1)$

and

$$\langle \sigma_t^{(k)},\sigma_{t'}^{(k)}\rangle^C \to K(t,t')\,. \tag{5.2}$$

The equivalence of these statements and of the statements of the theorem is due to the following general facts. Statement (5.1) is equivalent to the statement that the weak limit of $\sigma_t^{(k)}$ is Gaussian if it exists. Moreover if the limit exists then it is automodel and there exists only one isotropic automodel random field with asymptotics (1.5) (see Ref. 14).

We use the following asymptotic properties

$$\begin{aligned} &|\langle \sigma_t,\hat{\sigma}_{t'}\rangle_0^C| \leqslant \text{const}/|t-t'|^{\nu+1}\,, \\ &|\langle \hat{\sigma}_t,\hat{\sigma}_{t'}\rangle_0^C| \leqslant \text{const}/|t-t'|^{\nu+1}\,, \end{aligned} \tag{5.3}$$

which are quite evident.

We can write cluster expansion for

$$\langle \sigma_t,\sigma_{t'}\rangle = \langle \sigma_t,\sigma_{t'}\rangle_0 + \sum_{k\geqslant 1,\{x_1,\ldots,x_k\}\subset Z^\nu} \langle \sigma_t\sigma_{t'},F(x_1),\ldots,F(x_k)\rangle_0^C \times g_{\{x_1,\ldots,x_k\}}\,, \tag{5.4}$$

similar to the earlier cluster expansion. Convergence of this expansion follows from (5.3). From (5.3) it follows that

$$\frac{\langle \sigma_t,\sigma_{t'}\rangle^C}{\langle \sigma_t,\sigma_{t'}\rangle_0^C} \to 1 \text{ if } |t-t'|\to\infty\,, \tag{5.5}$$

because in the diagrammatic expansion of the second term in the rhs of (5.4) there is no line between vertices t and t'. Statement (5.1) can be proved quite similarly to the proof of the same statement in the preceding section. Theorem (5.1) is proved.

[1]B.S. Nakhapetian, Dokl. Akad. Nauk Arm. SSR **61**, 210 (1975).

[2]V.A. Malyshev, Dokl. Akad. Nauk SSSR **224**, 35 (1975).

[3]R. Minlos and A.M. Halfina, Izv. Akad. Nauk SSSR Ser. Mat. **34**, 1173 (1970).

[4]G. Del Grosso, Commun. Math. Phys. **37**, 141 (1974).

[5]R.L. Dobrushin and B. Tirozzi, Commun. Math. Phys. **54**, 173–192 (1977).

[6]Ya.G. Sinai, Teor. Veroyatn. Ee. Primen. **21**, 63 (1976).

[7]G. Gallavotti and G. Jona-Lasinio, Commun. Math. Phys. **41**, 301–307 (1975).

[8]P.M. Blekher, in *Mnogocomponentnie sluciaini sistemi* (Nauka, Moscow, 1978), p. 47.

[9]K.G. Wilson, Phys. Rev. B **4**, 3174 (1971).

[10]G. Gallavotti and A. Martin-Lof, Nuovo Cimento B **25**, 425 (1975).

[11]V.A. Malyshev, "Gibbsian perturbations of Gaussian random field with a slowly decreasing correlations," (preprint).

[12]D. Ruelle, *Statistical Mechanics, Rigorous Results* (Benjamin, New York, 1969), p. 82.

[13]G.S. Sylvester, "Continuous-spin Ising Ferromagnets," Ph. D. thesis, M.I.T.

[14]R.L. Dobrushin, in *Mnogocomponentnie sluciaini sistemi* (Nauka, Moscow, 1978), p. 179.

Diagonalizability of the longitudinal sector of the functional integral in massive SU (2) Yang–Mills theories via topological contributions

C. N. Ktorides[a]
Harvard University-Science Center, Cambridge, Massachusetts 02138

(Received 29 April 1980; accepted for publication 25 June 1980)

The problem of longitudinal sector diagonalizability of the functional integral in massive SU (2) Yang–Mills theories is revisited. A new decomposition law of a massive SU (2) vector field into transverse and longitudinal parts is proposed which takes into account the more recently discovered topological properties of SU (2) gauge theories. After establishing the credibility of this decomposition law, it is shown how it leads to the diagonalizability of the longitudinal sector of the functional integral. This result is related to the problem of the existence of the zero-mass limit in massive SU (2) Yang–Mills theories.

PACS numbers: 11.10.Ef, 11.10.Np, 11.30.Jw

I. INTRODUCTION

Consider a massive and a massless (mass inserted by hand) Yang–Mills field theories. They are described by the respective Langrangian densities

$$\mathscr{L}_{\text{massless}} = -\tfrac{1}{4}G_{\mu\nu}{}^{a}G^{a\mu\nu} \tag{1.1}$$

and

$$\mathscr{L}_{\text{massive}} = -\tfrac{1}{4}\widetilde{G}_{\mu\nu}{}^{a}\widetilde{G}^{a\mu\nu} - \tfrac{1}{2}m^2\widetilde{B}_{\mu}{}^{a}\widetilde{B}^{a\mu}, \tag{1.2}$$

where the symbol $\sim$ serves to distinguish the massive from the massless vector fields. The expression for $G_{\mu\nu}{}^{a}$ is well known and need not be explicitly displayed.

The question whether the quantum version of theory (1.2) yields theory (1.1) in the zero-mass limit has been examined 10 years ago by van Dam and Veltman[1] within the framework of perturbation theory. These authors had discovered that there exists a discrete gap between the two theories which cannot be bridged in any smooth way. In other words, it was determined that the zero-mass limit of theory (1.2) does not exist.

The relevant argument by van Dam and Veltman has its origin in one of the most fundamental realizations which pervades the whole of relativistic quantum field theory and which is the cornerstone of its particle interpretation. This is Wigner's characterization[2] of an elementary particle as a unitary irreducible representation of the (identity containing, connected component) Poincaré group. As well known, it follows from Wigner's prescription that an elementary particle with rest mass different from zero and spin s, has $2s+1$ independent polarization directions (they are often referred to as degrees of freedom for the field). On the other hand, zero-mass particles have only two polarization directions, irregardless of spin.

Theories (1.1) and (1.2) contain Yang–Mills vector fields the quanta of which describe spin-1 particles. In the massless case one can choose as polarization vectors two unit vectors pointing in the x- and y-direction respectively. According to this convention, the momentum of the particle is aligned with the z-direction. What is often needed in calculations is the sum over polarizations. Within the adopted scheme one determines

$$\sum_{i=1}^{2} e_{\mu}{}^{i}e_{\nu}{}^{i} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}. \tag{1.3}$$

The above sum can be expressed in terms of the 4-momentum of the (massless) vector particle, $p_\mu = (0,0,p,ip)$—where the usual imaginary rotation to Euclidean 4-space has been performed— and the 4-vector $\bar{p}_\mu = (0,0,-p,ip)$. The latter is obtained from p_μ via space reflection. Following van Dam and Veltman one sets

$$\sum_{i=1}^{2} e^{i}_{\mu}e^{i}_{\nu} = \delta_{\mu\nu} - \frac{\bar{p}_\mu p_\nu + p_\mu \bar{p}_\nu}{(p,\bar{p})}, \tag{1.4}$$

where $(p,\bar{p})$ denotes scalar product.

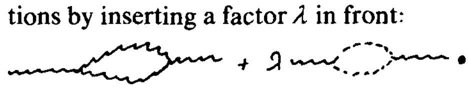
In the massive case, the three polarization unit vectors can be chosen to point along the x- and y- and z-direction respectively. One now determines

$$\sum_{i=1}^{3} e'_{\mu}{}^{i}e'_{\nu}{}^{i} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}. \tag{1.5}$$

The above sum can be conveniently expressed via the use of the 4-momentum vector in the rest frame of the particle, $p_\mu = (0,0,0,\text{im})$. One finds

$$\sum_{i=1}^{3} e'_{\mu}{}^{i}e'_{\nu}{}^{i} = \delta_{\mu\nu} + p_\mu p_\nu/m^2. \tag{1.6}$$

As is well known, sums over polarizations enter prominently into the unitarity diagrams of a field theory (sum over intermediate states). This fact was utilized by van Dam and Veltman in order to compare theories (1.1) and (1.2). In particular, they considered the vector field propagator at the one loop level. They parametrized the ghost loop contributions by inserting a factor λ in front:



(1.7)

(a) (b)

[a] Permanent address: University of Athens, Physics Dept., Div. of Mechanics, Athens, Greece.

 0022-2488/81/092026-08$01.00

Unitarity requires that the imaginary part of the sum of the above two diagrams equals the contribution of the diagram identical to (a) wherein the virtual particles are replaced by real ones (sum over polarizations times $\delta(q^2 + m^2)$ in the massive, $\delta(q^2)$ in the massless, case per internal vector meson line). The computations are straightforward and are given in Ref. 1. One finds that unitarity is satisfied for $\lambda = -1$ in the massive and $\lambda = -2$ in the massless case respectively. This means that if one were to start with theory (1.2), one would have to insert the factor $\lambda = -2$ in front of diagram (b) in order to ensure unitarity. But, then, in the limit $m \to 0$, the factor -2 would remain, resulting in a theory that does not satisfy unitarity. This is tantamount to saying that the limit theory is physically nonexistent.

Now, the above conclusion has been drawn strictly on the basis of perturbation theory arguments. On the other hand, more recent realizations have brought to surface the fact that a nonabelian gauge theory, of the type described by (1.1), posseses nontrivial classical solutions, instantons,[3] whose implications go beyond those of perturbation theory.[4,5] These solutions constitute the basis of nonperturbative effects, present in gauge theories. (Since instantons are best understood when the gauge group is SU (2), I shall assume this group from now on). It, then, becomes of interest to reconsider the van Dam–Veltman conclusion in the presence of instantons. This constitutes the underlying problem of the present investigation.

The direct study of such a problem would have been very straightforward to conduct, provided one knew the precise form of the full propagator of theory (1.1), i.e., the form of the propagator which incorporates instanton effects. [Note that theory (1.2) does not possess instanton solutions.] Then, one could go back and repeat the calculation by van Dam and Veltman. Unfortunately such a propagator is not exactly known at this time. This occurrence rules out the direct confrontation of the posed problem. At the same time, one should note that a first attempt to determine the modifications to the (massless) Yang–Mills propagator, brought about by the presence of instantons, *has* been made by Polyakov.[4] Even though he was unable to reach a final conclusion, Polyakov's viewpoint as well as the hints obtained by his work are especially relevant to the present problem.

Simply, Polyakov's objective had been to determine whether the presence of instantons in a Yang–Mills theory restores a seemingly broken symmetry. Any such lifting of vacuum degeneracy, Polyakov argues, implies that the theory ceases to possess bonafide massless vector fields. What happens, instead, is that *massive scalar fields* enter into the picture, modifying the propagator, *at the one-loop level*, according to

$$k^{-2}\left(\delta_{\mu\nu} - \frac{k_\mu k_\nu}{k^2}\right)$$
$$\to k^{-2}\left(\delta_{\mu\nu} - \frac{k_\mu k_\nu}{k^2} + \frac{k_\mu k_\nu}{k^2}\frac{M^2}{M^2 + k^2}\right)$$
$$= k^{-2}\left(\delta_{\mu\nu} - \frac{k_\mu k_\nu}{k^2 + M^2}\right), \tag{1.8}$$

where M is the mass of the scalar particles. These scalar particles become responsible for the restoration of the broken symmetry.

The particular form of the result, namely Eq. (1.8), was actually established by Polyakov for the (2 +1)-dimensional model of Georgi and Glashow, under certain approximations of semiclassical character. For the nonabelian Yang–Mills theory (1.1), with its instantons, only a *hint* was obtained regarding the presence of massive scalar particles.[4]

Let me, for a moment, take the above hint seriously and turn the argument around. Going back to theory (1.1), as well as expression (1.4), I ask what happens to the sum over polarizations if an additional massive scalar field (rest mass M) is actually present in the theory. Clearly, such an additional scalar field introduces an extra degree of freedom. Denoting the 4-momentum vector in the rest system of the additional scalar particle by $p'_\mu = (0,0,0,iM)$ and imitating (1.6), one arrives at the following sum over polarizations

$$\sum_{i=1}^{3} \hat e^i_\mu \hat e^i_\nu = \delta_{\mu\nu} - \frac{\bar p_\mu p_\nu + p_\mu \bar p_\nu}{(p,\bar p)} + \frac{p'_\mu p'_\nu}{M^2}$$
$$= \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}. \tag{1.9}$$

Notice the crucial difference between (1.9) and (1.5): The third polarization direction is timelike. It reminds one of the spin-zero daughter of a massive vector field before the imposition of the Proca condition.

Such a state of affairs serves notice to the fact that instanton contributions correspond to effects brought about by *pseudoparticles* in the theory. One suspects that pseudoparticle contributions do not amount to the emergence of an equivalent, fully physical vector particle—which would have been the case had one obtained the matrix in (1.5). In summary, it seems, from this turnabout argument, that the pseudonature of instantons lurks in the background as manifested by a timelike third polarization direction.

So one is now disappointed. The hints emerging from Polyakov's work do not seem to imply that the factor λ in (1.7) is -1 for the massless case as well. On the other hand, the possible presence of the additional scalar field at the one loop level, suggests that (1.7) be replaced, in the massless case, by

(1.10)

where the solid line stands for the additional massive scalar propagator. The question now obviously poses itself whether λ can be set to -1 with the additional contribution to unitarity coming from the third diagram. In such a case the whole van Dam–Veltman argument is upset since the third diagram appears only in the massless case.

The above line of reasoning is clearly speculative. Nevertheless, it strongly suggests that the zero-mass limit of massive Yang–Mills theories does, indeed, need reexamination. Obviously, the whole investigation has to start from scratch, adopting a more general methodological approach than the one employed in this section.

The approach to be adopted in this paper coincides with the one adopted by Boulware[6] in his attempt to solve the renormalization problem of massive Yang–Mills field theories. It is based on the path integral quantization of these theories. One specific aspect of Boulware's analysis, that will become the focal point of this investigation, is the diagonalization and decoupling of the longitudinal sector of the functional integral.

By reverting to Boulware's program the original objective will be somewhat lowered. Thus, the immediate problem to be dealt with in this paper will be the diagonalizability of the longitudinal sector of the Yang–Mills functional integral. The specification of this problem, as well as its implications on the zero-mass limit will be given in Sec. 2, where Boulware's work will be essentially reviewed. In Sec. 3 I perform my main task. I divide my work into three parts. First, I argue for a more general decomposition law of a vector field into transverse and longitudinal parts so as to take into account the nontrivial field topology. Second, I establish the credibility of the suggested decomposition law as well as make some interpretations from it. Lastly, I show how the new decomposition law can lead, in principle, to the diagonalizability of the longitudinal sector of the functional integral. In Sec. 4 I take a converse point of view. I show that if one a priori assumes the existence of the zero-mass limit in a nonabelian theory, then there *must* be present in the theory scalar fields whose vacuum expectation value does not vanish. Unless put in by hand, such scalar fields conceivably describe the effects of pseudoparticles. This is the closest I come, in this study, to addressing the zero-mass limit of Yang–Mills fields. A number of concluding remarks is made in the final section.

Two notes of interest: a) All functional integrals will be formulated in Euclidean space where they make sense. b) Given any Lie group G its corresponding Lie algebra will be symbolized by $\mathbf{g}$, i.e., the small letter(s) will be in bold face.

2. ON THE LONGITUDINAL SECTOR OF MASSIVE YANG–MILLS THEORIES

In this section I exhibit the diagonalizability problem of the longitudinal sector in massive SU (2) Yang–Mills theories by employing their functional integral formulation. No account is taken of the topological subtelties exhibited by SU (2) gauge theories. It will be shown that the diagonalizability (and, hence, longitudinal decoupling) problem becomes a question of a point transformation in the integration variables (field variables). It will be argued that the mathematical impass which arises and which implies the nondiagonalizability is inherent in the locality aspects of the gauge group and not in the particular method employed for the study of the longitudinal decoupling problem.

Let $B^a_\mu(x)$, $a = 1,2,3$, be a triplet of massive SU (2) Yang–Mills fields. I denote the corresponding algebra element by $\mathscr{B}_\mu(x)$. In other words, $\mathscr{B}_\mu(x) = B^a_\mu(x)t^a$, where the $\{t^a\}$ form a generator basis for the algebra **su (2)**. The decomposition of $\mathscr{B}_\mu(x)$ into transverse and longitudinal parts, ignoring topological complications, has been specified by Boulware.[6] It is given by the gauge transformation

$$\mathscr{B}_\mu(x) = S^{-1}(x)\mathscr{A}_\mu(x)S(x) + S^{-1}(x)\partial_\mu S(x)\,, \qquad (2.1)$$

where S(x) is an element of the gauge group satisfying appropriate conditions, such that the vector fields $\mathscr{A}_\mu(x) = A^a_\mu(x)t^a$ correspond to the transverse modes of the $B^a_\mu(x)$. These conditions have been spelled out by Boulware who has also shown the existence of the $A^a_\mu(x)$. The longitudinal modes $\xi(x) = \xi^a(x)t^a$ are contained in the gauge group element S(x) via the exponential mapping: S(x) $= e^{\xi(x)}$.

It should be emphasized from the outset that (2.1) will eventually become the focal point of the present effort. In particular, it will be asked whether (2.1) *does* take into account the full vacuum structure of SU (2) gauge theories and, if not, whether it should be replaced by a more appropriate decomposition law.

Given, now, (2.1) the question of interest becomes: Do the longitudinal components decouple in some limit? To formulate this problem let me consider the generating functional of the massive Yang–Mills system, given by

$$Z[\mathscr{S}] = Z^{-1}_{\mathscr{B}}\int[\mathscr{D}\mathscr{B}]\exp\Big[\int(\mathscr{L} + \mathscr{S}^\mu\mathscr{B}_\mu)d^4x\Big]\,, \qquad (2.2)$$

where $\mathscr{L}$ is the massive Yang–Mills Lagrangian and $\mathscr{S}_\mu = J^a_\mu t^a$ is an external source appropriate for the generation of the various Green functions.

Inserting transformation (2.1) into the generating functional (2.2) one accomplishes two things. First, the integration variables are split into transverse and longitudinal ones. Second, the very nature of (2.1) leads to the introduction of a Jacobian determinant identical to that of Faddeev and Popov.[7] This follows from the fact that decomposition (2.1) is nothing but a gauge transformation. For more details the reader can consult Boulware's paper.[6]

Accordingly, the transverse sector of the theory looks just like the corresponding gauge invariant theory, modulo a mass term in the propagator, i.e. the latter is of the form $\delta_{ab}(g_{\mu\nu} - p_\mu p_\nu/p^2)(p^2 - m^2)^{-1}$. In short, the transverse integration becomes

$$Z^{-1}_{\mathscr{A}}\int[\mathscr{D}\mathscr{A}][\mathscr{D}\vartheta\mathscr{D}\vartheta^*]\exp\{\int[\mathscr{L}(\mathscr{A}) + C(\mathscr{A}) + \mathscr{L}_{\text{ghost}}(\vartheta,\vartheta^*,\mathscr{A}) + \mathscr{S}^\mu\mathscr{A}_\mu]d^4x\}\,, \qquad (2.3)$$

where $\mathscr{L}(\mathscr{A}) = -\frac{1}{4}G^a_{\mu\nu}G^{a\mu\nu} - m^2\mathscr{A}^2_\mu$ but with the $\mathscr{A}_\mu$ a triplet of *transverse* fields. $C(\mathscr{A})$ is a gauge fixing term, inserted here in anticipation of its necessity in the zero-mass limit. It can be viewed as part of the integration measure. One now observes that in the limit $m \to 0$, (2.3) becomes identical to the generating functional of a proper, i.e., gauge invariant, Yang–Mills theory.

The longitudinal sector receives all of its contribution from the mass term $m^2\mathscr{B}^2_\mu$ in the original Langrangian. It has the form

$$G_S[\mathscr{S}'] = Z_S^{-1}\int [\mathscr{D}S]$$
$$\times\exp\left\{-\frac{m^2}{g^2}\int [\tfrac{1}{2}(S^{-1}\partial_\mu S)^2 + S^{-1}(\partial_\mu S)\mathscr{S}'^\mu]d^4x\right\}, \quad (2.4)$$
where
$$\mathscr{S}'_\mu = \mathscr{A}_\mu - (1/m^2)\mathscr{S}_\mu .$$
In terms of the actual scalar fields ξ^a (2.4) takes the following form
$$G_\xi[\mathscr{S}'] = Z_\xi^{-1}\int [\mathscr{D}\xi](\det G)^{1/2}$$
$$\times\exp\left\{-\frac{m^2}{g^2}\int [\tfrac{1}{2}\partial_\mu\xi_a G_{ab}(\xi)\partial^\mu\xi_b\right.$$
$$\left. + \partial_\mu\xi_a Q_{ab}(\xi)J'^\mu_b]d^4x\right\}, \quad (2.5)$$
where $(\det G)^{1/2}$ stands for the Jacobian determinant of the transformation and where the $G_{ab}(\xi)$, $Q_{ab}(\xi)$ are nonpolynomial expressions in the ξ_a's which will shortly take a considerable amount of our attention.

The problem in hand can now be stated clearly: If one were to succeed in showing that (2.5) takes the form
$$\widetilde{G}_\xi[\mathscr{S}'] = \widetilde{Z}_\xi^{-1}\int [\mathscr{D}\xi]$$
$$\times\exp\left\{-\frac{m^2}{g^2}\int [\tfrac{1}{2}(\partial_\mu\xi_a)^2 + \partial_\mu\xi_a J'^\mu_a]d^4x\right\}, \quad (2.6)$$
then the longitudinal sector would decouple. That is, the scalar modes will be integrated away. The resulting exponential can be easily accommodated in (2.3) via a redefinition of gauge. One now sees how the diagonalizability of the longitudinal sector is related to the existence of the zero-mass limit.

If follows that the next step should be to investigate the diagonalizability of matrices $Q_{ab}(\xi)$, $G_{ab}(\xi)$ through a point transformation $\xi\rightarrow\xi'(\xi)$. As $G = QQ^T$ one needs only to discuss the diagonalizability of, say, G. Now, G is by construction a symmetric 3×3 matrix built up, at each space-time point, from a *single* 3-vector ξ. It becomes instructive to actually generalize the argument, for a moment, to an arbitrary n-dimensional group. Assuming maximal rank (which holds true in this particular case) G contains, by virtue of its symmetry, $\frac{1}{2}n(n+1)$ independent elements constructed from the n-components of ξ. Consequently, its diagonalizability implies that $\frac{1}{2}n(n+1)$ conditions must be imposed on the aforementioned n-components. But $\frac{1}{2}n(n+1) > n$ for $n > 1$ and, hence, one has an overdetermined problem in hand, i.e. nonsolvable in general. And, indeed, a more exhaustive investigation of this problem confirms its nonsolvability.[6] Note, in passing, that for $n = 1$ one *does* get $\frac{1}{2}n(n+1) = n$ and thus recovers the well known result[8,9] for the $U(1)$ case.

It will now be shown that the above impass and, in particular, the presence of matrices G and Q is a direct consequence of the local character of the gauge group. To this end, consider, a triplet of real numbers $\xi^a(\in\mathbb{R})$, $a = 1,2,3$, which specify the components of an element in the linear, real vector space **su(2)**. As is well known the exponential mapping, which takes one from the Lie algebra to that connected component of the corresponding Lie group containing the identity element, provides the group manifold with a natural (normal) coordinate set. Therefore, the ξ^a provide a coordinate set for the (simply connected) group SU(2). Upon localizing the group with respect to space-time we are led to consider algebra elements of the form $\xi(x) = \xi^a(x)t^a$. The elevation of the ξ^a to function status, assigning group coordinates $\xi^a(x)$ at the space-time point x, naturally leads one to consider the functions $\partial_\mu\xi^a(x)$ as well. (I hold μ fixed so that I don't have to worry about Lorentzian aspects on top of everything else). Clearly, $\partial_\mu\xi(x) = \partial_\mu\xi^a(x)t^a$ *is* an element of **su(2)**. I shall now show that, unlike the $\xi^a(x)$, the $\partial_\mu\xi^a(x)$ cannot be used to parametrize the group manifold at each space-time point x.[10]

Let me consider the differential of the exponential mapping at the point ξ, to be denoted by $d\exp_\xi$. It maps **su(2)** onto itself since the differential of a given mapping among two manifolds acts between the respective tangent spaces and since **su(2)** is flat, i.e., it coincides with its own tangent space. Now the following holds true for the mapping $d\exp_X$ for any given Lie algebra $\mathbf{p}$[11]:
$$d\exp_X Y = [(1-e^{-adX})/adX]Y, \quad X,Y\in\mathbf{p},$$
where
$$\frac{1-e^{-adX}}{adX} = \sum_{n=0}^{\infty}\frac{(-adX)^n}{(n+1)!}$$
so that
$$d\exp_X Y = Y - \frac{1}{2!}[X,Y] + \frac{1}{3!}[X,[X,Y]]$$
$$-\frac{1}{4!}[X,[X,[X,Y]]] + \cdots .$$
In the above formulas X denotes the particular point of the Lie algebra $\mathbf{p}$ at which one considers the tangent space and Y is the point of $\mathbf{p}$ on which the differential of the exponential mapping acts.

I now examine the **su(2)** element, to be called the exponential derivative of ξ, defined as follows
$$\mathscr{D}_\mu\xi \equiv d\exp_\xi\partial_\mu\xi . \quad (2.7)$$
This is a nonpolynomial expression in the ξ_a's. Explicitly,
$$\mathscr{D}_\mu\xi^a = \partial_\mu\xi^a - \tfrac{1}{2}\epsilon_{abc}\xi^b\partial_\mu\xi^c + \frac{1}{3!}\epsilon_{abc}\epsilon_{cgh}\xi^b\xi^g\partial_\mu\xi^h$$
$$- \cdots .$$
The above expression when written compactly identifies the matrix Q_{ab} appearing in (2.5):
$$\mathscr{D}_\mu\xi^a = \partial_\mu\xi^b Q_{ba}(\xi) . \quad (2.8)$$
Similarly, $G_{ab}(\xi)$ in (2.5) is recovered from the inner product
$$\mathrm{Tr}\mathscr{D}_\mu\xi\mathscr{D}^\mu\xi = \partial_\mu\xi^a G_{ab}(\xi)\partial^\mu\xi^b . \quad (2.9)$$
It follows that, independently from decomposition (2.1) *and the ensuing analysis, matrices Q and G owe their existence to the local character of the symmetry group.*

It now becomes obvious that the nondiagonalizability of matrix $G_{ab}(\xi)$ through a point transformation $\xi\rightarrow\xi'(\xi)$ is equivalent to saying that given the **su(2)** element $\partial_\mu\xi$ there always exists another **su(2)** element, namely the exponential derivative $\mathscr{D}_\mu \xi$, which cannot be put in the form $\partial_\mu \xi'$. This also proves the inefficiency of the $\partial_\mu \xi^a, \mu$ fixed, to parametrize the gauge group manifold. In conclusion, Boulware's nondiagonalizability result can be seen as equivalent to the deficiency of the $\partial_\mu \xi^a$ to parametrize the gauge group manifold.

3. TOPOLOGICAL CONSIDERATIONS AND THE SOLUTION OF THE DIAGONALIZATION PROBLEM

A. Preliminary discussion; a generalization of decomposition law (2.1)

The analysis of the previous section has completely ignored the topological aspects of SU(2) gauge field theories.[3] It is the present intention to amend this situation. Let me start by giving a plausibility argument as to why it appears reasonable that the proper inclusion of SU(2) gauge field topological effects should be of importance for the problem in hand. Recall that the topological properties discovered in Ref. 3 arise in the Euclidean formulation of local SU(2) and stem from the homotopical classification of SU(2) on the 3-sphere in E_4. In juxtaposition, it was witnessed in Sec. 2 (cf. matrices Q and G owe their existence to the *local* character of the symmetry group) how all problems entering the diagonalization of the longitudinal sector can be traced to geometrical aspects of the gauge group. It, therefore, seems reasonable to expect that topological considerations, which lie at the root of the intimacy between the symmetry group and space-time, should play an important role in the final resolution of the diagonalization problem.

In this connection, it is worth noting the subtle point that Boulware overlooked in his analysis of massive Yang–Mills models and on which the present hope for a revision hinges. Boulware argued that the range of integration over the angular variables ξ^a in (2.5) could be carried from $-\infty$ to $+\infty$ instead of mod 2π. His thinking had been that, in this way, one covers the group a denumerably infinite number of times, the result being $\mathscr{N}_0$ times the integration over the compact region ($\mathscr{N}_0$ the usual infinite cardinal number). But, he argued, the same factor $\mathscr{N}_0$ also appears in the normalization factor Z_ξ^{-1} and, consequently, a cancellation results. However, he overlooked the fact that S_3 is the one-point compactification of E_3. This one extra point matters greatly as it becomes responsible for the existence of inequivalent homotopical mappings which, in turn, have an impact on the SU(2) gauge theory itself.

Another way of putting the same thing is the following. The normal coordinates are, in effect, *local* (in the sense of differential geometry) coordinates for the group. As one moves away from the origin of the algebra, viewed as a vector space, one covers more and more of the group. However, the compactness of SU(2) implies that it cannot be globally covered with the normal coordinates. A second coordinate patch is needed for the point on the manifold diametrically opposite the unit element. Now, the inclusion of this point is essential to establishing the topological properties of the gauge group. It, then, becomes obvious that the functional integral (2.5) will fail to include topological effects.

It may now be objected that there is no a priori reason to know that the massive Yang–Mills model under consideration possesses pseudoparticle solutions and, consequently, it may make no sense to study topological effects in such a model. This, however, is not the point. The real issue concerns decomposition law (2.1) and its longitudinal part $S^{-1}\partial_\mu S$ in particular. One *does* know that these pure gauge terms fall into topologically inequivalent classes characterized by a winding number. And, in fact, it is the longitudinal behavior of SU(2) Yang–Mills fields at infinity that separates them into inequivalent classes. The present objective, then, is not to include pseudoparticle contributions to the original massive Yang–Mills theory, in fact instanton solutions do not exist in this case, but to take into account the plethora of longitudinal vacua which are known to be available to an SU(2) Yang–Mills theory.

In different terms, one knows a *priori* that instanton effects are very real within a massless SU(2) gauge theory (cf. results of Refs. 4, and 5). It is disturbing, then, to operate with a decomposition law which does not keep track of these effects. It becomes obvious that what is needed is the formulation of an appropriate decomposition law, to replace (2.1), which takes into account the availability of the various inequivalent vacua to an SU(2) gauge theory.

In the search for such a law one must take into account the fact that if $\mathscr{G}_{\mu\nu} = 0$ ($\mathscr{G}_{\mu\nu} = G^a_{\mu\nu} t^a$), then $\mathscr{B}_\mu$ must be uniquely of the form $S^{-1}\partial_\mu S$, $S(x)$ an element of the gauge group, while its transverse component must be identically zero. As becomes apparent from Boulware's analysis of the nonabelian decomposition problem, this is tantamount to saying that the transverse component of $\mathscr{B}_\mu$ is unique.[12] It follows that any generalization of (2.1) should: a) Identify only one transverse field $\mathscr{A}_\mu$ as representing the corresponding modes of $\mathscr{B}_\mu$ and b) when $\mathscr{A}_\mu$ vanishes, $\mathscr{B}_\mu$ should be of the form $S^{-1}\partial_\mu S$.

Given the above guidelines I propose to consider the following generalization of (2.1)

$$\mathscr{B}_\mu(x) = \Big(\prod_j S_j\Big)^{-1} \mathscr{A}_\mu \Big(\prod_j S_j\Big) + \Big(\prod_j S_j\Big)^{-1} \partial_\mu \Big(\prod_j S_j\Big), \tag{3.1}$$

where the index j denotes topological class of the gauge group element in the sense of Jackiw and Rebbi.[13]

Obviously, what has been done through suggesting (3.1) is to resolve an SU(2) gauge group element into its topologically (better, Π_3-homotopically) inequivalent components. Formally, of course, if one sets $S = \prod_j S_j$ one recovers (2.1). However, a radically new ingredient now enters the analysis in view of the fact that $S_j, j \neq 0$, *cannot be built, within the gauge group, from elements of the type* S_0. On the other hand, the particular construction of Jackiw and Rebbi[13] shows that

$$S_j = (S_1)^j, \quad j \neq 0. \tag{3.2}$$

Therefore, it suffices to specialize (3.1) to $j = 0,1$. From now on I shall work with the decomposition law

$$\mathscr{B}_\mu(x) = (S_1 S_0)^{-1} \mathscr{A}_\mu(x)(S_1 S_0) + (S_1 S_0)^{-1} \partial_\mu (S_1 S_0), \tag{3.3}$$

which constitutes a minimal nontrivial extension of (2.1).

The crucial thing to note from (3.3) is that *six* independent scalar field variables enter the decomposition of $\mathscr{B}_\mu$. In particular, the original triplet $\{\xi_a\}$, which enters through $S_0 = e^\xi$, is supplemented by a second, independent, triplet $\{\eta_a\}$ that enters through $S_1 = e^\eta$. I stress the independence of the variables η_a from the variables ξ_a. In fact, if this were not so, i.e. if $S_1 = e^{f(\xi)}$, it would have followed that a winding-number-one element can be continuously built from zero-winding-number gauge group elements (since $e^{f(\xi)}$ involves continuous multiplications among gauge group elements of the type S_0). Finally, by invoking (3.2) one is convinced that no additional independent scalar variables enter through decomposition (3.1), thereby justifying one's working with the simplified version (3.3).

B. Interpretation and credibility of the new decomposition law

Since (3.3) is, in form, the same as (2.1) one expects that it properly describes a decomposition of $\mathscr{B}_\mu$ into transverse and longitudinal components. However, there are two points that call for clarification. To begin with, an interpretation is needed for the six independent scalar fields instead of the usual three. Secondly, because of the unorthodox character of the decomposition law one is compelled to explicitly show that it does indeed yield a transverse component $\mathscr{A}_\mu$ for every massive field $\mathscr{B}_\mu$.

To interpret the six scalar fields I go back to the discussion in the Introduction. What emerged from that discussion is that the scalar modes which represent instanton effects in the massless Yang–Mills theory are of a different nature than the longitudinal-scalar modes of the massive Yang-Mills fields. I argue that the ξ's and η's of decomposition law (3.3) are certain superpositions of physical, i.e., longitudinal, scalar modes and of modes that correspond to the scalar fields that eventually make their appearance in the massless theory via instanton contributions.

The above is a strictly intuitive interpretation. I shall elaborate more on the question of the six scalar field by employing formal arguments. To this end, consider the scalar (pure gauge) term of decomposition law (3.3). It reads

$$(S_1S_0)^{-1}\partial_\mu(S_1S_0) = S_0^{-1}\partial_\mu S_0 + S_0^{-1}(S_1^{-1}\partial_\mu S_1)S_0 . \quad (3.4)$$

Using the terminology of Sec. 2 I write $S_1^{-1}\partial_\mu S_1 = \mathscr{D}_\mu\eta$ (recall the denotations $S_1 = e^\eta$ and $S_0 = e^\xi$). Now, the second term in (3.4) is of the form $e^{-\xi}Ae^\xi$, A an element of the (local) algebra **su(2)**. I shall now show that

$$e^{-\xi}Ae^\xi = \mathrm{ad}(e^{-\xi})A , \quad (3.5)$$

where $\mathrm{ad}(e^{-\xi})$ denotes the adjoint action of the Lie group element $e^{-\xi}$ on the corresponding Lie algebra element.

By definition, it holds that

$$\mathrm{ad}(S_0)A = A', \quad e^{A'} = e^{-\xi}e^Ae^\xi . \quad (3.6)$$

Once the operator identity $Be^AB^{-1} = e^{BAB^{-1}}$ is recalled, (3.5) follows immediately.

This result yields

$$e^{-\xi}\mathscr{D}_\mu\eta e^\xi = A'_\mu , \quad (3.7)$$

where

$$e^{A'_\mu} = e^{-\xi}e^{\mathscr{D}_\mu\eta}e^\xi . \quad (3.8)$$

But[11]

$$e^{-\xi}e^{\mathscr{D}_\mu\eta}e^\xi = e^{\mathscr{D}_\mu\eta + (1/2)[\xi,\mathscr{D}_\mu\eta] + \cdots} . \quad (3.9)$$

Working in the neighborhood of the identity element, where (3.9) makes most sense, one finds, to a first approximation,

$$A'_\mu \simeq \mathscr{D}_\mu\eta = S_1^{-1}\partial_\mu S_1 , \quad (3.10)$$

whence (3.4) becomes

$$(S_1S_0)^{-1}\partial_\mu(S_1S_0) \simeq S_0^{-1}\partial_\mu S_0 + S_1^{-1}\partial_\mu S_1 . \quad (3.11)$$

In the Appendix it is also shown that

$$(S_{j+1}S_j)^{-1}\partial_\mu(S_{j+1}S_j) = S_j^{-1}\partial_\mu S_j + S_{j+1}^{-1}\partial_\mu S_{j+1} , \quad (3.12)$$

for $j \neq 0$. It follows that the longitudinal term in decomposition law (3.1) is of the form $\Sigma_j S_j^{-1}\partial_\mu S_j$, i.e. it is identical with the ϑ-vacuum superposition in nonabelian gauge theories, with $\vartheta = 0$[14]. Accordingly, the longitudinal part of proposed decomposition law (3.1) reflects, in a natural way, the possibility that *the longitudinal modes of the original massive vector fields do not necessarily "settle" into the winding-number-zero sector but into all available and topologically inequivalent longitudinal vacua.*

I am now left to show that (3.3) does indeed yield a triplet of transverse vector fields $A^a_\mu(x)$ for every massive vector field triplet $B^a_\mu(x)$. For this all I have to do is follow the strategy employed by Boulware in Appendix A of his paper. The problem is to find a gauge group element $S(x)$ such that the decomposition law yields a transverse field $\mathscr{A}_\mu(x)$. Actually, since S in (3.3) is of the form $S = S_0S_1$ I am presently looking for *two* gauge group elements S_0 and S_1. For the moment, however, let me keep S and proceed a certain distance before having to introduce its two components. Boulware's strategy is the following: Assume a solution has been found. Also assume that an arbitrary but small variation over this solution yields a new one. In this way one obtains a differential equation for the sought for group element(s).

Now, as shown by Boulware, the transversality condition leads to the equation

$$\partial^\mu(S^{-1}\mathscr{B}_\mu S) = -\partial^\mu(S^{-1}\partial_\mu S) . \quad (3.13)$$

Assuming a solution $(\mathscr{A}_\mu, S)$ has been found, one considers the variation $\mathscr{B}_\mu \to \mathscr{B}_\mu + \delta\mathscr{B}_\mu$. Owing to its exponential character, the corresponding variation in S is of the form $\delta S = S\delta\chi$. Upon substituting in (3.4) one obtains the following elliptic equation for $\delta\chi$ (I am constantly working in Euclidean space):

$$\partial^2\delta\chi(x) + \partial_\mu[\mathscr{A}^\mu,\delta\chi] = \partial_\mu(S^{-1}\delta\mathscr{B}^\mu S) . \quad (3.14)$$

The above equation represents, for the case of decomposition (2.1) wherein S is of the form $S_0 = e^\xi$, a simple system of elliptic equations of first degree solvable by standard Green function techniques. As shown by Boulware, the solvability of (3.14) establishes the existence of $\mathscr{A}_\mu$. Let me turn to decomposition law (3.3). There are now six variations $\delta\chi^a_1, \delta\chi^a_0$. Explicitly,

$$\delta S = \delta S_1S_0 + S_1\delta S_0 = S_1\delta\chi_1S_0 + S_1S_0\delta\chi_0 , \quad (3.15)$$

where $\delta_\chi = \delta\chi^a t^a$. One must now substitute (3.15) into the variation of (3.13). Actually, the same answer is obtained if one substitutes in (3.14) the formal relation

$$\delta\chi = S_0^{-1}\delta\chi_1 S_0 + \delta\chi_0\,, \tag{3.16}$$

which results after one sets $\delta S = S\delta\chi = S_1 S_0 \delta\chi$ in (3.15). The following systems of elliptic equations are thereby arrived at

$$\partial^2\delta\chi_0 + \partial^\mu[\mathscr{A}_\mu,\delta\chi_0] = 0\,, \tag{3.17a}$$

$$\partial^2[S_0^{-1}\delta\chi_1 S_0] + \partial^\mu[\mathscr{A}_\mu, S_0^{-1}\delta\chi_1 S_0] = \partial^\mu(S_0^{-1}S_1^{-1}\delta\mathscr{B}_\mu S_1 S_0)\,. \tag{3.17}$$

Setting

$$\delta\chi_1' = S_0^{-1}\delta\chi_1 S_0, \tag{3.18}$$

which means

$$e^{\delta\chi_1'} = e^{\xi}e^{\delta\chi_1}e^{-\xi} = e^{\delta\chi_1 + [\xi,\delta\chi_1] + [\xi,[\xi,\delta\chi_1]] + \cdots} \tag{3.19}$$

(and, alternatively, $e^{\delta\chi_1} = e^{-\xi}e^{\delta\chi_1'}e^{\xi}$), then (3.17b) becomes

$$\partial^2\delta\chi_1' + \partial^\mu[\mathscr{A}_\mu,\delta\chi_1'] = \partial^\mu\{(S_1S_0)^{-1}\delta\mathscr{B}_\mu(S_1S_0)\}\,. \tag{3.17b'}$$

The systems of Eqs. (3.17a) and (3.17b') are no more complicated than the system (3.14) obtained by Boulware (in fact, the set of equations specified by (3.17a) is simpler, i.e. homogeneous). This establishes the existence of the $\mathscr{A}_\mu$, given the decomposition (3.3). I shall not attempt in this paper to prove uniqueness for the $\mathscr{A}_\mu$.[15]

C. Diagonalizability of the longitudinal sector

Recall, now, the diagonalizability impass encountered in Sec. 2: One has to fulfill $\frac{1}{2}n(n+1)$ conditions with only n variables in one's disposal, namely the components of ξ. However, it has just been argued that field topology considerations introduce a second [in the case of SU(2) at least] scalar field triplet η which will also partake into the construction of G and Q. But, now, with $2n$ variables in one's disposal one has to fulfill the condition

$$\tfrac{1}{2}n(n+1) = 2n\,. \tag{3.20}$$

Apart from the trivial case $n = 0$, this condition is uniquely fulfilled for $n = 3$, i.e., it exactly suits SU(2).

To witness how the actual diagonalization procedure works let me first consider the following approximation

$$e^{\xi}e^{\eta} \simeq e^{\xi+\eta} \tag{3.21}$$

valid if one works very close to the identity element.

In this case, (3.3) induces, in comparison with (2.1), the replacement

$$\mathscr{D}_\mu\xi \rightarrow \mathscr{D}_\mu(\xi+\eta) = d\exp_{\xi+\eta}(\partial_\mu\xi + \partial_\mu\eta)\,. \tag{3.22}$$

The above generalization implies

$$\mathrm{Tr}\mathscr{D}_\mu\xi\mathscr{D}^\mu\xi \rightarrow \mathrm{Tr}\mathscr{D}_\mu(\xi+\eta)\mathscr{D}^\mu(\xi+\eta) = \partial_\mu(\xi_a+\eta_a)G_{ab}(\xi+\eta)\partial^\mu(\xi_b+\eta_b)\,. \tag{3.23}$$

Consider, now, the redefinition

$$\xi \rightarrow \xi(\xi') \quad \eta \rightarrow \eta(\eta')$$

and set

$$f_b^a(\xi') = \delta\xi_a/\delta\xi_b', \quad g_b^a(\eta') = \delta\eta_a/\delta\eta_b'\,.$$

Then, the diagonalization conditions become

$$\{f_a^c(\xi')f_b^d(\xi') + g_a^c(\eta')g_b^d(\eta') + f_a^c(\xi')g_b^d(\eta') + g_a^c(\eta')f_b^d(\xi')\}G_{cd}(\xi'+\eta') = \delta_{ab}\,. \tag{3.24}$$

The above relation represents six algebraic conditions on the six variables ξ_a', η_a' at each spacetime point. It is, in general, solvable. Consequently, the nonpolynomial Lagrangian expression entering the longitudinal sector of the functional integral now becomes

$$[\partial_\mu(\xi'+\eta')]^2 + \partial^\mu(\xi'+\eta')\mathscr{S}_\mu'\,. \tag{3.25}$$

Reinterpreting $\xi'+\eta'$ as a certain mixture of scalar modes, the desired diagonalization of the longitudinal sector has been finally achieved.

Note that approximation (3.21) is hardly necessary for arriving at the above result. Indeed, in the general case, one has

$$e^{\xi}e^{\eta} = e^{\xi+\eta+(1/2)[\xi,\eta]+(1/3!)[\xi,[\xi,\eta]]+\cdots}\,. \tag{3.26}$$

By setting

$$\zeta = \eta + \tfrac{1}{2}[\xi,\eta] + (1/3!)[\xi,[\xi,\eta]] + \cdots$$

one is in position to repeat the steps following (3.21), with η replaced by ζ.

The final form taken by the longitudinal sector of the functional integral is

$$G_{\xi',\eta'}[\mathscr{S}'] = Z_{\xi',\eta'}^{-1}\int[\mathscr{D}_\xi'][\mathscr{D}_\eta'],$$

$$\times\exp\left\{-\frac{m}{g^2}\int[\tfrac{1}{2}(\partial_\mu(\xi'+\eta')^2 + \partial^\mu(\xi'+\eta')\mathscr{S}_\mu']d^4x\right\}.$$

Note the absence of a Jacobian determinant factor. Its diagonalization is simultaneous with that of $G(\xi+\eta)$, hence it yields unity.

4. ON THE NECESSITY OF NONZERO WINDING NUMBER CONTRIBUTIONS

In the present section I shall focus on zero-mass limit aspects of the diagonalization-decoupling problem and take an opposite point of view. I shall *assume* that the zero-mass limit of a massive SU(2) Yang–Mills theory *does* yield a massless, gauge invariant SU(2) Yang–Mills theory and study the consequences.

Consider the equations of motion corresponding to theories (1.1) and (1.2):

$$\partial^\nu G_{\mu\nu}^a + \epsilon_{abc}B^{c\nu}G_{\mu\nu}^b = 0 \tag{4.1}$$

and

$$\partial^\nu \widetilde{G}_{\mu\nu}^a + m^2\widetilde{B}_\mu^a + \epsilon_{abc}\widetilde{B}^{c\nu}\widetilde{G}_{\mu\nu}^b = 0\,, \tag{4.2}$$

where the coupling constant has been set equal to 1.

Consider, next, the infinitesimal version of decomposition (2.1). It reads

$$\widetilde{B}_\mu^a = A_\mu^a + \epsilon_{abc}\xi_b A_\mu^c + \partial_\mu\xi_a + \mathscr{O}(\xi^2)\,. \tag{4.3}$$

Substituting (4.3) into (4.2) (I denote by $G_{\mu\nu}'^a$ the usual antisymmetric tensor constructed from the transverse fields A_μ^a). I obtain

$$\partial^\nu G_{\mu\nu}'^a - m^2A_\mu^a - m^2\epsilon_{abc}\xi_b A_\mu^c - m^2\partial_\mu\xi_a + \epsilon_{abc}A^{c\nu}G_{\mu\nu}'^b$$

$$+\epsilon_{abc}\epsilon_{cde}\xi_d A^{e\nu} G'^{b}_{\mu\nu} + \epsilon_{abc}\partial^\nu \xi_c G'^{b}_{\mu\nu} = 0 . \quad (4.4)$$

Suppose, now, one *identifies*, in the limit $m \rightarrow 0$, the transverse fields A^a_μ with the massless (transverse) gauge fields B^a_μ. The following identity then results, on account of (4.1),

$$-m^2 A^a_\mu - m^2 \epsilon_{abc}\xi_b A^c_\mu - m^2 \partial_\mu \xi_a + \epsilon_{abc}\epsilon_{cde}\xi_d A^{e\nu} G^b_{\mu\nu} + \epsilon_{abc}\partial^\nu \xi_c G^b_{\mu\nu} \equiv 0 , \quad (4.5)$$

where the prime notation has been dropped in view of the assumed $\lim G'^a_{\mu\nu} \rightarrow G^a_{\mu\nu}$.

Consider now infinitesimal quantum oscillations about the classical equilibrium points for the fields A^a_μ and ξ^a, i.e.,

$$A^a_\mu \rightarrow \langle A^a_\mu \rangle + \delta A^a_\mu, \quad \xi^a \rightarrow \langle \xi^a \rangle + \delta \xi^a . \quad (4.6)$$

As the A^a_μ are transverse fields it follows that $\langle A^a_\mu \rangle = 0$. Substituting (4.6) into identity (4.5) one obtains, to first order in δA^a_μ and $\delta \xi^a$,

$$-m^2 \delta A^a_\mu - m^2 \epsilon_{abc}\langle \xi_b \rangle \delta A^c_\mu - m^2 \langle \partial_\mu \xi_a \rangle - m^2 \partial_\mu \delta \xi_a + \epsilon_{abc}\langle \partial^\nu \xi_c \rangle \delta G^b_{\mu\nu} \equiv 0 . \quad (4.7)$$

In particular, for a translationally invariant vacuum, (4.7) becomes

$$m^2(\delta A^a_\mu + \epsilon_{abc}\langle \xi_b \rangle \delta A^c_\mu + \partial_\mu \delta \xi_a) \equiv 0 . \quad (4.8)$$

Alternatively, without the assumption of a translationally invariant vacuum and by the fact alone that δA^a_μ is an arbitrary variation, it follows, from (4.7),

$$-m^2 \delta A^a_\mu - m^2 \epsilon_{abc}\langle \xi_b \rangle \delta A^c_\mu + \epsilon_{abc}\langle \partial^\mu \xi_c \rangle \delta G^b_{\mu\nu} \equiv 0 . \quad (4.9)$$

From either (4.8) or (4.9) and excluding the trivial case $m^2 = 0$ it becomes obvious that compatibility between theories (1.1) and (1.2) is only possible, in the limit, if the scalar modes ξ_a have a nonvanishing classical value. But this is exactly the case with nonzero-winding number longitudinal "modes". Thus, replacing the ξ_a by $\Sigma_j \xi^j_a$, j the winding number index, and since $\langle \xi^j_a \rangle \neq 0$ for $j \neq 0$, one is now in position to consistently fulfill the identity for $m \neq 0$.

In summary, the work in this section has explicitly displayed the necessity of the inclusion of scalar fields whose expectation value does not vanish at infinity if one insists on achieving compatibility between the massive theory (1.2) and the massless theory (1.1). One way of achieving this compatibility is to insert, by hand, scalar fields with nonvanishing vev. This corresponds to the well known connection between gauge invariant theories and those of the spontaneously broken variety via Higgs fields (see, e.g., Ref. 10). In the absence of Higgs scalars the present analysis shows that either the van Dam–Veltman result persists, or the required scalar modes with nonvanishing vacuum value are furnished by way of instanton effects. The findings of the previous section show that the possibility cannot be ruled out.

5. CONCLUDING REMARKS

Consider the following analogy. The well known axial anomaly which arises in perturbation theory (triangle diagrams) can be analyzed via a suitable parametrization. Specifically, one takes up the dilemma gauge invariance vs conservation of axial current. By employing a certain parameter it is determined that for a certain value one gains gauge invariance while for another, different, value one recovers axial current conservation. Naturally, one chooses gauge invariance. With the advent of instantons, 't Hooft solved this problem[5] bridging the gap precisely via the contributions of these classical solutions. Now, the van Dam–Veltman analysis, presented in the Introduction, shows that a certain gap exists between massive and massless Yang–Mills theories within the framework of perturbation theory. The present work shows that this gap can, in principle, be bridged via instanton considerations.

A complete solution of the zero-mass limit problem could have an important bearing on the infrared behavior of Yang–Mills theories. This remains the least understood, albeit the most interesting, region of these theories.

APPENDIX

I shall presently show that

$$(S_{k+1}S_k)^{-1}\partial_\mu(S_{k+1}S_k) = S^{-1}_{k+1}\partial_\mu S_{k+1} + S_k^{-1}\partial_\mu S_k \quad (A1)$$

for $k \neq 0$.

I start from

$$(S_{k+1}S_k)^{-1}\partial_\mu(S_{k+1}S_k) = S_k^{-1}S^{-1}_{k+1}(\partial_\mu S_{k+1})S_k + S_k^{-1}\partial_\mu S_k . \quad (A2)$$

Using that[13] $S_{k+1} = (S_1)^{k+1}$ one easily obtains

$$S^{-1}_{k+1}\partial_\mu S_{k+1} = (k+1)S_1^{-1}\partial_\mu S_1 . \quad (A3)$$

Inserting (A3) into (A2) and using $S_k^{-1}S_1 = S_{k+1}$ as well as

$$(\partial_\mu S_1)S_1^k = [1/(k+1)]\partial_\mu (S_1)^{k+1} = [1/(k+1)]\partial_\mu S_{k+1},$$

one finally arrives at (A1).

[1] H. van Dam and M. Veltman, Nucl. Phys. B **22**, 397 (1970); see also D. G. Boulware and S. Deser, Phys. Rev. D **6**, 3368 (1972).
[2] E. P. Wigner, Ann. Math. **40**, 149 (1939).
[3] A. Belavin *et al.*, Phys. Lett. B **59**, 85 (1975).
[4] A. M. Polyakov, Nucl. Phys. B **120**, 429 (1977).
[5] G. 't Hooft, Phys. Rev. D **14**, 3432 (1976).
[6] D. G. Boulware, Ann. Phys. (NY) **56**, 140 (1970).
[7] L. P. Faddeev and V. N. Popov, Phys. Lett. B **25**, 29 (1967).
[8] F. Coester, Phys. Rev. **83**, 798 (1951).
[9] D. G. Boulware and W. Gilbert, Phys. Rev. **126**, 1563 (1962).
[10] C. N. Ktorides, Phys. Rev. D **13**, 2811 (1976).
[11] S. Helgason, *Differential Geometry and Symmetric Spaces* (Academic, New York, 1962).

[12] That is, unique within a gauge transformation which remains within the transverse sector. We ignore the ambiguities associated with such transformations (Gribov ambiguities).

[13] R. Jackiw and C. Rebbi, Phys. Rev. Lett. **37**, 172 (1976).

[14] A decomposition prescription corresponding to $\vartheta \neq 0$ is conceivable.
[15] See Ref. 6 for the methodology aimed at proving the uniqueness of $\mathscr{A}_\mu$ (Appendix A).

Toward exact solutions of the nonlinear Heisenberg–Pauli–Weyl spinor equation

Eckehard W. Mielke
International Center for Theoretical Physics, Trieste, Italy

(Received 22 January 1980; accepted for publication 24 June 1980)

In *color geometrodynamics* fundamental spinor fields are assumed to obey a $GL(2f,\mathbb{C}) \otimes GL(2c,\mathbb{C})$ gauge-invariant *nonlinear spinor equation* of the Heisenberg–Pauli–Weyl–type. Quark confinement, assimilating a scheme of Salam and Strathdee, is (partially) mediated by the tensor "gluons" of *strong gravity*. This hypothesis is incorporated into the model by considering the nonlinear Dirac equation in a curved space-time of hadronic dimensions. Disregarding internal degrees of freedom, it is then feasible, for a peculiar background space-time, to obtain *exact* solutions of the spherical bound-state problem. Finally, these solutions are tentatively interpreted as droplet-type *solitons* and remarks on their interrelation with Wheeler's *geon* construction are made.

PACS numbers: 11.10.Lm, 11.10.Np, 12.40.Bb, 04.50. + h

I. INTRODUCTION

Recent speculations[1] on a new geometrodynamical model of *extended* particles draws attention to the possibility of describing composite baryons by a $G \equiv GL(2f,\mathbb{C}) \otimes GL(2c,\mathbb{C})$ gauge-invariant nonlinear spinor equation is curved space-time. In accordance with the *strong gravity* hypothesis[2] the curving-up of the *internal space* is expected to occur in dimensions characterized by the modified *Planck length*

$$\ell^* = (8\pi\hbar G_s/c^3)^{1/2} = (8\pi)^{1/2}\hbar/cM^* \approx 10^{-13}\text{cm} \quad (1.1)$$

of the order of one Fermi.

The Poincaré–invariant gauge theory of gravity with spin and torsion analyzed by Hehl *et al.*[3] may be generalized[4] to one incorporating the flavor and color generating group $U(f) \otimes U(c)$. Then a *nonlinear spinor equation of the Heisenberg–Pauli–Weyl-type*[5,6]

$$\{iL^\mu\nabla_\mu - \tfrac{3}{8}\ell^{*2}\bar{\psi}L^5L_\mu\psi L^5L^\mu - \mu c/\hbar\}\psi = 0 \quad (1.2)$$

emerges which is G-gauge invariant. Unlike Heisenberg's unified field theory[6,7] in *color geometrodynamics* (CGMD)[8] the $f \times c$ fundamental spinor fields

$$\psi \equiv \{\psi^{(q_f q_c)} | q_f = 1,\dots,f; \quad q_c = 1,\dots,c\} \quad (1.3)$$

may be interpreted as quark fields distinguished by f flavor and c color degrees of internal freedom. In this gauge-theory[4] the matrices L^μ, L^5 are space-time dependent generalizations of the familiar Dirac matrices γ^μ tensored with $U(f) \otimes U(c)$ vector operators λ_j (generalized Gell-Mann matrices).

Whereas the nonlinearities induced into (1.2) by Cartan's *torsion tensor*[9] are suspected to yield (classical) bound-states of quarks, their (partial) confinement is conjectured[1,10] to result from curvature barriers produced by tensor gluon fields $f_{\mu\nu}$. This "role of (strong) gravitation in the building-up of elementary particles" has already been envisioned by Weyl[5] who also gave the prior construction of a generally covariant and $SL(2,\mathbb{C})$ gauge-invariant spinor equation.

The fundamental Heisenberg–Pauli–Weyl spinor equation has already been the subject of considerable work, the internal $U(f) \otimes U(c)$ symmetry usually being dropped for simplicity. Then, in two space-time dimensions the renormalizable theory described by the *massless* universal field equation[6] of Heisenberg is commonly referred to as the Thirring model.[11,12] Regarding ψ as interacting quantum fields, Glaser[13] has obtained explicit solutions (see also Ref. 14). More recently, classical instanton- and meron-type solutions of (1.2) have also been found[15] in two dimensions.

In the real world it is notoriously difficult to obtain solutions even for the *semiclassical problem*. Finkelstein *et al.*[16,17] Soler,[18] and more recently Rañada[19] as well as Takahashi[20] have obtained *radially localized* solutions of a nonlinear Dirac equation similar to (1.2). However, their analysis rests upon numerical calculations and in addition to that, is restricted to *flat* space-time. On the other hand, it is known[21,22] that a nonlinear scalar field theory having a dynamics which is related[23,24] to the squared form of (1.2) admits *exactly* solvable radial solutions in an Einstein universe.

In this paper, the freedom in the choice of the background space-time again is instrumental for the construction of *exact* radial solutions of a nonlinear spinor equation having a [compared to (1.2)] algebraically simplified self-interaction.

The paper proceeds as follows: In Sec. II the setup of a generally covariant nonlinear spinor theory in curved space-time is reviewed. Emphasis is laid on the case of a spherically symmetric background.

Section III deals with the spherical spinor formalism familiar from the bound-state problem of the hydrogen atom.[25,26] This formalism allows one to reduce the spinor equation to a system of two nonlinear differential equations which are of first order with respect to the radial coordinate r. Employing a stationary *Ansatz*, exact spherical spinor waves are explicitly constructed in Sec. IV. A peculiar class of spherically symmetric space-times is specifically adjointed to allow for solutions in closed form. Although unlikely, the possible physical significance of these exact solutions and connections with Wheeler's geon concept[27] are discussed in Sec. V.

 0022-2488/81/092034-06$01.00

II. THE MODEL

Disregarding from now on internal degrees of freedom, the self-interacting spinor model of CGMD is given[4] by the Lagrangian density

$$\mathscr{L}_{\text{D-W}} = \sqrt{|f|}\left\{i\bar{\psi}L^{\mu}\nabla_{\mu}\psi - \frac{3\epsilon}{16}\ell^{2}(\bar{\psi}L^{5}L^{\mu}\psi)^{2} - (\mu c/\hbar)\bar{\psi}\psi\right\}, \quad \epsilon = \pm 1 \tag{2.1}$$

defined on a curved space-time with pseudo-Riemannian metric and connection. [The basic field equation (1.2) may be derived by varying the Dirac–Weyl Lagrangian (2.1) for $\delta\mathscr{L}_{\text{D.-W.}}/\delta\bar{\psi}$ if $\epsilon = +1$ and $\ell = \ell^{*}$.]

In terms of the familiar Dirac matrices γ_{α} (conventions are those of Bjorken and Drell[25]) satisfying

$$\gamma_{\alpha}\gamma_{\beta} + \gamma_{\beta}\gamma_{\alpha} = 2\eta_{\alpha\beta}, \quad \text{diag}(\eta_{\alpha\beta}) = (1, -1, -1, -1), \tag{2.2}$$

space-time dependent matrices

$$L_{\mu} \equiv L^{\ \alpha}_{\mu}\gamma_{\alpha}, \tag{2.3}$$

$$L^{5} \equiv \frac{i}{4!}\frac{1}{\sqrt{|f|}}\epsilon^{\alpha_1\cdots\alpha_4}L_{\alpha_1}\wedge\cdots\wedge L_{\alpha_4}, \quad L^{5}L^{5} = 1 \tag{2.4}$$

are introduced in terms of the *vierbein* field $L^{\ \alpha}_{\mu}$. These matrices are via

$$f_{\mu\nu} = \tfrac{1}{4}\text{Tr}(L_{\mu}L_{\nu}) \tag{2.5}$$

related to the metric tensor $f_{\mu\nu}$ of the curved manifold (of hadronic dimensions).

The Dirac operator has been generalized to a curved space-time by using the SL(2,$\mathbb{C}$) gauge-covariant derivative[28,29]

$$\nabla_{\mu} \equiv \partial_{\mu} + i\Gamma_{\mu}, \tag{2.6}$$

Following Brill and Wheeler[30] the spinor connection Γ_{μ} can be explicitly expressed in terms of the symmetric Christoffel symbols

$$\Gamma^{\ \tau}_{\mu\nu} \equiv \tfrac{1}{2}f^{\tau\sigma}(\partial_{\nu}f_{\sigma\mu} + \partial_{\mu}f_{\sigma\nu} - \partial_{\sigma}f_{\mu\nu}) \tag{2.7}$$

(metrical connection coefficients) as follows:

$$\Gamma_{\mu} = \tfrac{1}{4}(\Gamma^{\ \tau}_{\mu\nu}L^{\nu\alpha}L^{\ \beta}_{\tau} - L^{\ \alpha}_{\nu}\partial_{\mu}L^{\nu\beta})\sigma_{\alpha\beta}. \tag{2.8}$$

Here

$$\sigma_{\alpha\beta} \equiv \tfrac{1}{2}i[\gamma_{\alpha}, \gamma_{\beta}] \tag{2.9}$$

denote the infinitesimal generators of the covering group SL(2,$\mathbb{C}$) of the Lorentz group.

According to well-known relations[17] between scalar products of bilinear forms containing *identical* spinors the self-interaction in (2.1), which is of the axial-vector type, may be replaced by a scalar-minus-a-pseudoscalar type self-coupling, i.e.,

$$(\bar{\psi}iL^{5}L_{\mu}\psi)(\bar{\psi}iL^{5}L^{\mu}\psi) = (\bar{\psi}L_{\mu}\psi)(\bar{\psi}L^{\mu}\psi) = (\bar{\psi}\psi)^{2} - (\bar{\psi}L^{5}\psi)^{2}. \tag{2.10}$$

Therefore we are justified to a certain extent in considering instead of (1.2) the nonlinear Dirac equation

$$\{iL^{\mu}\nabla_{\mu} + \tfrac{3}{8}\epsilon\ell^{2}\bar{\psi}\psi - \mu c/\hbar\}\psi = 0 \tag{2.11}$$

having an algebraically simplified self-interaction. (The inclusion of the pseudoscalar term would lead to a more intricate model.)

Besides zero, Eq. (2.11) admits for $\epsilon = +1$ the constant solution

$$\psi_c = \begin{cases} \psi_c^{(q_f, q_i)} = 0, & q = 1,\dots,f-1; \\ \psi_c^{(f, q_i)} = (2/\ell)(2\mu c/3\hbar)^{1/2}e^{i\delta} \end{cases} \tag{2.12}$$

Since *radially localized* "bound-state" configurations may (classically) describe *extended* particles it is consistent to base our search for solutions of (2.11) on a spherically symmetric background space-time. In the "isotropic" presentation the corresponding general line element reads

$$ds^{2} \equiv f_{\mu\nu}dx^{\mu}dx^{\nu} = e^{\nu}c^{2}dt^{2} - e^{\lambda}(dr^{2} + r^{2}d\theta^{2} + r^{2}\sin^{2}\theta d\varphi^{2}), \tag{2.13}$$

where the functions $\nu = \nu(\rho)$ and $\lambda = \lambda(\rho)$ depend solely on the dimensionless radial coordinate

$$\rho \equiv [(2\pi)^{1/2}/\ell^{*}]r = (M^{*}c/2\hbar)r. \tag{2.14}$$

The spherical coordinates $r \equiv |\mathbf{x}|$, θ, and φ may be expressed in terms of the usual Cartesian coordinates $\mathbf{x}$.

In order to facilitate the following analysis it will be assumed that the background metric $f_{\mu\nu}$ is conformally related to another metric via

$$f_{\mu\nu} = e^{\lambda}\bar{f}_{\mu\nu}. \tag{2.15}$$

Obviously, the corresponding *vierbein* then can be conformally related in a similar way:

$$L^{\ \alpha}_{\mu} = e^{\lambda/2}\bar{L}^{\ \alpha}_{\mu}. \tag{2.16}$$

With respect to the conformal change (2.15) the (symmetric!) Christoffel symbols are related via

$$\Gamma^{\ \tau}_{\mu\nu} = \bar{\Gamma}^{\ \tau}_{\mu\nu} + \tfrac{1}{2}(\delta^{\tau}_{\mu}\partial_{\nu}\lambda + \delta^{\tau}_{\nu}\partial_{\mu}\lambda - f_{\mu\nu}\partial^{\tau}\lambda). \tag{2.17}$$

[See, e.g., Ref. 31, Appendix (A2)]. Employing also (2.16) a short calculation reveals that

$$\Gamma_{\mu} = \bar{\Gamma}_{\mu} + \tfrac{1}{8}(\delta^{\beta}_{\mu}\partial^{\alpha}\lambda - \delta^{\alpha}_{\mu}\partial^{\beta}\lambda)\sigma_{\alpha\beta}. \tag{2.18}$$

Thus, as a byproduct, it has been shown that the nonlinear Dirac equation (2.11) on conformally related pseudo–Riemannian manifolds takes the form

$$\{i\bar{L}^{\mu}(\partial_{\mu} + i\bar{\Gamma}_{\mu} + \tfrac{3}{4}\partial_{\mu}\lambda) + \tfrac{3}{8}\epsilon\ell^{2}e^{\lambda/2}\bar{\psi}\psi - e^{\lambda/2}\mu c/\hbar\}\psi = 0. \tag{2.19}$$

In order to proceed further, the spinor connection $\bar{\Gamma}_{\mu}$ in terms of a metric $\bar{f}_{\mu\nu}$ which is via (2.15) conformally related to the background (2.13) has to be ascertained. The comparison with the results (30) of Ref. 30 obtained for a similar case shows that

$$\begin{aligned} &\bar{\Gamma}_{0} = \tfrac{1}{4}ie^{(\nu-\lambda)/2}\partial_{r}(\nu - \lambda)\gamma_{0}\gamma_{1}, \\ &\bar{\Gamma}_{1} = 0, \quad \bar{\Gamma}_{2} = \tfrac{1}{2}i\gamma_{2}\gamma_{1}, \\ &\bar{\Gamma}_{3} = \tfrac{1}{2}i(\sin\theta\gamma_{3}\gamma_{1} + \cos\theta\gamma_{3}\gamma_{1}). \end{aligned} \tag{2.20}$$

The insertion of these expressions into (2.19) finally leads to

$$\{i\gamma^{0}\partial_{0} - e^{(\nu-\lambda)/2}[i\boldsymbol{\gamma}\cdot\boldsymbol{\partial} + i\gamma^{r}\partial_{r}(\tfrac{1}{2}\lambda + \tfrac{1}{4}\nu)] + \tfrac{3}{8}\epsilon\ell^{2}e^{\nu/2}\bar{\psi}\psi - e^{\nu/2}\mu c/\hbar\}\psi = 0. \tag{2.21}$$

For later convenience, the spatially flat Dirac operator $i\boldsymbol{\gamma}\cdot\boldsymbol{\partial}$ which occurs has been expressed in terms of Cartesian coordinates.

III. SEPARATION *ANSATZ* FOR SPHERICAL WAVES

It may be noticed that the curved background occurs in the conformal spinor equation (2.21) only in a multiplicative manner, except for the third term. However, this term may be absorbed by the factor $\exp(-\lambda/2 - \nu/4)$ in the following *Ansatz*:

$$\psi = \frac{\psi}{\ell}\left(\frac{2\pi\mu c}{3\hbar}\right)^{1/2} e^{-\lambda/2-\nu/4} e^{-i\omega t\mu c^2/\hbar} \times \begin{bmatrix} iG(\rho)\chi^m_{\mathscr{K}} \\ F(\rho)\dfrac{\boldsymbol{\sigma}\cdot\mathbf{x}}{|\mathbf{x}|}\chi^m_{\mathscr{K}} \end{bmatrix} \tag{3.1}$$

Following essentially the notation of Rose,[26] the spin-weighted spherical harmonics $\chi^m_{\mathscr{K}}$ of parity $P = (-1)^l$ are defined by

$$\chi^m_{\mathscr{K}} \equiv \sum_{\bar{m}=\pm 1/2} C(l\tfrac{1}{2}j;m-\bar{m},\bar{m}) Y_l^{m-\bar{m}}(\theta,\varphi)\chi^{\bar{m}}. \tag{3.2}$$

Here

$$\mathscr{K} \equiv \mp (j+\tfrac{1}{2}) \quad \text{for } j = l \pm \tfrac{1}{2}, \tag{3.3}$$

and $C(l\frac{1}{2}j;m-\bar{m},\bar{m})$ are Clebsch–Gordan coefficients. These spherical 2-spinors are known to satisfy the eigenvalue equations

$$\mathbf{J}^2\chi^m_{\mathscr{K}} = j(j+1)\chi^m_{\mathscr{K}}, \tag{3.4}$$

$$\boldsymbol{\sigma}\cdot\mathbf{L}\chi^m_{\mathscr{K}} = -(\mathscr{K}+1)\chi^m_{\mathscr{K}}, \tag{3.5}$$

for the operators $\mathbf{J}$ and $\mathbf{L}$ of total and orbital angular momentum and also

$$(\boldsymbol{\sigma}\cdot\mathbf{x}/|\mathbf{x}|)\chi^m_{\mathscr{K}} = -\chi^m_{-\mathscr{K}}. \tag{3.6}$$

Moreover, it can be shown[26] that

$$i\gamma^0\boldsymbol{\gamma}\cdot\boldsymbol{\partial}\begin{bmatrix} iG(\rho) & \chi^m_{\mathscr{K}} \\ -F(\rho) & \chi^m_{-\mathscr{K}} \end{bmatrix} = \frac{\boldsymbol{\sigma}\cdot\mathbf{x}}{|\mathbf{x}|}\left(\partial_r - \frac{1}{r}\boldsymbol{\sigma}\cdot\mathbf{L}\right)\begin{bmatrix} iF(\rho) & \chi^m_{-\mathscr{K}} \\ G(\rho) & \chi^m_{\mathscr{K}} \end{bmatrix}. \tag{3.7}$$

with respect to the *Ansatz* (3.1) the self-coupling term in (2.21) takes the form[17]

$$\tfrac{3}{8}\ell^2\bar{\psi}\psi = 4\pi\frac{\mu c}{\hbar}e^{-\lambda-\nu/2} \times (G^2-F^2)\left|Y^{|\mathscr{K}|-\frac{1}{2}}_{|\mathscr{K}|-1}(\theta,\varphi)\right|^2. \tag{3.8}$$

This self-interaction potential has to be spherically symmetric in order to ensure separability. This is the case for $|\mathscr{K}| = 1$, only. Then, the quantum numbers for the spin and angular momentum of admissible solutions are restricted to $j = 1/2$, $l = 0$, 1, and $m = \pm 1/2$.

With all this information at hand the insertion of (3.1) into (2.21) yields

$$\partial_{\rho^*}G + \frac{1+\mathscr{K}}{\rho}e^{(\nu-\lambda)/2}G = \frac{1}{\beta}[\omega + e^{\nu/2} - \epsilon e^{-\lambda}(G^2-F^2)]F, \tag{3.9}$$

$$\partial_{\rho^*}F + \frac{1-\mathscr{K}}{\rho}e^{(\nu-\lambda)/2}F = \frac{1}{\beta}[-\omega + e^{\nu/2} - \epsilon e^{-\lambda}(G^2-F^2)]G \tag{3.10}$$

for the remaining *radial equations*. It is convenient, but not really necessary in our case, to employ Wheeler's "tortoise coordinate"[27] ρ^* defined by the differential form

$$d\rho^* = e^{(\lambda-\nu)/2}d\rho \tag{3.11}$$

in the above equations. In (3.9) and (3.10) the ratio

$$\beta = M^*/2\mu \tag{3.12}$$

measures how much the "bare mass" μ of a fermion contributes to the (strong) gravitational mass M^*. The nonlinear term $\sim(G^2-F^2)$ suggest we utilize the *Ansätze*

$$G(\rho) = e^{\vartheta(\rho)}\cosh\phi(\rho), \tag{3.13}$$

$$F(\rho) = e^{\vartheta(\rho)}\sinh\phi(\rho), \tag{3.14}$$

in order to achieve a simplification of the equations. (Similar Ansätze have already been used by Finkelstein *et al.*[17] and van der Merve[32] in their search for asymptotic solutions at spatial infinity.) Solving for the first derivatives yields:

$$\partial_\rho\phi = \frac{\mathscr{K}}{\rho}\sinh 2\phi + \frac{1}{\beta}(e^{\nu/2} - \epsilon e^{2\vartheta-\lambda} - \omega\cosh 2\phi)e^{(\lambda-\nu)/2}, \tag{3.15}$$

and

$$\partial_\rho\vartheta = \frac{\omega}{\beta}e^{(\lambda-\nu)/2}\sinh 2\phi - \frac{1}{\rho}(1+\mathscr{K}\cosh 2\phi). \tag{3.16}$$

IV. POSSIBLE EXACT SOLUTION

As the resulting radial equations (3.15) and (3.16) form an underdetermined system, the freedom in the choice of the background space-time may be used to absorb the nonlinear contribution $e^{2\vartheta}$ in Eq. (3.15). In particular, the assumption

$$e^{\nu/2} - \epsilon e^{2\vartheta-\lambda} = \omega\cosh 2\phi \tag{4.1}$$

will cause equation (3.15) and (3.16) to separate. Then, (3.15) reduces to

$$d\phi/\sinh 2\phi = \mathscr{K}d\rho/\rho. \tag{4.2}$$

The left-hand side can be integrated by employing formula 2.423.1 of Ref. 33:

$$\int dz/\sinh z = \ln\tanh(z/2). \tag{4.3}$$

After solving (4.2) this leads to the initial information that

$$\cosh\phi = (1 - \Lambda^2\rho^{4\mathscr{K}})^{-1/2}, \tag{4.4}$$

$$\sinh\phi = \Lambda\rho^{2\mathscr{K}}\cosh\phi, \tag{4.5}$$

where Λ denotes an integration constant.

To proceed further, the differential equation

$$\frac{d\vartheta}{d\phi} = \frac{\omega}{\beta\mathscr{K}}\rho e^{(\lambda-\nu)/2} - \frac{1}{\mathscr{K}\sinh 2\phi} - \coth 2\phi \tag{4.6}$$

expressing the implicit dependence of $\vartheta(\rho)$ on $\phi(\rho)$, in view of condition (4.1) can now be deduced from (3.16) and (3.15).

Under the further assumption that

$$e^{(\nu-\lambda)/2} = \rho\tanh 2\phi \equiv \frac{2\rho\sinh\phi\cosh\phi}{\sinh^2\phi + \cosh^2\phi} \tag{4.7}$$

holds Eq. (4.6) can be easily integrated with the aid of (4.3) and

$$\int(\cosh z/\sinh z)dz = \ln \sinh z \tag{4.8}$$

(formula 2.423.33 of Ref. 33). The result can be expressed in terms of the already known functions (4.4) and (4.5) as

$$e^{\vartheta} = \frac{C}{2\Lambda}\left(\frac{1}{\Lambda}\sinh\phi\right)^{(\omega-\beta)/2\beta\mathscr{S}-1/2}(\cosh\phi)^{(\omega+\beta)/2\beta\mathscr{S}-1/2}, \tag{4.9}$$

where C is a second integration constant.

The remaining task is to determine the metric functions e^{ν} and e^{λ}. The substitution of (4.7) into the other *subsidiary condition* (4.1) leads merely to an algebraic equation

$$y^3 + 3py + 2q = 0 \tag{4.10}$$

for

$$y \equiv e^{\lambda/2} - \sqrt{|p|}, \tag{4.11}$$

With respect to the parameters

$$p = -\left[\frac{\omega(\sinh^2\phi + \cosh^2\phi)^2}{6\rho\sinh\phi\cosh\phi}\right]^2 \leqslant 0, \tag{4.12}$$

and

$$q = -|p|^{3/2} - \frac{\epsilon e^{2\vartheta}}{2\rho\tanh 2\phi} \tag{4.13}$$

it takes the normal form (4.10) of a *cubic* equation. If $p<0$ (4.10) has one or three *real* roots depending whether or not the discriminant

$$D \equiv q^2 + p^3 = \left(\frac{e^{\vartheta}}{\rho\tanh 2\phi}\right)^2\left[\tfrac{1}{4}e^{2\vartheta} + \tfrac{1}{3}\epsilon|p|\omega\cosh 2\phi\right] \tag{4.14}$$

is positive or not (see Ref. 34, Sec. 59). By introducing the complex variable

$$\begin{aligned}\alpha &\equiv \tfrac{1}{3}\operatorname{arcosh}(|q|\,|p|^{-3/2})\\ &= \tfrac{1}{3}\operatorname{arcosh}\left(\left|1 - 2\epsilon\frac{3^3e^{2\vartheta}\rho^2\sinh^2\phi\cosh^2\phi}{\omega^3(\sinh^2\phi+\cosh^2\phi)^5}\right|\right)\\ &= \tfrac{1}{3}\operatorname{arcosh}\left(\left|1 - \frac{\epsilon 3^3C^2}{2\omega^3}\rho^{2\mathscr{S}+2\omega/\beta}\right.\right.\\ &\qquad\left.\left.\times(1+\Lambda^2\rho^{4\mathscr{S}})^{-5}(1-\Lambda^2\rho^{4\mathscr{S}})^{(4\beta\mathscr{S}-\omega)/\beta\mathscr{S}}\right|\right)\end{aligned} \tag{4.15}$$

with respect to the principal value of the inverse hyperbolic function, it can be inferred from the trigonometric indentity

$$4\cosh^3\alpha = 3\cosh\alpha + \cosh 3\alpha \tag{4.16}$$

(Ref. 33, formula 1.324.2) that

$$y_1 = -2\,\mathrm{sgn}(q)\sqrt{|p|}\begin{cases}\cos\alpha & \text{for } D\leqslant 0, \text{ or}\\ \cosh\alpha & \text{for } D>0\end{cases} \tag{4.17}$$

is a solution of (4.10) for $p<0$. Independent of the sign of D this solution is always real and therefore acceptable for the components of a (pseudo-) Riemannian metric.

Using this root in (4.11) and employing also (4.12) and (4.13) yields (written down below for $D>0$ only)

$$\begin{aligned}e^{\lambda/2} &= \frac{\omega(\sinh^2\phi+\cosh^2\phi)^2}{6\rho\sinh\phi\cosh\phi}\left[1 - 2\,\mathrm{sgn}(q)\cosh\alpha\right]\\ &= \frac{\omega}{6\Lambda}\frac{(1+\Lambda^2\rho^{4\mathscr{S}})^2}{\rho^{2\mathscr{S}+1}(1-\Lambda^2\rho^{4\mathscr{S}})}\\ &\quad\times\left[1 - 2\,\mathrm{sgn}(q)\cosh\alpha\right]\end{aligned} \tag{4.18}$$

as a result for one of the metric functions, whereas $e^{\nu/2}$ may now be derived from (4.7).

Since $F = A\rho^{2\mathscr{S}}G$ it is sufficient to record the *radial dependence* of the *upper* spinor component of the Ansatz (3.1). After employing (4.4), (4.5), (4.9), (3.13), (4.18), and (4.7) its explicit expression reads:

$$\begin{aligned}\tilde G &\equiv Ge^{-\lambda/2-\nu/2} = C\left(\frac{3}{\omega}\right)^{3/2}\rho^{\mathscr{S}+\omega/\beta}\\ &\times(1+\Lambda^2\rho^{4\mathscr{S}})^{-5/2}(1-\Lambda^2\rho^{4\mathscr{S}})^{(3\beta\mathscr{S}-\omega)/2\beta\mathscr{S}}\\ &\times\left\{1 - 2\,\mathrm{sgn}(q)\cosh\left[\tfrac{1}{3}\operatorname{arcosh}\left(\left|1 - \frac{\epsilon 3^3C^2}{2\omega^3}\rho^{2\mathscr{S}+2\omega/\beta}\right.\right.\right.\right.\\ &\left.\left.\left.\left.\times(1+\Lambda^2\rho^{4\mathscr{S}})^{-5}(1-\Lambda^2\rho^{4\mathscr{S}})^{(4\beta\mathscr{S}-\omega)/\beta\mathscr{S}}\right|\right)\right]\right\}^{-3/2}.\end{aligned} \tag{4.19}$$

A *time-independent* solution can be obtained from the *Ansatz* (3.1) by putting $\omega = 0$. This corresponds to the case $p = 0$ which, for a cubic equation, has to be treated separately.

However, then (4.10) admits the obvious solution

$$\begin{aligned}e^{\lambda/2} &= \left(\frac{\epsilon e^{2\vartheta}}{\rho\tanh 2\phi}\right)^{1/3}\\ &= \frac{C^{2/3}}{2\Lambda\rho^{1+4\mathscr{S}/3}}\left[\frac{1+\Lambda^2\rho^{4\mathscr{S}}}{\epsilon(1-\Lambda^2\rho^{4\mathscr{S}})}\right]^{1/3}.\end{aligned} \tag{4.20}$$

In the deduction of the second part of this relation use has been made of the fact that the expression (4.9) is still valid for $\omega = 0$. The radial dependence of the upper spinor component turns out to be comparatively simple:

$$\tilde G = [\epsilon(1-\Lambda^2\rho^{4\mathscr{S}})]^{-1/2}. \tag{4.21}$$

For $\rho < \rho_0 \equiv \Lambda^{-1/2\mathscr{S}}$ the metric function e^{λ} is real as required, whereas the spinor solutions are real for $\mathscr{S} = \epsilon = \pm 1$, only. In the domain $\rho > \rho_0$ both spinor components $\tilde G$ and $\tilde F$ become imaginary. This, in effect, amounts to a change of the sign of the self-interaction of our model (2.1). Therefore, only the interior (or exterior) solution (their domains being separated by a pole of (4.21) and (4.20) at ρ_0) would commonly be acceptable for $\epsilon = +1$ (or $\epsilon = -1$).

V. INTERPRETATION AND PROSPECTS

Some insights into the mathematical structure of these solutions may be gained by studying their asymptotic behavior. From the explicit expressions (4.15) it can be deduced that the term

$$|q|\,|p|^{-3/2}-1\sim\begin{cases}\rho^{2\mathscr{L}+2\omega/\beta} & \text{for } \rho\to 0,\\ \rho^{-(2\mathscr{L}+2\omega/\beta)} & \text{for } \rho\to\infty\end{cases}\tag{5.1}$$

tends to zero with the same degree at the origin and at infinity. Therefore, the radial part (4.19) of the stationary solution behaves as

$$\tilde{G}\sim\begin{cases}\rho^{\mathscr{L}+\omega/\beta}[1-2\operatorname{sgn}(q)]^{-3/2} & \text{for } \rho\to 0,\\ \rho^{-(3\mathscr{L}+\omega/\beta)}[1-2\operatorname{sgn}(q)]^{-3/2} & \text{for } \rho\to\infty.\end{cases}\tag{5.2}$$

For $q<0$ these asymptotic expressions are real. For the lower spinor component $\tilde{F}\equiv\Lambda\rho^{2\mathscr{L}}\tilde{G}$ to be asymptotically vanishing, it has to be required that $\omega/\beta > -\mathscr{L}$. Furthermore, only in the special case[35]

$$\omega = 3\beta\mathscr{L}\tag{5.3}$$

(4.19) and $\tilde{F}$ would constitute a real, *completely regular* and *localized* solution, if the underlying space-time manifold *were* flat. After all, it follows from (4.7) that the line element (2.13) is via

$$ds^2 = e^{\lambda} ds_0^2\tag{5.4}$$

conformally[31] related to the nonsingular line element

$$ds_0^2 = \frac{4\Lambda^2\rho^{4\mathscr{L}+2}}{(1+\Lambda^2\rho^{4\mathscr{L}})^2}c^2dt^2 - \frac{\ell^{*2}}{2\pi}(d\rho^2+\rho^2d\theta^2+\rho^2\sin^2\theta d\varphi^2)\tag{5.5}$$

which is degenerate at the origin (for every $\mathscr{L}$).

However, the conformal function e^{λ} given by (4.18) in the stationary case and by (4.20) for a time-independent solution is *singular* at the origin, at $\rho=\rho_0$, as well as at infinity. This fact could spoil altogether the regularity of the solution (4.19) in the case (5.3) if everything is expressed in a different coordinate system. For these (and other) reasons the *classical* field energy[30] associated with these solutions may turn out not to be finite.

In the time-independent case, the solution (4.21) and a similar expression for the lower spinor component are valid for a finite radius $\rho<\rho_0$ for $\epsilon=+1$. However, the obtained background geometries given by (5.4) and (4.20) likewise are of finite extent and, therefore, may remotely resemble the de Sitter spaces,[36] which have been recently considered for confining models.[10]

Nevertheless, the physical meaning of the obtained exact solutions so far remains rather obscure. The direction of future work may be indicated by the following related, but speculative remarks: For a massive nonlinear spinor theory with a peculiar self-interaction of polynomial degree k, $0<k<1$, Werle[37] found exact radial solutions in flat space-time which are also confined to the interior of a sphere. By continuing with $\psi\equiv 0$ outside this sphere, these solutions are claimed to have many features of droplets of "*bags*".[38] In the case of the massive Thirring model exact solutions owning the properties of bound states have been constucted by Chang *et al.*[39] and studied as an example of *chiral confinement*.

Furthermore, the construction scheme for our exact solution should also be commented upon. Recall, that the method of Sec. IV would not work in flat space-time. The freedom in the choice of a spherically symmetric background is essential for the method presented in this paper. However, in a *self-consistent* approach the metric functions $\nu(\rho)$ and $\lambda(\rho)$ would have to be determined by the stress-energy content (see Sec. 10 of the Brill and Wheeler paper[30]) of the spinor solution via Einstein's field equations. This could be regarded as a *geon-type*[27] construction of spinor solitons in CGMD.

Moreover, the internal symmetries inherent in the G-gauge-invariant equation (1.2) should be properly dealt with. A possible approach in this direction has been undertaken by Takahashi[40] who obtained numerically nontopological soliton solutions with vanishing (total) "color" charge.

ACKNOWLEDGMENTS

This work was supported by a grant of the Deutsche Forschungsgemeinschaft, Bonn. In connection with the renewal of this grant, I would like to express my sincere gratitude to Professors E. Bagge, F. W. Hehl, Abdus Salam, V. Weidemann, and J. A. Wheeler for kindly providing letters of recommendation. Thanks are also due to Professor Abdus Salam, the International Atomic Energy Agency, and UNESCO for hospitality at the International Center for Theoretical Physics, Trieste.

[1]E. W. Mielke, Phys. Rev. Lett. **39**, 530 (1977); **39**, 851 (E) (1977).

[2]C. J. Isham, A. Salam, and J. Strathdee, Phys. Rev. D **3**, 867 (1971); **9**, 1702 (1974).

[3]F. W. Hehl, P. von der Heyde, G. D. Kerlick, and J. M. Nester, Rev. Mod. Phys. **48**, 393 (1976).

[4]E. W. Mielke, "Gauge-Theoretical Foundation of Color Geometrodynamics" (to be published in the Proceedings of the 1978 Clausthal meeting, edited by H. D. Doebner).

[5]H. Weyl, Proc. Natl. Acad. Sci. USA **15**, 323 (1929); Phys. Rev. **77**, 699 (1950).

[6]W. Heisenberg, *Introduction to the Unified Field Theory of Elementary Particles* (Wiley, London, 1966).

[7]W. Heisenberg, Naturwissenschaften **61**, 1 (1974).

[8]E. W. Mielke, "The Eightfold Way to Color Geometrodynamics," Int. J. Theor. Phys. **19**, 189 (1980).

[9]F. W. Hehl, and B. K. Datta, J. Math. Phys. **12**, 1334 (1971).

[10]A. Salam and J. Strathdee, Phys. Rev. D **18**, 4596 (1978).

[11]W. Thirring, Ann. Phys. (Germany) **9**, 91 (1958).

[12]W. Thirring, Nuovo Cimento **9**, 1007 (1958).

[13]V. Glaser, Nuovo Cimento **9**, 990 (1958).

[14]H. Yamamoto, Prog. Theor. Phys. **58**, 1014 (1977).

[15]K. G. Akdeniz and A. Smailagić, Nuovo Cimento A **51**, 345 (1979).

[16]R. Finkelstein, R. LeLevier, and M. Ruderman, Phys. Rev. **83**, 326 (1951).

[17]R. Finkelstein, C. Fronsdal, and P. Kaus, Phys. Rev. **103**, 1571 (1956).

[18]M. Soler, Phys. Rev. D **1**, 2766 (1970).

[19]A. F. Rañada, J. Phys. A **11**, 341 (1978).

[20]K. Takahashi, J. Math. Phys. **20**, 1232 (1979).

[21]J. A. Okolowski, J. Math. Phys. **19**, 2253 (1978).

[22]E. W. Mielke, J. Math. Phys. **21**, 543 (1980).

[23]L. Vazquez, J. Phys. A **10**, 1361 (1977).

[24]W. Deppert and E. W. Mielke, Phys. Rev. D **20**, 1303 (1979).

[25]J. D. Bjorken and S. D. Drell, *Relativistic Quantum Mechanics* (McGraw-Hill, San Francisco, 1964).

[26]E. M. Rose, *Relativistic Electron Theory* (Wiley, New York, 1961).

[27]J. A. Wheeler, *Geometrodynamics* (Academic, New York, 1962).

[28]E. Schrödinger, Sitzungsber. Preuss. Akad. Wiss. Phys.-Math. Kl. XI, 105 (1932).

[29]V. Bargmann, Sitzungsber, Preuss. Akad. Wiss. Phys.-Math. Kl. XI, 346 (1932).

[30]D. R. Brill and J. A. Wheeler, Rev. Mod. Phys. **29**, 465 (1957); (E) **33**, 623 (1961).

[31]E. W. Mielke, Gen. Relativ. Gravit. **8**, 321 (1977).
[32]P. du T. van der Merve, Nuovo Cimento A **41**, 621 (1977).
[33]I. S. Gradshteyn and I. M. Ryzhik, *Table of Integrals, Series, and Products*, 4th ed. (Academic, New York, 1965).
[34]B. L. van der Waerden, *Algebra* I., Grundlehren der mathematischen Wissenschaften Bd. 33 (Springer–Verlag, Berlin, 1955).
[35]A deviation of the effective mass $\omega\mu$ appearing in the phase of the *Ansatz* (3.1) from the "bare mass" μ of the fundamental spinors may be attributed to a possible *Archimedes effect* [10] of the hadronic environment. In CGMD the latter may originate from nonlinear self-interactions and the gluonic curvature barriers of strong gravity.
[36]E. W. Mielke, Fortschr. Phys. **25**, 401 (1977).
[37]J. Werle, Phys. Lett. B **71**, 357 (1977); Lett. Math. Phys. **2**, 109 (1977).
[38]P. Hasenfratz and J. Kuti, Phys. Rep. C **40**, 75 (1978).
[39]S.-J. Chang, S. D. Ellis, and B. W. Lee, Phys. Rev. D **11**, 3572 (1975).
[40]K. Takahashi, Prog. Theor. Phys. **61**, 1251 (1979).

Algebraically special Yang–Mills solutions without sources

R. O. Fulp and Paul Sommers
Department of Mathematics, North Carolina State University, Raleigh, North Carolina 27650
L. K. Norris
Departments of Mathematics and Physics, North Carolina State University, Raleigh, North Carolina 27650

(Received 13 January 1981; accepted for publication 6 March 1981)

We consider source-free Yang–Mills solutions for which the curvature is *decomposable* in the sense that the curvature 2-form is the product of a single Lie-algebra-valued function and a real 2-form. If the curvature is everywhere nonnull (or null with twisting rays), then the solution is a connection in a principal fiber bundle, which is reducible to a source-free Maxwell principal bundle. All such solutions are therefore readily obtained, locally or globally, from Maxwell solutions. Our analysis uses the Ambrose–Singer theorem to show that the holonomy group is one-dimensional. A principal bundle-with-connection is reducible to the holonomy subbundle of any point, and, in this case, since the holonomy group is one-dimensional, the reduced bundle has the structure of a Maxwell bundle. On the other hand, if the curvature is null and twist-free on a full neighborhood of some point, then the bundle need not be reducible. The holonomy group is generally the entire gauge Lie group. The solutions can still be constructed locally from Maxwell solutions, there being extra freedom in the construction over regions where the Maxwell field is algebraically null with twist-free rays. We extend all these results to the class of solutions for which the *self-dual* curvature is (complex) decomposable. These are all the solutions of type D and type N in the classification of Anandan and Tod.

PACS numbers: 11.10.Np

I. INTRODUCTION

There now exist persuasive arguments[1] to support the view that gauge theories provide a useful description of fundamental physical interactions. These theories, like general relativity, admit an elegant geometrical formulation. Attention focuses on a connection in a principal fiber bundle over spacetime.[2] The gauge field is the curvature of that connection.[3] The Yang–Mills[4] field equations relate the currents of source fields to certain derivatives of the curvature. When there are no sources present, the field equations simply require that the dual of the curvature be D-closed, where D is the covariant exterior derivative defined by the bundle connection.[3] These field equations governing the connection are nonlinear equations, and little is known about properties of generic solutions. In this paper we show that there is, however, a large class of source-free Yang–Mills solutions whose properties are easily studied. Although they are too special to yield inferences about generic Yang–Mills solutions, the class is large enough to provide a substantial reservoir of solutions from which to draw examples and counterexamples.

Throughout this paper, reference to a Yang–Mills solution (or a Yang–Mills bundle-with-connection) should be understood as meaning a *source-free* Yang–Mills solution. Similarly, a Maxwell field (or a Maxwell bundle) will always refer to a *source-free* solution. The gauge group of a Yang–Mills solution is allowed to be any Lie group except where we make restrictions. By a *Maxwell bundle* we shall mean a Yang–Mills principal bundle-with-connection whose gauge group is one-dimensional. The group need not be U(1), nor even a connected group. It is only necessary that its Lie algebra be the real number line. The base space for any Yang–Mills bundle is assumed to be a four-dimensional spacetime, but it need not be Minkowski's spacetime.

We begin by recalling a well-known local construction of a class of Yang–Mills solutions: Let $\mathscr{A}$ be a real 1-form on spacetime whose curl is a (source-free) Maxwell field f. If $\mathbf{b}$ is a fixed element in the Lie algebra of some group, then $\mathbf{A} := \mathbf{b}\mathscr{A}$ is a local potential for a (source-free) Yang–Mills field. This can be easily verified by showing that $\mathbf{A}$ and its curvature $\mathbf{F}$ satisfy Eq. (2.5). The curvature is $\mathbf{F} = \mathbf{b}f$. If an arbitrary gauge transformation is made, the potential $\mathbf{A}$ can no longer be expected to be simply a product of a Lie-algebra-valued function with a real 1-form. The covariant curvature $\mathbf{F}$, however, is necessarily still a product of a Lie-algebra-valued function with f. The above construction, which has been described with reference to a local gauge, produces a Yang–Mills solution with *decomposable curvature*, a property which is gauge-independent.

The question now arises as to how general this simple construction is. Given a Yang–Mills solution with decomposable curvature, does it arise from a Maxwell solution in this special manner? The answer is a much qualified yes. The above construction is a local construction, employing a local gauge. To answer the question globally, we will need to use the geometry of connections in principal fiber bundles. If we restrict attention to solutions such that the field f occurring in the curvature is algebraically nonnull or null with twisting rays, then the answer is yes. If f is null with twist-*free* rays, then the strongest conclusion one can draw is that there exists a local gauge in which the potential has the form $\mathbf{A} = \mathbf{b}(u)\mathscr{A}$, where $\mathscr{A}$ is a potential for a Maxwell field f. Here the potential $\mathscr{A}$ has been chosen so that $\mathscr{A} \wedge du = 0$, where[5] u satisfies $f^{+} \wedge du = 0$, and $\mathbf{b}$, rather than being constant, can be an arbitrary Lie-algebra-valued function of u. Conversely, if $f = d\mathscr{A}$ is a null–twist-free Maxwell solution satisfying $\mathscr{A} \wedge du = 0$, where $f^+ \wedge du = 0$, then it is easily checked that $\mathbf{A} \equiv \mathbf{b}(u)\mathscr{A}$ is a solution of the Yang–Mills equation (2.5) for an arbitrary function $\mathbf{b}(u)$ of u.

Let us now regard a Yang–Mills solution as a connection in a principal fiber bundle, and suppose that its curvature is decomposable:

$$\mathbf{\Omega} = \boldsymbol{\beta}\omega.$$

Here $\mathbf{\Omega}$ is the Lie-algebra-valued curvature 2-form, $\boldsymbol{\beta}$ a Lie-algebra-valued function, and ω a real-valued 2-form. At any point of the bundle there is defined by $\boldsymbol{\beta}$ a one-dimensional subspace of the Lie algebra, and the curvature $\mathbf{\Omega}$ maps all pairs of tangent vectors at the point into the one-dimensional Lie algebra subspace. If the source-free Yang–Mills equations are satisfied (and $\mathbf{\Omega}$ is not null[6] with twist-free rays), then the Lie algebra subspace is constant along horizontal curves in the bundle. This means that if ξ is a point in the bundle and η is some other point which can be reached from ξ by a horizontal curve, then the curvature at η maps all pairs of tangent vectors to the same one-dimensional Lie algebra subspace as is defined by the curvature at ξ. Now, according to the theorem of Ambrose and Singer,[3] the Lie algebra of the holonomy group at ξ is spanned by the Lie-algebra vectors obtained by the action of the curvature 2-form on all pairs of vectors at all points in the holonomy bundle of ξ. But the holonomy bundle of ξ is the set of points η which can be reached by horizontal curves from ξ, and the curvature maps into the same one-dimensional subspace at all such points. So the holonomy Lie algebra is one-dimensioanl. Since any bundle-with-connection is reducible to the holonomy bundle of any point, in this case it can be reduced to a principal bundle-with-connection whose fibers are one-dimensional. The reduced connection satisfies the Yang–Mills equations, and so the reduced bundle-with-connection can be regarded as a Maxwell bundle. The construction of Yang–Mills solutions with decomposable curvature is therefore straightforward. Given a Maxwell bundle, it is not difficult to construct the bundles which are reducible to it.

In the simple picture just presented we tacitly assumed that the curvature was everywhere nonzero, thereby defining a nontrivial Lie algebra subspace at each point. If there are open regions of zero curvature, the global structure can be quite complicated. In particular, the holonomy group need not be one-dimensional. See Sec. III for further details on this.

The holonomy group may have more than one dimension also when there is an open region on which the curvature is null with twist-free rays. In such a region the Yang–Mills field equations do not require the direction of $\boldsymbol{\beta}$ to be constant along *all* horizontal curves, but only along horizontal curves whose projections lie in the hypersurfaces orthogonal to the rays. In this case we offer a different bundle description of how to construct local Yang–Mills solutions out of null–twist-free Maxwell solutions. See Sec. IV for details.

Whenever the gauge Lie group admits a two-dimensional Abelian subgroup, there is room for generalizing the decomposable solutions we are discusssing. Fix two independent Lie algebra elements, $\mathbf{b}$ and $\mathbf{c}$, which commute. Locally, suppose $d\mathscr{A} = f$ and $d\mathscr{B} = {}^*f$, where f is a Maxwell field. Then $\mathbf{A} := \mathbf{b}\mathscr{A} + \mathbf{c}\mathscr{B}$ satisfies the Yang–Mills equations (2.5), as can be easily checked. The curvature is $\mathbf{F} = \mathbf{b}f + \mathbf{c}{}^*f$, which is not decomposable. But the self-dual curvature is $\mathbf{F}^+ = (\mathbf{b} + i\mathbf{c})f^+$, which *is* the product of a single (complex) Lie-algebra-valued function and a complex 2-form. Such a solution has *decomposable self-dual curvature*. Any solution with decomposable curvature has decomposable self-dual curvature, but not conversely. The property of decomposable self-dual curvature is gauge invariant and thus is also a property of the principal bundle connection. When a decomposable self-dual curvature is everywhere nonnull or null-twisting, then the Yang–Mills bundle-with-connection has two-dimensional abelian holonomy groups (one complex dimension), and it is reducible to any holonomy subbundle. The construction of these solutions, locally or globally, is straightforward. The structure of the local solutions wih null–twist-*free* decomposable self-dual curvature will also be given in Sec. V. The entire class of solutions with decomposable self-dual curvature is the same as the class of solutions denoted as type D or type N by Anandan and Tod.[7] The type-N solutions have *null* decomposable self-dual curvature.

II. GEOMETRICAL BACKGROUND AND PRELIMINARY RESULTS

In this section we develop the preliminary background material utilized throughout the remainder of the paper. Generally our arena is that of a principal fiber bundle P whose structure group is a general Lie group G. The base space M of a bundle is assumed to be a four-dimensional spacetime. We denote the projection of P onto M by π. The reader may consult Trautman[2] or Kobayashi and Nomizu[3] for details regarding elementary properties of fiber bundles.

To say that s is a local section of P means that s is a mapping from an open subset U of M into P such that $\pi(s(x)) = x$ for every $x \in U$. In particular $s(U)$ will be a four-dimensional surface lying in the bundle in such a way that if $y \in U$ then the fiber $\pi^{-1}(y)$ over y intersects $s(U)$ in precisely one point. We will deliberately blur the distinction between s and the surface $s(U)$ depending on the context.

Following Trautman,[2] we will identify a local section with a local gauge. Moreover,we will often refer to a local section s as a local gauge. If s_1 and s_2 are local gauges and if their respective domains U_1 and U_2 intersect, then there is a smooth mapping from $U_1 \cap U_2$ into G, $x \rightarrow g(x)$, such that

$$s_1(x) = s_2(x)g(x)$$

for all $x \in U_1 \cap U_2$. This mapping will be referred to as a *gauge transformation*.

The paragraphs above describe the general setting which serves as background for the various fields which are the objects of our study. For notational convenience, functions and forms defined on bundles will be denoted by Greek letters while those on the base space will be denoted by Latin letters. Moreover, Lie-algebra-valued maps will be distinguished by the fact that they will always occur in bold type.

In the setting of fiber bundles both a connection and its curvature are Lie-algebra-valued forms on the bundle P. Thus, in conformity with our notational convention, $\mathbf{\Phi}$ will denote a connection on P and $\mathbf{\Omega}$ will denote its curvature. Recall[2,3] that the connection $\mathbf{\Phi}$ is related to its curvature $\mathbf{\Omega}$ by the equation

$$\mathbf{\Omega} = D\,\mathbf{\Phi} = \mathrm{hor}(d\,\mathbf{\Phi}) = d\,\mathbf{\Phi} + \tfrac{1}{2}[\mathbf{\Phi},\mathbf{\Phi}], \tag{2.1}$$

where $\mathrm{hor}(d\,\mathbf{\Phi})$ denotes the horizontal part[8] of $d\,\mathbf{\Phi}$. In Eq. (2.1) we have utilized the notation $d\,\mathbf{\Phi}$ and $[\mathbf{\Phi},\mathbf{\Phi}]$, which we now explain. Generally, if $\mathbf{T}$ is a Lie-algebra-valued p-form and $\{\mathbf{e}_i\}$ is a basis of the Lie algebra, then $dT \equiv (dT^i)\mathbf{e}_i$, where $T = T^i\mathbf{e}_i$. Similarly if $\mathbf{T} = T^i\mathbf{e}_i$ and $\mathbf{N} = N^i\mathbf{e}_i$, then $[\mathbf{T},\mathbf{N}] \equiv (T^i \wedge N^j)\,[\mathbf{e}_i,\mathbf{e}_j]$, where $[\mathbf{e}_i,\mathbf{e}_j]$ is the Lie bracket of $\mathbf{e}_i$ with $\mathbf{e}_j$ in the Lie algebra.

The reader may be more familiar with a local gauge formulation of the equation (2.1). Let

$$\mathbf{A} = s^*\mathbf{\Phi}, \quad \mathbf{F} = s^*\mathbf{\Omega}$$

in a local gauge s. Equation (2.1) becomes

$$\mathbf{F} = d\,\mathbf{A} + \tfrac{1}{2}[\mathbf{A},\mathbf{A}]. \tag{2.2}$$

In order to discuss the Yang-Mills equation, we need the concept of the dual curvature. To define this concept, let ${}^\star\mathbf{F}$ denote the usual dual of the 2-form $\mathbf{F}$ relative to the volume defined by the spacetime metric on M. Then define *the dual of* $\mathbf{\Omega}$, denoted ${}^\star\mathbf{\Omega}$, by the requirement that if $\xi \in P$ and Γ and Δ are tangent vectors to P at ξ, then

$${}^\star\mathbf{\Omega}_\xi(\Gamma,\Delta) = {}^\star\mathbf{F}_{\pi\xi}(d\pi(\Gamma),d\pi(\Delta)). \tag{2.3}$$

Observe that this definition is independent of which gauge s was used to obtain $\mathbf{F}$. It is not difficult to verify that ${}^\star\mathbf{\Omega}$ is a tensorial 2-form[9] on P and thus that $D{}^\star\mathbf{\Omega} = d{}^\star\mathbf{\Omega} + [\mathbf{\Phi},{}^\star\mathbf{\Omega}]$.

The equation $D\,\mathbf{\Omega} = 0$ is an identity and is called the Bianchi identity.[3] On the other hand, $D{}^\star\mathbf{\Omega} = 0$ is not always valid, but when it is, $\mathbf{\Omega}$ is referrred to as a *solution of the source–free Yang–Mills equation*. Since, however, we deal only with source-free equations we *will refer to* $D{}^\star\mathbf{\Omega} = 0$ *as the Yang–Mills equation, and in this case we refer to* $\mathbf{\Omega}$ *as a Yang–Mills curvature*. In a local gauge s the Bianchi identity becomes

$$D\,\mathbf{F} = d\,\mathbf{F} + [\mathbf{A},\mathbf{F}] = 0,$$

while the Yang–Mills equation reduces to

$$D{}^\star\mathbf{F} = d{}^\star\mathbf{F} + [\mathbf{A},{}^\star\mathbf{F}] = 0.$$

A field $\mathbf{\Omega}^+$ may be defined by

$$\mathbf{\Omega}^+ = \mathbf{\Omega} - i{}^\star\mathbf{\Omega}.$$

It follows that in this notation the two equations $D\,\mathbf{\Omega} = 0$ and $D{}^\star\mathbf{\Omega} = 0$ reduce to a single equation

$$D\,\mathbf{\Omega}^+ = d\,\mathbf{\Omega}^+ + [\mathbf{\Phi},\mathbf{\Omega}^+] = 0, \tag{2.4}$$

or, in a local gauge s,

$$D\,\mathbf{F}^+ = d\,\mathbf{F}^+ + [\mathbf{A},\mathbf{F}^+] = 0. \tag{2.5}$$

The field $\mathbf{\Omega}^+$ has the property that ${}^\star\mathbf{\Omega}^+ = i\mathbf{\Omega}^+$ and is called the *self-dual curvature*.

A curvature $\mathbf{\Omega}$ is said to be *decomposable* if, at each point ξ of P, there exists a neighborhood Υ of ξ such that

$$\mathbf{\Omega} = \boldsymbol{\beta}\omega, \tag{2.6}$$

where $\boldsymbol{\beta}$ is a smooth Lie-algebra-valued function on Υ and where ω is a smooth real-valued 2-form on Υ. Note that even though $\mathbf{\Omega}$ is smooth, one cannot infer the existence of such a $\boldsymbol{\beta}$ and ω if one merely assumes that, at each point $\xi \in P$, $\mathbf{\Omega}_\xi$ is the product of a Lie-algebra element and a real-valued skew-symmetric bilinear map on $T_\xi P$. Appendix A clarifies this point by giving an appropriate example.

Recall that if $\boldsymbol{\gamma}$ is a Lie-algebra-valued function defined on a bundle P such that

$$(R_g^*\boldsymbol{\gamma})(\xi) = \boldsymbol{\gamma}(\xi g) = \mathrm{Ad}(g^{-1})\boldsymbol{\gamma}(\xi)$$

for all $\xi \in P$ and g in the structure group G of P, then $\boldsymbol{\gamma}$ is called a *Higgs field*.[9] Also any real-valued p-form τ on P such that

$$R_g^*\tau = \tau$$

for all $g \in G$ will be called an *invariant p-form* on P. Observe that a p-form τ on P is both horizontal and invariant if and only if there is a p-form t on the base manifold M such that $\tau = \pi^*t$. In that case we sometimes refer to τ as a *pullback form on P*.

An open subset of Υ of a principal fiber bundle P will be called a *bundle neighborhood* of P if $\pi(\Upsilon)$ is open in M and $\Upsilon = \pi^{-1}(\pi(\Upsilon))$.

Observation: If $\mathbf{\Omega}$ *is a decomposable curvature then each point* $\xi \in P$ *is contained in a bundle neighborhood* Q *of* P *such that* $\mathbf{\Omega}$ *is the product of a Higgs field* $\boldsymbol{\gamma}$ *and a pullback 2-form* τ *on* Q.

To see that this is so, choose a local section s of P through ξ such that s lies in the neighborhood Υ on which $\mathbf{\Omega} = \boldsymbol{\beta}\omega$. If V is the domain of s let $Q = \pi^{-1}(V)$ and define $\boldsymbol{\gamma}$ on Q by requiring that $\boldsymbol{\gamma}$ agree with $\boldsymbol{\beta}$ on $s(V)$ and by forcing $\boldsymbol{\gamma}$ to have the Higgs field property on each fiber which intersects $s(V)$; thus $\boldsymbol{\gamma}(s(x)g) \equiv \mathrm{Ad}(g^{-1})\boldsymbol{\beta}(s(x))$ for $x \in V$, $g \in G$. It is clear that $\boldsymbol{\gamma}$ is smooth and is a Higgs field on Q. Define a 2-form τ on Q as follows: At each point of $s(V)$ let τ agree with ω, but for a pair of vectors (Γ,Δ) at a point $s(x)g$ on a fiber through $s(V)$ let

$$\tau_{s(x)g}(\Gamma,\Delta) = \omega_{s(x)}(dR_{g^{-1}}(\Gamma),dR_{g^{-1}}(\Delta)).$$

It is not difficult to show that τ is a smooth invariant 2-form on Q. We have that

$$\mathbf{\Omega}_{s(x)} = \boldsymbol{\beta}(s(x))\omega_{s(x)} = \boldsymbol{\gamma}(s(x))\tau_{s(x)}.$$

Since $\mathbf{\Omega}$ is tensorial[9], $\boldsymbol{\gamma}$ is a Higgs field, and τ is invariant, it follows that $\mathbf{\Omega} = \boldsymbol{\gamma}\tau$. It does not yet follow that τ is horizontal. Indeed, if $\xi \in P$ is a point where both $\mathbf{\Omega}$ and $\boldsymbol{\beta}$ vanish, then the equation $\mathbf{\Omega} = \boldsymbol{\beta}\omega$ offers no information about ω_ξ, and, consequently, τ_ξ might fail to be horizontal. On the other hand, $\mathbf{\Omega} = \boldsymbol{\gamma}\tau$ implies that $\mathbf{\Omega} = \mathrm{hor}\mathbf{\Omega} = \boldsymbol{\gamma}(\mathrm{hor}\tau)$ and $\mathrm{hor}\tau$ is an invariant horizontal form and thus is a pullback form. The observation follows.

Our results require a rather careful analysis of $\mathbf{\Omega}$ in the vicinity of those points of P where $\mathbf{\Omega}$ and certain of its Lie derivatives vanish. With this in mind let $\Sigma_{\mathbf{\Omega}}$ denote the set of points $\xi \in P$ such that $\mathbf{\Omega}(\xi) \equiv 0$ and let Σ denote those $\xi \in \Sigma_{\mathbf{\Omega}}$ such that

$$[(\mathscr{L}_{A_k}\cdots\mathscr{L}_{A_1})\mathbf{\Omega}](\xi) \equiv 0$$

for *every* sequence $\Lambda_1, \Lambda_2, \ldots, \Lambda_s$ of horizontally lifted vector fields on P. The set Σ will be called the *flat set* of $\mathbf{\Omega}$.

Observe that if Λ is a horizontally lifted vector field on P and $\boldsymbol{\beta}$ is a local Higgs field, then $\mathscr{L}_\Lambda \boldsymbol{\beta}$ is a local Higgs field. Similarly if τ is a pullback form, so is $\mathscr{L}_\Lambda \tau$. It follows that the Lie derivative of the product $\boldsymbol{\beta}\tau$ is the sum of two terms each of which is the product of a local Higgs field and a pullback form:

$$\mathscr{L}_\Lambda(\boldsymbol{\beta}\tau) = (\mathscr{L}_\Lambda \boldsymbol{\beta})\tau + \boldsymbol{\beta}(\mathscr{L}_\Lambda \tau).$$

Thus, if $\mathbf{\Omega}$ is decomposable, then, for every sequence of horizontally lifted vector fields $\Lambda_1, \Lambda_2, \ldots, \Lambda_s$, $(\mathscr{L}_{\Lambda_s} \cdots \mathscr{L}_{\Lambda_1})\mathbf{\Omega}$ is tensorial. One consequence of this fact is that Σ is "bundle-like," i.e., $\pi^{-1}(\pi\Sigma) = \Sigma$. Obviously $\Sigma_{\mathbf{\Omega}}$ also has this property.

We need to distinguish three classes of decomposable Yang–Mills curvatures. Let $\xi \in P$. A curvature $\mathbf{\Omega}$ is said to be *null* at ξ if $(\mathbf{\Omega} \wedge {}^*\mathbf{\Omega})(\xi) = 0$ and otherwise it is *nonnull* at ξ. In a local gauge s the null condition translates into $(\mathbf{F} \wedge {}^*\mathbf{F})(\pi(\xi)) = 0$. It is well known that if $\mathbf{F}$ is null in a neighborhood of $x \in M$ and $\mathbf{F}_x \neq 0$, then there is an open set U about x and a 1-form l (unique up to scale factor) on U such that $\mathbf{F}^+ \wedge l = 0$. The vector field corresponding to the 1-form l gives the "repeated principal null direction" of $\mathbf{F}$. Moreover, l can be pulled back to the bundle neigborhood $\pi^{-1}(U)$ of P. Thus we obtain a horizontal invariant 1-form $\lambda = \pi^* l$ on $\pi^{-1}(U)$ such that

$$\mathbf{\Omega}^+ \wedge \lambda = 0. \tag{2.7}$$

Observe that if $\mathbf{\Omega}$ is nonnull at a point ξ, then it is nonnull throughout a neighborhood of ξ. If $\mathbf{\Omega}$ is null and nonzero throughout a neighborhood of ξ, then there exists a pullback 1-form λ satisfying (2.7) above. In case $(d\lambda \wedge \lambda)(\xi) \neq 0$ we see that $d\lambda \wedge \lambda \neq 0$ in a neighborhood of ξ, and we say that $\mathbf{\Omega}$ is *null-twisting* at ξ. Thus if we say that $\mathbf{\Omega}$ is null-twisting at a point, we will mean that λ exists in an entire neighborhood of the point and that $d\lambda \wedge \lambda \neq 0$ on the neighborhood; in particular, $\mathbf{\Omega}$ is nonzero throughout the neighborhood. Finally, if $\mathbf{\Omega}$ is nonzero and null on a neighborhood of $\xi \in P$ and if $d\lambda \wedge \lambda = 0$ at ξ, then we say that $\mathbf{\Omega}$ is *null–twist-free* at ξ.

III. DECOMPOSABLE CURVATURES: THE NONNULL AND NULL-TWISTING CASES

In this section we will assume that the Yang–Mills curvature is decomposable and is not null–twist-free on any open subset of $P - \Sigma$. We will show that in the presence of certain topological restrictions on the flat set Σ of $\mathbf{\Omega}$ the holonomy subbundle $P(\xi)$ of P through $\xi \in P$ has fiber dimension 1. Utilizing this fact, we will show that the curvature is a Maxwell curvature on $P(\xi)$ and that the Yang–Mills bundle arises as an "extension" of a Maxwell bundle.

Lemma 3.1: Assume that $\mathbf{\Omega}$ is a decomposable Yang–Mills curvature which is not null–twist-free on any open subset of $P - \Sigma$. If $\xi \in P - \Sigma$, then there is a bundle neighborhood Q_ξ of ξ in $P - \Sigma$ such that $\mathbf{\Omega} = \boldsymbol{\beta}\omega$ on Q_ξ with ω an invariant horizontal form on Q_ξ and $\boldsymbol{\beta}$ a nonvanishing Higgs field on Q_ξ such that

$$D\boldsymbol{\beta} = \boldsymbol{\beta}\nu$$

for an appropriate 1-form ν.

Proof: Let $\xi \in P - \Sigma$ and let Υ be a bundle neighborhood of ξ in $P - \Sigma$ such that $\mathbf{\Omega} = \boldsymbol{\beta}\omega$ for some Higgs field $\boldsymbol{\beta}$ and some pullback 2-form ω on Υ. The Yang–Mills equation implies

$$\begin{aligned} 0 = D\mathbf{\Omega}^+ &= \operatorname{hor} d(\beta\omega^+) \\ &= \operatorname{hor}[(d\boldsymbol{\beta} \wedge \omega^+) + \boldsymbol{\beta} d\omega^+] \\ &= D\boldsymbol{\beta} \wedge \omega^+ + \boldsymbol{\beta}\operatorname{hor}(d\omega^+) \\ &= D\boldsymbol{\beta} \wedge \omega^+ + \boldsymbol{\beta} d\omega^+. \end{aligned}$$

Let $\cdot$ denote an arbitrary positive definite inner product on the Lie algebra G' and define a 1-form ν on $\Upsilon\text{-}\Sigma_{\mathbf{\Omega}}$ by $\nu = [\boldsymbol{\beta}\cdot D\boldsymbol{\beta}]/[\boldsymbol{\beta}\cdot\boldsymbol{\beta}]$. The identity $(D\boldsymbol{\beta} \wedge \omega^+) + \boldsymbol{\beta} d\omega^+ = 0$ implies that

$$d\omega^+ + (\nu \wedge \omega^+) = 0 \tag{3.1}$$

and hence

$$(D\boldsymbol{\beta} - \boldsymbol{\beta}\nu) \wedge \omega^+ = 0. \tag{3.2}$$

If $\mathbf{\Omega}$ is nonnull at $\xi' \in \Upsilon$, then it follows from (3.2) that $D\boldsymbol{\beta} = \boldsymbol{\beta}\nu$ on a bundle neighborhood of ξ' in Υ. Moreover, if $\mathbf{\Omega}$ is null and nonzero at $\xi' \in \Upsilon$, then (3.2) implies that

$$D\boldsymbol{\beta} = \boldsymbol{\beta}\nu + \boldsymbol{\eta}\lambda, \tag{3.3}$$

where λ is defined by the principal null direction of $\mathbf{\Omega}$ and where $\boldsymbol{\eta}$ is an appropriate Lie-algebra-valued function defined in a bundle neighborhood of ξ'.

We now argue that it is possible to modify $\boldsymbol{\eta}$ and ν in (3.3) to obtain a local Higgs field $\hat{\boldsymbol{\eta}}$ and a pullback form $\hat{\nu}$ such that $D\boldsymbol{\beta} = \boldsymbol{\beta}\hat{\nu} + \hat{\boldsymbol{\eta}}\lambda$. Because $D\boldsymbol{\beta}$ is tensorial, so is $\boldsymbol{\beta}\nu + \boldsymbol{\eta}\lambda$, and, since $\boldsymbol{\beta}$ is a local Higgs field and λ is a pullback form, it follows from the tensorial property that ν is a pullback form modulo addition of some multiple of λ. Suppose σ is a function such that $\hat{\nu} := \nu + \sigma\lambda$ is a pullback form. Then $D\boldsymbol{\beta} - \boldsymbol{\beta}\hat{\nu}$ is tensorial and by (3.3) is equal to $(\boldsymbol{\eta} - \boldsymbol{\beta}\sigma)\lambda$. Since λ is a pullback form it follows that $\hat{\boldsymbol{\eta}} := \boldsymbol{\eta} - \boldsymbol{\beta}\sigma$ is a local Higgs field. Thus $D\boldsymbol{\beta} = \boldsymbol{\beta}\hat{\nu} + \hat{\boldsymbol{\eta}}\lambda$, where $\hat{\nu}$ and $\hat{\boldsymbol{\eta}}$ have the desired properties.

We now show that if $\mathbf{\Omega}$ is null, nonzero, and not twist-free at ξ', then $D\boldsymbol{\beta} = \boldsymbol{\beta}\nu$ on a bundle neighborhood of ξ'. To accomplish this, we will show that $\hat{\boldsymbol{\eta}}(d\lambda \wedge \lambda) = -\boldsymbol{\beta}(d\hat{\nu} \wedge \lambda)$. It will then follow that $\hat{\boldsymbol{\eta}}$ is a multiple of $\boldsymbol{\beta}$ and thus that $D\boldsymbol{\beta} = \boldsymbol{\beta}\hat{\nu} + \hat{\boldsymbol{\eta}}\boldsymbol{\lambda} = \boldsymbol{\beta}\mu$ for an appropriate μ. We will then see that $\mu = \nu$. First observe that $D^2\boldsymbol{\beta} = D(D\boldsymbol{\beta}) = [\mathbf{\Omega}, \boldsymbol{\beta}] = [\boldsymbol{\beta}\omega, \boldsymbol{\beta}] = 0$. Since $\hat{\nu}$ is a horizontal invariant form $D\hat{\nu} = d\hat{\nu}$ and (3.3), appropriately modified for $\hat{\boldsymbol{\eta}}$ and $\hat{\nu}$, yields

$$\begin{aligned} 0 = D^2\boldsymbol{\beta} &= D(\boldsymbol{\beta}\hat{\nu}) + D(\hat{\boldsymbol{\eta}}\lambda) \\ &= (D\boldsymbol{\beta} \wedge \hat{\nu}) + \boldsymbol{\beta}(D\hat{\nu}) + D\hat{\boldsymbol{\eta}} \wedge \lambda + \hat{\boldsymbol{\eta}}(D\lambda) \\ &= \hat{\boldsymbol{\eta}}(\lambda \wedge \hat{\nu}) + \boldsymbol{\beta} d\hat{\nu} + (D\hat{\boldsymbol{\eta}}) \wedge \lambda + \hat{\boldsymbol{\eta}} d\lambda. \end{aligned}$$

If we wedge this equation with λ, we get

$$\boldsymbol{\beta}(d\hat{\nu} \wedge \lambda) = -\hat{\boldsymbol{\eta}}(d\lambda \wedge \lambda). \tag{3.4}$$

By transvecting this equation with an appropriate triple of vector fields we see that $\hat{\boldsymbol{\eta}}$ is a multiple of $\boldsymbol{\beta}$ and thus that $D\boldsymbol{\beta} = \boldsymbol{\beta}\mu$ for some μ. Dotting both sides of the this last equation with $\boldsymbol{\beta}$ and dividing by $\boldsymbol{\beta}\cdot\boldsymbol{\beta}$, we obtain

$$\mu = [\beta\cdot D\,\beta]/[\beta\cdot\beta] = \nu$$

and consequently $D\,\beta = \beta\nu$ on a bundle neighborhood of $\xi' \in \Upsilon$. It follows that $D\,\beta = \beta\nu$ on $\Upsilon - \Sigma_\Omega$.

If $\beta(\xi) \neq 0$, then there is an open set about ξ on which β does not vanish, and, since β is a Higgs field, there is a bundle neighborhood Q_ξ on which β does not vanish. In this case Q_ξ and β have all the properties required in the lemma.

If $\beta(\xi) = 0$, we must find another decomposition of Ω near ξ whose Lie algebra part is not zero at ξ. Since $\xi \notin \Sigma$, it follows that there exist horizontally lifted vector fields $\Lambda_1, \Lambda_2, \ldots, \Lambda_s$ such that $[(\mathscr{L}_{\Lambda_s} \cdots \mathscr{L}_{\Lambda_2} \mathscr{L}_{\Lambda_1})\Omega](\xi) \neq 0$. If $s = 1$, it follows that $(\mathscr{L}_{\Lambda_1}\beta)(\xi) \neq 0$ since $\mathscr{L}_{\Lambda_1}\Omega = \beta(\mathscr{L}_{\Lambda_1}\omega) + (\mathscr{L}_{\Lambda_1}\beta)\omega$ and $\beta(\xi) = 0$. More generally, it follows from the Leibniz formulae that there exist $k \geqslant 0$ such that all Lie derivatives of β with order less than k vanish at ξ, but there exists a sequence of horizontally lifted fields $\Lambda_1, \Lambda_2, \ldots, \Lambda_k$ such that $[(\mathscr{L}_{\Lambda_k} \cdots \mathscr{L}_{\Lambda_2} \mathscr{L}_{\Lambda_1})\beta](\xi) \neq 0$. We know that $D\,\beta = \beta\nu$ on $\Upsilon - \Sigma_\Omega$. But $\mathscr{L}_{\Lambda_1}\beta = d\,\beta(\Lambda_1) = d\,\beta(\mathrm{hor}\Lambda_1) = D\,\beta(\Lambda_1) = \beta\nu(\Lambda_1)$ and

$$\begin{aligned}\mathscr{L}_{\Lambda_2}(\mathscr{L}_{\Lambda_1}\beta) &= \mathscr{L}_{\Lambda_2}(\beta\nu(\Lambda_1))\\ &= \beta\mathscr{L}_{\Lambda_2}(\nu(\Lambda_1)) + (\mathscr{L}_{\Lambda_2}\beta)\nu(\Lambda_1)\\ &= \beta\left[\mathscr{L}_{\Lambda_2}(\nu(\Lambda_1)) + \nu(\Lambda_2)\nu(\Lambda_1)\right].\end{aligned}$$

An inductive argument shows that on $\Upsilon - \Sigma_\Omega$

$$(\mathscr{L}_{\Lambda_k} \cdots \mathscr{L}_{\Lambda_1})\beta = \beta\phi$$

for some smooth real-valued ϕ. Since $[(\mathscr{L}_{\Lambda_k} \cdots \mathscr{L}_{\Lambda_1})\beta](\xi) \neq 0$, there exists an open subset of Υ on which $\alpha := (\mathscr{L}_{\Lambda_k} \cdots \mathscr{L}_{\Lambda_1})\beta$ does not vanish. Moreover, α is a Higgs field and thus there is a bundle neighborhood Q_ξ of ξ in $P - \Sigma$ on which it does not vanish. Thus ϕ does not vanish on $Q_\xi - \Sigma_\beta$ and on this set $1/\phi = (\alpha\cdot\beta)/(\alpha\cdot\alpha)$. The right-hand side of the last equation is meaningful on all of Q_ξ, and, if we define $\theta = (\alpha\cdot\beta)/(\alpha\cdot\alpha)$ on Q_ξ, we have that $\Omega = \beta\omega = (\alpha\theta)\omega = \alpha(\theta\omega)$. Since $\alpha = (\mathscr{L}_{\Lambda_k} \cdots \mathscr{L}_{\Lambda_1})\beta$ is a Higgs fields, we see that $\theta\omega$ is a pullback form (it is a horizontal invariant 2-form on Q_ξ). If $\mu = (\alpha\cdot D\alpha)/(\alpha\cdot\alpha)$, one can show as before that $D\alpha = \alpha\mu$ on $Q_\xi - \Sigma_\Omega$. But μ is defined on all of Q_ξ and Σ_Ω contains no open subsets of Q_ξ. Thus it follows from continuity that $D\alpha = \alpha\mu$ on all of Q_ξ. The lemma follows.

Theorem 3.2: Assume that Ω is a decomposable Yang–Mills curvature which is not null–twist-free on any open subset of $P - \Sigma$. If Σ is the flat set of Ω and Q is any connected component of $P - \Sigma$, then Q is an open bundle neighborhood of P such that the holonomy bundle $Q(\xi)$ of Q through $\xi \in Q$ has one-dimensional fibers. Moreover, on the holonomy bundle $Q(\xi)$ there is an $\alpha_0 \in G'$ such that $\Omega = \alpha_0\omega$ for some smooth 1-form ω on $Q(\xi)$.

Remark: It should be noted that the Lie-algebra constant α_0 need not be a Higgs field on $Q(\xi)$ and ω may not be a pullback form.

Proof of Theorem 3.2: Let $\xi \in Q$ and let γ: $[0,1] \to Q$ be an arbitrary horizontal curve in Q such that $\gamma(0) = \xi$. Let $\Omega = \beta\omega$ be a decomposition of Ω on some bundle neighborhood Q_ξ of ξ such that β is a nonvanishing Higgs field on Q_ξ and $D\,\beta = \beta\nu$ for some ν. For each $0 \leqslant t \leqslant 1$ let $Q_t, \beta_t, \omega_t, \nu_t$ be similarly defined at $\gamma(t)$ with $Q_0 = Q_\xi$, $\beta_0 = \beta$, $\omega_0 = \omega$, $\nu_0 = \nu$. Since $\gamma([0,1])$ is compact, there exists

$$0 = t_0 < t_1 < t_2 < \cdots < t_n = 1$$

such that $\gamma[t_i, t_{i+1}] \subseteq Q_{t_i}$. Let $\beta_i = \beta_{t_i}$ and $\nu_i = \nu_{t_i}$. We have, for $t_i \leqslant t \leqslant t_{i+1}$,

$$\frac{d}{dt}[\beta_i(\gamma(t))] = d\,\beta_i\left(\frac{d\gamma}{dt}\right) = D\,\beta_i\left(\frac{d\gamma}{dt}\right) = \beta_i(\gamma(t))\nu_i\left(\frac{d\gamma}{dt}\right)$$

Thus, for $t_i \leqslant t \leqslant t_{i+1}$,

$$\beta_i(\gamma(t)) = \exp\left[\int_0^t \nu_i\left(\frac{d\gamma}{dt}(s)\right)ds\right]\beta_i(\gamma(t_i)),$$

and $\beta_i(\gamma(t))$ is a multiple of $\beta_i(\gamma(t_i))$. It follows that $\beta_i(\gamma(t)) \propto \beta(\xi)$ for $t \in [t_i, t_{i+1}]$. If $\alpha_0 = \beta(\xi)$, then Ω_η has α_0 as a factor for each $\eta \in Q(\xi)$. It follows that $\Omega = \alpha_0\omega$ for some smooth real-valued 2-form ω on $Q(\xi)$. From this and the Ambrose–Singer theorem[3] we see that the fibers of $Q(\xi)$ are one-dimensional. The theorem follows.

Observation 1: In general, decomposable Yang–Mills solutions do not reduce to subbundles of P having one-dimensional fibers. Indeed, suppose $P - \Sigma$ is disconnected with disjoint components Q_1 and Q_2. If $\xi_1 \in Q_1$ and $\xi_2 \in Q_2$, then one could have $\Omega = \alpha_1\omega_1$ on $Q_1(\xi_1)$ and $\Omega = \alpha_2\omega_2$ on $Q_2(\xi_2)$ with α_1 and α_2 independent elements of G'. If there exist horizontal curves from ξ_1 to ξ_2 in P, then the fibers of the holonomy bundle $P(\xi_1) = P(\xi_2)$ would be at least two-dimensional. This follows from the Ambrose–Singer theorem and the fact that α_1 and α_2 are independent.

Observation 2: Note that even in case $P - \Sigma$ has a single component Q, one generally still does not attain reduction of Ω to subbundles with one-dimensional fibers. Indeed, if Σ°, the interior of Σ, is multiply connected, then Σ° can have a nontrivial reduced holonomy group as a bundle over $\pi(\Sigma^\circ)$.[3] This is true even though the curvature of the connection Φ is obviously zero in Σ°. It follows that if $\xi \in Q$, then there may exist horizontal curves γ such that $\gamma(0) = \xi$, but γ enters the zero curvature region Σ°, "picks up holonomy" in Σ°, and reenters Q at a subbundle $Q(\eta) \neq Q(\xi)$. In such a case $P(\xi)$ would contain both $Q(\eta)$ and $Q(\xi)$ and thus its holonomy group would necessarily contain both the groups of $Q(\eta)$ and $Q(\xi)$.

Observation 3: We indicate one final obstruction to reducing Ω to some subbundle with one-dimensional fibers. It is possible that the boundary of the flat set Σ of Ω is not connected and that, for $\xi \in Q$, this disconnection of $\partial\Sigma$ could introduce additional holonomy in $P(\xi)$ which is not present in $Q(\xi)$. As an indication of how this could happen let the spacetime M be Minkowski space M_0 with a 2-plane S through the origin removed, and let the connection Φ be defined on an appropriate bundle P over M. Choose Φ such that the flat set Σ of the curvature of Φ projects onto an open half-space H of M_0 such that the boundary of H contains the 2-plane S. By modifying Φ it is possible to introduce extra holonomy in $P(\xi)$ by requiring that certain noncontractible curves be horizontal.

Other than the three restrictions pointed out in the observations above, the only other obstructions to reducing Ω to a subbundle of P with one-dimensional fibers are somewhat minor. In fact in order to reduce the connection to such subbundles the following hypothesis is sufficient.

Hypothesis: *The flat set Σ of Ω is a manifold with boundary and the boundary of Σ is a connected orientable submanifold of P with codimension* 1. *Also $Q = P - \Sigma$ is connected and the holonomy group of the bundle $\Sigma^\circ \rightarrow \pi(\Sigma^\circ)$ is trivial.*

Observe that if Σ° is simply connected, then the holonomy group of Σ° is trivial since Ω is zero on Σ°. Also notice that even if P is simply connected, deletion of the closure of Q from P may introduce multiconnectivity depending on how Q sits in P.

Most of the remainder of this section will be devoted to a proof of the following theorem.

Theorem 3.3: Let Ω be a decomposable Yang–Mills curvature which is not null–twist-free on any open subset of $P - \Sigma$. If the flat set Σ is such that the above hypothesis is satisfied, then the holonomy bundle $P(\xi)$ through any $\xi \in P$ has one-dimensional fibers.

Proof: In order to show that, for $\xi \in Q$, $P(\xi)$ has one-dimensional fibers, it suffices to show that, on $P(\xi)$, Ω is a multiple of a single Lie algebra direction; i.e., we show there exists $\alpha_0 \in G'$ such that, for every $\eta \in P(\xi)$ and $\Gamma, \Delta \in T_\eta P$, $\Omega_\eta(\Gamma,\Delta)$ is a multiple of α_0. To facilitate the proof of this statement, define an equivalence relation $\sim$ on $G' - \{0\}$ by $\beta_1 \sim \beta_2$ if β_1 and β_2 are linearly dependent. We denote the equivalence class of a nonzero $\beta \in G'$ by $[\beta]$ and the set of equivalence classes by ProjG'. It is easily seen that ProjG' is a manifold and that Ω defines a mapping ϕ from $Q - \Sigma_\Omega$ to ProjG' via $\phi(\eta) = [\Omega_\eta(\Gamma,\Delta)]$ for $\Gamma,\Delta \in T_\eta P$ such that $\Omega_\eta(\Gamma,\Delta) \neq 0$. To see that ϕ is well defined, recall that Lemma 3.1 implies that there is a bundle neighborhood Υ of an arbitary point $\eta \in Q$ and a decomposition of Ω on Υ as a product of a *nonvanishing* Higgs field β_0 and a real-valued 1-form ω_0. It follows that, for $\eta \in Q - \Sigma_\Omega$, $\phi(\eta) = [\beta_0(\eta)]$ independent of Γ and Δ. It is also apparent from the decomposability property that ϕ has a unique smooth extension to all of Q since Σ_Ω contains no open subsets in Q.

We will show how to extend ϕ to all of P such that ϕ is constant on every holonomy subbundle of P. Assuming for the moment that this is possible, we show how to complete the proof of the theorem.

Let $\xi \in Q$ and choose $\alpha_0 \in G'$ such that $\phi(\eta) = [\alpha_0]$ for all $\eta \in Q(\xi)$. Since ϕ is constant on every holonomy subbundle of P, ϕ is constant on $P(\xi)$ which contains $Q(\xi)$. Thus $\phi(\eta) = [\alpha_0]$ on $P(\xi)$. On $P - \Sigma_\Omega = Q - \Sigma_\Omega$ we know that $\phi(\eta) = [\Omega_\eta(\Gamma,\Delta)]$ for arbitrary Γ,Δ. It follows that there is a real-valued 2-form ω_0 on $P(\xi) - \Sigma_\Omega$ such that $\Omega = \alpha_0\omega_0$. If $\cdot$ is a positive-definite inner product on G', we see that $\omega_0 = (\alpha_0 \cdot \Omega)/\alpha_0^2$. Since this equation makes sense on all of $P(\xi)$, we see that ω_0 is well defined on $P(\xi)$ and $\Omega = \alpha_0\omega_0$ on $P(\xi)$. It follows from the Ambrose–Singer theorem that the fibers of $P(\xi)$ are one-dimensional.

Thus it suffices to show that ϕ has an appropriate extension to all of P. First choose a tubular neighborhood U of $\partial(\pi(\Sigma))$ in M. (Recall that such a neighborhood exists[10] and is diffeomorphic to the normal bundle of $\partial(\pi(\Sigma))$; under this diffeomorphism $\partial(\pi(\Sigma))$ is identified with the zero section of the normal bundle.) Since $\partial(\pi\Sigma)$ is orientable and the fibers of the normal bundle are one-dimensional, we can assume the fibers have an ordering on them such that the positive elements of a "fiber" of U are in $\pi(\Sigma^\circ)$ while negative elements are in $\pi(Q)$. Now if $\xi \in \partial\Sigma$, then $\pi(\xi) \in \partial(\pi\Sigma)$ and we can horizontally lift the fiber of U through $\pi(\xi)$ to a horizontal curve δ_ξ through ξ. Now let $\mathscr{T} = \cup_{\xi\in\partial\Sigma}\delta_\xi$. We see that $\mathscr{T}$ is a tubular neighborhood of $\partial\Sigma$. Since ϕ is constant on horizontal curves in Q, it is natural to extend ϕ to $\mathscr{T}$ by letting $\phi(\delta_\xi)$ be the constant $\phi(\delta_\xi \cap Q)$. Thus ϕ has a smooth extension to $\mathscr{T}$. We wish to extend ϕ to all of P. Since the holonomy bundle of Σ° over $\pi(\Sigma^\circ)$ is trivial by the Hypothesis, it follows that there is a global horizontal section $s: \pi(\Sigma^\circ) \rightarrow \Sigma^\circ$ of the bundle Σ°. Moreover, each point of Σ° lies on the image of some such section and, for each such s, $s(\pi(\Sigma^\circ))$ intersects $\mathscr{T}$. If we can show that ϕ is constant on $\mathscr{T} \cap s(\pi(\Sigma^\circ))$, then it will be natural to extend ϕ to all of $s(\pi(\Sigma^\circ))$ by requiring that ϕ be constant on $s(\pi(\Sigma^\circ))$. Thus we wish to show that ϕ is constant on $\mathscr{T} \cap s(\pi(\Sigma^\circ))$. Note that if $\eta_1, \eta_2 \in \mathscr{T} \cup s(\pi(\Sigma^\circ))$, then there exist $\xi_1, \xi_2 \in \partial(\Sigma)$ such that $\eta_1 \in \delta_{\xi_1}$ and $\eta_2 \in \delta_{\xi_2}$. A rather tedious argument can be utilized to show that there exists a horizontal curve γ from ξ_1 to ξ_2 in the boundary of Q. If we can show that $\phi \circ \gamma$ is constant, then it will follow that $\phi(\eta_1) = \phi(\delta_{\xi_1}) = \phi(\xi_1) = \phi(\gamma(0)) = \phi(\gamma(1)) = \phi(\delta_{\xi_2}) = \phi(\eta_2)$ and that ϕ is constant on $\mathscr{T} \cap s(\pi(\Sigma^\circ))$. Thus, to see that ϕ has an appropriate extension to all of P, we need to show that $\phi \circ \gamma$ is locally constant. Let $x = \pi \circ \gamma$ and let (x^1,x^2,x^3,x^4) denote local coordinates at a point $x(t_0)$ of x. Choose the coordinates in such a manner that $x^4 > 0$ defines points of $\pi(\Sigma^\circ)$, $x^4 < 0$, define points of $\pi(Q)$, and $x^4 = 0$ is contained in $\pi(\partial\Sigma) = \pi(\partial Q)$. Now define curves $\{x_\epsilon\}_{\epsilon<0}$ in $\pi(Q)$ by $x^i(x_\epsilon(t)) = x^i(x(t))$ for $i = 1,2,3$ but $x^4(x_\epsilon(t)) = \epsilon$. It is now possible to find horizontal lifts of the cuves x_ϵ to γ_ϵ in Q which have γ as a limit: $\lim_{\epsilon\to 0}\gamma_\epsilon(t) = \gamma(t)$. Since ϕ is constant on horizontal curves in Q (by Theorem 3.2), we see that $(\phi \circ \gamma)(t) = \lim_{\epsilon\to 0}\phi(\gamma_\epsilon(t))$ is independent of t. Thus $\phi \circ \gamma$ is constant and, as we have seen above, ϕ is well defined on all of P.

Finally, we show that ϕ is constant on $P(\xi)$ by showing that ϕ is constant on horizontal curves in P. Let γ be any horizontal curve in P. The interval $[a,b]$ on which γ is defined is the union of subintervals $[a_i,b_i]$ such that, for each i, $\gamma(a_i,b_i)$ lies completely in Q, or in Σ°, or in $\partial\Sigma = \partial Q$. If $\gamma(a_i,b_i) \subseteq \Sigma^\circ$, it follows that it is a subset of a single horizontal global section s of the bundle $\Sigma^\circ \rightarrow \pi(\Sigma^\circ)$, and, since ϕ was defined to be constant on such sections, $\phi(\gamma(a_i,b_i))$ is constant. If $\gamma(a_i,b_i) \subseteq Q$, we know that ϕ is constant on horizontal curves in Q and thus $\phi(\gamma(a_i,b_i))$ is constant. In the previous paragraph it was shown that ϕ is constant on horizontal curves in $\partial Q = \partial\Sigma$. Thus if $\gamma(a_i,b_i) \subseteq \partial Q$, $\phi(\gamma(a_i, b_i))$ is constant. Therefore, $\phi \circ \gamma$ is constant on each of the subintervals (a_i,b_i), and, since ϕ is continuous, it follows that $\phi \circ \gamma$ is constant on $[a,b]$. The theorem follows.

In the introduction we mentioned a well-known result which shows how to build solutions of the Yang–Mills equations from given solutions of Maxwell's equations. The result can be formulated in fiber bundle language and is described below for the convenience of the reader.

We postulate a Maxwell bundle $E(M,G_0)$ over a space-time M with a connection Φ_0 on E such that the Maxwell curvature is $\mathbf{\Omega}_0 = D\,\mathbf{\Phi}_0$. We also assume that the one-dimensional structure group G_0 of the Maxwell bundle can be identified with some Lie subgroup H of a more general Lie group G. We show that the bundle E can be "embedded" in a principal bundle $P(M,G)$ and that the connection $\mathbf{\Phi}_0$ can be "extended" to a connection $\mathbf{\Phi}$ on P with $\mathbf{\Phi}$ satisfying the Yang–Mills equations and having decomposable curvature.

First we "extend" E. Since E is a bundle, we know that there exists a family $\{(U_\alpha, s_\alpha)\}$ of local gauges of E such that the $\{U_\alpha\}$ cover M. If (α,β) is a pair such that $U_\alpha \cap U_\beta \neq \emptyset$, then we know that there is a gauge transformation $g_{\alpha\beta}$: $U_\alpha \cap U_\beta \rightarrow G_0$ such that $s_\beta = s_\alpha g_{\alpha\beta}$. Since G_0 is identified with $H \subseteq G$, we see that the $\{g_{\alpha\beta}\}$ have their values in G, and thus by Proposition 5.2 of Ref. 3 there is a principal fiber bundle $P(M,G)$ with transition functions $\{g_{\alpha\beta}\}$. It also follows from Proposition 5.3 of Ref. 3 that P reduces to a subbundle i: $E \rightarrow P$ of P. By Theorem 6.1 of Ref. 3 there now exists a connection form $\mathbf{\Phi}$ on P such that $i^*\mathbf{\Phi} = \mathbf{\Phi}_0$ and the curvature $\mathbf{\Omega}$ of $\mathbf{\Phi}$ is related to the curvature $\mathbf{\Omega}_0$ of $\mathbf{\Phi}_0$ by $i^*\mathbf{\Omega} = \mathbf{\Omega}_0$.

We now show that $\mathbf{\Omega}$ is decomposable. To see this, first observe that the Lie algebra of $G_0' \cong H'$ is one-dimensional and can be realized as a one-dimensional subalgebra $H' \subseteq G'$ of the Lie algebra of G. Let $\boldsymbol{\beta}_0$ be a nonzero generator of this subalgebra. Since $\mathbf{\Omega}_0$ has its values in $G_0' = H'$, every value of $\mathbf{\Omega}_0$ is a multiple of $\boldsymbol{\beta}_0$. Thus, if $\xi \in E \cong i(E) \subseteq P$ and Γ and Δ are tangent to E in P, then $\mathbf{\Omega}_\xi(\Gamma,\Delta)$ is a multiple of $\boldsymbol{\beta}_0$. Since $\mathbf{\Omega}$ is horizontal, we see that $\mathbf{\Omega}_\xi(\Gamma,\Delta)$ is a multiple of $\boldsymbol{\beta}_0$ for every pair of vectors Γ and Δ at $\xi \in E$. Since $\mathbf{\Omega}$ has the property that $R_g^*\mathbf{\Omega} = \mathrm{Ad}(g^{-1})\mathbf{\Omega}$, we see that $\mathbf{\Omega}_{\xi g}$ is a multiple of $\mathrm{Ad}(g^{-1})\boldsymbol{\beta}_0$ for each $g \in G$. Thus, if we *define* $\boldsymbol{\beta}$ by $\boldsymbol{\beta}(\xi g) = \mathrm{Ad}(g^{-1})\boldsymbol{\beta}_0$ for every $\xi \in E$ and $g \in G$, then $\boldsymbol{\beta}$ is a well-defined global Higgs field on P. Moreover, if $\omega \equiv [\boldsymbol{\beta}\cdot\mathbf{\Omega}]/[\boldsymbol{\beta}\cdot\boldsymbol{\beta}]$ for an arbitrary positive definite inner product $\cdot$ on G' then we see that $\mathbf{\Omega} = \boldsymbol{\beta}\omega$. Thus $\mathbf{\Omega}$ is globally decomposable on P.

Finally we show that $\mathbf{\Omega}$ is a Yang–Mills curvature. If Γ,Δ,Λ are tangent to E, we have

$$\begin{aligned}
0 &= (D\,\mathbf{\Omega}_0^+)(\Gamma,\Delta,\Lambda) = D\,(i^*\mathbf{\Omega}^+)(\Gamma,\Delta,\Lambda)\\
&= \mathrm{hor}(d\,(i^*\mathbf{\Omega}^+)(\Gamma,\Delta,\Lambda))\\
&= i^*(d\,\mathbf{\Omega}^+)(\mathrm{hor}\Gamma,\mathrm{hor}\Delta,\mathrm{hor}\Lambda)\\
&= d\,\mathbf{\Omega}^+(di(\mathrm{hor}\Gamma),di(\mathrm{hor}\Delta),di(\mathrm{hor}\Lambda))\\
&= D\,\mathbf{\Omega}^+(di(\Gamma),di(\Delta),di(\Lambda)).
\end{aligned}$$

But if Γ_0, Δ_0, Λ_0 are tangent to P at a point of $i(E)$, then $\mathrm{hor}\Gamma_0$, $\mathrm{hor}\Delta_0$, $\mathrm{hor}\Lambda_0$ are tangent to E and thus are of the form $di(\Gamma),di(\Delta),di(\Lambda)$ for appropriate Γ,Δ,Λ, respectively. Thus $D\,\mathbf{\Omega}^+(\Gamma_0,\Delta_0,\Lambda_0) = 0$. Since $D\,\mathbf{\Omega}^+$ is tensorial, it follows that $D\,\mathbf{\Omega}^+ = 0$ everywhere and $\mathbf{\Omega}$ is a Yang–Mills curvature.

Theorem 3.4: If (E,M,G_0) is a Maxwell bundle and $\mathbf{\Omega}_0$ is a Maxwell curvature on E, then for each Lie group G containing G_0 as a subgroup there exists a principal bundle (P,M,G) containing (E,M,G_0) and a globally decomposable Yang–Mills curvature $\mathbf{\Omega}$ on P such that P reduces to E and $\mathbf{\Omega}$ agrees with $\mathbf{\Omega}_0$ on E.

Observe that this theorem shows how to construct Yang–Mills fields from given Maxwell fields even in the case when the Maxwell field is null–twist-free. At this point already we obtain a partial converse to this theorem in case the given field is not null–twist-free on any open subset of P.

First recall that a connection $\mathbf{\Phi}$ reduces to a subbundle i: $E \rightarrow P$ of P if and only if the horizontal subspaces of P at points of $i(E)$ are tangent to $i(E)$. Thus any connection $\mathbf{\Phi}$ of a bundle P reduces to any one of its holonomy subbundles.[3] If $\mathbf{\Phi}'$ is such a reduced connection, then $\mathbf{\Phi}' = i^*\mathbf{\Phi}$ and the curvature $\mathbf{\Omega}'$ of $\mathbf{\Phi}'$ satisfies $\mathbf{\Omega}' = i^*\mathbf{\Omega}$. Moreover,

$$D'\mathbf{\Omega}^+ = \mathrm{hor}'(d\,(i^*\mathbf{\Omega})^+) = \mathrm{hor}(di^*(\mathbf{\Omega}^+)) = i^*(D\,\mathbf{\Omega}^+) = 0.$$

Thus $\mathbf{\Omega}'$ is a Yang–Mills curvature on the holonomy bundle. In the presence of the Hypothesis it follows from Theorem 3.3 that each holonomy subbundle has one-dimensional fibers. Consequently, each holonomy subbundle of P is a Maxwell bundle with $\mathbf{\Omega}'$ as its Maxwell curvature. From these remarks we have the following:

Theorem 3.5: Assume that $\mathbf{\Omega}$ is an arbitrary decomposable Yang–Mills curvature which is not null–twist-free on any open subset of P. If the flat set of $\mathbf{\Omega}$ satisfies the Hypothesis, then $\mathbf{\Omega}$ arises from a Maxwell curvature as in Theorem 3.4.

In the next section we give a local characterization of null Yang–Mills curvatures with twist-free rays and obtain a local converse similar to Theorem 3.5 above.

IV. DECOMPOSABLE CURVATURES: THE NULL–TWIST-FREE CASE

In the last section where the decomposable curvature was assumed to be nonnull or null with twisting rays, we were able to give, under suitable topological restrictions, a simple global characterization of the bundle structure—the bundle and connection reduce to a subbundle with one-dimensional fibers. In this section we show that if the decomposable curvature is null with twist-free rays, then this bundle reduction does not occur in general, not even locally. However, we will see that the curvature 2-form can be related to a null Maxwell field. The following theorem demonstrates that, locally, null decomposable curvatures with twist-free rays are readily constructible from null–twist-free Maxwell fields on the base space M.

Theorem 4.1: Let $P(M,G)$ be an arbitrary principal bundle over M, and let f be a null–twist-free Maxwell field on M. For each $p \in M$ there exists a bundle neighborhood Q_p of $\pi^{-1}(p)$ such that on Q_p f induces a class of decomposable null–twist-free curvatures $\mathbf{\Omega} = \boldsymbol{\beta}\pi^*(f)$ satisfying the Yang–Mills equation. The Lie-algebra element $\boldsymbol{\beta} = \boldsymbol{\beta}(u)$ is an arbitrary function of one variable.

Proof: Let $p \in M$. Since f is null–twist-free on M, there is a neighborhood U of p and a 1-form l on U such that $f^+ \wedge l = 0$ and such that $l = du$ for some real-valued function u on U. Moreover, since P is locally trivial and f is a Maxwell field, we can assume without loss of generality that $\pi^{-1}(U)$ is trivial and that on U $f = da$ for some 1-form a. Let $Q_p = \pi^{-1}(U)$ and note that $da \wedge du = 0$ on U. The potential a thus satisfies the conditions of the lemma given in Appendix D and therefore may be written in the form

$$a = H\,du + dK$$

for some functions H and K. We may choose the Maxwell "gauge" such that $a = H\,du$.

Choose an arbitrary section s: $U \to Q_p$ and let $\hat{\Phi}$ be the flat connection on Q_p determined by requiring $\hat{\Phi}$ to vanish on vectors tangent to $s(U)$. Thus, for any $\xi \in Q_p$ the horizontal subspace $\mathscr{H}_\xi \subseteq T_\xi Q_p$ is defined to be

$$\mathscr{H}_\xi = dR_g T_{s(\pi\xi)} \tilde{U},$$

where $\tilde{U} = s(U)$ and $\xi = R_g(s(\pi(\xi)))$.

Next choose an arbitrary Lie-algebra-valued function $\mathbf{b}$: $U \to G'$ that is constant on each "$u =$ const" hypersurface in U. Thus $\mathbf{b} = \mathbf{b}(u)$. We use $\mathbf{b}$ to define a Higgs field $\boldsymbol{\beta}$: $Q_p \to G'$ by requiring it to agree with $\mathbf{b}$ on the gauge s and then by extending it to all of the Q_p by right translation. For each $\xi \in Q_p$ there is a $g \in G$ such that $\xi = s(\pi(\xi)) \cdot g$. Define $\boldsymbol{\beta}$ by the equation

$$\boldsymbol{\beta}(\xi) = \mathrm{Ad}(g^{-1})\mathbf{b}(\pi(\xi)).$$

We now define a Lie-algebra-valued 1-form $\boldsymbol{\Phi}$ on Q_p by

$$\boldsymbol{\Phi} = \hat{\boldsymbol{\Phi}} + \boldsymbol{\beta}\pi^*(a). \tag{4.1}$$

It is easy to check that $\boldsymbol{\Phi}$ defines a connection 1-form on Q_p. The curvature $\boldsymbol{\Omega}$ of $\boldsymbol{\Phi}$ can be evaluated using the structure equation (2.1):

$$\begin{aligned}\boldsymbol{\Omega} &= d\boldsymbol{\Phi} + \tfrac{1}{2}[\boldsymbol{\Phi},\boldsymbol{\Phi}]\\ &= d(\hat{\boldsymbol{\Phi}} + \beta\pi^*(a)) + \tfrac{1}{2}[\hat{\boldsymbol{\Phi}} + \boldsymbol{\beta}\pi^*(a), \hat{\boldsymbol{\Phi}} + \beta\pi^*(a)]\\ &= \hat{\boldsymbol{\Omega}} + \boldsymbol{\beta}\pi^*(da) + (d\boldsymbol{\beta} + [\hat{\boldsymbol{\Phi}} + \boldsymbol{\beta}\pi^*(a), \boldsymbol{\beta}]) \wedge \pi^*(a).\end{aligned}$$

Since $\hat{\boldsymbol{\Phi}}$ is flat the curvature $\hat{\boldsymbol{\Omega}}$ of $\hat{\boldsymbol{\Phi}}$ is zero and thus

$$\boldsymbol{\Omega} = \boldsymbol{\beta}\pi^*(f) + D\boldsymbol{\beta} \wedge \pi^*(a), \tag{4.2}$$

where $D\boldsymbol{\beta} \equiv d\boldsymbol{\beta} + [\boldsymbol{\Phi},\boldsymbol{\beta}]$. In the local gauge s used above to define $\hat{\boldsymbol{\Phi}}$ the connection $\boldsymbol{\Phi}$ takes the form

$$\mathbf{A} = s^*\boldsymbol{\Phi} = \mathbf{b}a.$$

Thus

$$\begin{aligned}s^*(D\boldsymbol{\beta}) = D\mathbf{b} &= d\mathbf{b} + [\mathbf{A},\mathbf{b}]\\ &= d\mathbf{b} + a[\mathbf{b},\mathbf{b}]\\ &= d\mathbf{b}.\end{aligned}$$

Furthermore, as $\mathbf{b} = \mathbf{b}(u)$, $d\mathbf{b} = \mathbf{c}\, du$ for some Lie-algebra-valued function $\mathbf{c}$. Hence

$$D\mathbf{b} = \mathbf{c}\, du. \tag{4.3}$$

As $D\boldsymbol{\beta}$ is tensorial this equation implies that on Q_p there exists a Lie-algebra-valued function $\boldsymbol{\gamma}$, with $s^*\boldsymbol{\gamma} = \mathbf{c}$, such that

$$D\boldsymbol{\beta} = \boldsymbol{\gamma}\pi^*(du).$$

Hence $D\boldsymbol{\beta} \wedge \pi^*(du) = 0$, and, since $\pi^*(a) = \pi^*(H\, du)$, the last term in (4.2) above vanishes. The curvature $\boldsymbol{\Omega}$ reduces to the form

$$\boldsymbol{\Omega} = \boldsymbol{\beta}\omega, \tag{4.4}$$

where $\omega \equiv \pi^*(f)$. The curvature $\boldsymbol{\Omega}$ is thus a null decomposable curvature with twist-free rays. To verify that $\boldsymbol{\Omega}$ satisfies the Yang–Mills equation, we must show that $D^*\boldsymbol{\Omega} = 0$. From (4.4) we have

$$D^*\boldsymbol{\Omega} = D\boldsymbol{\beta} \wedge {}^*\omega + \boldsymbol{\beta} d^*\omega. \tag{4.5}$$

As shown above, $D\boldsymbol{\beta} \wedge \pi^*(du) = 0$. Since f is a null Maxwell field, ${}^*\omega$ will contain a factor $\pi^*(du)$ and thus the first term on the right-hand side of (4.5) will vanish. Moreover, $d^*\omega = 0$ since ${}^*\omega = \pi^*({}^*f)$ and $d^*f = 0$. The curvature $\boldsymbol{\Omega}$ thus satisfies the Yang–Mills equation. The theorem follows.

Remark: in the above construction the Lie-algebra-valued function $\mathbf{b}$ was choosen to be constant on each $u =$ const hypersurface, but otherwise $\mathbf{b}$ could be any smooth function $\mathbf{b} = \mathbf{b}(u)$. Such a $\mathbf{b}$ was shown to satisfy Eq. (4.3), namely $D\mathbf{b} = d\mathbf{b} = \mathbf{c}\, du$ in a local gauge. The Lie-algebra-valued function $\mathbf{c}$ is thus also arbitrary to the extent that it is defined by $d\mathbf{b} = \mathbf{c}\, du$. In fact, one could choose $\mathbf{c} = \mathbf{c}(u)$ arbitrarily and then determine $\mathbf{b}$ locally by integrating the equation $d\mathbf{b} = \mathbf{c}\, du$. On the bundle Q_p, $\boldsymbol{\beta}$ will satisfy $D\boldsymbol{\beta} = \boldsymbol{\gamma}\pi^*(du)$ where $\boldsymbol{\gamma}$ enjoys the same arbitrariness that $\mathbf{c}$ does on $\pi(Q_p)$. These facts will prove useful in interpreting Theorem 4.2 given below.

Now suppose that $\boldsymbol{\Omega}$ is a null decomposable curvature on a principal bundle $P(M,G)$ over spacetime. Let $\Sigma_{\boldsymbol{\Omega}}$ denote the set of points of P where $\boldsymbol{\Omega}$ vanishes and set $Q = P - \Sigma_{\boldsymbol{\Omega}}$. We shall show that if $\boldsymbol{\Omega}$ is a null–twist-free Yang–Mills curvature on Q, then for each $\xi \in Q$ there exists a bundle neighborhood Q_ξ of ξ on which the curvature $\boldsymbol{\Omega}$ is related to a Maxwell field in the manner described in the proof of the Theorem 4.1.

Theorem 4.2: Let $\boldsymbol{\Omega}$ be a null decomposable Yang–Mills curvature with twist-free rays on the principal bundle Q. For each $\xi \in Q$ there exists a bundle neighborhood Q_ξ of ξ such that $\boldsymbol{\Omega} = \boldsymbol{\beta}\pi^*(f)$ on Q_ξ, where f is a null–twist-free Maxwell field on $\pi(Q_\xi)$ and where $D\boldsymbol{\beta} = \boldsymbol{\gamma}\pi^*(du)$ for some Lie-algebra-valued function $\boldsymbol{\gamma}$.

Proof: Let $\xi \in Q$. As discussed in Sec. II, we may assume that there is an open bundle neighborhood Q'_ξ of ξ such that on Q'_ξ the curvature $\boldsymbol{\Omega}$ can be written as the product $\boldsymbol{\Omega} = \boldsymbol{\alpha}\omega$. In this case $\boldsymbol{\alpha}$ is a smooth nonvanishing Lie-algebra-valued Higgs field on Q'_ξ and $\omega = \pi^*(h)$ is the pullback under π of a real-valued 2-form on $\pi(Q'_\xi)$. We choose Q'_ξ small enough so that it is trivial.

The field equations $D\boldsymbol{\Omega}^+ = 0$ now imply

$$D\boldsymbol{\alpha} \wedge \omega^+ + \boldsymbol{\alpha}\, d\omega^+ = 0. \tag{4.6}$$

As in the proof of Lemma 3.1 an arbitrary positive definite inner product can be choosen in G' so that (4.6) can be rewritten in the form

$$d\omega^+ + \mu \wedge \omega^+ = 0, \tag{4.7}$$

$$D\boldsymbol{\alpha} = \boldsymbol{\alpha}\mu + \boldsymbol{\eta}\lambda. \tag{4.8}$$

Without loss of generality we may assume that $\boldsymbol{\eta}$ is a Lie-algebra-valued Higgs field and μ is a pullback 1-form on Q'_ξ. Again as in the proof of Lemma 3.1, we may use Eq. (4.8) together with the identity $D^2\boldsymbol{\alpha} = [\boldsymbol{\Omega},\boldsymbol{\alpha}]$ to show [see the calculations leading up to Eq. (3.4)]

$$\boldsymbol{\alpha}(d\mu \wedge \lambda) + \boldsymbol{\eta}(d\lambda \wedge \lambda) = 0.$$

This equation implies that $d\mu \wedge \lambda = 0$ since $d\lambda \wedge \lambda = 0$ and $\boldsymbol{\alpha} \neq 0$ on Q'_ξ.

As λ is twist-free, there is an open bundle neighborhood $Q''_\xi \subseteq Q'_\xi$ such that on Q''_ξ the 1-form λ can be rescaled so that $\lambda = dv$. The 1-form μ now satisfies the equation $d\mu \wedge dv = 0$ and thus satisfies the conditions of the lemma given in Appendix D. Hence there exist functions θ' and ψ'

such that

$$\mu = \psi' d\nu + d\theta' \tag{4.9}$$

on a possibly smaller bundle neighborhood $Q_\xi \subseteq Q''_\xi$. Now, although μ and $\lambda = d\mu$ are pullback 1-forms on Q_ξ, we cannot conclude from (4.9) that ψ' and θ' are invariant functions. If they are not, we may replace them by invariant functions in the following way. Choose an arbitrary local gauge s: $\pi(Q_\xi) \to Q_\xi$. From (4.9) we have

$$s^*\mu = (s^*\psi')(s^*d\nu) + s^*(d\theta').$$

Pulling this 1-form back to Q_ξ, using π^*, we have

$$\mu = \pi^*(s^*\psi')d\nu + \pi^*s^*(d\theta'), \tag{4.10}$$

where we have used the fact that μ and $d\nu$ are invariant. Define new invariant functions θ and ψ by

$$\psi = \pi^*(s^*\psi'), \quad \theta = \pi^*(s^*\theta').$$

Equation (4.10) may now be rewritten as

$$\mu = \psi d\nu + d\theta. \tag{4.11}$$

Returning now to the field equations, we substitute (4.11) into (4.7) and (4.8) to obtain

$$d\omega^+ + d\theta \wedge \omega^+ = 0, \tag{4.12}$$

$$D\boldsymbol{\alpha} = \boldsymbol{\alpha} d\theta + (\boldsymbol{\alpha}\psi + \boldsymbol{\eta})d\nu. \tag{4.13}$$

Now θ is invariant and we write it as $\pi^*(t)$ for some smooth function t on $\pi(Q_\xi)$. Using θ and $\omega^+ \equiv \pi^*(h^+)$, we define a new field f^+ by the equation

$$f^+ = e^t h^+.$$

Hence

$$e^\theta \omega^+ = \pi^*(f^+). \tag{4.14}$$

The exterior derivative of $e^\theta\omega^+$ is

$$d(e^\theta\omega^+) = e^\theta(d\theta \wedge \omega^+ + d\omega^+),$$

which vanishes by (4.12). The field f^+ is thus a null–twist-free Maxwell field on $\pi(Q_\xi)$. Define also a new Higgs field $\boldsymbol{\beta}$ by

$$\boldsymbol{\beta} = e^{-\theta}\boldsymbol{\alpha}. \tag{4.15}$$

The exterior covariant derivative of $\boldsymbol{\beta}$ is

$$D\boldsymbol{\beta} = e^{-\theta}(-\boldsymbol{\alpha} d\theta + D\boldsymbol{\alpha}).$$

Substituting (4.13) for $D\boldsymbol{\alpha}$ in this last equation, we obtain

$$D\boldsymbol{\beta} = \boldsymbol{\gamma}\pi^*(du),$$

where $\boldsymbol{\gamma} \equiv e^{-\theta}(\boldsymbol{\alpha}\psi + \boldsymbol{\eta})$ and $\pi^*(du) \equiv d\nu$.

Finally observe that from (4.14) and (4.15) we have

$$\boldsymbol{\beta}\pi^*(f^+) = e^{-\theta}\boldsymbol{\alpha} e^\theta\omega^+ = \boldsymbol{\alpha}\omega^+.$$

Hence on Q_ξ the curvature takes the form

$$\boldsymbol{\Omega} = \boldsymbol{\beta}\pi^*(f),$$

where $\boldsymbol{\beta}$ and f have the required properties. The theorem follows.

Remark: In Sec. III it was shown that if a decomposable Yang–Mills curvature is nonnull or null-twisting and if the bundle and curvature are sufficiently regular, then the holonomy subbundles $P(\xi)$ through each point $\xi \in P$ have one-dimensional fibers. This result followed from the Ambrose–Singer theorem and rested on the fact that the Higgs field $\boldsymbol{\beta}$ in $\boldsymbol{\Omega} = \boldsymbol{\beta}\omega$ satisfied $D\boldsymbol{\beta} = \boldsymbol{\beta}\nu$. This latter equation was used to show that, along horizontal curves in P, $\boldsymbol{\beta}$ maps into a single Lie-algebra direction. In the null–twist-free case we have shown in Theorem 4.2 that $\boldsymbol{\beta}$ satisfies a different equation, namely

$$D\boldsymbol{\beta} = \boldsymbol{\gamma}\pi^*(du) \tag{4.16}$$

for some Lie-algebra-valued Higgs field $\boldsymbol{\gamma}$. Let $x(t)$ be a curve lying entirely within a "$u =$ const" submanifold $N \subseteq \pi(Q_\xi)$. For each t the tangent $\dot{x}(t)$ will satisfy $du(\dot{x}(t)) = 0$. Let $\tilde{x}(t)$ be a horizontal lift of $x(t)$ to a curve in Q_ξ. Then from (4.16) we have

$$\begin{aligned} D\boldsymbol{\beta}(\dot{\tilde{x}}(t)) &= \boldsymbol{\gamma}(\tilde{x}(t))\pi^*(du)(\dot{\tilde{x}}(t)) \\ &= \boldsymbol{\gamma}(\tilde{x}(t))du(d\pi(\dot{\tilde{x}}(t))) \\ &= \boldsymbol{\gamma}(\tilde{x}(t))du(\dot{x}(t)) \\ &= 0. \end{aligned}$$

Thus $\boldsymbol{\beta}$ is constant on horizontal curves over each "$u =$ const" submanifold N. However, suppose that $y(t)$ is a curve transverse to the $u =$ const submanifolds so that for each t the tangent $\dot{y}(t)$ satisfies $du(\dot{y}(t)) \neq 0$. Denote the horizontal lift of $y(t)$ to a point of Q_ξ by $\tilde{y}(t)$. Then from (4.16) we have

$$\begin{aligned} D\boldsymbol{\beta}(\dot{\tilde{y}}(t)) &= \boldsymbol{\gamma}(\tilde{y}(t))\pi^*(du)(\dot{\tilde{y}}(t)) \\ &= \boldsymbol{\gamma}(\tilde{y}(t))du(\dot{y}(t)), \end{aligned} \tag{4.17}$$

which does not vanish in general. From the remarks following Theorem 4.1 we know that $\boldsymbol{\gamma}$ can be almost any Lie-algebra-valued Higgs field. Equation (4.17) now shows that a generic $\boldsymbol{\beta}$ will trace out a general curve in the Lie algebra G' as we follow a horizontal curve that projects to a curve transverse to the "$u =$ const" submanifolds in M. Now $\boldsymbol{\Omega} = \boldsymbol{\beta}\omega$ and the Ambrose–Singer theorem states that the Lie algebra of the holonomy group of a point $\xi \in P$ is generated by $\boldsymbol{\Omega}_\eta(\Delta,\psi)$ at all points η connected to ξ by horizontal curves. One would expect, then, that in the generic null–twist-free decomposable case the holonomy group of the connection would be the full gauge group G. In this case the holonomy bundle through any point $\xi \in P$ would coincide with P.

Suppose, however, that the Lie-algebra-valued Higgs field $\boldsymbol{\gamma}$ in (4.16) above is proportioanl to $\boldsymbol{\beta}$, that is, $\boldsymbol{\gamma} = \sigma\boldsymbol{\beta}$ for some smooth function σ. By using a replica of the proof of Theorem 3.2, one can show in this case that the holonomy subbundles $Q_\xi(\eta)$ for each $\eta \in Q_\xi$ have one-dimensional fibers. Since $\boldsymbol{\Omega} = \boldsymbol{\beta}\pi^*(f)$ on all of Q_ξ and since f is a Maxwell field on M, these holonomy subbundles are Maxwell bundles. We state this result in the following theorem as a local converse to Theorem 3.4 in the null–twist-free case.

Theorem 4.3: If the Lie-algebra-valued Higgs field $\boldsymbol{\gamma}$ defined in Theorem 4.2 is proportional to $\boldsymbol{\beta}$, then the holonomy subbundles $Q_\xi(\eta)$ for each $\eta \in Q_\xi$ are Maxwell bundles.

Returning now to the general null–twist-free case, we show that the induced bundle $P(N,G)$ over each "$u =$ const" submanifold N is a bundle with a flat connection.

Let $Q(U,G)$ be a bundle neighborhood of the type defined in the statement of Theorem 4.2. Let $I \subseteq \mathbb{R}$ be such that for each $t \in I$ there is a three-dimensional submanifold N_t

obtained as the intersection of U with the hypersurface $u = t$. By rechoosing U and u if necessary we may assume that each N_t is connected. Denote the inclusion map by $i_t: N_t \to U$, and for each $t \in I$ let $P(N_t, G)$ denote the bundle over N_t induced[3] by i_t.

Theorem 4.4: For each $t \in I$ the induced connection 1-form Φ_t' on the bundle $P(N_t, G)$ is flat. Consequently, $Q(U,G)$ is partitioned into the disjoint union of flat $(3+n)$-dimensional bundles, where $n = \dim(G)$.

Proof: Recall[3] that for each t the induced bundle $P(N_t, G)$ is the subset of $N_t \times Q$ such that, for $(y,\xi) \in N_t \times Q$, $i_t(y) = \pi(\xi)$. There is a natural homomorphism $j_t: P(N_t, G) \to Q(U,G)$ given by $j_t(y,\xi) = \xi$, and the corresponding group homomorphism $G \to G$ is the identity automorphism of G. Given this structure, one can show[3] that there is a unique connection in $P(N_t, G)$ whose horizontal subspaces are mapped into the horizontal subspaces of the given connection in $Q(U,G)$. Moreover, the curvature 2-form $\mathbf{\Omega}_t'$ of this connection in $P(N_t,G)$ is related to the curvature 2-form $\mathbf{\Omega} = \boldsymbol{\beta}\pi^*(f)$ in $Q(U,G)$ by

$$\mathbf{\Omega}_t' = j_t^* \mathbf{\Omega}.$$

We show that $\mathbf{\Omega}_t'$ vanishes for each t. Let Γ and Δ be two vectors tangent to $P(N_t,G)$ at $(y,\xi) \in P(N_t,G)$. Then

$$\mathbf{\Omega}_t'(\Gamma,\Delta) = \mathbf{\Omega}(dj_t\Gamma, dj_t\Delta).$$

A vector Γ tangent to $P(N_t,G)$ at (y,ξ) is of the form $(X,\theta) \in T_yN_t \times T_\xi Q$, where $di_t(X) = d\pi(\theta)$. Hence $dj_t\Gamma = \theta$. Similarly, writing $\Delta = (Y,\psi) \in T_yN_t \times T_\xi Q$ with $di_t(Y) = d\pi(\psi)$, we have $dj_t(\Delta) = \psi$. Hence

$$\begin{aligned}\mathbf{\Omega}_t'(\Gamma,\Delta) &= \mathbf{\Omega}(dj_t\Gamma, dj_t\Delta)\\ &= \mathbf{\Omega}(\theta,\psi)\\ &= \boldsymbol{\beta}(\xi)(\pi^*f)(\theta,\psi)\\ &= \boldsymbol{\beta}(\xi)f(d\pi\theta, d\pi\psi)\\ &= \boldsymbol{\beta}(\xi)f(di_tX, di_tY).\end{aligned} \tag{4.18}$$

Now di_tX and di_tY are both tangent to N_t, and hence $du(di_tX) = du(di_tY) = 0$. Since $f \wedge du = 0$, we see that $f(di_tX, di_tY) = 0$ and thus $\mathbf{\Omega}_t'(\Gamma,\Delta) = 0$ by Eq. (4.18). As Γ and Δ were arbitrary vectors, it follows that $\mathbf{\Omega}_t' = 0$. The connection 1-form Φ_t' of $\mathbf{\Omega}_t'$ is thus a flat connection.

Finally, observe that $i_t(N_t) \cap i_{t'}(N_{t'}) = \emptyset$ for $t \neq t'$, and $\pi(Q) = \cup_{t\in I} i_t(N_t)$. Identifying $P(N_t,G)$ with $j_t(P(N_t,G))$ for each t, we see that Q is partitioned into the disjoint union of flat $(3+n)$-dimensional bundles, where $n = \dim(G)$ and $3 = \dim(N_t)$ for each $t \in I$. The theorem follows.

V. COMPLEX-DECOMPOSABLE SELF-DUAL CURVATURES

In our analysis so far we have been concerned with the decomposability of the curvature 2-form $\mathbf{\Omega}$. The concept of a decomposable curvature can be generalized to the case where only the self-dual curvature 2-form $\mathbf{\Omega}^+$ is assumed decomposable.

We say that a self-dual curvature $\mathbf{\Omega}^+$ *is complex-decomposable* if, at each point ξ of P, there is a neighborhood Υ of ξ such that on Υ $\mathbf{\Omega}^+ = \boldsymbol{\alpha}\omega^+$, where $\boldsymbol{\alpha} = \boldsymbol{\beta} + i\boldsymbol{\gamma}$ for smooth Lie-algebra-valued functions $\boldsymbol{\beta}$ and $\boldsymbol{\gamma}$ and where ω^+ is a smooth complex-valued 2-form on Υ. Observe that since ${}^*\mathbf{\Omega}^+ = i\mathbf{\Omega}^+$ it follows that ${}^*\omega^+ = i\omega^+$. In this case we can define $\omega = \frac{1}{2}(\omega^+ + \overline{\omega}^+) = \operatorname{Re}(\omega^+)$ and thus $\omega^+ = \omega - i^*\omega$. It follows that

$$\mathbf{\Omega} = \boldsymbol{\beta}\omega + \boldsymbol{\gamma}\,{}^*\omega \tag{5.1}$$

and

$${}^*\mathbf{\Omega} = -\boldsymbol{\gamma}\omega + \boldsymbol{\beta}\,{}^*\omega.$$

If $\boldsymbol{\beta}$ and $\boldsymbol{\gamma}$ are linearly dependent, then $\mathbf{\Omega}$ itself is real decomposable, and the analysis of the previous sections applies. Here we are particularly interested in curvatures for which $\boldsymbol{\beta}$ and $\boldsymbol{\gamma}$ are linearly independent.

It is not hard to show that if $\mathbf{\Omega}^+$ is complex-decomposable, then at each point $\xi \in P$ there is a factorization $\mathbf{\Omega}^+ = \boldsymbol{\alpha}\omega^+$, where $\boldsymbol{\alpha}$ is a Higgs field[11] and ω^+ is a complex-valued invariant horizontal 2-form on some bundle neighborhood of ξ. This follows as in the Observation of Sec. II.

A self-dual complex decomposable Yang–Mills curvature has many properties that are analogous to the properties obtained in the preceding sections in the decomposable case. As before, one may analyze the structure of self-dual complex-decomposable curvatures by considering the nonnull, null-twisting, and null–twist-free cases. In fact, the anslysis is so similar that we are contented with a statement of the main results along with some comments regarding appropriate modifications of the proofs of the results of Sec. III and IV.

The most important difference occurs when one attempts to prove a result analogous to Lemma 3.1. This result should be modified to read as follows:

Lemma 5.1: Let $\mathbf{\Omega}^+$ be a self-dual Yang–Mills curvature which is not null–twist-free on any open subset of some bundle neighborhood Υ of P. If $\mathbf{\Omega}^+ = \boldsymbol{\alpha}\omega^+$ is complex-decomposable on Υ with $\boldsymbol{\alpha} = \boldsymbol{\beta} + i\boldsymbol{\gamma}$ and $\omega^+ = \pi^*(f^+)$, then

$$D\boldsymbol{\alpha} = \boldsymbol{\alpha}\nu$$

on $\Upsilon - \Sigma_{\mathbf{\Omega}^+}$ and

$$[\boldsymbol{\beta},\boldsymbol{\gamma}] = 0.$$

If, on the other hand, $\mathbf{\Omega}^+ = \boldsymbol{\alpha}\omega^+$ is null–twist-free at each point of Υ then on $\Upsilon - \Sigma_{\mathbf{\Omega}^+}$ the Higgs field $\boldsymbol{\alpha}$ satisfies

$$D\boldsymbol{\alpha} = \boldsymbol{\alpha}\mu + \boldsymbol{\eta}\lambda$$

for some Higgs field $\boldsymbol{\eta}$ and some pullback complex-valued 1-form μ such that

$$d\mu \wedge \lambda = 0.$$

Here λ is the null 1-form of $\mathbf{\Omega}$.

The main difference in the proof of the first part of this Lemma and its counterpart in Sec. III occurs in the calculation of $D^2\boldsymbol{\alpha}$ [see the calculations just prior to Eq. (3.4)]. For the present case

$$\begin{aligned}D^2\boldsymbol{\alpha} = [\mathbf{\Omega},\boldsymbol{\alpha}] &= [\boldsymbol{\beta}\omega + \boldsymbol{\gamma}\,{}^*\omega, \boldsymbol{\beta} + i\boldsymbol{\gamma}]\\ &= [\boldsymbol{\gamma},\boldsymbol{\beta}]\,{}^*\omega + [\boldsymbol{\beta},\boldsymbol{\gamma}]i\omega\\ &= i[\boldsymbol{\beta},\boldsymbol{\gamma}]\omega^+.\end{aligned} \tag{5.2}$$

On the other hand, the field equations $D\mathbf{\Omega}^+ = 0$ imply, as in Sec. III, that on $\Upsilon - \Sigma_{\mathbf{\Omega}^+}$

$$d\omega^+ = -\mu \wedge \omega^+,$$
$$D\alpha = \alpha\mu + \eta\lambda$$

for some 1-form μ. Moreover, as in Sec. III it can be shown that it is no loss of generality to assume that η is a Higgs field and μ is a pullback form. Note that the term $\eta\lambda$ does not appear when Ω is nonnull. Computing $D^2\alpha$ from the last equation and setting the result equal to the right-hand side of (5.2), we obtain

$$i[\beta,\gamma]\omega^+ = \alpha\, d\mu + (D\eta \wedge \lambda) + \eta((\lambda \wedge \mu) + d\lambda). \quad (5.3)$$

In the null cases one wedges this equation with the null 1-form λ to obtain:

$$0 = \alpha(d\mu \wedge \lambda) + \eta(d\lambda \wedge \lambda). \quad (5.4)$$

If λ is twist-free, then $d\lambda \wedge \lambda = 0$, and Eq. (5.4) implies that $d\mu \wedge \lambda = 0$. The assertions about the twist-free case now follow.

We claim that in both the nonnull and the null-twisting cases (5.3) reduces to

$$i[\beta,\gamma]\omega^+ = \alpha\, d\nu. \quad (5.5)$$

for an appropriate 1-form ν. The latter statement is obvious in case Ω is nonnull since $\eta = 0$ for this case and (5.5) is an obvious consequence of (5.3) if we take $\nu = \mu$. On the other hand, if Ω is null and λ is twisting, then (5.4) implies that η is a multiple of α and consequently $D\alpha = \alpha\nu$ for an appropriate 1-form ν. Thus $D^2\alpha = \alpha\, d\nu + D\alpha \wedge \nu = \alpha\, d\nu + \alpha(\nu \wedge \nu) = \alpha\, d\nu$. This result along with (5.2) implies equation (5.5)

We now utilize Eq. (5.5) to show that $[\beta,\gamma] = 0$. Suppose $[\beta,\gamma] \neq 0$. Then $d\nu \neq 0$ and thus there exist vector fields Γ,Δ such that $d\nu(\Gamma,\Delta) \neq 0$. Transvecting (5.5) with these fields shows that $\alpha = \sigma[\beta,\gamma]$ for some nonvanishing complex-valued function σ. But this equation implies that the real and imaginary parts of α are linearly dependent, i.e., $\beta = \rho\gamma$ for some real-valued function ρ. Hence $[\beta,\gamma] = [\rho\gamma,\gamma] = 0$ contrary to assumption. Thus $[\beta,\gamma] = 0$. The remainder of the proof follows as in the proof of Lemma 3.1.

Remark: We can now pretty well follow the proofs of Sec. III to obtain results analogous to those of that section. For example, if Ω is not null–twist-free on any open subset of $P - \Sigma$ and γ is a horizontal curve from some point ξ in a connected component Q of $P - \Sigma$, then $[\alpha \circ \gamma]$ is constant for every α which arises via a decomposition of Ω^+: $\Omega^+ = \alpha\omega^+$. It follows, as before, that on the holonomy bundle $Q(\xi)$ of Q through ξ one has $\Omega^+ = \alpha_0\omega^+$ for some *constant* α_0 in the complexified Lie algebra $G' \oplus iG'$. If $\alpha_0 = \beta_0 + i\gamma_0$, then it follows from the Ambrose–Singer theorem that the fibers of the holonomy bundle are either one or two dimensional depending on whether or not β_0 and γ_0 are dependent.

We have:

Theorem 5.2: Assume that Ω^+ is a self-dual complex-decomposable Yang–Mills curvature which is not null–twist-free on any open subset of $P - \Sigma$. Let Q be a connected component of $P - \Sigma$ and $\xi \in Q$. Finally, let $\Omega^+ = \alpha\omega^+$ with $\alpha = \beta + i\gamma$ in a neighborhood of ξ:

(1) If $\beta(\xi)$ and $\gamma(\xi)$ are linearly dependent then the holonomy bundle $Q(\xi)$ of Q through ξ has one real fiber dimension. If, on the other hand, $\beta(\xi)$ and $\gamma(\xi)$ are independent, then $Q(\xi)$ has two real fiber dimensions. Moreover, in both cases, if $\alpha_0 = \beta_0 + i\gamma_0$ with $\beta_0 = \beta(\xi)$ and $\gamma_0 = \gamma(\xi)$, then on $Q(\xi)$ we have $\Omega^+ = \alpha_0\bar{\omega}^+$ for some pull back form $\bar{\omega}^+$.

(2) If Ω satisfies the Hypothesis of Sec. III, then the holonomy bundle $P(\xi)$ of P through ξ either has one fiber dimension over each point of M or it has two fiber dimensions over each point of M.

By a similar modification of the results of Sec. IV we have:

Theorem 5.3: Let Ω^+ be a self-dual complex-decomposable Yang–Mills curvature which is null–twist-free at each point of $P - \Sigma$.

(1) For each $\xi \in P - \Sigma$ there exists a bundle neighborhood Q_ξ of ξ such that on Q_ξ

$$\Omega^+ = \alpha\pi^*(f^+),$$

where f^+ is a null twist-free self-dual Maxwell field on $\pi(Q_\xi)$ and where

$$D\alpha = \rho\pi^*(du)$$

for some Higgs field ρ.

(2) In case ρ is a multiple of α on the neighborhood Q_ξ the Lie-algebra function $\alpha = \beta + i\gamma$ is constant on Q_ξ and, for each $\eta \in Q_\xi$, the fiber dimension of the holonomy subbundle $Q(\eta)$ of Q_ξ is precisely the dimension of the Lie algebra generated by β and γ (generally, $[\beta,\gamma] \neq 0$).

Finally we obtain the local gauge expressions for the potentials of a self-dual complex-decomposable Yang–Mills curvature referred to in the Introduction. In the following we will only be concerned with a connected component Q of $P - \Sigma$ since the components of the curvature will be zero in every local gauge lying in the flat set Σ.

Let Ω^+ be a self-dual complex-decomposable Yang–Mills curvature which is not null–twist-free on any open subset of $P - \Sigma$. For each $\xi_0 \in Q$ let $Q(\xi_0)$ be the holonomy bundle through ξ_0 and let $\Omega^+ = \alpha_0\omega^+$ be the decomposition of Ω^+ on Q referred to in Theorem 5.2 (1). By the same theorem we know that the fiber dimension of Q is either 1 or 2. For brevity we consider just the second case so that if $\alpha_0 = \beta_0 + i\gamma_0$, then β_0 and γ_0 are independent. By Eq. (5.1) the curvature Ω on Q is thus

$$\Omega = \beta_0\omega + \gamma_0{}^*\omega. \quad (5.6)$$

The connection Φ on $Q(\xi_0)$ may also be written in terms of β_0 and γ_0 as

$$\Phi = \beta_0\phi + \gamma_0\psi \quad (5.7)$$

for some real 1-forms ϕ and ψ.

Now by Lemma 5.1 and Theorem 5.2 the holonomy group of $Q(\xi_0)$ is commutative. Calculating the curvature using Eqs. (2.1) and (5.7) and equating the result to (5.6) shows that

$$\omega = d\phi, \quad {}^*\omega = d\psi. \quad (5.8)$$

The 2-form $\omega = \pi^*(f)$ is a pullback form on $Q(\xi_0)$, and by (5.8) f is a Maxwell field on $\pi(Q(\xi_0))$. Locally we may therefore write f and its dual in terms of potentials as

$$f = d\mathscr{A}, \quad {}^*f = d\mathscr{B}.$$

The 1-forms ϕ and ψ of the connection Φ may hence be written as

$$\phi = \pi^*(\mathscr{A}) + d\chi_1$$

and

$$\psi = \pi^*(\mathscr{B}) + d\chi_2.$$

Substituting these expressions into (5.7) yields

$$\mathbf{\Phi} = \{\beta_0\pi^*(\mathscr{A}) + \gamma_0\pi^*(\mathscr{B})\} + \{\beta_0 d\chi_1 + \gamma_0 d\chi_2\}.$$

Since the connection $\mathbf{\Phi}$ sends the fundamental vertical vector field δ^* corresponding to an arbitrary $\boldsymbol{\delta}\in G'$ to $\boldsymbol{\delta}$, it follows that a submanifold given by $\chi_1 = k_1$ and $\chi_2 = k_2$ can have no vertical tangent vector. Since such a submanifold is transverse to the fibers, it can be used to define a local gauge s: $U \rightarrow Q(\xi_0)$. In this local gauge the potential $\mathbf{A} = s^*\mathbf{\Phi}$ is equal to

$$\mathbf{A} = \mathbf{b}_0\mathscr{A} + \mathbf{c}_0\mathscr{B}, \tag{5.9}$$

where $\mathbf{b}_0$ and $\mathbf{c}_0$ are constant Lie algebra vectors.

VI. DISCUSSION

We have considered the source-free Yang–Mills solutions with decomposable self-dual curvature. These are the solutions of type D and type N in the Anandan–Tod classification. Any such solution is *locally* intimately related to a Maxwell solution. Indeed, there always exists a local guage in which the Yang–Mills potential takes the form of a Lie-algebra-valued function multiplied by a Maxwell potential. Because of this relationship, one can use his knowledge of local properties of Maxwell solutions to infer properties of these special Yang–Mills solutions. For example, these special Yang–Mills fields propagate at the speed of light.

Globally, on the other hand, these algebraically special Yang–Mills solutions are not so simply related to global Maxwell solutions. That is to say, even if null–twist-free regions are disallowed, it is not true that a Yang–Mills solution with decomposable curvature is necessarily globally reducible to a Maxwell bundle. Global reducibility requires more than just decomposable curvature. It is necessary that the holonomy groups be one-dimensional, and, as stated in Theorem 3.4, this will be the case provided the conditions of our Hypothesis are met.

A part of the Hypothesis is the asusmption that the flat set does not divide the region of curvature into disconnected sets. If the spacetime is Minkowski space, this assumption is superfluous since it is not possible to have two Maxwell fields in Minkowski space with nonoverlapping supports. In more general spacetimes, however, the assumption is not trivial. Consider, for example, a spacelike hyperplane in Minkowski space and Cauchy data for two Maxwell fields such that the two Cauchy data sets have non overlapping supports on the hyperplane. By deleting suitable regions from the Minkowski space, the two developed Maxwell solutions may have nonoverlapping supports on the resulting spacetime. We might say, therefore, that the Hypothesis is reasonable for special relativistic physics, but in the context of general relativity one should not expect global reducibility of Yang–Mills solutions with decomposable curvature.

Remark: If a global Yang–Mills solution with decomposable curvature is *analytic* and not null–twist-free, then it necessarily reduces to a Maxwell bundle. (A Yang–Mills solution is analytic if the bundle and connection are real analytic.) Analyticity forces the restricted holonomy group to agree with infinitesimal holonomy group,[3] which, for these solutions, is one-dimensional. Thus the holonomy group itself is one-dimensional, and consequently the reduced holonomy bundle is a Maxwell bundle. Alternatively, one can see that a global analytic solution can have no flat set (unless it is everywhere flat), and so the conditions of the Hypothesis are necessarily met.

Remark: Yasskin[12] showed that a local Einstein–Maxwell solution can be used to generate local Einstein–Yang–Mills solutions with decomposable curvature. The method is to multiply the Maxwell potential by a constant Lie algebra element to produce the local gauge components of a Yang–Mills connection. Provided the Lie algebra element has unit length with respect to the Killing metric, the stress–energy tensor of the Yang–Mills field will be identical to that of the Maxwell field. Our analysis shows that, except for null–twist-free curvatures, *any* local Yang–Mills solution with decomposable curvature arises from a Maxwell field as in the Yasskin construction. In any Einstein–Yang–Mills solution with decomposable curvature which is not null–twist-free; therefore, the local spacetime geometry agrees with that of some Einstein–Maxwell spacetime. This result needs some qualification if the Killing metric is not positive definite. For then it is possible to have Einstein–Yang–Mills solutions of this type which do not confirm to the positive energy conditions which Maxwell solutions satisfy.

Remark: Einstein–Yang–Mills solutions with null–twist-free curvature are excluded from the previous remark because there is additional freedom in generating Yang–Mills solutions from Maxwell solutions when the Maxwell fields are null–twist-free. Examples of null–twist-free decomposable Einstein–Yang–Mills solutions which are not of the Yasskin type can be found in the work of Güven.[13]

Remark: From the close relationship between algebraically special Maxwell solutions and Yang–Mills solutions of type N, we infer a simple extension of Robinson's theorem[14]: If a spacetime admits a type-N Yang–Mills solution, then it has a shear-free geodetic null congruence of curves. Conversely, given a shear-free geodetic null congruence, one may construct Yang–Mills solutions of type N.

ACKNOWLEDGMENTS

The authors thank M. Mostow and S. Schecter for the example presented in Appendix A.

APPENDIX A

The notion of a "decomposable curvature" is based primarily on an algebraic criterion: At each point it must be possible to factor the curvature as a product of a Lie algebra vector and a real 2-form. In addition, however, we have required that in a neighborhood of each point it be possible to choose the factors so that each factor is C^∞. We present an example, due to Mostow and Schecter,[15] of a connection whose curvature is everywhere C^∞ and which can be factored at each point, but at some point ξ_0 it is impossible to choose factors of the curvature which are C^∞ throughout a

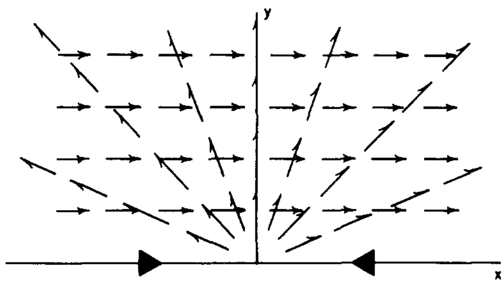


FIG. 1. Here a cone is represented by a half-plane with the positive and negative x axes identified. The ordinary arrows represent a locally constant electric field, while the single-barbed arrows indicate a value in the Lie algebra so (3). In this figure these fields are represented in a neighborhood of the y axis.

neighborhood of ξ_0.

Let the gauge group be an abelian 2-dimensional group with Lie algebra basis $\{\mathbf{e}_1,\mathbf{e}_2\}$. Let the spacetime M be Minkowski space with global coordinates (x,y,z,t). The bundle P is a trivial bundle, and we express the connection, curvature, etc., in terms of a fixed global gauge of the bundle. Thus let the potential 1-form (connection) be given by $\mathbf{A} = A^i\mathbf{e}_i$, where

$$A^1 = -xy\ dx + k(y)\ dz,$$

$$A^2 = xh(-y)\ dy.$$

Here k is any antiderivative of h and h is any C^∞ function from $\mathbb{R}$ to $\mathbb{R}$ which is zero at each negative number and nonzero at each positive number. The essential property is that $h(y)h(-y) = 0$ for every $y \in \mathbb{R}$. The curvature $\mathbf{F} = F^i\mathbf{e}_i$ of $\mathbf{A}$ has components

$$F^1 = x(dx\wedge dy) + h(y)(dy\wedge dz),$$

$$F^2 = h(-y)(dx\wedge dy).$$

The curvature defined by $\mathbf{F}$ does not admit smooth factors on any neighborhood of any point on the fiber over the origin of M. On the other hand, the curvature can be factored at each point since $F^i = b^i f$, where:

(1) for $x\neq 0$, $b^1 = 1$, $b^2 = (1/x)h(-y)$, $f = F^1$;

(2) for $x = 0$ and $y\geqslant 0$, $b^1 = 1$, $b^2 = 0$, $f = h(y)(dy\wedge dz)$;

(3) for $x = 0$ and $y\leqslant 0$, $b^1 = 0$, $b^2 = 1$,

$f = h(-y)(dx\wedge dy)$.

APPENDIX B

We mentioned in the text that a Yang–Mills solution could have decomposable curvature everywhere and yet not admit *global* forms $\boldsymbol{\beta}$ and ω such that $\boldsymbol{\Omega} = \boldsymbol{\beta}\omega$ everywhere. We shall illustrate this with a simple example.

This Yang–Mills solution is a connection in a trivial principal bundle with sructure group SO (3). The spacetime is flat, with topology $S^1\times R^3$. Let us regard it as a cone translated in one space dimension and one time dimension. The Yang–Mills solution is also invariant under those translations, so the solution will be described only on the two-dimensional cone. Since the bundle is trivial, we may specify the solution by describing it on one (global gauge) cross section.

Figure 1 represents the cone by a half-plane with the negative x axis identified with the positive x axis as indicated by the solid arrowheads. On this cone is a (locally) constant electric field, shown with ordinary arrows. The local electric field defines a local Maxwell 2-form f. We emphasize that the field is local because, as drawn in Fig. 1, it is discontinuous at points of the x axis where it has two limits differing by a factor of -1. Also shown in Fig. 1 is a field of radially pointing (single-barbed) arrows. These are intended to indicate a local Lie-algebra-valued function $\mathbf{b}$. At each point on the cone is a single-barbed arrow which represents a value in the Lie algebra so (3). In the gauge we are using, these values all fall in a single plane of so (3), and by identifying points of that plane with points of the (cone's) xy plane we have indicated a $\mathbf{b}$ field. Note that, since the positive and negative x axes in so (3) space are *not* identified, the $\mathbf{b}$ field is also discontinuous on the cone's seam (the x axis) because it too has two limits differing by a factor of -1. The product $\mathbf{F} = \mathbf{b}f$, on the other hand, is smooth, having an unambiguous limit on the seam. Figure 2 shows a neighborhood of the seam, with a smooth $\hat{\mathbf{b}}$ and a smooth $\hat{f}$ yielding the same smooth product $\mathbf{F} = \hat{\mathbf{b}}\hat{f}$. Note that the single-barbed arrows representing the value of $\hat{\mathbf{b}}$ at identified points of the x axis define the same vector in so (3).

This curvature 2-form arises from a Yang–Mills solution, as can be seen by considering a neighborhood of any point. By a suitable point-dependent SO (3) rotation (to a different gauge) the barbed arrows can be made locally constant, say $\mathbf{b}'$. If a is a local potential for the electric field, so that $f = da$, then $\mathbf{A} := \mathbf{b}'a$ satisfies the Yang–Mills equation (2.5).

On the global cross section we employed in Figs. 1 and 2, there is no global smooth $\mathbf{b}$ or global smooth f such that $\mathbf{F} = \mathbf{b}f$. Since a global decomposition is not possible on that section, it is certainly not possible on the entire SO (3) bundle.

APPENDIX C

The example presented in Appendix B suggests the following question: When does a decomposable Yang–Mills

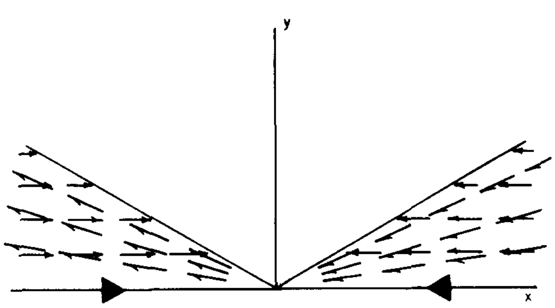


FIG. 2. Here a cone is represented by a half-plane with the positive and negative x axes identified. The ordinary arrows represent a locally constant electric field, while the single-barbed arrows indicate a value in the Lie algebra so (3). In this figure these fields are represented in a neighborhood of the x axis.

curvature admit a global decomposition? A partial answer to this question is given by the following:

Theorem: Let $\mathbf{\Omega}$ be a decomposable Yang–Mills curvature which is not null–twist-free on any open subset of P. If $H^2(M-\pi(\Sigma), \mathbb{R}^*)=0$, then there is a nonvanishing Higgs field $\boldsymbol{\beta}$ and a pullback form ω on $P-\Sigma$ such that $\mathbf{\Omega}|(P-\Sigma)=\boldsymbol{\beta}\omega$.

Remark: In the proof below we utilize Čech cohomology since in that case it is easy to describe the cocycles and coboundaries as follows. If $\mathscr{U}$ is an open cover of $M-\pi(\Sigma)$ which has the property that for $U, V\in\mathscr{U}$ either $U\cap V=\emptyset$ or $U\cap V$ is not empty and contractible, then a 2-cocycle is a function which associates with each pair (U, V) such that $U\cap V\neq\emptyset$ an element c_{UV} of the group $\mathbb{R}^*$. The function $c=\{c_{UV}\}$ must satisfy the identity

$$c_{UW}=c_{UV}c_{VW}$$

whenever $U\cap V\cap W\neq\emptyset$. A 2-coboundary is a function b from $\mathscr{U}$ to $\mathbb{R}^*$. To say that $H^2(M-\pi(\Sigma), \mathbb{R}^*)=0$ means every 2-cocycle can be obtained from a 2-coboundary as follows: Given c there is a b such that $c_{UV}=b_U b_V^{-1}$ whenever $U\cap V\neq\emptyset$. A reference where this is all spelled out in some detail is Ref. 16.

Proof of the Theorem: First choose an open cover $\{U_\alpha\}$ of $M-\pi(\Sigma)$ such that

(1) for (α,γ) such that $U_\alpha\cap U_\gamma\neq\emptyset$, $U_\alpha\cap U_\gamma$ is contractible,

(2) if $\Upsilon_\alpha=\pi^{-1}(U_\alpha)$, then $\mathbf{\Omega}|\Upsilon_\alpha=\boldsymbol{\beta}_\alpha\omega_\alpha$ for some nonvanishing Higgs field $\boldsymbol{\beta}_\alpha$ and some pullback form ω_α on Υ_α. We may also assume, by Lemma 3.1, that $D\boldsymbol{\beta}_\alpha=\boldsymbol{\beta}_\alpha\mu_\alpha$ for some 1-form μ_α on Υ_α.

If $U_\alpha\cap U_\gamma\neq 0$ let $g_{\alpha\gamma}=(\boldsymbol{\beta}_\alpha\cdot\boldsymbol{\beta}_\gamma)/\boldsymbol{\beta}_\alpha^{\,2}$, where $\cdot$ is an arbitrary positive-definite inner product on G'. We note that $\mathbf{\Omega}=\boldsymbol{\beta}_\alpha\omega_\alpha=\boldsymbol{\beta}_\gamma\omega_\gamma$ implies $\omega_\alpha=g_{\alpha\gamma}\omega_\gamma$ and thus that $\boldsymbol{\beta}_\gamma=g_{\alpha\gamma}\boldsymbol{\beta}_\alpha$. It follows that $g_{\alpha\gamma}$ is an invariant function on $\Upsilon_\alpha\cap\Upsilon_\gamma$, which is never zero. Now

$$\begin{aligned}D\boldsymbol{\beta}_\gamma&=(Dg_{\alpha\gamma})\boldsymbol{\beta}_\alpha+g_{\alpha\gamma}(D\boldsymbol{\beta}_\alpha)\\&=(Dg_{\alpha\gamma})(g_{\alpha\gamma}^{-1}\boldsymbol{\beta}_\gamma)+g_{\alpha\gamma}(\boldsymbol{\beta}_\alpha\mu_\alpha)\\&=[(Dg_{\alpha\gamma})g_{\alpha\gamma}^{-1}+\mu_\alpha]\boldsymbol{\beta}_\gamma.\end{aligned}$$

Since $D\boldsymbol{\beta}_\gamma=\boldsymbol{\beta}_\gamma\mu_\gamma$, we have that

$$\mu_\gamma\boldsymbol{\beta}_\gamma=[(Dg_{\alpha\gamma})g_{\alpha\gamma}^{-1}+\mu_\alpha]\boldsymbol{\beta}_\gamma.$$

Dotting both sides with $\boldsymbol{\beta}_\gamma$ and cancelling $\boldsymbol{\beta}_\gamma^{\,2}$ yields

$$dg_{\alpha\gamma}/g_{\alpha\gamma}=\mu_\gamma-\mu_\alpha. \tag{C1}$$

Since $U_\alpha\cap U_\gamma$ is connected and simply connected, we may choose an arbitrary point x_0 in $U_\alpha\cap U_\gamma$ and, for each $x\in U_\alpha\cap U_\gamma$, an arbitrary curve $c(x)$ from x_0 to x in $U_\alpha\cap U_\gamma$. We then integrate both sides of Eq. (C1) to get

$$g_{\alpha\gamma}(x)=c_{\alpha\gamma}\exp\left[\int_{c(x)}(\mu_\gamma-\mu_\alpha)\right].$$

for some constant $c_{\alpha\gamma}$. Now $c_{\alpha\gamma}$ exists for every pair (U_α,U_γ) such that $U_\alpha\cap U_\gamma$ is not empty. We claim that $\{c_{\alpha\gamma}\}$ is a cocycle. To see this, observe that if $\Upsilon_\alpha\cap\Upsilon_\gamma\cap\Upsilon_\delta\neq\emptyset$, then $\boldsymbol{\beta}_\delta=g_{\gamma\delta}\boldsymbol{\beta}_\gamma=g_{\gamma\delta}g_{\alpha\gamma}\boldsymbol{\beta}_\alpha$ and

$$g_{\alpha\delta}=g_{\gamma\delta}g_{\alpha\gamma}.$$

This implies that

$$\begin{aligned}&c_{\alpha\delta}\exp\left[\int_{c(x)}(\mu_\delta-\mu_\alpha)\right]\\&\quad=c_{\gamma\delta}\exp\left[\int_{c(x)}(\mu_\delta-\mu_\gamma)\right]c_{\alpha\gamma}\exp\left[\int_{c(x)}(\mu_\gamma-\mu_\alpha)\right]\end{aligned}$$

and thus that $c_{\alpha\delta}=c_{\alpha\gamma}c_{\gamma\delta}$. It follows that $\{c_{\alpha\delta}\}$ is a cocycle. Choose a coboundary $\{b_\alpha\}$ defined by the cover $\{U_\alpha\}$ such that $c_{\alpha\delta}=b_\alpha b_\delta^{-1}$. Then we have

$$g_{\alpha\delta}(x)=\left[b_\alpha\exp\left(-\int_{c(x)}\mu_\alpha\right)\right]\left[b_\delta\exp\left(-\int_{c(x)}\mu_\delta\right)\right]^{-1}.$$

If we define

$$\boldsymbol{\beta}_\alpha'(x)=b_\alpha\exp\left(-\int_{c(x)}\mu_\alpha\right)\boldsymbol{\beta}_\alpha(x),$$

$$\omega_\alpha'(x)=b_\alpha^{-1}\exp\left(\int_{c(x)}\mu_\alpha\right)\omega_\alpha(x),$$

then we see that on $U_\alpha\cap U_\gamma\neq\emptyset$ we have $\boldsymbol{\beta}_\alpha'=\boldsymbol{\beta}_\gamma'$ and $\omega_\alpha'=\omega_\gamma'$. Thus there exists globally defined $\boldsymbol{\beta}$ and ω on $P-\Sigma$ such that $\boldsymbol{\beta}$ agrees with $\boldsymbol{\beta}_\alpha'$ on Υ_α and ω agrees with ω_α' on Υ_α for each α. We have that

$$\mathbf{\Omega}=\boldsymbol{\beta}\omega$$

and the theorem follows.

APPENDIX D

Lemma[17]: If $d\omega\wedge du=0$ for a 1-form ω and a function u, then there exist functions H and K such that $\omega=H\,du+dK$.

Proof: If $d\omega\wedge du=0$ for a 1-form ω and function u, then $d\omega$ must contain a factor du. Hence we may write

$$d\omega=\theta\wedge du$$

for some 1-form θ. Using the Frobenius theorem, one can show that this equation implies that ω is of the form

$$\omega=f(x,u)dx+g\,du$$

for some functions $f=f(x,u)$, x and g. Thus

$$d\omega=\psi\wedge du, \tag{D1}$$

where

$$\psi\equiv-\frac{\partial f}{\partial u}dx+dg. \tag{D2}$$

Now the 1-form ψ satisfying (D1) is fixed only to within addition of a 1-form proportional to du. Thus, if we define

$$\bar\psi=\psi+h\,du, \tag{D3}$$

then $\bar\psi$ also satisfies

$$d\omega=\bar\psi\wedge du. \tag{D4}$$

We show that the function h in (D3) can be chosen so that $d\bar\psi=0$. From (D2) and (D3) we have

$$\begin{aligned}d\bar\psi&=d\psi+dh\wedge du\\&=-\frac{\partial^2 f}{\partial u^2}du\wedge dx+dh\wedge du\\&=\left(\frac{\partial^2 f}{\partial u^2}dx+dh\right)\wedge du.\end{aligned} \tag{D5}$$

Choose $h=h(x,u)$ such that

$$\frac{\partial^2 f}{\partial u^2} + \frac{\partial h}{\partial x} = 0.$$

Then by (D5) we have $d\bar{\psi} = 0$. Hence locally

$$\bar{\psi} = dH$$

for some function H. Using $\bar{\psi} = dH$ in (D4), we obtain

$$d\omega = dH \wedge du = d\,(H\,du),$$

or equivalently

$$d\,(\omega - H\,du) = 0.$$

The 1-form $(\omega - H\,du)$ is thus closed and locally may be written as dK for some function K. We then have

$$\omega = H\,du + dK.$$

[1] E. S. Abers and B. W. Lee, Phys. Rep. **9**, 1 (1973); A. Salam, Rev. Mod. Phys. **52**, 525 (1980).

[2] See the following for a clear account of fiber bundles in physics: A. Trautman, Rep. Math. Phys. **1**, 29 (1970); Czech. J. Phys. B **29**, 107 (1979).

[3] See the following for details about principal bundles and connections: S. Kobayashi and K. Nomizu, *Foundations of Differential Geometry* (Interscience, New York and London, 1963), Vol. I.

[4] C. N. Yang and R. L. Mills, Phys. Rev. **96**, 191 (1954).

[5] When f is null, there exists a single principal null direction l such that $l \wedge f = 0 = l \wedge {}^*f$, where ${}^*f$ is the Hodge dual of f. Writing $f^+ = f - i\,{}^*f$ the condition becomes $l \wedge f^+ = 0$. When l is twist-free, $l \wedge dl = 0$, and it can be scaled so that $l = du$.

[6] As discussed in Sec. II, the curvature has the form $\mathbf{\Omega} = \beta\pi^*(f)$ where f is a real 2-form on spacetime. The curvature $\mathbf{\Omega}$ is null with twist-free rays if f has a twist-free repeated principal null direction $(du \wedge f^+ = 0)$.

[7] J. Anandan and K. P. Tod, Phys. Rev. D **18**, 1144 (1978).

[8] At each point ξ of a principal fiber bundle P there exists a subspace $V_\xi \subseteq T_\xi P$, the vertical subspace at ξ, consisting of all vectors tangent to the fiber through ξ. A connection in P can be defined as a choice for each $\xi \in P$ of a complimentary subspace $H_\xi \subseteq T_\xi P$, the horizontal subspace at ξ, satisfying certain conditions.[3] Once a connection is specified each vector $\Delta \in T_\xi P$ can be decomposed uniquely into a sum of a vertical part and a horizontal part: $\Delta = \mathrm{ver}(\Delta) + \mathrm{hor}(\Delta)$. If ψ is a Lie-algebra-valued p-form on P, then the horizontal part of ψ, denoted $\mathrm{hor}(\psi)$, is defined by $(\mathrm{hor}\psi)(\Delta_1,\ldots,\Delta_p) = \psi(\mathrm{hor}\Delta_1,\ldots,\mathrm{hor}\Delta_p)$.

[9] A *pseudotensorial* p-form on P of type Ad is a Lie-algebra-valued p-form ψ on P satisfying $R_g{}^*\psi = \mathrm{Ad}(g^{-1})\psi$. Here $R_g{}^*$ is the pullback map on forms on P induced by the right "translation" map $R_g: P \to P$ defined by $R_g\xi = \xi g$. A p-form ψ that is pseudotensorial of type Ad and which in addition satisfies $\psi(\Delta_1,\ldots,\Delta_p) = 0$ if any one of the vectors $\Delta_1,\ldots,\Delta_p$ is vertical is termed a tensorial p-form of type Ad. For a more complete discussion of pseudotensorial forms see Ref. 3. Where no confusion would arise, we will refer to a pseudotensorial p-form of type Ad as simply a pseudotensorial p-form. Moreover, for brevity we will, following Trautman,[2] often refer to a tensorial zero-form as a Higgs field. The exterior covariant derivative of a pseudotensorial form ψ, denoted $D\psi$, is defined by[8] $D\phi = \mathrm{hor}(d\psi)$. Note in particular that a connection 1-form $\mathbf{\Phi}$ is pseudotensorial while the curvature of ϕ, $\mathbf{\Omega} \equiv D\,\mathbf{\Phi}$. is a tensorial 2-form.

[10] S. Lang, *Differentiable Manifolds* (Addison-Wesley, Reading, Mass., 1972).

[11] Note that the adjoint action of G on G' extends to $G' \oplus iG'$ via $\mathrm{Ad}(g)(\beta + i\gamma) = \mathrm{Ad}(g)(\beta) + i\mathrm{Ad}(g)(\gamma)$. It follows that $\mathbf{\Omega}^+$ and $D\,\mathbf{\Omega}^+$ are tensorial forms with respect to this extended action.

[12] P. B. Yasskin, Phys. Rev. D **12**, 2212 (1975).

[13] R. . Güven, Phys. Rev. D **19**, 471 (1979).

[14] I. Robinson, J. Math. Phys. **2**, 290 (1961).

[15] M. Mostow and S. Schecter (Department of Mathematics, North Carolina State University), private communication (1980).

[16] B. Kostant, in *Lectures in Modern Analysis and Applications III*, edited by C. T. Taam (Springer-Verlag, Berlin, Heidelberg, and New York, 1970).

[17] This lemma is implicit in the work of I. Robinson.[14]

Generalized Schrödinger representation and its application to gauge field theories

Hiroaki Arisue, Toshiaki Fujiwara, Takuo Inoue, and Kaku Ogawa
Department of Physics, Kyoto University, Kyoto 606, Japan

(Received 12 September 1980; accepted for publication 18 March 1981)

A generalization of the Schrödinger representation is presented and applied to theories with indefinite metric as well as those with positive definite metric. It enables us to construct the wavefunctional formalism of the quantum electromagnetic field equivalent to the operator formalism in covariant gauges. The theoretical foundation of the Wick rotation of the dynamical variable $A_0(x)$ is established.

PACS numbers: 11.10.Np, 03.70. + k, 03.65.Db

1. INTRODUCTION

In 1926 Schrödinger presented the wave functional formalism of quantum mechanics equivalent to the algebraic formalism given by Heisenberg in 1925. There the arguments of the wavefunctions are real numbers and it cannot describe a theory with indefinite metric.

On the other hand, the third formalism, path integral, was proposed by Feynman in 1948. In the last several years the structure of the QCD vacuum has been investigated extensively using the path integral. In these works the prescription of the Wick rotation ($A_0 = -i\varphi$, $t = -i\beta$) is used frequently.[1] The theoretical footing of the Wick rotation of the time axis ($t = -i\beta$) is given by Matsubara in his study of finite temperature Green's functions.[2] However, the Wick rotation of the dynamical variable ($A_0 = -i\varphi$) is still lacking its theoretical foundation.[3]

In this paper we will give the theoretical ground to the Wick rotation of dynamical variables $A_0(x)$. For this purpose we present a generalization of the Schrödinger representation. In this generalized Schrödinger representation (GSR) the arguments of the wavefunctions are complex in general.

It has been believed that the eigenvalue of a Hermitian operator is real in a theory with definite metric. The usual argument proceeds as follows: Let $\hat{\alpha}$ be a Hermitian operator ($\hat{\alpha}^+ = \hat{\alpha}$) and $\langle\alpha|$ be its eigenstate ($\langle\alpha|\hat{\alpha} = \alpha\langle\alpha|$). Then

$$\{\langle\alpha_1|\hat{\alpha}|\alpha_2\rangle\}^* = \rangle\alpha_2|\hat{\alpha}|\alpha_1\rangle$$

gives

$$(\alpha_1^* - \alpha_2)\langle\alpha_2|\alpha_1\rangle = 0. \qquad (1.1)$$

Putting $\alpha_2 = \alpha_1$ we have

$$(\alpha_1^* - \alpha_1)\langle\alpha_1|\alpha_1\rangle = 0.$$

Since $\langle\alpha_1|\alpha_1\rangle \neq 0$, we get $\alpha_1^* = \alpha_1$, i.e., α_1 is real. However, this argument does not hold for a Hermitian operator with continuous spectrum. In this case $\langle\alpha_2|\alpha_1\rangle$ is not a function but a distribution. In general the numerical value of a distribution is meaningless by itself. In fact we will give an example which satisfies Eq. (1.1) as a distribution for complex eigenvalues. We can generalize consistently the theory in the Schrödinger representation on the real axis ($\alpha\in\mathbb{R}$) to that on a curve Γ in the complex plane ($\alpha\in\Gamma\subset\mathbb{C}$).[4]

Although we are interested in gauge field theories, the essential point of GSR and Wick rotation of the dynamical variable will be clarified by studying harmonic oscillators with positive definite and indefinite metrics. In Sec. II, GSR for the harmonic oscillator with definite metric is presented, which is equivalent to the Fock representation in the operator formalism. The completeness and orthogonality relations are examined, where a nontrivially generalized δ function is introduced. In Sec. III it is shown that, using this GSR, the indefinite metric harmonic oscillator can be treated on the same footing as the positive metric one. In Sec. IV, the wavefunctional formalism of the electromagnetic field is developed using the GSR.

It is shown that the GSR provides the foundation of the Wick rotation of the dynamical variable $A_0(x)$.

2. GENERALIZED SCHRÖDINGER REPRESENTATION

In this section we investigate the harmonic oscillator

$$H = \tfrac{1}{2}(\hat{p}^2 + \hat{q}^2), \qquad (2.1)$$

with the commutation relation

$$[\hat{p},\hat{q}] = -i. \qquad (2.2)$$

We take the Fock space as our starting point. The definition of the annihilation and creation operators (a and a^+) is given as follows;

$$\hat{p} = (a^+ + a)/\sqrt{2},$$
$$\hat{q} = (a^+ - a)/(\sqrt{2})i, \qquad (2.3)$$

and their commutation relations are

$$[a,a^+] = 1,$$
$$[a,H] = a, \qquad (2.4)$$
$$[a^+,H] = -a^+.$$

The positivity of energy eigenvalues and the above commutation relations require us to define the vacuum as follows:

$$a|0\rangle = 0, \qquad (2.5)$$

with

$$\langle 0|0\rangle = 1.$$

The Fock space is spanned by the basis vectors

$$|n\rangle = (1/\sqrt{n!})\,(a^+)^n|0\rangle. \qquad (2.6)$$

Now we construct the eigenstate $\langle q|$ of the operator $\hat{q}$

 0022-2488/81/092055-05$01.00

from the Fock space. It is defined by the following equations:

$$\langle q|\hat{q} = q\langle q|, \tag{2.7}$$

$$\langle q|\hat{p} = -i\frac{d}{dq}\langle q|. \tag{2.8}$$

Expanding $\langle q|$ by the basis $\langle n|$,

$$\langle q| = \sum_n \langle n|[(i)^n/(\sqrt{\pi}2^n n!)^{1/2}]f_n(q); \tag{2.9}$$

the coefficient $f_n(q)$ satisfies the recursion equations from Eqs. (2.3), (2.7), and (2.8),

$$f_{n+1}(q) - 2qf_n(q) + 2nf_{n-1}(q) = 0, \tag{2.10}$$

$$f_{n+1}(q) + 2\frac{d}{dq}f_n(q) - 2nf_{n-1}(q) = 0. \tag{2.11}$$

Equation (2.10) implies that $f_n(q)$ is proportional to the Hermite polynomial $H_n(q) \equiv (-1)^n e^{q^2}(d^n/dq^n)e^{-q^2}$,

$$f_n(q) = f(q)H_n(q). \tag{2.12}$$

Then, since $(d/dq)H_n(q) = 2nH_{n-1}(q)$, Eq. (2.11) leads to

$$\frac{d}{dq}f(q) = -qf(q), \tag{2.13}$$

so that

$$f_n(q) = H_n(q)e^{-q^2/2}, \tag{2.14}$$

where the overall factor is taken to be unity for later convenience. Thus we get

$$\langle q| = \sum_n \langle n|[(i)^n/(\sqrt{\pi}2^n n!)^{1/2}]H_n(q)e^{-q^2/2}, \tag{2.15}$$

and its conjugate

$$|q\rangle = \sum_n [(-i)^n/(\sqrt{\pi}2^n n!)^{1/2}]H_n(q^*)e^{-q^{*2}/2}|n\rangle, \tag{2.16}$$

where we have used the relation $[H_n(q)]^* = H_n(q^*)$. Notice that, in the above construction of the state $\langle q|$, we have not made the assumption that the eigenvalue q is real. In fact, the above argument holds for a general complex number q.[5]

Next we analyze the structure of the space $\{\langle q| \,|\, q\in\mathbb{C}\}$. For this purpose we define a set of connected curves in the complex plane as follows:

Definition: A curve Γ is called "real-like" if there exists a parametrization of Γ $[z = z(s)]$ such that

$$\lim_{s\to\pm\infty} |z(s)| = \infty,$$

$$\lim_{s\to-\infty} |\pi - \arg z(s)| < \pi/4 - \epsilon \quad (\epsilon > 0),$$

$$\lim_{s\to+\infty} |\arg z(s)| < \pi/4 - \epsilon' \quad (\epsilon' > 0).$$

From now on we denote $\Gamma\sim\mathbb{R}$ if Γ is real-like.

It is important to note that the orthogonal property of $H_n(q)$ holds for an arbitrary real-like curve Γ.

$$\int_{\Gamma\sim\mathbb{R}} dq\; H_m(q)H_n(q)e^{-q^2} = \delta_{mn}(\pi)^{1/2}2^n n!. \tag{2.17}$$

It is precisely this property that enables us to generalize the Schrödinger representation.

Now we are at the stage to discuss the structure of $\{\langle q|\}$. First, by virtue of Eqs. (2.15)–(2.17), we obtain

$$\begin{aligned}\int_{\Gamma\sim\mathbb{R}} dq\,|q^*\rangle\langle q| &= \sum_{m,n}|m\rangle\langle n|\frac{(i)^{n-m}}{(\pi 2^{m+n}m!n!)^{1/2}} \\ &\quad\times\int_{\Gamma\sim\mathbb{R}} dq\,H_m(q)H_n(q)e^{-q^2} \\ &= \sum_n |n\rangle\langle n| = \mathbb{1}.\end{aligned} \tag{2.18}$$

This equation shows that the set $\{\langle q| \,|\, q\in{}^\forall\Gamma\sim\mathbb{R}\}$ forms a complete set. Observe that $\int_{\Gamma\sim\mathbb{R}} dq|q\rangle\langle q| \neq \mathbb{1}$ except for $\Gamma = \mathbb{R}$, and that only the combination $|q^*\rangle\langle q|$ leads to $\mathbb{1}$. From Eq. (2.18) the inner product of $|f\rangle$ and $|g\rangle$ becomes

$$\langle f|g\rangle = \int_{\Gamma\sim\mathbb{R}} dq\,\langle f|q^*\rangle\langle q|g\rangle = \int_{\Gamma\sim\mathbb{R}} dq\,[f(q^*)]^* g(q). \tag{2.19}$$

Second, according to the orthogonality of the Fock bases $\langle m|n\rangle = \delta_{mn}$, we have

$$\langle q_1|q_2^*\rangle = \sum_n \frac{1}{\sqrt{\pi}2^n n!}H_n(q_1)H_n(q_2)e^{-\frac{1}{2}q_1^2-\frac{1}{2}q_2^2}. \tag{2.20}$$

We call this the generalized delta function $\delta_G(q_1 - q_2)$.[6] This is, of course, not a function and it has meaning only as a distribution. Let us restrict the functional space to the Hilbert space $\{f\,|\,f(q) = \Sigma_n a_n[1/(\sqrt{\pi}2^n n!)^{1/2}]H_n(q)e^{-q^2/2}, \Sigma_n|a_n|^2 < \infty\}$. For any element of this space the generalized δ function acts as follows:

$$\int_{\Gamma\sim\mathbb{R}} dq_2\,\delta_G(q_1 - q_2)f(q_2) = f(q_1). \tag{2.21}$$

This also means

$$\int_{\Gamma\sim\mathbb{R}} dq_2[(q_1 - q_2)\delta_G(q_1 - q_2)]f(q_2) = 0. \tag{2.22}$$

Surprisingly, Eq. (2.21) holds for any $q_1\in\mathbb{C}$, even for $q_1\notin\Gamma$; i.e., the "support" of $\delta_G(z)$ is not concentrated on the origin $(z = 0)$. Further, in spite of the fact just mentioned, $z\delta_G(z) = 0$ as a distribution from Eq. (2.22). (Recall the arguments in Sec. 1).

Finally, we check the Hermiticity of the operators $\hat{q}$ and $\hat{p}$;

$$\begin{aligned}\langle f|\hat{q}g\rangle &\equiv \int_{\Gamma\sim\mathbb{R}} dq[f(q^*)]^* q\,g(q) = \int_{\Gamma\sim\mathbb{R}} dq\,[q^* f(q^*)]^* g(q) \\ &= \langle\hat{q}f|g\rangle,\end{aligned} \tag{2.23}$$

and

$$\begin{aligned}\langle f|\hat{p}g\rangle &\equiv \int_{\Gamma\sim\mathbb{R}} dq\,[f(q^*)]^*\Big(-i\frac{d}{dq}g(q)\Big) \\ &= \int_{\Gamma\sim\mathbb{R}} dq\,i\frac{d}{dq}[f(q^*)]^* g(q) \\ &= \int_{\Gamma\sim\mathbb{R}} dq\left[-i\frac{d}{dq^*}f(q^*)\right]^* g(q) \equiv \langle\hat{p}f|g\rangle.\end{aligned} \tag{2.24}$$

In conclusion, GSR of the harmonic oscillator has been constructed, where the wavefunction defined on arbitrary real-like Γ is used. It would be appropriate to call it "the Schrödinger representation (SR) on Γ." The SR on each Γ is equivalent to the Fock representation.[7] The transformation matrix relating the SR on Γ_1 with that on Γ_2 is the generalized δ function

$$\langle q_1|q_2^*\rangle = \delta_G(q_1 - q_2), \tag{2.25}$$

where

$$q_1 \in \Gamma_1 \sim \mathbb{R}, \quad q_2 \in \Gamma_2 \sim \mathbb{R}.$$

3. SCHRÖDINGER REPRESENTATION OF INDEFINITE METRIC THEORY

In this section we consider the indefinite metric harmonic oscillator

$$H = -\tfrac{1}{2}(\hat{p}^2 + \hat{q}^2) \quad \text{with } [\hat{p},\hat{q}] = -i \tag{3.1}$$

in the Schrödinger representation. This is a system with indefinite metric, and is unavoidable, to introduce complex eigenvalues for $\hat{q}$ because if the theory were constructed only with real eigenvalues the inner product would be positive definite

$$\langle f|f\rangle = \int_{\mathbb{R}} dq\, [f(q)]^* f(q) > 0. \tag{3.2}$$

Hence, the generalized Schrödinger representation developed in Sec. 2 is crucial.

As in Sec. 2 we start from the Fock space. The annihilation and creation operators (a and a^+) are defined as follows:

$$\hat{p} = (a^+ + a)/\sqrt{2}, \tag{3.3}$$
$$\hat{q} = (a - a^+)/\sqrt{2}i,$$

and their commutation relations are

$$[a,a^+] = -1,$$
$$[a,H] = a, \tag{3.4}$$
$$[a^+,H] = -a^+.$$

To guarantee the positivity of energy eigenvalues we must define the vacuum as

$$a|0\rangle = 0, \tag{3.5}$$

with

$$\langle 0|0\rangle = 1.$$

The basis vectors of the Fock space are given by

$$|n\rangle = (1/\sqrt{n!})\,(a^+)^n|0\rangle. \tag{3.6}$$

Then we have the indefinite metric

$$\langle m|n\rangle = (-1)^n\delta_{m,n}. \tag{3.7}$$

Let us construct the eigenstate $\langle q|$ of the operator $\hat{q}$ from the Fock space, which is defined by

$$\langle q|\hat{q} = q\langle q|. \tag{3.8}$$

$$\langle q|\hat{p} = -i\frac{d}{dq}\langle q|. \tag{3.9}$$

The eigenstate $\langle q|$ is expanded by the bases $\langle n|$ as

$$\langle q| = \sum_n \langle n|\frac{1}{(\sqrt{\pi}2^n n!)^{1/2}} g_n(q). \tag{3.10}$$

From Eqs. (3.3), (3.8), and (3.9), we have the recursion equations for the coefficient g_n,

$$g_{n+1}(q) - 2iqg_n(q) + 2ng_{n-1}(q) = 0, \tag{3.11}$$

$$g_{n+1}(q) + 2dg_n(q)/d(iq) - 2ng_{n-1}(q) = 0. \tag{3.12}$$

Comparing these equations with Eqs. (2.10) and (2.11), it is easily seen that the solution of the above equations is given by

$$g_n(q) = H_n(iq)e^{q^2/2}. \tag{3.13}$$

So we have

$$\langle q| = \sum_n \langle n|\frac{1}{(\sqrt{\pi}2^n n!)^{1/2}} H_n(iq)e^{q^2/2}, \tag{3.14}$$

and its conjugate

$$|q\rangle = \sum_n \frac{1}{(\sqrt{\pi}2^n n!)^{1/2}} H_n(-iq^*)e^{q^{*2}/2}|n\rangle$$
$$= \sum_n [(-1)^n/(\sqrt{\pi}2^n n!)^{1/2}] H_n(iq^*)e^{q^{*2}/2}. \tag{3.15}$$

To investigate the nature of these states we define a set of connected curves in the complex plane.

Definition: A curve Γ is called "imaginary-like" if there exists a parametrization of Γ such that

$$\lim_{s \pm \infty} |z(s)| = \infty,$$

$$\lim_{s\to -\infty} |\tfrac{1}{2}\pi - \arg z(s)| < \tfrac{1}{4}\pi - \epsilon \quad (\epsilon > 0),$$

$$\lim_{s\to +\infty} |\tfrac{3}{2}\pi - \arg z(s)| < \tfrac{1}{4}\pi - \epsilon' \quad (\epsilon' > 0).$$

We denote $\Gamma \sim \mathbb{I}$ if Γ is imaginary-like. The orthogonal property of $H_n(iq)$ is given by

$$\mathrm{i}\int_{\Gamma\sim\mathbb{I}} dq\, H_m(iq)H_n(iq)e^{q^2} = \delta_{mn}\pi^{1/2}2^n n!. \tag{3.16}$$

Using the above property, the completeness of the set $\{\langle q|\,|q\in{}^\forall\Gamma\sim\mathbb{I}\}$ can be shown as follows:

$$i\int_{\Gamma\sim\mathbb{I}} dq\, |q^*\rangle\langle q|$$
$$= \sum_{m,n} |m\rangle\langle n|\frac{(-1)^m i}{(\pi 2^{m+n} m!n!)^{1/2}}\int_{\Gamma\sim\mathbb{I}} dq\, H_m(iq)H_n(iq)e^{q^2}$$
$$= \sum_n (-1)^n|n\rangle\langle n| = \mathbb{1}. \tag{3.17}$$

From Eq. (3.17) the inner product of $|f\rangle$ and $|g\rangle$ becomes

$$\langle f|g\rangle = i\int_{\Gamma\sim\mathbb{I}} dq\, \langle f|q^*\rangle\langle q|g\rangle = i\int_{\Gamma\sim\mathbb{I}} dq\, [f(q^*)]^* g(q). \tag{3.18}$$

The generalized δ function in this case is given by

$$\delta_G(q_1 - q_2) \equiv \langle q_1|q_2^*\rangle$$
$$= \sum_n \frac{1}{\sqrt{\pi}2^n n!} H_n(iq_1)H_n(iq_2)e^{\frac{1}{2}q_1^2 + \frac{1}{2}q_2^2}. \tag{3.19}$$

Thus in the case of the indefinite metric space, the Schrödinger representation on each $\Gamma\sim\mathbb{I}$ is equivalent to the Fock representation. In this way, we are ready to study the indefinite metric quantum field theories in the Schrödinger representation.

4. ELECTROMAGNETIC FIELD

The free electromagnetic field is described by the action

$$S = \int d^4x \left\{ -\tfrac{1}{4}F_{\mu\nu}F^{\mu\nu} - \frac{1}{2\alpha}(\partial_\mu A^\mu)^2 \right\}. \tag{4.1}$$

In the following we choose the gauge $\alpha = 1$ (Feynman gauge) for simplicity. Dropping the total divergence $-\frac{1}{2}\partial_\mu(A_\nu \partial^\nu A^\mu - A^\mu \partial_\nu A^\nu)$, the action becomes

$$S = \int d^4x \left(-\tfrac{1}{2}\partial_\mu A^\nu \partial^\mu A_\nu \right). \tag{4.2}$$

The Hamiltonian in terms of the field A^μ and the momenta $\pi^\mu (\equiv \delta \mathscr{L}/\delta \dot{A}_\mu) = -\dot{A}^\mu$ is given by

$$H = \int d^3x \left(-\tfrac{1}{2}\pi_\mu \pi^\mu - \tfrac{1}{2}\partial_i A_\nu \partial_i A^\nu \right) \quad (i = 1,2,3), \tag{4.3}$$

and we take the following canonical commutation relations

$$[\pi^\mu(x), A_\nu(y)]_{x_0 = y_0} = -i\delta^\mu_\nu \delta^3(\mathbf{x} - \mathbf{y}). \tag{4.4}$$

This system consists of infinitely many four-component harmonic oscillators. For positive metric harmonic oscillators $[A_i(\mathbf{x}),\ i = 1,2,3]$ the GSR developed in Sec. 2 is applicable, and for indefinite metric harmonic oscillators $[A_0(x)]$ the GSR exhibited in Sec. 3 can be adopted.

From the above discussion, the set of states $C(\Gamma_\mu) \equiv \{\langle A_\mu| \,|A_0 \in \Gamma_0 \sim \mathbb{I},\ A_i \in \Gamma_i \sim \mathbb{R}\}$ (we omit the space index $\mathbf{x}$ for brevity) forms a complete set for the indefinite metric space. It should be noted, however, that the set of representations $\{C(\Gamma_\mu) | \Gamma_0 \sim \mathbb{I},\ \Gamma_i \sim \mathbb{R}\}$ is not closed under the Lorentz transformation, which mixes A_i and A_0. In fact, the components of the Lorentz transformed field $A'_\mu = \Lambda_\mu^{\ \nu} A_\nu$ $(A_0 \in \Gamma_0 \sim \mathbb{I}, A_i \in \Gamma_i \sim \mathbb{R})$ lie neither on an imaginary-like line nor on a real-like one. However, as will be shown in the Appendix, the set of states

$$C(\Gamma_\mu; \Lambda_\mu^{\ \nu}) \equiv \{\langle A'_\mu| \,|A'_\mu = \Lambda_\mu^{\ \nu} A_\nu,\ A_\mu \in \Gamma_\mu,\ \Gamma_0 \sim \mathbb{I},\ \Gamma_i \sim \mathbb{R}\}$$

also forms a complete set of states. This leads us to a further generalized representation $C(\Gamma_\mu; \Lambda_\mu^{\ \nu})$, and the set of representations $\{C(\Gamma_\mu; \Lambda_\mu^{\ \nu}) | \Gamma_0 \sim \mathbb{I},\ \Gamma_i \sim \mathbb{R},\ \Lambda_\mu^{\ \rho}\Lambda_{\rho\nu} = g_{\mu\nu}\}$ is closed under the Lorentz transformation.

Now we proceed to discuss the path integrals. The propagation kernel is decomposed as

$$\langle A_f | e^{-iHt} | A_i^* \rangle = \langle A_f | e^{-iH_\Delta t} \mathbb{1} e^{-iH_\Delta t} \mathbb{1} e^{-iH_\Delta t} \cdots e^{-iH_\Delta t} \mathbb{1} e^{-iH_\Delta t} | A_i^* \rangle, \tag{4.5}$$

where

$$\mathbb{1} = \int_{\substack{\Gamma \sim \mathbb{R} \ \text{for } A_i(\mathbf{x})\ (i = 1,2,3) \\ \Gamma \sim \mathbb{I} \ \text{for } A_0(\mathbf{x})}} \mathscr{D}A_\mu(\mathbf{x})\, |A_\mu(\mathbf{x})^*\rangle \langle A_\mu(\mathbf{x})|. \tag{4.6}$$

We obtain

$$\langle A_f | e^{-iHt} | A_i^* \rangle = \int_{\{A_i\}}^{\{A_f\}} \mathscr{D}A_\mu(x) \exp\left(i\int_0^t L\, dt\right), \qquad \begin{matrix} \Gamma \sim \mathbb{R} \ \text{for } A_i(x)\ (i = 1,2,3) \\ \Gamma \sim \mathbb{I} \ \text{for } A_0(x) \end{matrix} \tag{4.7}$$

where $L = \int d^3x\left(-\frac{1}{2}\partial_\mu A^\nu \partial^\nu A_\mu\right)$. Note that the contour of integration of $A_0(x)$ is an imaginary-like curve.

Finally we mention the density matrix, which is written as

$$\langle A_F | e^{-\beta H} | A_I^* \rangle = \int_{\substack{\{A_I\} \\ \Gamma \sim \mathbb{R} \ \text{for } A_i(x)\ (i = 1,2,3) \\ \Gamma \sim \mathbb{I} \ \text{for } A_0(x)}}^{\{A_F\}} \mathscr{D}A_\mu(x) \exp\left[\int_0^\beta dx_4 \int d^3x \tfrac{1}{2}(\partial_i A^\nu \partial_i A_\nu + \partial_4 A^\nu \partial_4 A_\nu)\right],$$

$$= \int_{\substack{\{A_I\} \\ \Gamma \sim \mathbb{R} \ \text{for } \tilde{A}_\mu(x)\ (\mu = 1,2,3,4)}}^{\{A_F\}} \mathscr{D}\tilde{A}_\mu \exp\left[-\int_0^\beta dx_4 \int d^3x \mathscr{L}_E\right], \tag{4.8}$$

where $\tilde{A}_\mu = (A_i, iA_0)$ and $\mathscr{L}_E = \frac{1}{2}\Sigma^4_{\mu,\nu=1}(\partial_\mu \tilde{A}_\nu)(\partial_\mu \tilde{A}_\nu)$, which is precisely the same as the Lagrangian used in the usual Euclidean field theory. In other words, the density matrix of Electromagnetics is given by the Euclidean path integral. This establishes the prescription of Wick rotation in the path integral expression.

5. SUMMARY AND DISCUSSION

In this paper we have constructed GSR and have shown that the GSR on each Γ is equivalent to the Fock representation. For the positive metric harmonic oscillator Γ is a real-like curve, and for the indefinite metric harmonic oscillator

Γ is an imaginary-like curve. The transformation matrix between GSR on Γ and Γ' is given by the generalized δ function. In the wave functional formalism of the electromagnetic field, GSR does enable us to construct a theory equivalent to that in the Fock representation. Even in the path integral form of the propagation kernel, the integral contour is an imaginary-like curve for $A_0(x)$. Furthermore, the path integral expression for the density matrix is correctly given by the Euclidean path integral.

We have developed GSR for the abelian gauge field theory, but the extension to the nonabelian gauge field theory in perturbative sense is trivial. We hope that nonabelian gauge theories would be described correctly by the GSR with $\Gamma\sim\mathbb{R}$ for A_i^a and $\Gamma\sim\mathbb{I}$ for A_0^a also in the nonperturbative treatment. In that case, calculations based on the Euclidean path integral are meaningful. Of course, there remains a possibility that GSR on another set of Γ's is needed in these treatments.

ACKNOWLEDGMENTS

The authors would like to express their sincere thanks to Professor T. Kugo for fruitful discussions. They also thank their colleagues at Kyoto University.

APPENDIX

In this appendix, we show that the set of states

$$C(\Gamma_\mu;A_\mu^\nu)$$
$$\equiv\{\langle A'_\mu|A'_\mu=A_\mu^\nu A_\nu,\ A_\mu\in\Gamma_\mu\ \ \Gamma_0\sim\mathbb{I},\ \Gamma_i\sim\mathbb{R}\}$$

forms a complete set for the indefinite metric space. We consider, for simplicity, the Lorentz boost along the third axis,

$$A'_0=A_0\cosh\tau+A_3\sinh\tau,$$
$$A'_3=A_0\sinh\tau+A_3\cosh\tau. \tag{A1}$$

To prove the completeness of the set of states $C(\Gamma_\mu;A_\mu^\nu)$ it is sufficient to show that

$$\int_{\substack{\Gamma\sim\mathbb{I}\text{ for }A_0\\ \Gamma\sim\mathbb{R}\text{ for }A_i}} dA_0\,dA_3\langle f|A_0^{*\prime},A_3^{*\prime}\rangle\,\langle A'_0,A'_3|g\rangle
=\int_{\substack{\Gamma\sim\mathbb{I}\text{ for }A_0\\ \Gamma\sim\mathbb{R}\text{ for }A_i}} dA_0\,dA_3\langle f|A_0^*,A_3^*\rangle\,\langle A_0,A_3|g\rangle, \tag{A2}$$

for arbitrary states $\langle f|$ and $|g\rangle$. (We have omitted the irrelevant components A_1 and A_2.) We first note that the wavefunctions take the form

$$P(A_0,A_3)\exp\alpha(A_0^2-A_3^2)\quad(\alpha>0),$$

where $P(A_0,A_3)$ is a polynomial in A_0 and A_3. Notice that the exponential factor, which is Lorentz invariant, guarantees the convergence of the integral in both sides of Eq. (A2). The proof of Eq. (A2) proceeds as follows. The right-hand side of Eq. (A2) can be written as

$$\begin{aligned}
&\int_{\substack{\Gamma\sim\mathbb{I}\text{ for }A_0\\ \Gamma\sim\mathbb{R}\text{ for }A_3}} dA_0\,dA_3 f(A_0^*,A_3^*)^*g(A_0,A_3),=i\int_{\Gamma=\mathbb{R}\text{ for }A_4\text{ and }A_3} dA_4\,dA_3 f(-iA_4,A_3)^*g(iA_4,A_3),\\
&=i\int_{\Gamma=\mathbb{R}\text{ for }A'_4\text{ and }A'_3} dA'_4\,dA'_3 f(-i(A'_4\cos\theta-A'_3\sin\theta),A'_4\sin\theta+A'_3\cos\theta)^*\\
&\quad g(i(A'_4\cos\theta-A'_3\sin\theta),A'_4\sin\theta+A'_3\cos\theta),
\end{aligned}\tag{A3}$$

where we have performed a rotation in the real A_3-A_4 plane in the last equality. Since the last expression in Eq. (A3) is convergent and the result of integration is independent of θ, we can make an analytic continuation $\theta\to i\tau$, and Eq. (A3) becomes

$$\begin{aligned}
&i\int_{\Gamma=\mathbb{R}\text{ for }A_4\text{ and }A_3} dA_4\,dA_3 f(-i(A_4\cosh\tau+iA_3\sinh\tau),-iA_4\sinh\tau+A_3\cosh\tau)^*g(i(A_4\cosh\tau-iA_3\sinh\tau),iA_4\sinh\tau+A_3\cosh\tau)\\
&=\int_{\substack{\Gamma\sim\mathbb{I}\text{ for }A_0\\ \Gamma\sim\mathbb{R}\text{ for }A_3}} dA_0\,dA_3 f(A_0^*\cosh\tau+A_3^*\sinh\tau,A_0^*\sinh\tau+A_3^*\cosh\tau)^*g(A_0\cosh\tau+A_3\sinh\tau,A_0\sinh\tau+A_3\cosh\tau),
\end{aligned}\tag{A4}$$

which is the left-hand side of Eq. (A2). This completes the proof of Eq. (A2).

[1]G. 't Hooft, Phys, Rev. D **14**, 3432 (1976); A. M. Polyyakov, Nucl. Phys. B **121**, 429 (1977); C. G. Callan, R. Dashen, and D. Gross, Phys. Rev. D **17**, 2717 (1978); *ibid.* **19**, 1826 (1979).

[2]T. Matsubara, Prog. Theor. Phys. **14**, 351 (1955).

[3]T. Barnes and G. I. Ghandour, Nucl. Phys. B **166**, 125 (1980).

[4]In another context, the complexification of eigenvalues of an Hermitian operator is considered in E. C. G. Sudarshan, C. B. Chiu, and Vittorio Gorini, Phys. Rev. D **18**, 2914, (1978).

[5]The derivative in Eqs. (2.11) and (2.13) is well defined as differentiation of analytic functions.

[6]By formal calculation using the integral representation $H_n(q)=\frac{2^n}{\sqrt{\pi}}\int_{-\infty}^{\infty}d\lambda e^{-\lambda^2}(q+i\lambda)^n$, it yields

$\langle q_1|q_2^*\rangle=\exp[-\frac{1}{2}(q_1-q_2)^2]\frac{1}{2\pi}\int_{-\infty}^{\infty}d\lambda\exp[-i\lambda(q_1-q_2)].$

[7]Strictly speaking, we define the SR on Γ only for a pre-Hilbert space, i.e., the space generated from a free vacuum by applying polynomials of creation operators.

Path integral quantization of the supersymmetric model of the Dirac particle

A. T. Ogielski and J. Sobczyk
Institute for Theoretical Physics, University of Wrocław, 50–205 Wrocław, Poland

(Received 7 April 1980; accepted for publication 25 June 1980)

We investigate in detail the representation of the Dirac propagator as a path integral over virtual trajectories in the phase space with anticommuting variables. Using the reparametrization and supergauge invariant action integral proposed by Berezin and Marinov [Ann. Phys. **104**, 336 (1977)], we analyze the relation of causality to reparametrization invariance, and we construct the Faddeev–Popov measure for the symbol of the evolution operator. We present a precise definition of the path integral as a limit of finite dimensional integrals, and we explicitly perform the integration obtaining the familiar result. We also analyze another approach, where reparametrization and supergauge invariance of the action is preserved with the help of independent *einbein* variables. This method, as opposed to Faddeev–Popov technique, dispenses with the gauge fixing problem and allows us to construct the correct functional measure by explicit factorization of the volume of reparametrization group and supergauge group; the former is infinite, the latter identically vanishes. A sequence of finite dimensional approximations to the functional measure is given also in this case together with explicit calculations.

PACS numbers: 11.10.Qr, 03.70. + k, 11.30.Pb

1. INTRODUCTION

Recent years brought a series of papers devoted to the relativistic classical mechanics of spinning particles with spin degress of freedom described by Grassmannian variables.[1-5] The canonical quantization of these systems has been extensively studied and the appearance of relativistic wave equations as constraint conditions is well understood.

Here we examine in detail the alternative quantization procedure based on phase space path integrals which, to the best of our knowledge, has not been successfully carried out in the case of Grassmannian "pseudomechanics" of relativistic spinning particles. To describe classically the relativisitic spin $\frac{1}{2}$ particle we adopt the framework proposed by Berezin and Marinov.[1] Althoug a consistent interpretation at the classical level is not possible, this model correctly reproduces the Bargman–Michel–Telegdi equation,[6] and thus provides a good justification of this choice of variables prior to quantization. It is an amusing question, therefore, how to construct the path integral over virtual trajectories in their model of "classical" phase space to obtain explicitly the Feyman kernel of the Dirac equation:

$$S(x,x') = \int \frac{d^4p}{(2\pi)^4}\, e^{ip(x-x')}\, \frac{\gamma^\mu p_\mu + m}{p^2 - m^2 + i\epsilon}$$
$$-(i\gamma^\mu \partial_\mu + m)S(x,x') = \delta(x-x'). \qquad (1.1)$$

The point is that there is no intrinsic distinction between canonical momenta and coordinates in the phse space of anticommuting variables.[7] So it is important to realize what quantity should be defined by a path integral, and to find its relation to the position space propagator (1.1). It follows from the general discussion in Ref. 1 that once phase space variables are treated on an equal footing it is natural to formulate the quantum theory in terms of symbols of operators,[8] in contradistinction to the representation of operators by their kernels, which presupposes the difference between positions and momenta. Since we want to construct a phase space path integral for the propagator (1.1), which is expressed in terms of position variables and γ-matrices, we found it convenient to consider a hybrid quantity. A kernel in position space with respect to space-time variables, and simultaneously a symbol with respect to fermionic variables.

The crucial feature of relativisitc mechanics of spinning particles is the reparametrization and supergauge invariance of the action. In the canonical quantization it immediately leads to the appearance of relativisitc wave equations as constraint conditions on physical states. In the functional language, however, it leads to the modification of the measure on the space of virtual trajectories. We present two different solutions of this problem here. One is the application of the Faddeev–Popov technique to our case, another is based on the reformulation of the model in terms of independent Lagrange multipliers (*einbein* variables of Ref. 5), which after some manipulations allows to find a proper measure by explicit factorization of the volume of reparametrization and supergauge group. It should be noted that the latter approach does not require us to fix the gauge, and is suitable for calculation of gauge-invariant quantities. In the case of a spinless particles we obtain in this way the path integral derived earlier from the Fock's proper time formalism[9]

$$\Delta_F(x-x') = \int_0^\infty dL\, e^{-im^2L} \int \prod_{[0,L]} dx^\mu \Big|_{x'}^{x} e^{-i\int_0^L ds\, \dot{x}_\mu \dot{x}^\mu}$$

We should say here that many different approaches to the path integral representation of the Dirac propagator have been proposed in the past.[10] It is difficult to trace all references to the subject, and we apologize for not giving the complete list. Generally speaking, two principal ideas were involved: The first, already suggested by Feynman[11] extends the proper time formalism, here however, the space-time position $x_\mu(s)$ and the spin part are treated asymmetricaly, the

 0022-2488/81/092060-06$01.00

latter being represented by γ-matrices already on the classical level. Another method emerged from consideration of relativistic spinning top,[12] however here usually all values of spin are obtained and one must use subsidiary conditions to select a particular spin component. The organization of the paper is the following: First, in Sec. 2 we recapitulate briefly the pseudoclassical model of Ref. 1 and its canonical quantization mainly for the convenience of the reader. Section 3 begins with a discussion of a relation between reparametrization invariance and causality, and contains the construction of a Faddeev–Popov measure as a limit of finite-dimensional approximations. We show that the unphysical parameter labelling the points on the trajectory of a particle completely disappears from the final result. In Sec. 4 was reformulate the problem in terms of *einbein* variables, and we construct another path integral by explicitly factoring out the gauge group volume.

Throughout the paper we treat the case of a massive particle. It is only the *einbein* formulation that has the zero mass limit),[5] and the integration procedure of Sec. 4 can be carried through also in this case.

We use the system of units $\hbar = c = 1$, and our Minkowski metric has the signature $(+\ -\ -\ -)$.

2. RESUME OF THE CANONICAL FORMALISM

To describe the massive spin $\frac{1}{2}$ particle we follow Berezin and Marinov[1] who supersymmetrized the standard reparametization invariant action $\int ds\,(\dot{x}_\mu^2)^{1/2}$ of a spinless particle by introducing five anticommuting generators ξ_μ and ξ_5 of a real Grassmann algebra. The ξ_μ transform as components of a pseudovector, and ξ_5 as a pseudoscalar under the full Lorentz grop (before quantization one could equally well require them to transform as vector and scalar, respectively). The following action integral is postulated:

$$\mathscr{A} = \int ds \left\{ -\tfrac{1}{2}i\xi_5\dot{\xi}_5 - \tfrac{1}{2}i\xi_\mu\dot{\xi}^\mu - m\left[\left(\dot{x}_\mu - \frac{i}{m}\xi_\mu\dot{\xi}_5\right)^2\right]^{1/2}\right\}. \tag{2.1}$$

The action is invariant under reparametrizations of the time s

$$s \rightarrow s' = f(s), \tag{2.2}$$

where f is a strictly monotonic function on the interval $[0, L]$, and (up to a total derivative) under supergauge transformations parametrized by an anticommuting "gauge function" $\alpha(s)$:

$$\begin{aligned} &\xi_\mu \rightarrow \xi_\mu + (1/m)\alpha p_\mu, \\ &\xi_5 \rightarrow \xi_5 + \alpha, \\ &x_\mu \rightarrow x_\mu + (i/m^2)\alpha p_\mu \xi_5. \end{aligned} \tag{2.3}$$

Because the derivatives of Grassmann variables ξ_μ and ξ_5 enter the Lagrangian linearly, they can be treated as phase variables from the beginning. Introducing the momentum conjugate to position x_μ

$$p_\mu = \frac{\delta L}{\delta \dot{x}^\mu} = -\frac{m[\dot{x}_\mu - (i/m)\xi_\mu\dot{\xi}_5]}{[(\dot{x}_\nu - (i/m)\xi_\nu\dot{\xi}_5)^2]^{1/2}} \tag{2.4}$$

and the appropriate Poisson Bracket one finds that nonvanishing PB's of the phase space variables are

$$\begin{aligned} &\{\xi_\mu, \xi_\nu\} = -ig_{\mu\nu}, \\ &\{\xi_5, \xi_5\} = i, \\ &\{x_\mu, p_\nu\} = g_{\mu\nu}. \end{aligned} \tag{2.5}$$

Invariance under (2.2) and (2.3) results in first class constraints

$$\varphi_1 = p^2 - m^2, \quad \varphi_2 = p\xi - m\xi_5. \tag{2.6}$$

Following the quantization rule $\{\cdot,\cdot\} \rightarrow (1/i)[\cdot,\cdot]_\pm$ it is seen that the classical generators of the Grassmann algebra are replaced after quanitzation by the generators of the corresponding Clifford algebra. Its irreducible representation is finite-dimensional, and within this approach there is no more reference to Grassmann variables on the quantum level. They are replaced by the appearance of spinor indices of the state vectors. Quantum operators $\hat{\xi}_\mu, \hat{\xi}_5$ can be represented by

$$\hat{\xi}_\mu = (\sqrt{\tfrac{1}{2}})\gamma_5\gamma_\mu, \quad \hat{\xi}_5 = (\sqrt{\tfrac{1}{2}})\gamma_5 \tag{2.7}$$

and the constraints (2.6) which are interpreted as conditions on physical states give the mass-shell condition and the Dirac equation up to multiplication by γ_5 from the left. This deviation from the ordinary Dirac operator (1.1) is important, because the propagator obtained from the path integral constructed with the action (2.1) would then yield the propagator $G(x - x')$ corresponding to the equation

$$\gamma_5(i\hat{p} - m)G(x - x') = \delta(x - x') \tag{2.8}$$

rather than to (1.1). It is clear that these two functions are related by $G(x - x') = S(x - x')\gamma_5$.

3. THE PATH INTEGRAL : FADDEEV–POPOV MEASURE

There is a subtlety in defining the time evolution of the system (2.1). The physical time variable is x_0, the time component of a position four-vector x_μ of a particle relative to the Lorentz frame in which the motion is described–one may be tempted to call it "time of external observer". Moreover, energy of a particle is related to translations of x_0. In the canonical framework, however, x_0 appears as one of phase space variables and it is the arbitrary parameter s that plays the role of time in classical equations of motion. It follows that the Hamiltonian, governing the evolution in s, vanishes on the physical subspace of the phase space: The canonical Hamiltonian identically vanishes, and evolution in parameter s is generated by the extended Hamiltonian in the sense of Dirac.

$$H = \alpha(p^2 - m^2) + \beta(p\xi - m\xi_5) \tag{3.1}$$

in which each particular choice of Lagrange multipliers α and β as functions of phase space variables corresponds to definite parametrization of trajectories. It is evident now that the quantum evolution kernel, expressed by matrix element of the type

$$\langle\psi|e^{-is\hat{H}}|\phi\rangle \tag{3.2}$$

will be dependent of s for physical states which are annihilated by constraints.

Related to the above remark is the problem of causality: The phase space path integral must be defined in such way that it gives the Feynman propagator (1.1). We can disregard the spin variables for a while, for they are essential for this problem and concentrate on the spinless particle described by the action $\mathscr{A} = -m\int(\dot{x}_\mu^2)^{1/2}$. The constraint $p^2 - m^2 = 0$ splits the physical subspace of the phase space of varibles x_μ, p_μ (in a relativistically invariant manner) into two disjoint connected components characterized by $p_0 > 0$ and $p_0 < 0$, respectively. Recalling that momentum p_0 conjugate to x_0 is

$$p_0(s) = -m\dot{x}_0(s)/(\dot{x}^2)^{1/2},$$

we see that for classical motion from point x' to x'' the trajectory will lie on the sheet $p_0 < 0$ if $x_0'' > x_0'$, and on the sheet $p_0 > 0$ if $x_0'' < x_0'$. We will show that this also holds for virtual trajectories from x' to x'', because they are subjected to mass-shell constraint. The correct phase space path integral for the Feynman propagator is therefore a sum of two terms, where the momentum integration is carried over two sheets of the physical phase space, and the choice of sheet in which the virtual trajectories lie corresponds to the causal ordering of points x' and x''.

This reasoning, needless to say, is mostly a conveneint reformulation of ideas contained in Feynman's papers of 1948–9, especially in Ref. 13, and may be summarized in the well known identity

$$\Delta_F(x) = \Theta(x_0)\Delta^{(+)}(x) + \theta(-x^0)\Delta^{(-)}(x),$$

which we could take as a starting point showing that it is $\Delta_{(+)}(x)$ that has a simple and unambiguous phase space path integral representation.

After this introductory exposition of ideas we return to the problem of the Dirac particle defined by (2.1). We remind the reader that we will consider a hybrid quantity closely related to the Dirac propagator, namely a kernel of the evolution operator with respect to space-time variables and a symbol of evolution operator with respect to Grassmann spin variables. For unconstrained systems with Grassmann variables symbols of evolution operators were introduced in Ref. 1. We consider therefore the formal expression

$$G(x'',x',\xi_\mu,\xi_5) = \theta(x_0'' - x_0')G^{(+)} + \theta(x_0' - x_0'')G^{(-)} \tag{3.3}$$

using a condensed notation

$$\xi = (\xi_5\xi_\mu); \quad \xi^0\eta = \xi_5\eta_5 - \xi_\mu\eta^\mu$$

$$\begin{aligned} G^{(\pm)}(x'',x',\xi) &= \int\prod_s dx_\mu(s)dp_\mu(s)d\eta(s) \\ &\times \prod_{a=1}^{2}\prod_s \delta(\varphi_a)\delta(\chi_a) \\ &\times\theta(\mp p^0)|\text{sdet}\{\chi_a,\varphi_b\}| \\ &\times\exp\left\{i\int_0^L ds[p\dot{x} - \tfrac{1}{2}i\eta\dot{\eta} \right. \\ &+ \tfrac{1}{2}[2\xi(\eta(0) - \eta(L)) + \eta(0)\eta(L)]\}, \end{aligned} \tag{3.4}$$

with the boundary conditions $x(0) = x'$, $x(L) = x''$ while the remaining variables are entirely integrated.

We immediately observe that with the suitable choice of gauge conditions χ_a the integrations need not be complicated (althogh not Gaussian), for the canonical Hamiltonian which ordinarily appears in the exponent identically vanishes. We are left essentially with an integral over a chain of delta-functions. The crucial point, however, is the choice of a sequence of finite dimensional integrals which gives meaning to the formal expression (3.4) especially the finite dimensional (lattice) approximation of individual terms in F-P measure.

We remark in passing that we deliberately did not put the L- dependence in the LHS of (3.4) because, as was announced earlier, with the correct choice of finite dimensional approximations this dependence completely drops out.

The well known gauge conditions fixing the parameter s are the laboratory time gauge

$$\chi_1 = q_0 + s \tag{3.5}$$

and the proper time gauge

$$\chi_1 = (1/m)xp + s. \tag{3.6}$$

To fix the supergauge freedom we impose the simplest possible constraint

$$\chi_2 = \eta_5, \tag{3.7}$$

which could equally well be replaced by the condition that any others of the η_μ vanishes. We remind the reader that for Grassmann variables $\delta(\eta) = \eta$, for $\int\eta\, d\eta = 1$. We have performed explicit calculations in both gauges (3.5) and (3.6). Below we will limit the discussion to the laboratory time gauge (3.5), which is a little simpler in actual calculations, together with (3.7).

Using a standard prescription for calculation of the superdeterminant[14] we find that

$$\text{sdet}\{\chi_a,\varphi_b\} = \text{sdet}\begin{bmatrix} 2p_0 & \eta_0 \\ 0 & -im \end{bmatrix} = \frac{2ip_0}{m}. \tag{3.8}$$

We turn now to the main result of this section, that is, the correct finite dimensional approximations to (3.4) that would eventually lead to the Dirac propagator. As usual, we replace the interval $[0,L]$ by a linear lattice $\{s_1,s_2,\ldots,s_N,s_{N+1}\}$ with uniform spacing $\epsilon = s_{k+1} - s_k = L/N$. The integrand in (3.4) is approximated according to the Weyl quantization rule, the so-called "midpoint prescription" consisting in the replacement[13]

$$f(x(s_k),p(s_k)) \to f(\tfrac{1}{2}(x_k + x_{k+1}),p_k).$$

We found that it is necessary then to extend the midpoint rule also to Grassmann variables. Summarizing, we consider (3.4) as a limit

$$G^{(\pm)}(x'',x',\xi) = \lim_{N\to\infty} G_N^{(\pm)}(x'',x',\xi), \tag{3.9}$$

with (including suitable normalization constants)

$$G_N^{(\pm)}(x'',x',\xi) = \int \prod_{k=2}^{N} dx_k^\mu \prod_{L=1}^{N} \frac{dp_L^\mu}{(2\pi)^3} \prod_{m=1}^{N} d\eta_m^\mu d\eta_m^5 (2)^{3/2}$$
$$\times \left\{ \prod_{k=1}^{N} \delta(p_k^2 - m^2)\theta(\mp p_k^0) 2\delta\left(\tfrac{1}{2} p_k(\eta_k + \eta_{k+1}) - \frac{m}{2}(\eta_k^5 - \eta_{k+1}^5)\right)\right\}$$
$$\times \left\{ \prod_{k=2}^{N} \delta(\tfrac{1}{2}(x_k^0 + x_{k+1}^0) + \tfrac{1}{2}(s_k + s_{k+1}))\delta[\tfrac{1}{2}(\eta_k^5 + \eta_{k+1}^5)] \right\}$$
$$\times \left\{ \prod_{k=2}^{N} \frac{2p_k^0}{m} \right\} \times \exp\left\{ i \sum_{l=1}^{N} p_l(x_{l+1} - x_l) - \tfrac{1}{2} \sum_{L=1}^{N} \eta_l \eta_{l+1} \right.$$
$$+ \frac{1}{2} \sum_{l=1}^{N} \eta_l^5 \eta_{l+1}^5 - \frac{1}{2} [\eta_1 \eta_{N+1} - \eta_1^5 \eta_{N+1}^5 + 2\xi(\eta_1 - \eta_{N+1})$$
$$\left. - 2\xi^5(\eta_1^5 - \eta_{N+1}^5)] \right\}. \tag{3.10}$$

Despite the complicated form, (3.10) is relatively easy to integrate in a straightforward manner.

First, however, we would like to comment on some relevant features of (3.10).

1. It is crucial to have the number of delta functions in the gauge term one less than then the number of delta functions corresponding to constraints. The reason for this will become evident in the next section.

2. Correspondingly, there are $N-1$ terms in the approximation of functional determinant.

3. $x_{N+1} = x''$ $\quad x_1 = x'$.

4. On the contrary, the seemingly free variables η_{N+1}^μ, η_{N+1}^5 are completely spurious and were introduced for the sake of notation, because they can be eliminated by translation.

Evaluation of the multiple integral (3.10), which in practice reduces to the stepwise elimination of all integrations save one due to delta functions, can be conveniently performed in the following order: first we integrate over p_k^0 and x_k^0, and in the next step over $\bar{x}_k$ so that after disentangling the delta functions only one integration over $\bar{p}_1$ remains.

In the next step we integrate out the Grassmann variables in descending order, from η_N^5 to η_2^5 and from η_N^0 to η_1^0, and consequently over $\bar{\eta}_k$, obtaining $\delta(p\xi - m\xi_5) = p\xi - m\xi_5$ from the last integration over η_1^5. The result

$$G_N^{(\pm)}(x'',x',\xi) = \frac{1}{(2\pi)^3} \int \frac{d\bar{p}}{2\omega(\bar{p})} \times \exp i\{\mp \omega(\bar{p})(x_0'' - x_0') - \bar{p}(\bar{x}'' - \bar{x}')\} \times (p\xi - m\xi_5)|_{p^0 = \mp\omega(\bar{p})} \tag{3.11}$$

is independent of N and L, as expected. It may be worthwhile to mention that in proper time gauge also no dependence on the internal parameter—the proper time—remains (see Ref. 2).

Putting both frequencies, $G^{(+)}$ and $G^{(-)}$ together according to (3.3) we obtain the familiar expression

$$G(x'',x',\xi) = i \int \frac{d^4 p}{(2\pi)^4} e^{ip(x''-x')} \frac{p\xi - m\xi_5}{p^2 - m^2 + i\epsilon}. \tag{3.12}$$

To relate the last formula—which is a symbol of the Dirac propagator with respect to anticommuting ξ variables—to the Feynman kernel (1.1) we have to refer to the general argument presented in Ref. 1. It is simpler to notice, however, that since we employed the midpoint rule in the evaluation of the path integral which corresponds to Weyl (or rather "anti-Weyl") quantization, the transition from a symbol $\mathscr{A}(\xi)$ of arbitrary operator to its representation $\hat{A}(\hat{\xi})$ in terms of γ-matrices is effected by a substitution of $\hat{\xi}$ given by (2.7) followed by the antisymmetrization of all products. The expression (3.12), however, is linear in anticommuting variables and it suffices to make the replacement $\xi \to \hat{\xi}$ of (2.7) to transform it to the Dirac propagator multiplied by γ_5. The appearance of this additional γ_5 is necessary for consistence, as explained at the end of Sec. 2.

4. THE PATH INTEGRAL: *EINBEIN* FORMALISM

An alternative approach to quantization of gauge invariant system, which was widely used in lattice field theories, is to calculate the functional integral ignoring gauge fixing terms completely. This technique applied for calculation of gauge-invariant quantities allows us to factor out the gauge group volume from the final result, which then can be dropped by a suitable normalization of the measure. For successful calculations, however, we have to rewrite the action (2.1) in a more tractable form using the einbein variables, introduced in this model by Brink, di Vecchia, and Howe.[5] On the classical level the action (2.1) is completely equivalent to the following one:

$$\mathscr{A} = -\tfrac{1}{2}\int_0^L ds \left\{ i(\xi_\mu \dot{\xi}^\mu - \xi_5 \dot{\xi}_5) + i\psi\left(m\xi_5 - \frac{1}{e}\dot{x}\xi\right) + em^2 + \frac{1}{e}\dot{x}_\mu^2 \right\}, \tag{4.1}$$

since after elimination of e and ψ from (4.1) by their Euler–Lagrange equations we reobtain the original form (2.1). The Lagrangian in (4.1) is invariant under reparametrizations and supergauge transformations, provided that e and ψ transform as

$$e'(s') = \frac{df}{ds} e(f(s)),$$
$$\psi'(s') = \frac{df}{ds} \psi(f(s)), \tag{4.2}$$

under reparameterization (2.2), and if we consider the following supergauge trnasformations[5]:

$$\begin{aligned}
\delta x_\mu &= i\alpha\xi_\mu \\
\delta\xi_\mu &= \alpha[(i/2e)\psi\xi_\mu - \dot{x}_\mu/e] \\
\delta e &= -i\alpha\psi \\
\delta\psi &= 2\dot{\alpha} \\
\delta\xi_5 &= m\alpha + (i/me)\alpha\xi_5(\tfrac{1}{2}m\psi - \dot{\xi}_5).
\end{aligned} \tag{4.3}$$

It is important to remark that the Lagrangian in (4.1) is not degenerate any more, because e and ψ are independent Lagrange multipliers and not functions of dynamical variables.

We could attempt now to set up a path integral of the type

$$\int \mathscr{D}e\mathscr{D}\psi\mathscr{D}x_\mu\mathscr{D}\eta_\mu\mathscr{D}\eta_5 e^{i\mathscr{A}}$$

which is a configuration space path integral with respect to space-time variables. Although not strictly Gaussian, this integral is calculable provided we relegate the integration over *einbeins* to the last stage of calculations. Its true virtue, however, is that it is the integral that can be calculated in Euclidean space, and the such result can then be Wick rotated back to Minkowski metric.

We have performed the explicit calculations in the manner described above, obtaining the correct result (3.12). However, configuration space path integrals require extra normalizations, which are automatically included in phase space integrals, and because we feel that phase space treatment is simpler we will describe the latter approach in more detail.

There is a fine point in defining functional integration over the *einbein* variable $e(s)$, due to its relation to reparametrizations. From the geometric viewpoint $e(s)$ for $s\in[0,L]$ defines the orientation of the one-dimensional "parameter manifold" $[0,L]$. Its sign therefore may not change on each particular trajectory, and thus to avoid double counting we should integrate only over positive values of $e(s)$. The causal character of the propagator calculated by a method described in this section will emerge thus in a way different from that of Sec. 3.

We introduce the momenta conjugate to x_μ from (4.1)

$$p_\mu = -(1/e)\dot{x}_\mu + (1/2e)\psi\xi_\mu \tag{4.4}$$

to rewrite (4.1) in a Hamiltonian form

$$\mathscr{A} = \int_0^L ds\,\{p_\mu\dot{x}^\mu + \tfrac{1}{2}i\xi_5\dot{\xi}_5 - \tfrac{1}{2}i\xi_\mu\dot{\xi}^\mu - H\}, \tag{4.5}$$

with

$$H = \tfrac{1}{2}e(m^2 - p^2) + \tfrac{1}{2}i\psi(p\xi - m\xi_5). \tag{4.6}$$

The decomposition (3.8) need to be used now and we have instead

$$G(x'',x',\xi) = (1/V)\int\prod_s \mathscr{D}e(s)\mathscr{D}\psi(s)\mathscr{D}p(s)\mathscr{D}x(s)\mathscr{D}\eta(s)\mathscr{D}\eta^5(s) \times\exp\{i\mathscr{A} + \tfrac{1}{2}[\xi^0(\eta(0) - \eta(L)) + \eta(0)\eta(L)]\}. \tag{4.7}$$

The integration is to be performed only over positive values of $e(s)$, while the momentum integrations are not restricted. The only nonintegrated variables are $x(L) = x''$ and $x(0) = x'$. V is a normalization factor.

Our goal is to show that one may explicitly factor out the reparametrization and supergauge group volume from (4.7) by a suitable change of variables, obtaining essentially the functional measure of the Fock s proper time formation, which within our approach clearly shows its relation to gauge invariances, and by no means appears as related to the *ad hoc* introduction of additional time parameter conjugate to mass as an earlier discussions.[16]

Precisely, we necessarily have to provide a finite dimensional approximation to (4.7). First, we must notice that the very geometric interpretation of *einbein* variables $e_k = e(s_k)$, and $\psi_k = \psi(s_k)$ implies that they are not sitting on *sites* s_k of the linear lattice, but rather on *links* (s_k,s_{k+1}); and therefore the midpoint rule used in construction of finite dimensional approximation of the integrand should not be used for these variables (this is similar to the rules of momentum integration).

We consider therefore the following integrals:

$$G_N(x'',x',\xi) = \frac{1}{V_N}\int_0^\infty\prod_{k=1}^N de_k\int\prod_{k=1}^N d\psi_k d\eta_k d\eta_k^5\int\prod_{k=1}^N d^4p_k\prod_{l=2}^N d^4x_l \exp\Big\{\sum_{k=1}^N ip_k(x_{k+1} - x_k) + \tfrac{1}{2}\eta_k\eta_{k+1} - \tfrac{1}{2}\epsilon e_k(m^2 - p_k^2) - \tfrac{1}{2}\psi_k[p_k(\eta_k + \eta_{k+1}) - m(\eta_k^5 + \eta_{k+1}^5)]\epsilon + \tfrac{1}{2}((\eta_1\eta_{N+1} + 2\xi(\eta_1 - \eta_{N=1}))\Big\}. \tag{4.8}$$

We will show that after integration over p_k, x_k and η_k the only dependence of the integrand on the *einbein* variables is through the combinations

$$L[e] = \sum_{k=1}^N e_k \tag{4.9}$$

and

$$\lambda[\psi] = \sum_{k=1}^N \psi_k. \tag{4.10}$$

From formula (4.8) it is already seen how the causal property of G is obtained: the momentum integration is not well defined, and has to be improved in a standard way by the introduction of a small imaginary part in the exponent. It is simpler, however, to perform first the integral over x_k, which is unambiguous and produces the momentum delta function eliminating all momentum integrations save the last one, and to postpone the necessary $i\epsilon$ prescription till the next stop. We obtain

$$G_N = (1/V_N)\int \prod_{k=1}^{N} (de_k d\psi_k) \times \int d^4p\, e^{i(p^2 - m^2)L} e^{\lambda(p\xi - m\xi_5)} e^{ip(x'' - x')}, \tag{4.11}$$

which suggests the following change of variables

$$\bar{e}_1 = L, \quad \bar{e}_k = e_k \quad \text{for } k = 2,\ldots,N$$

and (4.12)

$$\bar{\psi}_1 = \lambda, \quad \bar{\psi}_k = \psi_k \quad \text{for } k = 2,\ldots,N.$$

The Jacobian is equal to unity. Here it is clearly seen that only $N-1$, and not N integrations over *einbeins* correspond to the gauge freedom of the system, which explains why we had one fewer delta function in the gauge fixing term in the finite-dimensional approximations to F-P measure (3.10). Factorizing the redundant integration

$$V_N^G = \int_0^\infty \prod_{k=2}^{N} d\bar{e}_k, \tag{4.13}$$

$$V_N^{SG} = \int \prod_{k=2}^{N} d\bar{\psi}_k, \tag{4.14}$$

which in the lim $N \to \infty$ correspond to the "volume" of reparametrization group and supergauge group, respectively, and putting $V_N = V_N^G \cdot V_N^{SG}$ we have

$$\begin{aligned} G_N &= \int_0^\infty dL\, d^4p\, e^{ip(x'' - x')} e^{i(p^2 - m^2)L} (p\xi - m\xi^5) \\ &= \int d^4p\, e^{ip(x'' - x')} \frac{p\xi - m\xi^5}{p^2 - m^2 + i\epsilon}, \end{aligned} \tag{4.15}$$

where in order to define the last integral we had to use the standard prescription $m^2 \to m^2 - i\epsilon$. The relation of (4.15) to the Dirac propagator has been discussed in Sec. 3.

It should be said that the volume of the reparametization group (4.13) is already infinite for finite N, while the volume of the supergauge group vanishes identically. The latter property is typical for supergroups[17] and is a consequence of specific properties of integrals over anticommuting Grassmann variables.

CONCLUDING REMARKS

We have shown that supersymmetric relativistic "pseudomechanics" of the spinning particles, known to lead to the conventional wave equations for the spinorial wave functions in the canonical quantization scheme, can also be consistently quantized in the path integral framework. This requies us to consider the less familiar concept of the symbol of operator, whose connection with the standard Dirac propagator is fortunately very simple. Our motivation to investigate this time-honored problem was not, of course, to learn anything new about the spin $\frac{1}{2}$ particle itself, but the intrerest in finding the correct way of defining and calculating path integrals for constrained supersymmetric mechanical systems. In conclusion we would like to make several comments:

1. The polynomial form of time reparametrization invariant Lagrangians, which is so troublesome in attempts to directly calculate the configuration space path integrals[10,15,16] leads to strinkingly simple integrals in the phase space if the integration measure is suitable chosen. This is due to the general property of such system: the canonical Hamiltonian identically vanishes, and the phase space path integral essentially reduces to a chain of delta-functions.

2. We have not included in the discussion the problem of spinning particles interacting with external fields: electromagnetic, Yang–Mills, gravitational.[1,5,18] Since these problems do not appear to be exactly solvable at present, it would be interesting to try to formulate the saddle-point method for the approximate calculations of corresponding path integrals, and to compare the results with known WKB approximation for relativistic wave equations. In particular, this could show whether classical solutions for anticommuting variables have any relevance for the quantum theory.

ACKNOWLEDGMENT

We wish to thank Professor J. Lukierski for many stimulating discussions.

[1] F. A. Berezin and M. S. Marinov, Ann. Phys. **104**, 336 (1977).

[2] R. Casalbuoni, Nuovo Cimento A **33**, 115 (1976).

[3] R. Casalbuoni, Nuovo Cimento A **33**, (1976).

[4] A. Barducci, R. Casalbuoni, and L. Lusanna, Nuovo Cimento **35**, 377 (1976).

[5] L. Brink, R. Di Vecchia, and P. Howe, Nucl. Phys. B **118**, 76 (1977).

[6] V. Bargman, L. Michel, and V. Telegdi, Phys. Rev. Lett. **2**, 435 (1059).

[7] Even if one attempts to define the canonical formalism for Grassmann variables in close analogy to the ordinary mechanics, thus beginning with even-dimensional phase space, the number of second class constraints need not be necessarily even, so that odd-dimensional physical phase spaces arise quite naturally. Moreover, in general neither Poisson Brackets, nor the Hamilton equations may be used to distinguish "coordinate" and "momenta".

[8] F. A. Berezin, Theor. and Math. Phys. **6**, 194 (1971).

[9] R. P. Feynman, Phys. Rev. **84**, 108 (1951).

[10] See references given in Ref. 1.

[11] Fock's proper time formalism should not be confused with a "proper time gauge" of the canonical formulation.

[12] E.g., see references given in A. J. Hanson and T. Regge, Ann. Phys. **87**, 498 (1974).

[13] R. P. Feynman, Phys. Rev. **76**, 769 (1949).

[14] B. F. Pakhomov, Mat. Zametki **16**, 65 (1974).

[15] M. M. Mizrahi, J. Math. Phys. **16**, 2201 (1975).

[16] G. Garrod, Rev. Mod. Phys. **38**, 483 (1966).

[17] F. A. Berezin, Preprint ITEP, 1978.

[18] A. Barducci, R. Casalbuoni, and L. Lusanna, Nucl. Phys. B **124**, 93 (1977).

Decomposition of the natural representation of the U(2) color charge algebra in the three-quark sector

Nguyen Van Hieu
Institute of Physics, Nghia Do, Tu Liem, Hanoi, Vietnam

(Received 18 December 1980; accepted for publication 30 January 1981)

It was shown that the natural 16-dimensional representation $\psi_{a_1a_2a_3a_0}$, $a_n = 1,2$, $n = 1,2,3,0$, of the three-quark U(2) color charge algebra proposed by Adler, contains five irreducible representations (1,0), three irreducible representations $(\bar{3}, 0)$, and one irreducible representation $(1, \frac{1}{2})$ of the SU(3) $\oplus$ SU(2) subalgebra.

PACS numbers: 11.30.Jw, 12.40.Bb, 02.20.Rt, 02.20.Sv

I. INTRODUCTION

In an attempt to generalize the U(n) classical chromodynamics to the case of noncommuting source charges, Adler[1] suggested a new algebra with the elements being the noncommuting vectors

$$u = \{u^\alpha,\ \alpha = 1,2,...,n^2\},\quad v = \{v^\alpha,\quad \alpha = 1,2,...,n^2\},$$

and with the outer product

$$\underline{P}(u,v) = \{\underline{P}^\alpha(u,v),\quad \alpha = 1,2,...,n^2\},$$
$$\underline{P}^\alpha(u,v) = g^{\alpha\beta\gamma}(u^\beta v^\gamma - v^\beta u^\gamma),$$

where $g^{\alpha\beta\gamma}$ are determined from the multiplication rule of the generators λ^α of the U(n) group:

$$\lambda^\alpha\lambda^\beta = g^{\alpha\beta\gamma}\lambda^\gamma.$$

As the examples of this new algebra Adler introduced the quark color charge algebras whose elements are the operators acting on the product Hilbert space constructed by taking the direct product of the color Hilbert spaces for various quarks and antiquarks. Recently a new composite model for quarks and leptons was proposed by Harari[2] and Shupe.[3] The Adler color charge algebra may be relevant in constructing the mathematical tools for the composite models of this kind.

The color charge algebra for the U(2) group in the three quarks sector was studied in detail by many authors.[4–7] For the convenience, instead of the n^2-dimensional vectors u^α, v^α we can use the matrices

$$U = u^\alpha\lambda^\alpha,\quad V = v^\alpha\lambda^\alpha.$$

Then the outer product $\underline{P}^\alpha(u,v)$ will be substituted by the usual commutator

$$\underline{P}(U,V) = [U,V],$$

and the color charge algebra becomes a Lie algebra. In this language the three-quark color charge algebra for the U(2) group is isomorphic to the SU(3) $\oplus$ SU(2) $\oplus$ U(1) $\oplus$ U(1) algebra. In the present paper we study the decomposition of the natural representation of the three quarks U(2) color charge algebra into their irreducible representations. For this purpose we must consider in detail the SU(3) $\oplus$ SU(2) subalgebra.

II. GENERATORS OF THE SU(3) $\oplus$ SU(2) ALGEBRA

We shall use following notation for the Pauli matrices τ^i, $i = 1,2,3,0$ and the Gell-Mann matrices

λ^α, $\alpha = 1,2,...,8$:

$$\tau^1 = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix},\quad \tau^2 = \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix},$$
$$\tau^3 = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix},\quad \tau^0 = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix},$$
$$\lambda^1 = \begin{pmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix},\quad \lambda^2 = \begin{pmatrix} 0 & -i & 0 \\ i & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix},$$
$$\lambda^3 = \begin{pmatrix} 1 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & 0 \end{pmatrix},\quad \lambda^4 = \begin{pmatrix} 0 & 0 & 1 \\ 0 & 0 & 0 \\ 1 & 0 & 0 \end{pmatrix},$$
$$\lambda^5 = \begin{pmatrix} 0 & 0 & -i \\ 0 & 0 & 0 \\ i & 0 & 0 \end{pmatrix},\quad \lambda^6 = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{pmatrix},$$
$$\lambda^7 = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & -i \\ 0 & i & 0 \end{pmatrix},\quad \lambda_8 = \frac{1}{\sqrt{3}}\begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & -2 \end{pmatrix}.$$

According to the definition of the three-quark color charge algebra for the U(2) group, its elements are the operators acting on the 16-dimensional vector space direct product of four two-dimensional spinor spaces of the U(2) group. These operators themselves are those of the natural representation of the given color charge algebra. The basis vectors in this representation depend on four indices $a_n = 1,2$, $n = 1,2,3,0$ and will be denoted by $\psi_{a_1a_2a_3a_0}$.

From the Pauli matrices $\tau^i_{(n)}$ acting on the indices a_n, $n = 1,2,3,0$, we can construct 13 operators $H_1,...,H_6$, $I_1,...,I_4$, K_1, K_2, K_3 commuting with the operators

$$S^i = \tau^i_{(1)} + \tau^i_{(2)} + \tau^i_{(3)} + \tau^i_{(0)}. \tag{1}$$

The definitions of these 13 operators were given in Ref. 6. Introduce their linearly independent combinations

$$X^1 = (1/\sqrt{6})(\mathscr{B}_1 - (1/2\sqrt{2})\mathscr{D}_2),$$
$$X^2 = -(1/\sqrt{6})(\mathscr{B}_2 + (1/2\sqrt{2})\mathscr{D}_1),$$
$$X^3 = \mathscr{C}_3,$$
$$X^4 = \tfrac{1}{4}(\mathscr{C}_1 + \mathscr{C}_2) + (1/4\sqrt{3})(\mathscr{B}_1 + \mathscr{B}_2) + (1/4\sqrt{6})(\mathscr{D}_2 - \mathscr{D}_1),$$

 0022-2488/81/092066-03$01.00

$$
\begin{aligned}
X^5 &= \tfrac14(\mathscr{C}_2 - \mathscr{C}_1) + (1/4\sqrt{3})(\mathscr{B}_2 - \mathscr{B}_1) \\
&\quad - (1/4\sqrt{6})(\mathscr{D}_1 + \mathscr{D}_2), \\
X^6 &= -\tfrac14(\mathscr{C}_1 + \mathscr{C}_2) - (1/4\sqrt{3})(\mathscr{B}_1 + \mathscr{B}_2) \\
&\quad + (1/4\sqrt{6})(\mathscr{D}_2 - \mathscr{D}_1), \\
X^7 &= \tfrac14(\mathscr{C}_2 - \mathscr{C}_1) + (1/4\sqrt{3})(\mathscr{B}_1 - \mathscr{B}_2) \\
&\quad + (1/4\sqrt{6})(\mathscr{D}_1 + \mathscr{D}_2), \\
X^8 &= \mathscr{B}_3, \quad Z = \mathscr{G}, \quad W = \mathscr{F}, \\
Y^1 &= (1/4\sqrt{3})(\mathscr{A}_1 - \mathscr{D}_1), \\
Y^2 &= (1/4\sqrt{3})(\mathscr{A}_2 - \mathscr{D}_2), \\
Y^3 &= (1/8\sqrt{3})\mathscr{E},
\end{aligned}
\tag{2}
$$

where the operators $\mathscr{A}_i, \mathscr{B}_i, \mathscr{C}_i, \mathscr{D}_i, \mathscr{E}, \mathscr{F}, \mathscr{G}$ were defined by Eqs. (29)–(41) in Ref. 6. It is easy to verify that the operators $X^\alpha, \alpha = 1,2,\ldots,8$ and Y^i, $i = 1,2,3$ satisfy the same commutation relations as those for $\frac12\lambda^\alpha$ and $\frac12\tau^i$, resp.; X^α commute with Z, W commute with all these operators. For the fundamental representation 3 of the SU(3) algebra or the fundamental representation with $j = \frac12$ of the SU(2) algebra we have, resp., $x^\alpha \to \frac12\lambda^\alpha$, $\alpha = 1,2,\ldots,8$, $\quad Y^i \to \frac12\tau^i$, $i = 1,2,3$.

III. DECOMPOSITION OF THE NATURAL REPRESENTATION

Each irreducible representation of the SU(3) ⊕ SU(2) algebra can be characterized by two numbers (m,j), the integer $m = 1,3,\bar{3},6,\bar{6},8,10, \overline{10},27\cdots$ specifying the irreducible representation of the SU(3) subalgebra and the half-integer $j = 0,\frac12,1,\frac32$ determining that of the SU(2) subalgebra. The basis vectors in the irreducible representation (m,j) will be denoted by

${}^{(m,j)}\Phi^{(\sigma)}, {}^{(m,j)}\Phi'^{(\sigma)}, {}^{(m,j)}\psi_i^{(\sigma)}$, etc.,

where the meaning of the indices σ, i will be explained later.

We have calculated the matrix elements of the operators X^α, $\alpha = 1,2,3\ldots,8$ and Y^i, $i = 1,2,3$ in the natural basis $\psi_{a_1a_2a_3a_4}, a_n = 1,2$, of the 16-dimensional vector space of the natural representation of the given color charge algebra. It can be shown that this 16-dimensional space is split into the invariant subspaces with following basis vectors:

$$
\left\{
\begin{aligned}
{}^{(\bar{3},0)}\psi_1^{(1)} &= (e^{-i\frac14\pi}/\sqrt{3})\big[\psi_{1121} + (-\tfrac12 + i\tfrac12\sqrt{3})\psi_{2111} \\
&\qquad + (-\tfrac12 - i\tfrac12\sqrt{3})\psi_{1211}\big], \\
{}^{(\bar{3},0)}\psi_2^{(1)} &= (e^{i\frac14\pi}/\sqrt{3})\big[\psi_{1121} + (-\tfrac12 - i\tfrac12\sqrt{3})\psi_{2111} \\
&\qquad + (-\tfrac12 + i\tfrac12\sqrt{3})\psi_{1211}\big], \\
{}^{(\bar{3},0)}\psi_3^{(1)} &= (1/\sqrt{12})[3\psi_{1112} - (\psi_{1121} + \psi_{2111} + \psi_{1211})],
\end{aligned}
\right.
$$

$$
\left\{
\begin{aligned}
{}^{(\bar{3},0)}\psi_1^{(0)} &= (e^{-i\frac14\pi}/\sqrt{6})[(\psi_{2211} - \psi_{1122}) \\
&\qquad + (-\tfrac12 + i\tfrac12\sqrt{3})(\psi_{1221} - \psi_{2112}) \\
&\qquad + (-\tfrac12 - i\tfrac12\sqrt{3})(\psi_{2121} - \psi_{1212})], \\
{}^{(\bar{3},0)}\psi_2^{(0)} &= (e^{i\frac14\pi}/\sqrt{6})[(\psi_{2211} - \psi_{1122}) \\
&\qquad + (-\tfrac12 - i\tfrac12\sqrt{3})(\psi_{1221} - \psi_{2112}) \\
&\qquad + (-\tfrac12 + i\tfrac12\sqrt{3})(\psi_{2121} - \psi_{1212})], \\
{}^{(\bar{3},0)}\psi_3^{(0)} &= (1/\sqrt{6})[(\psi_{2211} - \psi_{1122}) + (\psi_{1221} - \psi_{2112}) \\
&\qquad + (\psi_{2121} - \psi_{1212})],
\end{aligned}
\right.
\tag{3}
$$

$$
\left\{
\begin{aligned}
{}^{(\bar{3},0)}\psi_1^{(-1)} &= (e^{-i\frac14\pi}/\sqrt{3})[\psi_{2212} \\
&\qquad + (-\tfrac12 + i\tfrac12\sqrt{3})\psi_{1222} + (-\tfrac12 - i\tfrac12\sqrt{3})\psi_{2122}], \\
{}^{(\bar{3},0)}\psi_2^{(-1)} &= (e^{i\frac14\pi}/\sqrt{3})[\psi_{2212} \\
&\qquad + (-\tfrac12 - i\tfrac12\sqrt{3})\psi_{1222} + (-\tfrac12 + i\tfrac12\sqrt{3})\psi_{2122}], \\
{}^{(\bar{3},0)}\psi_3^{(-1)} &= (1/\sqrt{12})[3\psi_{2221} - (\psi_{2212} + \psi_{1222} + \psi_{2122})],
\end{aligned}
\right.
$$

$$
\left\{
\begin{aligned}
{}^{(1,0)}\Phi^{(2)} &= \psi_{1111}, \\
{}^{(1,0)}\Phi^{(1)} &= \tfrac12[\psi_{2111} + \psi_{1211} + \psi_{1121} + \psi_{1112}], \\
{}^{(1,0)}\Phi^{(0)} &= (1/\sqrt{6})[\psi_{2211} + \psi_{1122} + \psi_{1221} \\
&\qquad + \psi_{2112} + \psi_{2121} + \psi_{1212}], \\
{}^{(1,0)}\Phi^{(-1)} &= \tfrac12[\psi_{2221} + \psi_{2212} + \psi_{2122} + \psi_{1222}], \\
{}^{(1,0)}\Phi^{(-2)} &= \psi_{2222},
\end{aligned}
\right.
\tag{4}
$$

$$
\left\{
\begin{aligned}
{}^{(1,\frac12)}\psi_1 &= (1/\sqrt{6})[(\psi_{2211} + \psi_{1122}) \\
&\qquad + (-\tfrac12 + i\tfrac12\sqrt{3})(\psi_{1221} + \psi_{2112}) \\
&\qquad + (-\tfrac12 - i\tfrac12\sqrt{3})(\psi_{2121} + \psi_{1212})], \\
{}^{(1,\frac12)}\psi_2 &= (1/\sqrt{6})[(\psi_{2211} + \psi_{1122}) \\
&\qquad + (-\tfrac12 - i\tfrac12\sqrt{3})(\psi_{1221} + \psi_{2112}) \\
&\qquad + (-\tfrac12 + i\tfrac12\sqrt{3})(\psi_{2121} + \psi_{1212})].
\end{aligned}
\right.
\tag{5}
$$

In each three-dimensional subspace with three basis vectors ${}^{(\bar{3},0)}\psi_1^{(\sigma)}, {}^{(\bar{3},0)}\psi_2^{(\sigma)}, {}^{(\bar{3},0)}\psi_3^{(\sigma)}$ for each given value of the index σ, the matrix elements of the operators X^α are identical to those of the matrices $-\frac12(\lambda^\alpha)^T$;

$$X^\alpha \to -\tfrac12(\lambda^\alpha)^T,$$

while the operators Y^i vanish,

$$Y^i \to 0.$$

Therefore we have three irreducible representations with $m = \bar{3}$ and $j = 0$.

In the two-dimensional subspace with the basis vectors ${}^{(1,\frac12)}\psi_1$ and ${}^{(1,\frac12)}\psi_2$, the matrix elements of the operators Y^i are identical to those of the matrices $\frac12\tau^i$,

$$Y^i \to \tfrac12\tau^i,$$

while those of X^α are equal to zero. This is the space of the irreducible representation with $m = 1, j = \frac12$.

Acting on each basis vector ${}^{(1,0)}\Phi^{(\sigma)}$ the operators X^α and Y^i give zero: Each of these five vectors corresponds to a one-dimensional representation (1,0) of the SU(3) ⊕ SU(2) subalgebra.

Let us consider the transformations of the basis vectors (3)–(5) under the SU(2) algebra generated by the operators S^i, $i = 1,2,3$, defined according to Eq. (1). Denote be s the half-integer characterizing each irreducible representation of this algebra. It is easy to verify that five vectors ${}^{(1,0)}\Phi^\sigma$, $\delta = -2, -1,0,1,2$, span the invariant subspace of the irreducible representation with $s = 2$ of the SU(2) algebra with generators S^i; each triplet ${}^{(3,0)}\psi_i^{(\sigma)}, \sigma = -1,0,1$ for each given $i = 1,2$, or 3 is an irreducible representation with $s = 1$ of this algebra, while ${}^{(1,\frac{1}{2})}\psi_1$ and ${}^{(1,\frac{1}{2})}\psi_2$ are two singlets with $s = 0$. The index σ is equal to the corresponding eigenvalue of the operator S^3.

IV. CONCLUDING REMARKS

In connection with the conclusion in Refs. 5,6 that the three-quark U(2) color charge algebra contains an SU(3) ⊕ SU(2) ⊕ U(1) subalgebra, it is natural to ask whether the given color charge algebra is related to the physical SU(3) ⊗ SU(2) ⊗ U(1) symmetry, SU(3) being the color group of QCD and SU(2) ⊗ U(1) being the broken symmetry group of the Weinberg–Salem theory of the electroweak interaction. To answer this question, it is necessary to study the decomposition of the natural representation of the color charge algebra into the irreducible representations of the SU(3) ⊕ SU(2) subalgebra.

We have shown that the natural 16-dimensional representation of the given color charge algebra contains five singlets, three SU(3) triplets, and one SU(2) doublet, each component of an SU(3) triplet being an SU(2) singlet and each component of the SU(2) doublet being an SU(3) singlet. If we identify the SU(3) subalgebra with that of the color symmetry group in QCD, then there are two alternatives: either the SU(2) ⊕ U(1) subalgebra cannot be identified with that of the Weinberg–Salem theory or we cannot assign the members of the triplets as the left handed u and d quarks. This means that the SU(3) ⊕ SU(2) ⊕ U(1) subalgebra of the given color charge algebra cannot be identified with that of the physical symmetry group SU(3) ⊗ SU(2) ⊗ U(1).

The physical implications of the particle assignments of above SU(3) ⊗ SU(2) multiplets will be discussed elsewhere.

[1]S. L. Adler, Phys. Rev. D **17**, 3212 (1978).
[2]H. Harari, Phys. Lett. B **86**, 83 (1979).
[3]M. A. Shupe, Phys. Lett. B **86**, 87 (1979).
[4]S. L. Adler, "Quaternionic Chromodynamics as a Theory of Composite Quarks and Leptons," preprint, Princeton University, December, 1979.
[5]P. Cvitanovic, R. J. Gonsalves, and D. E. Neville, Phys. Rev. D **18**, 3881 (1978).
[6]V. Rittenberg and D. Wyler, Phys. Rev. D **18**, 4806 (1978).
[7]S. C. Lee, Phys. Rev. D **20**, 1951 (1979).

Higgs multiplets in the adjoint representation of SU(*N*)

P. Mitra
Tata Institute of Fundamental Research, Homi Bhabha Road, Bombay 400 005, India

(Received 10 April 1980; accepted for publication 20 June 1980)

The general pattern of symmetry breaking due to two Higgs multiplets in the adjoint representation of SU(N) is studied. It is shown that when $N > 2$, the requirement of renormalizability raises the minimum number of multiplets required for a complete breakdown from two to three.

PACS numbers: 11.30.Qc, 11.30.Ly, 02.20.Qs, 11.10.Np

1. INTRODUCTION

Ever since the discovery of the Higgs mechanism for generating masses for gauge bosons[1] there has been a lot of interest in the ways in which various gauge groups can be broken by Higgs fields in various representations. The pioneering in this direction was done by Li,[2] who considered several groups and studied Higgs fields in quite a few irreducible representations. The results of this and related papers have been applied to gauge models attempting to unify different interactions between fundamental particles. While the weak and the electromagnetic interactions have now been definitely united, there is still no universally accepted gauge model encompassing all nongravitational phenomena. Thus it is that interest still survives in general questions of symmetry breaking by Higgs fields.[3]

Recently, O'Raifeartaigh *et al.*[4] have shown that Higgs fields in the direct sum of any number of adjoint representations in a renormalizable SU(N) gauge theory can acquire vacuum expectation values only in certain special directions. We may point out in this connection that the adjoint representation of SU(N), where $N > 2$, is not isotropic. In other words, the representation admits vectors having different little-groups. The special directions chosen by Higgs fields correspond to rather large little-groups. We shall refer to vectors in such directions as H-vectors (precise definition in Sec. 2). Apart from the physical significance of H-vectors, they are mathematically interesting in their own right. We shall not go into their geometric properties, however (see Michel and Radicati[5]).

In this paper we shall be interested in the joint little-group (i.e., the intersection of the little-groups) of two or three H-vectors. For two vectors in the adjoint representation of SU(N), the joint little-group can, in general, be trivial.[6] We shall show that if the vectors happen to be H-vectors, and if $N > 2$, the joint little-group cannot be trivial. Indeed we shall derive the most general form of this group. Then we shall demonstrate that three H-vectors can have little-groups with a trivial common intersection. In other words, we shall study the most general symmetry breaking pattern with *two* Higgs multiplets in the adjoint representation of SU(N) and show that if $N > 2$, *three* are needed for a complete breakdown[6] of the symmetry. The requirement of renormalizability of the gauge theory, as imposed through our use of the results of O'Raifeartaigh *et al.*,[4] is crucial in drawing this conclusion. Without the requirement, *two* Higgs multiplets in the adjoint representation suffice to break SU(N) completely.

2. *H*-VECTORS IN THE ADJOINT REPRESENTATION OF SU(*N*)

The adjoint representation of SU(N) is conveniently defined on the $(N^2 - 1)$-dimensional real vector space of traceless, Hermitian $N \times N$ matrices. The action of an element g of the group on a vector of this space, say the matrix x, is given by

$$x \to U(g) x U(g)^{-1},$$

where $U(g)$ is the representation of g in a fundamental (N-dimensional) representation. The little group of x is determined by the degrees of degeneracy of its eigenvalues: If there are n_1 eigenvalues of one value, n_2 of another, and so on, with $n_1 + n_2 + \cdots = N$, the little-group of x is $S[U(n_1) \otimes U(n_2) \otimes \cdots]$. Thus it is convenient to classify the matrices x according to the degrees of degeneracy of their eigenvalues. We shall be interested in those matrices x which have exactly two distinct eigenvalues. These will be called H-vectors.[7] If one eigenvalue occurs n_1 times, and the other n_2 times ($n_1 + n_2 = N$), we shall say that the H-vector is of type (n_1, n_2), without distinguishing between different orderings of the two numbers.

The importance of H-vectors in the context of the Higgs mechanism is indicated by the following theorem.[4] When in a renormalizable SU(N) gauge theory with one or more Higgs multiplets in adjoint representations, the Higgs potential attains its absolute minimum, the Higgs fields become either zero or H-vectors. The residual symmetry group is simply the subgroup of SU(N) transformations which leave all these H-vectors invariant. We are thus led to the study of the joint little-group of a set of H-vectors.

For a single H-vector of type (n_1, n_2), the little-group is $S[U(n_1) \otimes U(n_2)]$. For two H-vectors, the form of the joint little-group depends on their relative direction. In the next section we discuss the general case. Section 4 contains some remarks about the case of three H-vectors.

3. JOINT LITTLE-GROUP OF TWO *H*-VECTORS

We start with the observation that a H-vector H, since it has two eigenvalues, must satisfy a quadratic equation, say

$$H^2 + aH + b = 0. \tag{1}$$

Let H' be a second H-vector, satisfying

 0022-2488/81/092069-04$01.00

$$H'^2 + a'H' + b' = 0. \tag{2}$$

Let H have the two distinct eigenvalues c_1 and c_2, occurring respectively n_1 and n_2 times, and let c'_1, c'_2, n'_1, and n'_2 be the corresponding quantities for H'. Of course c_1 and c_2 can be expressed in terms of a and b, but we do not need their expressions. It may be noted that $n_1 + n_2 = n'_1 + n'_2 = N$.

We construct the traceless Hermitian matrix

$$x = aH' + a'H + (HH' + H'H - 2N^{-1}\,\mathrm{tr}HH'). \tag{3}$$

It may vanish, but this will not affect any of our arguments. We can see that it always commutes with H and H': thus,

$$[x,H] = a[H',H] + [H',H^2] = [H', -b] = 0.$$

So x and H can be simultaneously diagonalized. Let Φ be a common nonzero eigenvector, with eigenvalues x_0 and c_0, say ($c_0 = c_1$ or c_2). We shall show that the space spanned by Φ and $\Phi' = H'\Phi$ is left invariant by H and H'. That the actions of H and H' on Φ do not produce anything outside this space is trivial to see. As regards Φ', we see, using (2), that

$$H'\Phi' = -a'\Phi' - b'\Phi,$$

while (3) gives

$$H\Phi' = (x_0 - a'c_0 + 2N^{-1}\,\mathrm{tr}HH')\Phi - (a + c_0)\Phi'. \tag{4}$$

Thus we have found a subspace of the N-dimensional representation space such that it is left invariant by H and H' and is of dimensionality at most two. Since H and H' are Hermitian, the orthogonal subspace too is left invariant by them. If this subspace is nontrivial, we can pick out a common eigenvector of x and H and repeat the above procedure. By going through the same steps a finite number of times, we can decompose the N-dimensional space into a direct sum of orthogonal subspaces each of which is left invariant by H and H' and is of at most two dimensions.

Consider any of the two-dimensional invariant subspaces obtained in this way, spanned as above by Φ and Φ'. The fact that it is two-dimensional means that Φ' is linearly independent of Φ. It is easy to see that the restriction of the matrix H' to this subspace has to have two distinct eigenvalues: Otherwise it would be proportional to the identity in this subspace and Φ' would be linearly dependent on Φ. In fact the restriction to this subspace of H too must have two distinct eigenvalues. If it had only the eigenvalue c_0, we would have

$$H\Phi' = c_0\Phi'$$

which, for consistency with (4), would require that

$$c_0 = -a/2.$$

But (1) shows that the sum of eigenvalues of H is $-a$, so that if one is $-a/2$, the other has to have the same value, contradicting the datum that H has two distinct eigenvalues. We also see that the restrictions of H, H' to the two-dimensional subspace cannot commute: If they did, Φ would be an eigenvector of H', so that Φ' would be linearly dependent on Φ.

Let m be the number of two-dimensional invariant subspaces in our decomposition of the N-dimensional space. Then there are $N - 2m$ one-dimensional invariant subspaces. Such subspaces can be of four kinds, corresponding to the four possible combinations of the eigenvalues of H and H'. The number of those where H has the value c_i and H' the value c'_j ($i = 1, 2; j = 1, 2$) will be denoted by n_{ij}. Since, the total number of eigenvectors of H corrsponding to the eigenvalue c_i is n_i, and m of these go into the two-dimensional invariant subspaces, we must have

$$n_i = m + n_{i1} + n_{i2}. \tag{5}$$

Similarly,

$$n'_i = m + n_{1i} + n_{2i}. \tag{6}$$

These four equations are not all independent. The sum of the two equations in (5) gives the same result as the sum of the two equations in (6). Between the five quantities m, n_{ij}, there are thus three relations, so that only two of them are independent. This, however, does not mean that two can be chosen completely arbitrarily: there are the obvious restrictions $m \geqslant 0$, $n_{ij} \geqslant 0$.

Now, we are well set to determine the most general forms of H and H', and then the structure of their joint little-group. We choose a basis for our N-dimensional representation space as follows. First we take n_{11} normalized vectors, one from each of the n_{11} one-dimensional invariant subspaces where H, H' have the respective values c_1, c'_1. These are followed by n_{12} normalized vectors taken similarly from the subspaces where the corresponding eigenvalues are c_1, c'_2. Next we take m normalized eigenvectors of H with the eigenvalue c_1, one being selected from each of the two-dimensional invariant subspaces. After these come normalized eigenvectors of H with the eigenvalue c_2 from the same subspaces but *in the reverse order*. Next we place n_{21} normalized vectors, one from each of the one-dimensional invariant subspaces where H, H' take the respective values c_2, c'_1, and finally n_{22} normalized vectors taken similarly from the one-dimensional spaces with the corresponding values c_2, c'_2. This completes our basis of N orthonormal vectors. If we write down H, H' using this basis, the matrices will look like the following:

$$H = \begin{pmatrix} H_1 & & & & & \\ & H_2 & & & & \\ & & H_3 & & & \\ & & & H_4 & & \\ & & & & H_5 & \\ & & & & & H_6 \end{pmatrix}$$

$$H' = \begin{pmatrix} H'_1 & & & & & \\ & H'_2 & & & & \\ & & H'_3 & E & & \\ & & E^\dagger & H'_4 & & \\ & & & & H'_5 & \\ & & & & & H'_6 \end{pmatrix}.$$

Here H_1, H_2, H_3, H_4, H_5, H_6 are square submatrices, with n_{11}, n_{12}, m, m, n_{21}, n_{22} rows, respectively. The corresponding primed matrices have the same sizes. H is totally diagonal and all diagonal elements in H_1, H_2, H_3 are c_1, while all diagonal elements in H_4, H_5, H_6 are c_2. H'_1, H'_2, H'_5, H'_6 are constant matrices with eigenvalues c'_1, c'_2, c'_1, c'_2 respectively. The matrices H'_3, H'_4 also are diagonal and real. E is an $m\times m$ matrix which has no vanishing element on the diagonal joining its top right-hand corner with the bottom left hand corner, and no nonvanishing element elsewhere. These properties of E follow easily from what we have said above about the two-dimensional invariant subspaces.

In order to determine what matrices commute with H and H', we apply the simple rule that they should not connect eigenvectors (of H or H') having unequal eigenvalues. Thus a matrix commuting with H must be of the form

$$\left(\begin{array}{c|c} P & \\ \hline & Q \end{array}\right),$$

where P, Q are respectively $(n_{11}+n_{12}+m)\times(n_{11}+n_{12}+m)$ and $(m+n_{21}+n_{22})\times(m+n_{21}+n_{22})$ submatrices. If the matrix is to commute with H' as well, it must have further restrictions, e.g., it must not connect vectors in the subspace corresponding to H'_1 (H'_5) with those in the subspace corresponding to H'_2 (H'_6). Before determining the other restrictions, it is convenient to write down the general form of a Hermitian matrix that obeys the ones so far found

$$\begin{pmatrix} L_1 & 0 & M_1 & & & \\ 0 & L_2 & M_2 & & & \\ M_1^\dagger & M_2^\dagger & L_3 & & & \\ & & & L_4 & M_3 & M_4 \\ & & & M_3^\dagger & L_5 & 0 \\ & & & M_4^\dagger & 0 & L_6 \end{pmatrix}.$$

Here, of course, $L_1, L_2, L_3, L_4, L_5, L_6$ are square Hermitian matrices with $n_{11}, n_{12}, m, m, n_{21}, n_{22}$ rows, respectively. M_1, M_2, M_3, M_4 will not be square matrices in general. The composite matrix clearly commutes with H but if it is to commute with H' as well, the following additional relations have to be satisfied:

$$M_1E=0,\quad M_2E=0,\quad EM_3=0,\quad EM_4=0, \tag{7}$$

$$M_1(H'_3-c'_1)=0,\quad M_2(H'_3-c'_2)=0,$$

$$(H'_4-c'_1)M_3=0,\quad (H'_4-c'_2)M_4=0, \tag{8}$$

$$L_3E=EL_4, \tag{9}$$

$$[H'_3,L_3]=0, \tag{10a}$$

$$[H'_4,L_4]=0. \tag{10b}$$

Now, E is invertible, so that (7) yields

$$M_1=0, M_2=0, M_3=0, M_4=0.$$

Then (8) is automatically satisfied. From (9) we obtain

$$L_4=E^{-1}L_3E. \tag{11}$$

(10a) is nontrivial, but we shall show that once it is satisfied, (11) enforces (10b). Consider the submatrix

$$\begin{pmatrix} H'_3 & E \\ E^\dagger & H'_4 \end{pmatrix},$$

and recall how it was constructed. The rr element of H'_3 ($1\leqslant r\leqslant m$) and the ss elements of H'_4, where $s=m+1-r$, are the two diagonal elements of the restriction of H' to a two-dimensional invariant subspace. The sum of these is therefore the sum of the eigenvalues of this 2×2 matrix, which is the sum of the two eigenvalues of H', i.e., $-a'$. Thus, H'_4 is obtained from H'_3 by subtracting from $-a'$ and reversing the order of the diagonal elements. Noting that if D is any diagonal matrix, $E^{-1}DE$ is what is obtained by reversing the order of the elements of D, we can write

$$H'_4=-a'-E^{-1}H'_3E,$$

which, together with (10a) and (11), implies (10b).

Hence the most general Hermitian matrix commuting with both H and H' is of the form

$$\begin{pmatrix} L_1 & & & & & \\ & L_2 & & & & \\ & & L_3 & & & \\ & & & L_4 & & \\ & & & & L_5 & \\ & & & & & L_6 \end{pmatrix} \tag{12}$$

where L_1, L_2, L_5, L_6 are arbitrary Hermitian matrices of dimensions $n_{11}\times n_{11}$, $n_{12}\times n_{12}$, $n_{21}\times n_{21}$, $n_{22}\times n_{22}$, respectively, L_3 is an $m\times m$ Hermitian matrix commuting with H'_3 and L_4 is another such matrix determined in terms of L_3 by E. In order to know what matrices L_3 are allowed, it is necessary to know the frequencies of repetition of the different diagonal elements of H'_3. In general, nothing more can be said about H'_3 than that it is diagonal and real. Correspondingly, all that can be said about the set of allowed Hermitian matrices L_3 is that it includes the set of real, diagonal $m\times m$ matrices. Let us denote by G the group of unitary $m\times m$ matrices which leave H'_3 unchanged, so that G is some subgroup of $U(m)$ containing a Cartan subgroup. Then it is clear that the group generated by traceless, Hermitian matrices commuting with H and H' is $S[U(n_{11})\otimes U(n_{12})\otimes G \otimes U(n_{21})\otimes U(n_{22})]$. This is the most general form of the joint little-group of two H-vectors. The numbers n_{11}, n_{12}, m, n_{21}, n_{22} have to satisfy the restrictions given above. Some of them could be zero, in which case the corresponding factor could be omitted. For example, if $[H,H']=0$, m vanishes and the factor G drops out. Obviously, all the factors cannot drop out simultaneously. Indeed, the rank of the joint little-group is $n_{11}+n_{12}+m+n_{21}+n_{22}-1$
$=N-m-1\geqslant N-1-\min(n_1,n_2,n'_1,n'_2)$, which sets a lower bound on the size of this group. For $N>2$, it has to be a nontrivial group.

4. *H*-VECTORS WITH TRIVIALLY INTERSECTING *N* LITTLE-GROUPS

It is well known that the intersection of the little-groups of two vectors in the adjoint representation of SU(N) can be trivial. That is to say, there exist two traceless Hermitian matrices such that no nonzero, traceless, Hermitian matrix commutes with both of them. For example, one can be taken to be diagonal, with N distinct eigenvalues, and the other to have all nondiagonal elements nonvanishing. Then any matrix which commutes with the former has to be diagonal, and if it is to commute with the latter, its diagonal elements must all be equal, so that it has to be proportional to the identity matrix; the tracelessness condition then forces it to be zero. Our results of the previous section show that if $N > 2$, one can always find nonzero traceless Hermitian matrices commuting with any two given H-vectors of SU(N). What, then, is the smallest number of traceless Hermitian matrices for which it is not always possible to find a common, commuting, nonzero, traceless, Hermitian matrix? The answer is three, as we shall now demonstrate.

We first look for two H-vectors, H, H', such that the only Hermitian matrices that commute with them are diagonal. Referring to the preceding section, we see that this happens when none of the numbers $n_{11}, n_{12}, n_{21}, n_{22}$ exceeds unity and H'_3 has m distinct eigenvalues. The first requirement makes the submatrices L_1, L_2, L_5, L_6 in (12) diagonal (or vanishing altogether) while the second one forces L_3 to be diagonal. Then it follows from (11) that L_4 too is diagonal.

If we can now find an H-vector H'' which is such that the only diagonal matrices which commute with it are the ones proportional to the identity, the only traceless Hermitian matrix commuting with H, H', and H'' will be zero. For this purpose, H'' has to have so many vanishing nondiagonal elements that given any two indices a,b ($1 \leqslant a, b \leqslant N$), there will exist numbers $c_1, c_2, \ldots, c_n$ in the same range such that the $(ac_1), (c_1c_2), \ldots, (c_{n-1}c_n), (c_nb)$ matrix elements of H'' are nonvanishing. Do H-vectors with such a property exist? The answer is yes. A trivial example is the $(N-1,1)$ H-vector which has all its diagonal elements equal to zero and all others equal to 1.

If one wishes to have all three H-vectors of the same type, and if $N \geqslant 4$, the H'' chosen above is not admissible: for $N \geqslant 4$, two $(N-1,1)$ H-vectors cannot have all the numbers $n_{11}, n_{12}, n_{21}, n_{22}$ less than 2. We can however make H, H', and H'' of the type (p,p) or $(p+1,p)$ depending on whether N is even $(=2p)$ or odd $(=2p+1)$. In the former case one can have $m = p$, $n_{11} = n_{12} = n_{21} = n_{22} = 0$, while in the latter case a possible combination is $m = p$, $n_{11} = 1$, $n_{12} = n_{21} = n_{22} = 0$. Of course, in addition to this it is necessary to ensure that H'_3 has all eigenvalues distinct; but this is easily done. As regards H'', one can start from a diagonal (p,p) or $(p+1,p)$ H-vector and make small SU(2) rotations in the $(1,2),(2,3),\ldots,(N-1,N)$ subspaces. It is straightforward to show that an H-vector can be constructed in this way with nonvanishing $(1,2),(2,3),\ldots,(N-1,N)$ matrix elements. Then the only traceless Hermitian matrix to commute with H, H', and H'' will be zero.

5. DISCUSSION

We have studied the general case of the spontaneous breakdown of an SU(N) gauge symmetry caused by two Higgs multiplets in the adjoint representation and have shown that for $N > 2$ three such multiplets are needed if the symmetry is to be broken completely. It must be emphasized that the result[4] that the minimum of the Higgs potential can occur only when the Higgs fields take zero or H-vector values ignored radiative effects of the kind studied by Coleman and Weinberg.[8] Consequently our conclusions too can be upset by such effects.

Our proof of the fact that three Higgs multiplets in the adjoint representation can break SU(N) fully does not show that they always do. Indeed, it is easily seen from our proof that there can also be a residual $U(1)$ symmetry. It should be possible to enumerate all the subgroups of SU(N) that can be left unbroken when there are three (or more) multiplets. We do not have any general result of this kind.

Our exclusive concern with the adjoint representation for Higgs multiplets is to be attributed to mathematical reasons rather than physical exigencies. Such a choice of Higgs fields can perhaps be motivated by the hope that subsequently these fields will be expressible as dynamical constructs of the gauge bosons. However for other reasons (e.g., for the purpose of generating masses for fermions) model builders have generally preferred to put some Higgs multiplets in fundamental representation. In particular, this is the case with the SU(5) unification[9] of the strong, electromagnetic, and weak interactions. Thus we can only hope that our results will be useful elsewhere.

[1]P. Higgs, Phys. Rev. **145**, 1156 (1966); T. Kibble, Phys. Rev. **155**, 1554 (1967).

[2]L.-F. Li, Phys. Rev. D **9**, 1723 (1974).

[3]See for example, Ph. Combe, A. Sciarrino, and P. Sorba, CERN Report Ref. TH 2609-CERN (1978); also F. Buccella, H. Ruegg, and C. A. Savoy, Nucl. Phys. B **169**, 68 (1980).

[4]L. O'Raifeartaigh, S. Y. Park, and K. C. Wali (to be published); see their Syracuse University Report SU-4217-146 (1979), based on a talk delivered at the University of Southern Illinois Einstein Commemoration Symposium, March 1979. These authors actually have stronger results than the one we quote in the text. Their result for the special case of a single Higgs multiplet in the adjoint representation of SU(N) was known earlier: cf. L.-F. Li, Ref. 2, and L. Michel, CERN Report Ref. TH. 2716-CERN (1979).

[5]L. Michel and L. A. Radicati, Ann. Inst. Henri Poincaré **18**, 185 (1973).

[6]We ignore discrete subgroups of SU(N). Thus when we say that certain subgroups intersect trivially, we mean that the intersection is a discrete subgroup of SU(N), and when we say that SU(N) is broken down completely, it is merely broken down to a discrete subgroup.

[7]In the adjoint representation of SU(3), traceless Hermitian matrices with just two distinct eigenvalues were called q-vectors by Michel and Radicati, Ref. 5. In the SU(N) case, they reserved this name for H-vectors of type $(N-2,2)$. So we had to invent a new name for the general case. Our choice was prompted by the fact that Higgs fields take H-vector values at the minimum of the symmetry breaking potential.

[8]S. Coleman and E. Weinberg, Phys. Rev. **D 7**, 1888 (1973).

[9]H. Georgi and S. L. Glashow, Phys. Rev. Lett. **32**, 438 (1974).

Propagation in a tropospheric duct with a single-step discontinuity in the refractive index in the direction of propagation

R. H. Ott

U.S. Department of Commerce,[a] *National Telecommunications and Information Administration, Institute for Telecommunication Sciences, Boulder, Colorado 80303*

(Received 24 June 1980; accepted for publication 9 January 1981)

A Green's function approach is used to derive an expression for the field inside a laterally inhomogeneous duct. The laterally inhomogeneous duct is assumed to have a single step, and convergence criteria for the step size and number of modes are discussed.

PACS numbers: 41.10.Hv, 84.40.Ed, 92.60. – e

I. INTRODUCTION

The problem of finding the electromagnetic field in a duct with a single-step lateral inhomogeneity is given, using the Green's function method. The extension of the solution for propagation in a uniform medium to a medium composed of steps, to represent the slow variation in refractive index with distance along the direction of propagation, allows a solution for propagation in a laterally inhomogeneous medium. The coupling between normal modes in each region is easily separated out of the solution in the present formulation. The theory could be used to study the problem of propagation of underwater sound waves in shallow water with slowly varying depth.[1] The problem of propagation in a laterally inhomogeneous duct was investigated by Bahar[2] using an interative solution to Maxwell's equations directly.

II. ANALYSIS

The geometry of the propagation problem is shown in Fig. 1. The term duct refers to the concept of the trapping of modes and the resulting propagation over long distances.

Our approach will be to find the Green's function for the bounded waveguide cross section in Fig. 1. In particular, we will analyze the effect of the step size, $(a_2 - a_1)$, and the refractive index contrast, $n_2 - n_1$.

The electric and magnetic fields will satisfy the two-dimensional, time-harmonic, Helmholtz equation

$$\frac{1}{r}\frac{\partial}{\partial r}\left(r\frac{\partial G}{\partial r}\right) + \frac{1}{r^2}\frac{\partial^2 G}{\partial \phi^2} + k^2 G = \frac{\delta(r - r_s)\delta(\phi - \phi_s)}{r}, \quad (1)$$

$$a_0 \leqslant r < \infty, \quad 0 \leqslant \phi \leqslant 2\pi,$$

where (r_s,ϕ_s) and (r,ϕ) are the source and observation coordinates, respectively, k is the wavenumber, and the time dependence is $e^{-i\omega t}$. Considering regions I and II separately, G must meet the periodicity requirement, and we have

$$\frac{1}{r}\frac{\partial}{\partial r}\left(r\frac{\partial G}{\partial r}\right) + \frac{1}{r^2}\frac{\partial^2 G}{\partial \phi^2} + k^2 G = \frac{\delta(r - r_s)}{r}\sum_{n=-\infty}^{\infty}\delta(\phi - \phi_s - 2n\pi). \quad (2)$$

We look for solutions to (2) with a singularity at r_s,ϕ_s on each "Riemann sheet" n as

$$G(\mathbf{r},\mathbf{r}') = \sum_{-\infty}^{\infty} G_\infty(\mathbf{r},\mathbf{r}'_n), \quad \mathbf{r}'_n = (\mathbf{r}',\phi_s + 2n\pi),$$

where G_∞ satisfies

$$\frac{1}{r}\frac{\partial}{\partial r}\left(r\frac{\partial G_\infty}{\partial r}\right) + \frac{1}{r^2}\frac{\partial^2 G_\infty}{\partial \phi^2} + k^2 G_\infty = \frac{\delta(r - r_s)\delta(\phi - \phi_s)}{r}, \quad (3)$$

$$a_0 \leqslant r < \infty, \quad -\infty < \phi < \infty.$$

The completeness relation is

$$G_\infty(\mathbf{r},\mathbf{r}'_n) = \frac{-1}{2\pi i}\oint g_r(r,r_s;\lambda)g_\phi(\phi,\phi_s;\lambda)\,d\lambda, \quad (4)$$

where the contour (counterclockwise) in (4) is selected to enclose all the singularities in the complex λ plane.

If we define

$$\nu = \sqrt{\lambda}, \quad \mathrm{Im}(\nu) > 0$$

$$d\nu = (1/2\sqrt{\lambda})\,d\lambda, \quad (5)$$

then the Green's function g_ϕ on an "infinite" angular transmission line is

$$g_\phi(\phi,\phi_s;\nu^2) = \exp[i\nu(\phi - \phi_s - 2n\pi)]/2i\nu \quad (6)$$

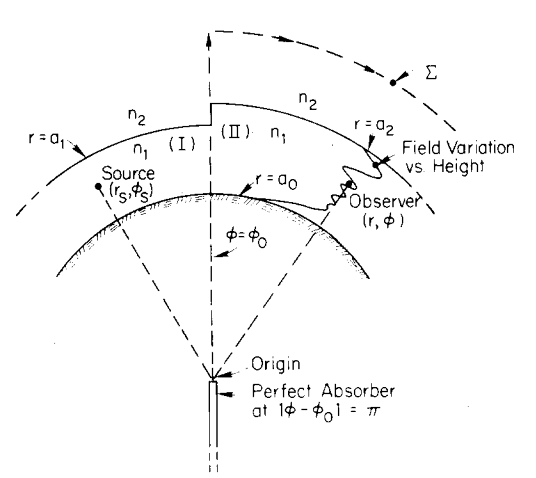


FIG. 1. Geometry for single step discontinuity in a tropospheric duct.

[a] The author is presently with the Consulative Committee for International Radio, International Telecommunications Union 1121 Geneva 20, Switzerland.

with $|g_\phi|\to 0$ as $|\nu|\to\infty$. Substituting (6) in (4) gives

$$G_\infty(\mathbf{r},\mathbf{r}_n') = \frac{1}{2\pi}\int_{-\infty}^{\infty} g_r(r,r_s;\nu^2) \times \exp[i\nu(\phi-\phi_s-2n\pi)]\,d\nu \tag{7}$$

with Fourier inversion

$$g_r(r,r_s;\nu^2) = \int_{-\infty}^{\infty} G_\infty(r,r';\phi,\phi_s+2n\pi)\exp[-i\nu(\phi-\phi_s-2n\pi)]\,d\phi. \tag{8}$$

From (3), g_r satisfies

$$\frac{1}{r}\frac{\partial}{\partial r}\left(r\frac{\partial g_r}{\partial r}\right)+k^2 g_r-\frac{\nu^2}{r^2}g_r=\frac{\delta(r-r_s)}{r}. \tag{9}$$

With a perfect absorber at $|\phi-\phi_s|=\pi$, Fig. 1 corresponds to the case $n=0$ for (7) and (8). Much of the discussion up to this point can be found in the literature.[3]

The field $E(r)$ in region II due to the aperture field in the plane $\phi=\phi_0$ in Fig. 1 is obtained from Green's theorem after integrating over the cylinder Σ at infinity and over the aperture plane $\phi=\phi_0$ yielding

$$E(r)=\int_{a_0}^{\infty}\frac{dr'}{r'}\left[E(r')\frac{\partial G_\infty^{(2)}}{\partial\phi'}-G_\infty^{(2)}\frac{\partial E}{\partial\phi'}\right], \tag{10}$$

where by analogy with the problem to the left of the aperture

$$G_\infty^{(2)}(r,r';\phi,\phi')=\frac{1}{2\pi}\int_{-\infty}^{\infty} g_r^{(2)}(r,r';\nu^2)e^{i\nu(\phi-\phi')}\,d\nu, \tag{11}$$

from which it follows that

$$\frac{\partial G_\infty^{(2)}}{\partial\phi'}=\frac{1}{2\pi i}\int_{-\infty}^{\infty} g_r^{(2)}(r,r';\nu^2)e^{i\nu(\phi-\phi')}\nu\,d\nu. \tag{12}$$

We now make the parabolic wave equation assumption

$$\frac{\partial E}{\partial\phi'}\sim ika_1E(r'). \tag{13}$$

Substituting (12) and (13) into (10) gives

$$E(r)=\int_{a_0}^{\infty}\frac{dr'}{r'}E(r')\left[\frac{1}{2\pi i}\int_{-\infty}^{\infty} g_r^{(2)}(r,r';\nu^2)e^{i\nu(\phi-\phi')}\nu\,d\nu + \frac{ka_1}{2\pi i}\int_{-\infty}^{\infty} g_r^{(2)}(r,r';\nu^2)e^{i\nu(\phi-\phi')}\,d\nu\right], \tag{14}$$

and, because we are interested in solutions when $\nu\cong ka_1$, (14) becomes

$$E(r)=\frac{ka_1}{\pi i}\int_{a_0}^{\infty}\frac{dr'}{r'}E(r')\int_{-\infty}^{\infty} g_r^{(2)}(r,r';\nu^2)e^{i\nu(\phi-\phi')}\,d\nu. \tag{15}$$

Now $E(r')$ is the field in the plane $\phi=\phi_0$ due to the point source at r_s,ϕ_s neglecting reflected fields and is given by

$$E(r')=\frac{1}{2\pi}\int_{-\infty}^{\infty} g_r^{(1)}(r',r_s;\nu^2)e^{i\nu(\phi_0-\phi_s)}\,d\nu. \tag{16}$$

Substituting (16) into (15) gives the solution

$$E(r)=\frac{ka_1}{2\pi^2}\int_{a_0}^{\infty}\frac{dr'}{r'}\int_{-\infty}^{\infty}d\nu_1\int_{-\infty}^{\infty}d\nu_2 \times g_r^{(1)}(r',r_s;\nu_1^2)g_r^{(2)}(r,r';\nu_2^2) \times\exp[i\nu_1(\phi_0-\phi_s)]\exp[i\nu_2(\phi-\phi_0)], \tag{17}$$

which is easily generalized to two-step discontinuities as

$$E(r)=\frac{ka_1}{2\pi^2}\int_{a_0}^{\infty}\frac{dr''}{r''}\int_{a_0}^{\infty}\frac{dr'}{r'}\int_{-\infty}^{\infty}d\nu_1\int_{-\infty}^{\infty}d\nu_2\int_{-\infty}^{\infty}d\nu_3 \times g_r^{(1)}(r',r_s;\nu_1^2)g_r^{(2)}(r',r'',\nu_2^2)g_r^{(3)}(r'',r;\nu_3^2) \times\exp[i\nu_1(\phi_1-\phi_s)]\exp[i\nu_2(\phi_2-\phi_1)] \times\exp[i\nu_3(\phi_2-\phi_0)]. \tag{18}$$

The Greens's function for the source r_s in the duct is given by the "broken" function

$$g(r,r_s;\nu^2)=\begin{cases}T(\nu)[H_\nu^{(1)}(kr_s)+R_1(\nu)H_\nu^{(2)}(kr_s)]H_\nu^{(1)}(\hat{k}r), & a_1\leqslant r<\infty,\\ \alpha(\nu)[H_\nu^{(1)}(kr)+R_2(\nu)H_\nu^{(2)}(kr)][H_\nu^{(1)}(kr_s)+R_1(\nu)H_\nu^{(2)}(kr_s)], & r_s\leqslant r\leqslant a_1,\\ \alpha(\nu)[H_\nu^{(1)}(kr_s)+R_2(\nu)H_\nu^{(2)}(kr_s)][H_\nu^{(1)}(kr)+R_1(\nu)H_\nu^{(2)}(kr)], & a_0\leqslant r\leqslant r_s.\end{cases} \tag{19}$$

In (19) $H_\nu^{(1)}$ and $H_\nu^{(2)}$ are Hankel functions of the first and second kind. a_1 is the height where the refractive index changes from k to $\hat{k}$ where

$$\hat{k}=k(1-\Delta n) \tag{20}$$

and a_0 is the radius of the earth and Δn is the refractive index contrast. The Green's function if the source is outside the duct is

$$\hat{g}(r,r_s;\nu^2)=\begin{cases}\hat{\alpha}(\nu)[H_\nu^{(1)}(\hat{k}r_s)+\hat{R}_2(\nu)H_\nu^{(2)}(\hat{k}r_s)]H_\nu^{(1)}(\hat{k}r), & r_s\leqslant r<\infty,\\ \hat{\alpha}(\nu)[H_\nu^{(1)}(\hat{k}r)+\hat{R}_2(\nu)H_\nu^{(2)}(\hat{k}r)]H_\nu^{(1)}(\hat{k}r_s), & a_2\leqslant r\leqslant r_s,\\ \hat{T}(\nu)[H_\nu^{(1)}(kr)+R_1(\nu)H_\nu^{(2)}(kr)]H_\nu^{(1)}(\hat{k}r_s), & a_0\leqslant r\leqslant a_2,\end{cases} \tag{21}$$

where the height of the duct is now labeled a_2, because, as we see in Fig. 2, the integration point for region II eventually lies outside the duct.

From the jump condition

$$\left.\frac{\partial g}{\partial r}\right|_{r=r_s+\epsilon}-\left.\frac{\partial g}{\partial r}\right|_{r=r_s-\epsilon}=\frac{-1}{2\pi r_s} \tag{22}$$

and the use of the following Wronskian

$$H_\nu^{(1)\prime}(kr_s)H_\nu^{(2)}(kr_s)-H_\nu^{(1)}(kr_s)H_\nu^{(2)\prime}(kr_s)=4i/\pi kr_s, \tag{23}$$

we obtain

$$\alpha(\nu)=-i/8[R_2(\nu)-R_1(\nu)] \tag{24}$$

and the "resonance" condition; i.e., $R_2(\nu)=R_1(\nu)$. From the

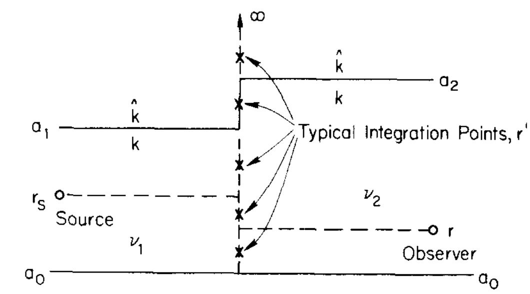


FIG. 2. Geometry for aperture integration. Modes in regions I and II are ν_1 and ν_2, respectively.

boundary conditions at $r = a_1$, i.e.,

$$g(r,r_s;\nu^2)\big|_{r=a_1-\epsilon} = g(r,r_s;\nu^2)\big|_{r=a_1+\epsilon} \tag{25}$$

and

$$\left.\frac{\partial g}{\partial r}\right|_{r=a_1-\epsilon} = \left.\frac{\partial g}{\partial r}\right|_{r=a_1+\epsilon}, \tag{26}$$

we find

$$R(\nu) = -\,[\hat{k}H_\nu^{(1)\prime}(\hat{k}a_1)H_\nu^{(1)}(ka_1) - kH_\nu^{(1)}(\hat{k}a_1)H_\nu^{(1)\prime}(ka_1)]/ \\ [\hat{k}H_\nu^{(1)\prime}(\hat{k}a_1)H_\nu^{(2)}(ka_1) - kH_\nu^{(1)}(\hat{k}a_1)H_\nu^{(2)\prime}(ka_1)], \tag{27}$$

a "reflection" coefficient at the boundary. Similarly, from the boundary condition at $r = a_0$, i.e.,

$$\left.\frac{\partial g}{\partial r}\right|_{r=a_0} = -\,ik\delta g\big|_{r=a_0}, \tag{28}$$

we find

$$R_1(\nu) = -\,[H_\nu^{(1)\prime}(ka_0) + i\delta H_\nu^{(1)}(ka_0)]/ \\ [H_\nu^{(2)\prime}(ka_0) + i\delta H_\nu^{(2)}(ka_0)], \tag{29}$$

where

$$\delta = \begin{cases} (\eta-1)^{1/2}/\eta, & \text{vertical polarization} \\ (\eta-1)^{1/2}, & \text{horizontal polarization} \end{cases} \tag{30}$$

and

$$\eta = \epsilon_r + i\sigma/\omega\epsilon_0, \tag{31}$$

where σ is the ground conductivity in Siemens/m and ϵ_r is the relative dielectric constant. Knowing $R_1(\nu)$, $R_2(\nu)$, and $\alpha(\nu)$ allows us to solve for $T(\nu)$ in (19), again using the continuity of g at $r = a_1$, and we obtain

$$T(\nu) = -\,(i/8)[H_\nu^{(1)}(ka_1) + R_2(\nu)H_\nu^{(2)}(ka_1)]/ \\ [R_2(\nu) - R_1(\nu)]H_\nu^{(1)}(\hat{k}a_1). \tag{32}$$

Similarly, the jump condition for $\hat{g}$ gives

$$\hat{\alpha}(\nu)\hat{R}_2(\nu) = (i/8). \tag{33}$$

From the boundary condition at $r = a_2$, we find

$$\begin{aligned}\hat{R}_2(\nu) = -\,\{&\hat{k}H_\nu^{(1)\prime}(\hat{k}a_2)\,[H_\nu^{(1)}(ka_2) + R_1(\nu)H_\nu^{(2)}(ka_2)] \\ &- kH_\nu^{(1)}(\hat{k}a_2)[H_\nu^{(1)\prime}(ka_2) + R_1(\nu)H_\nu^{(2)\prime}(ka_2)]\}/ \\ \{&\hat{k}H_\nu^{(2)\prime}(\hat{k}a_2)[H_\nu^{(1)}(ka_2) + R_1(\nu)H_\nu^{(2)}(ka_2)] \\ &- kH_\nu^{(2)}(\hat{k}a_2)[H_\nu^{(1)\prime}(ka_2) + R_1(\nu)H_\nu^{(2)\prime}(ka_2)]\}\end{aligned} \tag{34}$$

and

$$\hat{T}(\nu) = \hat{\alpha}(\nu)[H_\nu^{(1)}(\hat{k}a_2) + \hat{R}_2(\nu)H_\nu^{(2)}(\hat{k}a_2)]/ \\ [H_\nu^{(1)}(ka_2) + R_1(\nu)H_\nu^{(2)}(ka_2)]. \tag{35}$$

We now turn our attention to the aperture integration $a_0 \leqslant r' < \infty$ in (17). The geometry of the problem is given in Fig. 2, where we will confine our analysis to the case where $a_0 \leqslant r_s \leqslant a_1$ and $a_0 \leqslant r \leqslant a_2$.

Using the "broken" functions in (19) and (20) with appropriate interpretations for the source and observations points (i.e., in the aperture plane the integration point r' becomes the observation point r in region I while the integration point r' becomes the source point r_s for region II), we have

$$\begin{aligned}\int_{a_0}^{\infty} &\frac{dr'}{r'}\, g_r^{(1)}(r',r_s;\nu_1^2)g_r^2(r,r';\nu_2^2) \\ &= \alpha(\nu_1)\alpha(\nu_2)\psi_{\nu_1}(kr_s)\psi_{\nu_2}(kr)\int_{a_0}^{r}\frac{dr'}{r'}\,\phi_{\nu_1}(kr')\phi_{\nu_2}(kr') \\ &\quad + \alpha(\nu_1)\alpha(\nu_2)\psi_{\nu_1}(kr_s)\phi_{\nu_2}(kr)\int_{r}^{r_s}\frac{dr'}{r'}\,\phi_{\nu_1}(kr')\psi_{\nu_2}(kr') + \alpha(\nu_1)\alpha(\nu_2)\phi_{\nu_1}(kr_s)\phi_{\nu_2}(kr)\int_{r_s}^{a_1}\frac{dr'}{r'}\,\psi_{\nu_1}(kr')\psi_{\nu_2}(kr') \\ &\quad + \alpha(\nu_2)T(\nu_1)\phi_{\nu_1}(kr_s)\phi_{\nu_2}(kr)\int_{a_1}^{a_2}\frac{dr'}{r'}\,H_{\nu_1}^{(1)}(\hat{k}r')\psi_{\nu_2}(kr') + \hat{T}(\nu_2)T(\nu_1)\phi_{\nu_1}(kr_s)\phi_{\nu_2}(kr)\int_{a_2}^{\infty}\frac{dr'}{r'}\,H_{\nu_1}^{(1)}(\hat{k}r')H_{\nu_2}^{(1)}(\hat{k}r'),\end{aligned} \tag{36}$$

where

$$\begin{aligned}\phi_\nu(kr) &= H_\nu^{(1)}(kr) + R_1(\nu)H_\nu^{(2)}(kr), \\ \psi_\nu(kr) &= H_\nu^{(1)}(kr) + R_2(\nu)H_\nu^{(2)}(kr).\end{aligned} \tag{37}$$

From the differential equation (9) for the radial functions and linear combinations as given in (37), we have

$$\begin{aligned}\int_{a_0}^{r} &\frac{dr'}{r'}\,\phi_{\nu_1}(kr')\phi_{\nu_2}(kr') \\ &= \frac{r}{(\nu_1^2-\nu_2^2)}\left[\phi_{\nu_2}(kr)\left.\frac{\partial\phi_{\nu_1}}{\partial r'}\right|_{r'=r} - \phi_{\nu_1}(kr)\left.\frac{\partial\phi_{\nu_2}}{\partial r'}\right|_{r'=r}\right] - \frac{a_0}{(\nu_1^2-\nu_2^2)}\left[\phi_{\nu_2}(ka_0)\left.\frac{\partial\phi_{\nu_1}}{\partial r'}\right|_{r'=a_0} - \phi_{\nu_1}(ka_0)\left.\frac{\partial\phi_{\nu_2}}{\partial r'}\right|_{r'=a_0}\right] \\ &= \frac{r}{(\nu_1^2-\nu_2^2)}\left[\phi_{\nu_2}(kr)\left.\frac{\partial\phi_{\nu_1}}{\partial r'}\right|_{r'=r} - \phi_{\nu_1}(kr)\left.\frac{\partial\phi_{\nu_2}}{\partial r'}\right|_{r'=r}\right]\end{aligned} \tag{38}$$

because $\phi_\nu(kr)$ satisfies (28). Also

$$\int_r^{r_s} \frac{dr'}{r'} \phi_{\nu_1}(kr')\psi_{\nu_2}(kr')$$

$$= \frac{r_s}{(\nu_1^2 - \nu_2^2)} \left[\psi_{\nu_2}(kr_s) \frac{\partial\phi_{\nu_1}}{\partial r'}\bigg|_{r'=r_s} - \phi_{\nu_1}(kr_s) \frac{\partial\psi_{\nu_2}}{\partial r'}\bigg|_{r'=r_s}\right] - \frac{r}{(\nu_1^2 - \nu_2^2)} \left[\psi_{\nu_2}(kr) \frac{\partial\phi_{\nu_1}}{\partial r'}\bigg|_{r'=r} - \phi_{\nu_1}(kr) \frac{\partial\psi_{\nu_2}}{\partial r'}\bigg|_{r'=r}\right], \quad (39)$$

$$\int_{r_s}^{a_1} \frac{dr'}{r'} \psi_{\nu_1}(kr')\psi_{\nu_2}(kr')$$

$$= \frac{a_1}{(\nu_1^2 - \nu_2^2)} \left[\psi_{\nu_2}(ka_1) \frac{\partial\psi_{\nu_1}}{\partial r'}\bigg|_{r'=a_1} - \psi_{\nu_1}(ka_1) \frac{\partial\psi_{\nu_2}}{\partial r'}\bigg|_{r'=a_1}\right] - \frac{r_s}{(\nu_1^2 - \nu_2^2)}\left[\psi_{\nu_2}(kr_s) \frac{\partial\psi_{\nu_1}}{\partial r'}\bigg|_{r'=r_s} - \psi_{\nu_1}(kr_s) \frac{\partial\psi_{\nu_2}}{\partial r'}\bigg|_{r'=r_s}\right]. \quad (40)$$

The following integral involves both the wavenumbers k and $\hat{k}$; i.e.,

$$\int_{a_1}^{a_2} \frac{dr'}{r'} H_{\nu_1}^{(1)}(\hat{k}r')\psi_{\nu_2}(kr') = \frac{a_2}{(\nu_1^2 - \nu_2^2)} \left[\psi_{\nu_2}(ka_2) \frac{\partial H_{\nu_1}^{(1)}(\hat{k}r')}{\partial r'}\bigg|_{r'=a_2} - H_{\nu_1}^{(1)}(\hat{k}a_2)\frac{\partial\psi_{\nu_2}}{\partial r'}\bigg|_{r'=a_2}\right]$$
$$- \frac{a_1}{(\nu_1^2 - \nu_2^2)} \left[\psi_{\nu_2}(ka_1) \frac{\partial H_{\nu_1}^{(1)}(\hat{k}r')}{\partial r'}\bigg|_{r'=a_1} - H_{\nu_1}^{(1)}(\hat{k}a_1) \frac{\partial\psi_{\nu_2}}{\partial r'}\bigg|_{r'=a_1}\right]$$
$$+ \frac{2\Delta n k^2}{(\nu_1^2 - \nu_2^2)} \int_{a_1}^{a_2} r'\, dr'\, H_{\nu_1}^{(1)}(\hat{k}r')\psi_{\nu_2}(kr') \quad (41)$$

and finally,

$$\int_{a_2}^{\infty} \frac{dr'}{r'} H_{\nu_1}^{(1)}(\hat{k}r')H_{\nu_2}^{(1)}(\hat{k}r')$$

$$= -\frac{a_2}{(\nu_1^2 - \nu_2^2)} \left[H_{\nu_2}^{(1)}(\hat{k}a_2) \frac{\partial H_{\nu_1}^{(1)}(\hat{k}r')}{\partial r'}\bigg|_{r'=a_2} - H_{\nu_1}^{(1)}(\hat{k}a_2)\frac{\partial H_{\nu_2}^{(1)}(\hat{k}r')}{\partial r'}\bigg|_{r'=a_2}\right]. \quad (42)$$

Substituting (38)–(42) into (36) gives

$$\int_{a_0}^{\infty} \frac{dr'}{r'} g_r^{(1)}(r',r_s;\nu_1^2) g_r^{(2)}(r,r';\nu_2^2)$$

$$= \left(\frac{-1}{64}\right) \frac{1}{[R_2(\nu_2) - R_1(\nu_2)]\,[R_2(\nu_1) - R_1(\nu_1)](\nu_1^2 - \nu_2^2)} \Bigg\{\psi_{\nu_1}(kr_s)\phi_{\nu_1}(kr) r\left[\phi_{\nu_2}(kr) \frac{\partial\psi_{\nu_2}}{\partial r'}\bigg|_{r'=r} - \psi_\nu(kr) \frac{\partial\phi_{\nu_2}}{\partial r'}\bigg|_{r'=r}\right]$$
$$+ \phi_{\nu_2}(kr)\psi_{\nu_2}(kr_s) r\left[\psi_{\nu_1}(kr_s) \frac{\partial\phi_{\nu_1}}{\partial r'}\bigg|_{r'=r_s} - \phi_{\nu_1}(kr_s) \frac{\partial\psi_{\nu_1}}{\partial r'}\bigg|_{r'=r_s}\right]$$
$$+ \phi_\nu(kr)\phi_{\nu_1}(kr_s) \frac{\psi_{\nu_1}(ka_1)}{H_{\nu_1}^{(1)}(\hat{k}a_1)} 2\Delta n k^2 \int_{a_1}^{a_2} r'\, dr'\, H_{\nu_1}^{(1)}(\hat{k}r')\psi_{\nu_2}(kr')\Bigg\}. \quad (43)$$

Returning to (17), we have the ν_1 and ν_2 integrations remaining. These are easily performed using Cauchy's theorem; the integrands are analytic except at the simple poles which are solutions of[4]

$$R_1\binom{\nu_1}{\nu_2} = R_2\binom{\nu_1}{\nu_2}. \quad (44)$$

The residues at the simple poles are

$$a_1(t) = \frac{\pi}{2i}\left\{\frac{x_D[Wi^{(1)}(t + x_D)]^2}{[Wi^{(2)}(t)Wi^{(1)\prime}(t + x_D) - Wi^{(2)\prime}(t)Wi^{(1)}(t + x_D)]^2} - \frac{1}{[Wi^{(2)}(t + x_0)]^2}\right\}^{-1} \quad (45)$$

with

$$x_0 = k(a_{\substack{1\\2}} - a_0)/(ka_{\substack{1\\2}}/2)^{1/3}. \quad (46)$$

The ν_1 and ν_2 integrations yield the desired result

$$E(r) = (-1/32)\Big\{(ka/2)^{-1/3} \sum_m a_1(t_1^m)\phi_{t_1^m}(kr)\psi_{t_1^m}(kr_s)\exp[i(\phi - \phi_s)(ka_1 + i(ka_1/2)^{1/3}t_1^m]$$
$$+ (ka/2)^{-1/3} \sum_n a_1(t_2^n)\phi_{t_2^n}(kr)\psi_{t_2^n}(kr_s)\exp[i(\phi - \phi_s)(ka_2 + i(ka_2/2)^{1/3}t_2^n] + (\pi^2\Delta n/4)(ka/2)^{1/3} \sum_m \sum_n a_1(t_1^m)a_1(t_2^n)$$
$$\times \{\phi_{t_2^n}(kr)\phi_{t_1^m}(kr_s)/[(\nu_1^m)^2 - (\nu_2^n)^2]\}\,[\psi_{t_1^m}(ka_1)/H_{\nu_1^m}^{(1)}(\hat{k}a_1)]$$
$$\times \exp[ika_1 + (\phi_0 - \phi_s)t_1^m(ka_1/2)^{1/3} + ika_2 + (\phi - \phi_0)t_2^n(ka_2/2)^{1/3}]k\int_{a_1}^{a_2} dr'\, H_{t_1^m}^{(1)}(\hat{k}r')\psi_{t_2^n}(kr')\Big\} \quad (47)$$

with

$$[(\nu_1^m)^2 - (\nu_2^n)^2] \cong 2ka_1\{k(a_1 - a_2) + (ka_1/2)^{1/3}[t_1^m(a_2/a_1)^{4/3} - t_2^n]\}. \tag{48}$$

In (47), $E(r)$ is the field normalized by the source intensity; this result is used for numerical experimentation and has been further normalized as $W = |E(r)|(-1/32)$.

III. REMARKS

(1) The single sums over the modes in (47) represent the field, assuming uncoupled normal modes.[5,6] The assumption proceeds by taking the boundary conditions to be independent of the coordinate in the direction of propagation, but the boundary conditions in the normal direction are the same that would be applied for perfectly stratified media. It is also assumed that eigenfunctions corresponding to a particular normal mode are orthonormal; i.e.,

$$\int_{a_0}^{\infty} \frac{dr'}{r'} \phi_n(kr')\phi_m(kr') = \delta_{mn}.$$

Unfortunately, in many theories, it is often very difficult to justify when to neglect the coupling. The solution given in (58) allows a direct determination of the effect of coupling. In (58), if the coupling is negligible, the solution suggests the angular position of the step is unimportant, and one must only remember which medium he is in; e.g., source in medium I, observer in medium II.

(2) Cho and Wait[7] gave a derivation for the field in a stepped model for a nonuniform duct which employed the use of a non-Hilbert space inner product; i.e., $\langle\phi_n,\phi_m\rangle$, instead of the usual definition in terms of a complex-valued function or ordered pairs with inner product $\langle\phi_n,\phi_m^*\rangle$. The natural metric

$$\{x-y,x-y\}^{1/2}$$

is a real nonnegative quantity and represents the physical quantity power. Recalling that a metric space is complete if every Cauchy sequence is a convergent sequence, the usual definition of a Hilbert space is an inner product space which is complete with respect to its *natural* metric. The Cho and Wait result can be explained by the use of "biorthogonal" coordinates.[8] Let $\{\nu_n\}$ be the set of nonzero eigenvalues of the differential operator

$$\mathscr{L} = \frac{\partial}{\partial r}\left(r\frac{\partial}{\partial r}\right) + rk^2,$$

and let $\{\phi_n\}$ be the corresponding eigenfunctions. The nonzero eigenvalues of the adjoint (formal) operator $\overline{\mathscr{L}}$ are given by $\{\nu_m^*\}$ and the corresponding eigenfunctions will be denoted by $\{\psi_m\}$. Now, take for the set $\{\psi_m\}$

$$\{\psi_m\} = \{\phi_1^*\phi_2^*,\cdots\}.$$

Then, indeed, the inner product will satisfy

$$\langle\phi_n,\phi_m\rangle = \delta_{mn}.$$

In fact, Cho and Wait's result for $\langle\phi_n,\phi_m\rangle$ equals our result; i.e., $\langle\phi_n,\phi_m\rangle = a_1(t)$. This yields the interesting conclusion that the Cho–Wait inner product will equal zero if and only if $a_1(t)$ equals zero, which requires the existence of a double root! This can occur even in a single section duct.

Proof: Since the denominator of $a_1(t)$ is a rational function, the only singularities it can have in the entire complex plane are poles. A double root suggests degenerate modes in the two regions; i.e., the spatial distribution of sources across the aperture plane $\phi = \phi_0$ has the same wavlength as the normal modes being driven and a resonance occurs. The integral formulation used here in terms of the Green's function

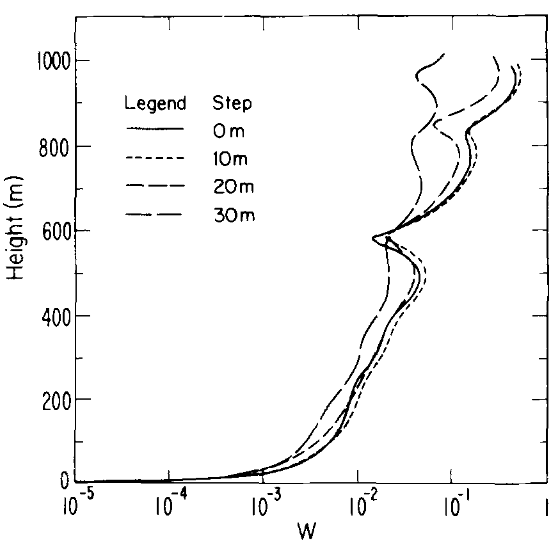


FIG. 3. Height-gain curves for a frequency of 100 MHz. Step height shown in legend. Refractive index contrast is 25 N-units. W is the normalized signal strength. The radial wavelength corresponding to the curve "20 m" is 0.463 radians.

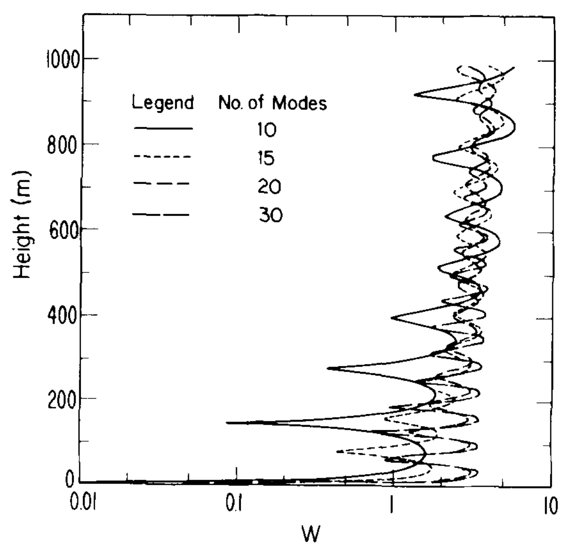


FIG. 4. Convergence of solution with increasing number of modes. Distance between source and step and step and observer is 10 km. The frequency is 100 MHz and the refractive index contrast 25 N-units.

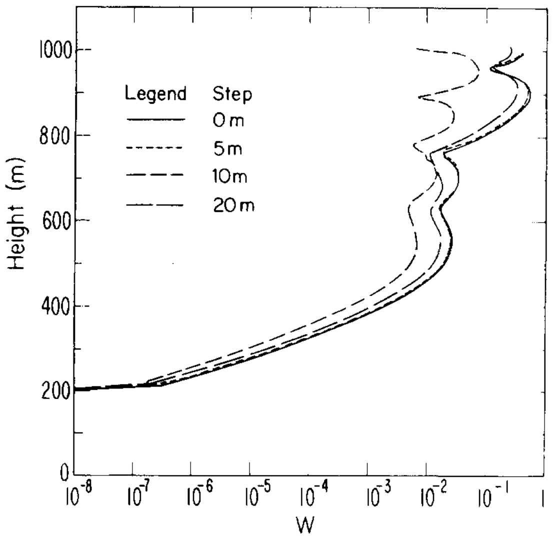


FIG. 5. Height-gain run for 300 MHz. W is the normalized signal strength. The refractive index contrast is 25 N-units. 100-km separation between source and aperture plane.

approach completely sidesteps the issue of what the normalized "eigenfunctions" should be in the biorthonormal coordinate approach. It may be that since the residue in the Green's function approach equals the inner product in the biorthonormal case and because the $H_{\nu_m}^{(1)}(kr)$ are dense in our Hilbert space, any function can be approximated to within $\epsilon > 0$ by $\Sigma_m\, a_m H_{\nu_m}^{(1)}(kr)$. The problem comes in finding how to express the a_m's. This is probably an example of a problem where the solution can be found by a Green's function method but only a generalization of the notion of eigenfunctions permit a solution in terms of the latter. The other point is that the residues come out naturally in the Green's function method.

(3) The double sum in (47) depends upon the location of the vertical step ϕ_0 and represents the coupling from mode m to mode n. The magnitude of this term depends, on the electrical step size, $k(a_2 - a_1)$, as seen in (47) and (48).

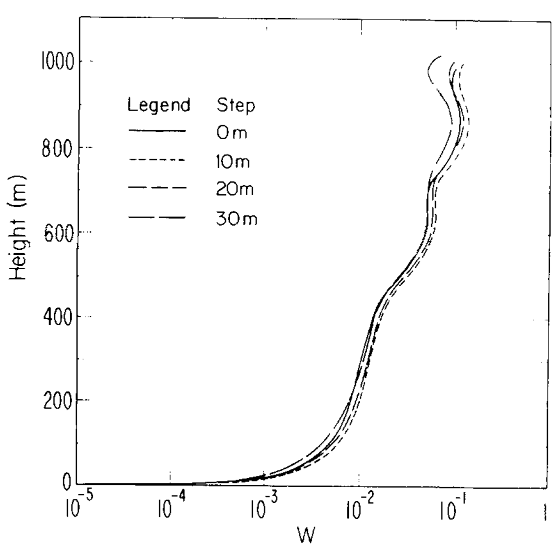


FIG. 6. Height-gain run for 60 MHz. W is the normalized signal strength. The refractive index contrast is 25 N-units. 100-m separation.

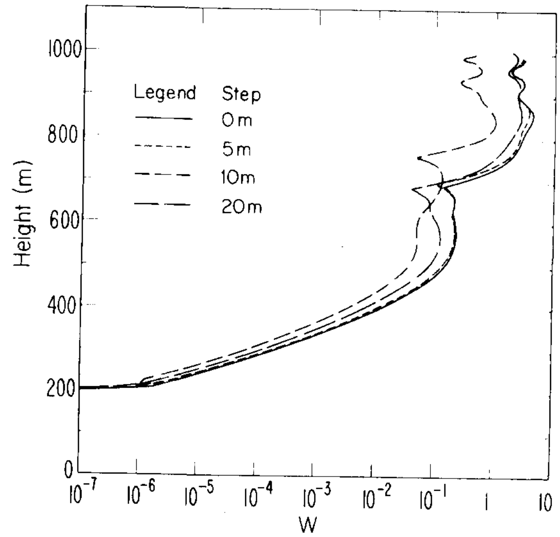


FIG. 7. Height-gain run for 300 MHz. W is the normalized signal strength. The refractive index contrast is 50 N-units.

EXAMPLE

The numerical results for the four "height-gain" curves in Fig. 3 correspond to the following choice of parameters:

$$a_0 = 6378 \text{ km},$$
$$a_1 = 6379 \text{ km},$$
$$a_2 = 6\,379\,000,\ 6\,379\,010,\ 6\,379\,020,\ 6\,379\,030 \text{ m},$$
$$r_s = 6379 \text{ km},$$
$$f = 100 \text{ MHz},$$
$$\Delta n = 25\ N\text{-units},$$
$$a_0(\phi_0 - \phi_s) \simeq 100 \text{ km},$$
$$a_0(\phi_s - \phi) \cong 100 \text{ km},$$
$$\delta = 0.3 + i4\times 10^{-2} \quad (\sigma = 0.001 \text{ Siemens/m}, \epsilon_r = 10).$$

From Fig. 3 it appears that a step size of about 20 m causes significant change in the height-gain pattern. We will refer to this as the "resonant" step size. This would correspond to a radial wavenumber of about 0.463 radians (i.e., about $\pi/8$). The second limiting criteria for our solution in (47) is the number of modes required for convergence of the series. For the example in Fig. 3, 10 modes gave two significant figures. The convergence of the series is dominated by the exponential terms in the series for small m and by the asymptotic decay of residues for large m; i.e.,

$$a_1(t_m) \sim \exp(-4/3t_m^{3/2})/4t_m^{3/2},$$

where, for $t_m < x_0$, the imaginary part of t_m becomes small. In Fig. 4, the effect of the number of modes is shown for a 10-km separation between source and step and step and observer. At this distance 30 modes are required for convergence.

In Figs. 5 and 6, the choice of the parameters is the same as in Fig. 3 except the frequency is 300 and 60 MHz, respectively. Figure 7 is the same as Fig. 4 except the refractive index contrast is 50 N-units. The limiting step size for this case is about 10 m. The resonant step size for 300 MHz is about 5 m for $\Delta n = 25$. In Figs. 5, 6, and 7, 10 modes provided adequate convergence of the sums.

In Figs. 3, 5, 6, and 7 the double sum in (47) representing the coupling term has magnitude of the same order as the single sums in (47) representing the uncoupled fields, and the effect of the coupling is included in these figures.

V. CONCLUSIONS

A Green's function approach is used to examine the effect of varying the step height in a tropospheric duct with a single step discontinuity. If the electrical height of the step is less than the "radial" separation,

$$k(a_2 - a_1)/(ka/2)^{1/3},$$

the vertical distribution of the field strength agrees with the fields in a duct with no discontinuity. This agrees with a result obtained by Wait and Spies for an ionospheric duct.[9] The location of the step in relation to the source and observer determines the number of modes required for convergence.

ACKNOWLEDGMENTS

I wish to thank my wife, Ms. Katherine Ott-Warner, for several helpful discussions. We also acknowledge Mr. Ray Wasky of the Fire Control Analysis Group of the Avionics Lab. of Wright-Patterson Air Force Base for supporting and closely following the analysis presented in this paper.

[1]A. Nagl, H. Uberall, G. L. Zarur, and A. J. Haug, "Adiabatic mode theory of underwater sound propagation in a range dependent environment," Acoust. Soc. Am. **63**, 739 (1978).

[2]E. Bahar, "Propagation of VLF radio waves in a model earth-ionosphere waveguide of arbitrary height and finite surface impedance boundary: Theory and experiment," Radio Sci. (New Ser.) **1**(8), 925–938 (1966).

[3]L. B. Felsen and N. Marcuvitz, *Radiation and Scattering of Waves* (Prentice-Hall, Englewood Cliffs, N.J., 1973), pp. 306–14.

[4]R. H. Ott, "Roots to the model equation for EM wave propagation in a tropospheric duct," J. Math. Phys. **21**, 1256–66 (1980).

[5]J. B. Keller, "Progress and prospects in the theory of linear wave apropagation," SIAM Rev. **21** (2), 234–5 (1979) (discussion of "Propagation and normal modes").

[6]Allan D. Pierce, "Extension of the method of normal modes to sound propagation in an almost-stratified medium," J. Acoust. Soc. Am. **37**, 19–27 (1965).

[7]S. H. Cho and J. R. Wait, "EM wave propagation in a laterally non-uniform troposphere," EM Report No. 1 (CIRES, Boulder, Colorado, June 1977). A summarized version is given in S. H. Cho and J. R. Wait, "Analysis of Microwave Ducting in an Inhomogeneous Troposphere," Pure Appl. Geophys. **116**, 1118–42 (1978).

[8]I. Stakgold, *Boundary Value Problems of Mathematical Physics* (MacMillan, London, 1964), Vol. 1, p. 257.

[9]J. R. Wait and K. P. Spies, "On the calculation of mode conversion at a graded height change in the earth-ionosphere waveguide at VLF," Radio Sci. (New Ser) **3** (8), 787–91 (1968).

Spectral estimates, stability conditions, and the rotating screw-pinch

Eliezer Hameiri
Courant Institute of Mathematical Sciences, New York University, New York, New York 10012

(Received 7 August 1980; accepted for publication 17 April 1981)

This article presents two sufficient conditions for the linear stability of rotating ideal plasmas, the first based on conservation of circulation and the second based on circle theorems applicable to linear Hamiltonian systems. The circle theorems also provide bounds on eigenmodes in the complex plane. All results are applied to the rotating screw-pinch which can be described by a single second-order ordinary differential equation.

PACS numbers: 52.35.Py, 52.30. + r, 52.55.Ez, 02.30.Hg

I. INTRODUCTION

Recently, there has been growing interest in the behavior of rotating plasma systems, mainly due to the fact that tokamaks exhibit macroscopic flows after being heated by neutral beams.[1] Plasmas in other experimental devices were known for many years to rotate, for example in θ pinches where the rotation is believed to be induced by end shorting of the electric field,[2,3] and also in field reversed plasmas, for reasons which are still controversial.[4,5]

Among the new questions that this phenomenon presents is the effect of rotation on the stability of the confined plasma. We will be concerned here with the linear stability of equilibrium states with flow, for plasmas described by the ideal magnetohydrodynamics model. The main difficulty in dealing with flow problems arises from the fact that they generally do not have a self-adjoint formulation like the static case,[6] and thus the powerful tool of an energy principle is not available. The most one can do is to use methods for Hamiltonian systems, as the plasma is known to be one.[7] Mainly due to this difficulty, flow problems were hardly treated until very recently and the few cases which were treated tended to be rather specialized, e. g., rotating θ pinches[8–10] or multipoles with purely toroidal rotation.[11] A result of a more general character is the extension of Suydam's stability criterion[12] to the rotating plasma.[13]

In this work we will be concerned with the linear stability of general rotating equilibria. Because of the nonexistence of an energy principle we cannot hope to find conditions which are both necessary and sufficient for stability. However, it will be possible to derive sufficient conditions for stability as well as bounds on the spectrum in the complex plane. The bounds are necessary for an effective numerical investigation since they are the main tool for restricting the search for eigenmodes to a reasonably limited domain in the complex plane. Alternative devices, like properties of the nodal structure of the eigenfunctions, are not available in the non-self-adjoint case. One of the sufficient conditions we obtain is based on the conservation of circulation for a moving plasma. [All other conservation laws are automatically incorporated by a special choice of the perturbed density and magnetic field (5)]. Clearly, every conserved quantity restricts the growth of a perturbation and implies some special property of the equations, which we utilize. The improvement in stability, however, is effective only for symmetry preserving modes. Another sufficient condition, as well as the bounds on the spectrum, are obtained through the use of circle theorems[13,14] which are a generalization of Howard's[15] estimate for incompressible fluids. This technique was recently utilized to obtain spectral bounds for a rotating θ pinch[10] and for other configurations in fluid dynamics.[16]

Another objective of this work is to introduce the rotating screw-pinch as an important model for the investigation of the effects of flows. Due to the axial and azimuthal symmetrics, the problem reduces to the study of an ordinary differential operator. It can actually be reduced to a single second order equation, even for magnetic fields and flows having both axial and poloidal components. This property was observed[17] in the static case, and helped make the screw-pinch the most fruitful case for studies of spectral properties of such equilibria.[18] We anticipate that the reduction of this problem to a manageable form will facilitate exhaustive analytical and numerical studies comparable to those carried out in the static case.[19,20]

In the next section we describe the equations of motion and reduce the rotating screw-pinch problem to a single second order equation. In Sec. III we derive a sufficient condition for stability based on the constants of the motion. The condition is then applied to the screw-pinch. Section IV describes an improved form of the circle theorem from which spectral bounds and another sufficient condition are obtained. Section V is devoted to a rather thorough application of the spectral bounds to the screw-pinch, commensurate with the importance we attach to this model. We show how these bounds can be computed by a minor modification of any numerical code which will be used in the investigation.

II. THE EQUATIONS OF MOTION

The plasma is described using the ideal magnetohydrodynamics equations

$$\begin{aligned}
&\rho \mathbf{u}_t + \rho \mathbf{u}\cdot\nabla \mathbf{u} + \nabla p = \mathbf{J}\times\mathbf{B},\\
&\rho_t + \operatorname{div}(\rho \mathbf{u}) = 0,\\
&\mathbf{B}_t + \operatorname{curl}(\mathbf{B}\times\mathbf{u}) = 0,\\
&s_t + \mathbf{u}\cdot\nabla s = 0,\\
&\operatorname{div}\mathbf{B} = 0,\quad \mathbf{J} = \operatorname{curl}\mathbf{B},\quad p = p(\rho,s),
\end{aligned}\tag{1}$$

where $\mathbf{B}$, $\mathbf{u}$, p, ρ, s are the magnetic field and the plasma velocity, pressure, density, and entropy, respectively. The

 0022-2488/81/092080-09$01.00

last equation is a prescribed equation of state. We also define

$$a^2 = \partial p(\rho,s)/\partial\rho, \quad \mathbf{B}_*^2 = \mathbf{B}^2 + \rho a^2,$$
$$\beta = \rho a^2/B_*^2, \quad p_* = p + \mathbf{B}^2/2, \tag{2}$$

a is the speed of sound. Given an equilibrium state with flow we linearize Eq. (1) about it. Frieman and Rotenberg[7] used the Lagrangian displacement vector $\boldsymbol{\xi}$ as the dependent variable to obtain the following linearized equation, which we describe in detail for later use,

$$\rho\boldsymbol{\xi}_{tt} + 2A\,\boldsymbol{\xi}_t - F\boldsymbol{\xi} = 0, \tag{3}$$

where the operators A and F are

$$A\,\boldsymbol{\xi} = \rho\mathbf{u}\cdot\nabla\boldsymbol{\xi},$$
$$F\boldsymbol{\xi} = \nabla(\rho a^2\nabla\cdot\boldsymbol{\xi} + \boldsymbol{\xi}\cdot\nabla p - \mathbf{B}\cdot\mathbf{Q}) + \mathbf{B}\cdot\nabla\mathbf{Q} + \mathbf{Q}\cdot\nabla\mathbf{B} + \nabla\cdot(\rho\boldsymbol{\xi}\mathbf{u}\cdot\nabla\mathbf{u} - \rho\mathbf{u}\mathbf{u}\cdot\nabla\boldsymbol{\xi}) \tag{4}$$

and $\mathbf{Q} = \mathrm{curl}(\boldsymbol{\xi}\times\mathbf{B})$. All coefficients are equilibrium quantities. Equation (3) is obtained by using the following relations between the perturbed Eulerian quantities (denoted by a subscript 1) and $\boldsymbol{\xi}$:

$$\mathbf{B}_1 = \mathrm{curl}(\boldsymbol{\xi}\times\mathbf{B}),$$
$$\rho_1 = -\,\mathrm{div}(\rho\boldsymbol{\xi}), \tag{5}$$
$$p_1 = -\,\boldsymbol{\xi}\cdot\nabla p - \rho a^2\mathrm{div}\boldsymbol{\xi},$$
$$\mathbf{u}_1 = \mathbf{u}\cdot\nabla\boldsymbol{\xi} - \boldsymbol{\xi}\cdot\nabla\mathbf{u} + \boldsymbol{\xi}_t.$$

$\boldsymbol{\xi}(\mathbf{r},t)$ expresses the displacement from the equilibrium trajectory of the fluid particle which would have been at the position $\mathbf{r}$ at time t in the unperturbed motion. It was observed[7] that A is anti-Hermitian and F is Hermitian with respect to the inner product

$$(\boldsymbol{\xi},\boldsymbol{\eta}) = \int \boldsymbol{\xi}\cdot\boldsymbol{\eta}^*\,d^3\mathbf{r} \tag{6}$$

for twice differentiable vector functions with a vanishing normal component at the boundary of the plasma (which is assumed to be a flux surface). Equation (3) has the form of a general linear Hamiltonian system, where the momentum coordinates were eliminated in favor of $\boldsymbol{\xi}_t$. The evolution of solutions in time can be found by considering a dependence $\boldsymbol{\xi}\sim\exp(-i\omega t)$. We obtain the nonlinear eigenvalue problem in ω,

$$L_\omega\boldsymbol{\xi} \equiv (\omega^2\rho + 2\omega iA + F)\boldsymbol{\xi} = 0. \tag{7}$$

Notice that by taking the inner product with $\boldsymbol{\xi}$ one obtains a quadratic equation for ω, and its solution is not complex if

$$(F\boldsymbol{\xi},\boldsymbol{\xi})(\rho\boldsymbol{\xi},\boldsymbol{\xi}) - (iA\,\boldsymbol{\xi},\boldsymbol{\xi})^2 \leqslant 0. \tag{8}$$

If this inequality holds for all $\boldsymbol{\xi}$ it can be concluded that the plasma is exponentially stable.[7]

A configuration of great simplicity is the rotating screw-pinch. The plasma is contained in a circular cylinder of length L (finite or infinite) and radius l. The two ends of the cylinder are identified to resemble a torus. All equilibrium profiles depend on the radial variable r only (in r, θ, z coordinates) and the magnetic field and velocity have vanishing radial components. Thus $\mathbf{B} = B_\theta\hat{\theta} + B_z\hat{z}$, $\mathbf{u} = v\hat{\theta} + w\hat{z}$ and we denote $v = r\Omega(r)$. Such an equilibrium state has to satisfy the single equation

$$(p + \mathbf{B}^2/2)' + B_\theta^2/r = \rho v^2/r, \tag{9}$$

where the prime denotes d/dr. Notice that the axial velocity w does non enter into Eq. (9) so that an arbitrary axial flow could be surperimposed on any equilibrium state to yield another such state. This of course is not true in a toroidal geometry where a centrifugal force associated with w will be present.

In the linearized system, the coordinates θ, z are ignorable and we can assume a dependence $\boldsymbol{\xi}\sim \exp i(m\theta + kz - \omega t)$. It was observed before[21] that in the linearized system (1) only two variables are differentiated in a direction normal to the equilibrium flux surfaces (the surfaces r = const. in the screw-pinch). Thus in our configuration only two equations contain derivatives while six others are algebraic. The algebraic equations can be used to eliminate six Eulerian variables and we obtain a 2×2 ordinary differential system

$$AC\begin{pmatrix}\xi\\ \phi\end{pmatrix}' = \Lambda\begin{pmatrix}\xi\\ \phi\end{pmatrix}, \tag{10}$$

where the elements of the matrix Λ are

$$\lambda_{11} = C\,[(2m/r^2)(B_\theta H + \rho\sigma v) - (1/r)A\,] + (2/r)\rho^2\sigma^2 B_\theta(\sigma H v + B_\theta\sigma^2) + (1/r)\rho^2\sigma^2 v^2 A,$$
$$\lambda_{12} = -\,[\rho^2\sigma^4 + C(k^2 + m^2/r^2)],$$
$$\lambda_{21} = -\,C\,[A^2 + Ar((1/r)p_*')' - (4/r^2)(B_\theta H + \rho\sigma v)^2] + [(1/r)\rho v^2 A + (2/r)\rho B_\theta(\sigma H v + B_\theta\sigma^2)]^2,$$
$$\lambda_{22} = -\,\lambda_{11} - (1/r)AC,$$

and where

$$H \equiv \mathbf{k}\cdot\mathbf{B} \equiv mB_\theta/r + kB_z,$$
$$\sigma \equiv \omega - \mathbf{k}\cdot\mathbf{u} \equiv \omega - (m\Omega + kw), \tag{11}$$
$$A = H^2 - \rho\sigma^2, \quad C = B_*^2(\beta\mathrm{H}^2 - \rho\sigma^2), \tag{12}$$

ϕ is the perturbed total pressure $p_1 + \mathbf{B}\cdot\mathbf{B}_1$ and ξ is ξ_r, the radial component of $\boldsymbol{\xi}$, and is related to the perturbed velocity by $u_r = -\,i\sigma\xi$.

For given k and m, Eq. (10) is an "eigenvalue" equation for ω, where the boundary conditions to be satisfied are regularity of $r\xi$ at $r=0$ and $\xi = 0$ at the wall $r = l$. Notice that the coefficients in (10) are real except for ω, and thus if ω is an eigenvalue so is ω^*. Also $-\,\omega$, $-\,\omega^*$ are eigenvalues for wavenumbers $-\,k$ and $-\,m$. For stability it is sufficient to consider ω in the upper half complex plane. Also observe the existence of a real continuous spectrum given by the range of ω for which $A = 0$ (the Alfvén continuum) or $C = 0$ (the cusp continuum) anywhere in the plasma. For these values of ω Eq. (10) becomes singular. This is analogous to the static case.[18]

System (10) can be further reduced to a single, second order equation in ξ. This can be achieved by multiplying the second equation in (10) by an integration function f such that it takes the form

$$(fAC\phi)' = f\lambda_{21}\xi, \tag{13}$$

and f should satisfy

$$(fAC)' = -\,f\lambda_{22}. \tag{14}$$

The first equation in (10) is now multiplied by fAC/λ_{12} and differentiated. Eqs. (13) and (14) are used to eliminate ϕ and f. The resulting equation has the form

$$\left(\frac{rAC}{\lambda_{12}}\xi'\right)' + r\left[\frac{\det A}{AC\lambda_{12}} - \left(\frac{\lambda_{11}}{\lambda_{12}}\right)'\right]\xi = 0, \qquad (15)$$

which has the associated quadratic form

$$Q(\xi) = \int_0^l (p|\xi'|^2 - q|\xi|^2)\, dr, \qquad (16)$$

where p is the coefficient of ξ'' and q the coefficient of ξ in (15). It is pointed out that the points where $\lambda_{12} = 0$ are not singular points of Eq. (15), as is clear from Eq. (10).

We now make the observation, to be used later on, that Eq. (15) can be derived formally in a variational way. Let ω be real for the time being, and consider the quadratic form associated with L_ω in (7). The reality of ω implies that L_ω is Hermitian. Vary the quadratic form with respect to admissible $\boldsymbol{\xi}$ and find the Euler equation. We have $\delta(L_\omega\boldsymbol{\xi},\boldsymbol{\xi}) = (L_\omega\boldsymbol{\xi},\delta\boldsymbol{\xi}) + (L_\omega\delta\boldsymbol{\xi},\boldsymbol{\xi}) = 2\mathrm{Re}(L_\omega\boldsymbol{\xi},\delta\boldsymbol{\xi}) = 0$. By taking $i\delta\boldsymbol{\xi}$ instead of $\delta\boldsymbol{\xi}$ we find $\mathrm{Im}(L_\omega\boldsymbol{\xi},\delta\boldsymbol{\xi}) = 0$. Thus $L_\omega\boldsymbol{\xi} = 0$. In the screw-pinch case $(L_\omega\boldsymbol{\xi},\boldsymbol{\xi})$ contains derivatives of $\xi_r \equiv \xi$ only, but not of ξ_θ,ξ_z. We find the stationary point of the quadratic form with respect to variations of ξ_θ,ξ_z only, but not of ξ. This yields ξ_θ and ξ_z as linear combinations of ξ and ξ'. For these values of ξ_θ,ξ_z the variation of $(L_\omega\boldsymbol{\xi},\boldsymbol{\xi})$ with respect to ξ will yield Eq. (15). Notice that although we used ω real in the derivation, the same equation for ξ holds for complex ω since the coefficients are known to be rational functions of ω. We now claim that for ξ_θ,ξ_z as found before,

$$(L_\omega\boldsymbol{\xi},\boldsymbol{\xi}) \equiv 2\pi Q(\xi). \qquad (17)$$

This is clear since the two forms generate the same differential equation in ξ and thus can differ only by a constant multiple. The constant can be found by looking at the leading term in ω^2 for large ω.

It will be useful for the implementation of the results described in Secs. IV and V, to write down the explicit form of $(L_\omega\boldsymbol{\xi},\boldsymbol{\xi})$. Letting $(\xi,\eta,\zeta) = (\xi_r, i\xi_\theta, i\xi_z)$ we get a quadratic form with real coefficients (except ω).

$$(L_\omega\boldsymbol{\xi},\boldsymbol{\xi}) = -\int (S_0 + S_1 + S_2)\, d^3\mathbf{r},$$

where for real ξ,η,ζ,

$$\begin{aligned}
S_0 &= B_*^2\xi'^2 + \xi^2[H^2 - \rho\sigma^2 - (B_*^2/r)' + r(B_\theta^2/r^2)'],\\
S_1 &= (2/r)[(m/r)\eta + k\zeta]\\
&\quad\times[\rho a^2(r\xi)' + \rho v^2\xi] + 2(\eta B_z - \zeta B_\theta)\\
&\quad\times[(kB_\theta + (m/r)B_z)\xi/r\\
&\quad + ((m/r)B_z - kB_\theta)\xi'] + 4\rho\Omega\sigma\xi\eta,\\
S_2 &= \rho a^2[(m/r)\eta + k\zeta]^2 + (m^2/r^2 + k^2)\\
&\quad\times(\eta B_z - \zeta B_\theta)^2 - \rho\sigma^2(\eta^2 + \zeta^2).
\end{aligned}$$

III. CIRCULATION INVARIANTS

In this section we assume that the equilibrium state consists of a family of closed flux surfaces, not necessarily nested, and labeled by a parameter ψ. We also require the entropy to be constant on each surface $s = s(\psi)$. (A form of the following results holds even if the entropy assumption is violated. See the remark at the end of the section.) Since entropy and flux are carried by the fluid, the same properties hold for the time-dependent motion and indeed, even for the perturbed system, as can be derived from (5). It is known[22] that for such configurations $\int_\psi \mathbf{u}\cdot\mathbf{B}\, d^3\mathbf{r}$ is a constant of the motion, where the integral is understood to be taken over a volume enclosed by the same moving flux tube ψ. This conservation law is an analog of the consevation of circulation in fluid dynamics. We have, from the definition of $\boldsymbol{\xi}$,

$$\int_\psi \mathbf{u}(\mathbf{r}+\boldsymbol{\xi})\cdot\mathbf{B}(\mathbf{r}+\boldsymbol{\xi})\, d^3(\mathbf{r}+\boldsymbol{\xi}) = \int_\psi \mathbf{u}_0(\mathbf{r})\cdot\mathbf{B}_0(\mathbf{r})\, d^3\mathbf{r} + \text{initial perturbation}, \qquad (18)$$

where here the subscript 0 denotes equilibrium quantities and the other quantities represent the nonlinear time dependent state. The initial perturbation term is time-independent and arises from the fact that perturbed states were not required to have the same $\int_\psi \mathbf{u}\cdot\mathbf{B}$ as the equilibrium state. To first order in $\boldsymbol{\xi}$,

$$\mathbf{u}(\mathbf{r}+\boldsymbol{\xi}) = \mathbf{u}(\mathbf{r}) + \boldsymbol{\xi}\cdot\nabla\mathbf{u}(\mathbf{r}) = \mathbf{u}_0(r) + \mathbf{u}_1(r) + \boldsymbol{\xi}\cdot\nabla\mathbf{u}_0(r)$$

and

$$d^3(\mathbf{r}+\boldsymbol{\xi}) = (1 + \mathrm{div}\boldsymbol{\xi})\, d^3\mathbf{r}.$$

It follows from (5) after expanding $\mathbf{B}(\mathbf{r}+\boldsymbol{\xi})$ and dropping the subscript 0,

$$\int_\psi [\mathbf{B}\cdot(\boldsymbol{\xi}_t + \mathbf{u}\cdot\nabla\boldsymbol{\xi}) + \mathbf{u}\cdot(\mathbf{B}\cdot\nabla\boldsymbol{\xi})]\, d^3\mathbf{r} = \text{initial perturbation}, \qquad (19)$$

where all coefficients of $\boldsymbol{\xi}$ are equilibrium quantities. Using the equilibrium equations and also $\mathbf{B}\cdot\nabla\psi = 0$ and $\mathbf{u}\cdot\nabla\psi = 0$ (the equilibrium flow is within a flux surface), the last term under the integral can be integrated by parts and replaced by $\mathbf{B}\cdot(\mathbf{u}\cdot\nabla\boldsymbol{\xi})$. The initial perturbation term can be eliminated by differentiating Eq. (19) by t. We get

$$\int_\psi \mathbf{B}\cdot(\boldsymbol{\xi}_{tt} + 2\mathbf{u}\cdot\nabla\boldsymbol{\xi}_t)\, d^3\mathbf{r} = 0. \qquad (20)$$

A comparison with Eq. (3) shows

$$\int_\psi \mathbf{B}\cdot F(\boldsymbol{\xi})/\rho\, d^3\mathbf{r} = 0 \qquad (21)$$

for every ψ and all admissible $\boldsymbol{\xi}$. A direct verification of property (21) will be comforting. We use the symmetry of F to write

$$\int_\psi \mathbf{B}\cdot F(\boldsymbol{\xi})/\rho\, d^3\mathbf{r} = \int_\psi \boldsymbol{\xi}\cdot F(\mathbf{B}/\rho)\, d^3\mathbf{r} + \text{boundary terms}.$$

The boundary terms may be present since the normal component of $\boldsymbol{\xi}$ does not necessarily vanish on ψ. It can be easily seen directly, though, that $F(\mathbf{B}/\rho) = 0$ and the boundary terms, which can be simply found from (4), also vanish. In fact, the same thing happens if $\mathbf{B}$ is replaced by any vector field $\mathbf{b}$ satisfying

$$\mathrm{div}\mathbf{b} = 0, \quad \mathbf{b}\cdot\nabla\psi = 0, \qquad (22)$$

$$\mathrm{curl}(\mathbf{b}\times\mathbf{u}) = 0, \quad \mathrm{curl}(\mathbf{b}\times\mathbf{B}/\rho) = 0, \qquad (23)$$

and we have, corresponding to Eq. (20),

$$\int_\psi \mathbf{b}\cdot(\xi_{tt} + 2\mathbf{u}\cdot\nabla\xi_t)\, d^3\mathbf{r} = 0. \quad (24)$$

This property is related to the existence[23] of additional circulation invariants in magnetohydrodynamics. Equatioh (22) means that $\mathbf{b}$ is a magnetic field-like vector field, having common flux surfaces with $\mathbf{B}$. To solve Eqs. (22) and (23), we first notice that $\mathbf{b} = \mathbf{B}$ and $\mathbf{b} = \rho\mathbf{u}$ are solutions. If $\mathbf{u}$ is not parallel to $\mathbf{B}$ we can write

$$\mathbf{b} = f\mathbf{B} + g\rho\mathbf{u}, \quad (25)$$

where f and g are scalar functions. However, from (23) we get $(\mathbf{B}\times\mathbf{u})\times\nabla f = 0$, hence $f = f(\psi)$. Likewise $g = g(\psi)$ and then $\mathbf{b}$ is also divergence free. If $\mathbf{u}$ is parallel to $\mathbf{B}$, the two equations in (23) have the same content. To find a general $\mathbf{b}$ we write $\mathbf{b} = f\mathbf{B} + g\mathbf{B}\times\nabla\psi$. Equation (23) implies that $g\mathbf{B}^2/\rho = h(\psi)$, some function of ψ. Since $\mathbf{b}$ is divergence free, we have $\mathbf{B}\cdot\nabla f = -h(\psi)\mathrm{div}(\rho\mathbf{B}\times\nabla\psi/\mathbf{B}^2)$, and f is determined up to a function of ψ (for ergodic field lines), i. e.,

$$\mathbf{b} = k(\psi)\mathbf{B} + h(\psi)\mathbf{b}_1, \quad (26)$$

where $\mathbf{b}_1$ is a particular solution for $\mathbf{b}$ which has a $\mathbf{B}\times\nabla\psi$ component and therefore is never parallel to $\mathbf{B}$. [Notice that $\rho\mathbf{u}$ also must have the form (26) and therefore if it is parallel to $\mathbf{B}$ at a point, it will be parallel over the entire flux surface.] We remark that if the equilibrium state is axisymmetric, say independent of a toroidal angle θ, $\mathbf{b} = r\rho\hat{\theta}$ is a solution, where r is the cylindrical radial coordinate. This solution is suggested by conservation of angular momentum. Also, for static equilibria $\mathbf{b} = \mathbf{J}$ is a solution. Our first conclusion is

Theorem 1: The point $\omega = 0$ is an eigenvalue of Eq. (7) with infinite multiplicity.

Proof: It was already pointed out that $F(\mathbf{b}/\rho) = 0$. The theorem was known for the static case but our treatment reveals its origin, namely conservation of circulation. The existence of an $\omega = 0$ eigenvalue does not therefore indicate bifurcation or marginal stability, at least within the present ideal plasma model.

It is highly suggestive that we try to eliminate these zeroes of F from consideration, that is, restrict ourselves to states which automatically satisfy the conservation of circulation. In order to do this it is convenient to introduce a new inner product

$$\langle\xi,\eta\rangle = \int \rho\,\xi\cdot\eta^*\, d^3\mathbf{r}. \quad (27)$$

With respect to (27) we have the Hermitian operators F', A', where

$$A' = (i/\rho)A, \quad F' = (1/\rho)F, \quad (28)$$

and the eigenvalue equation (7) can be rewritten as

$$(F' + 2\omega A' + \omega^2)\xi = 0. \quad (29)$$

Define now the subspace S which consist of all vectors $\mathbf{b}/\rho$, with $\mathbf{b}$ as in (26). Let P denote the orthogonal projection operator associated with S. Namely, $P\xi = f(\psi)\mathbf{B}/\rho + g(\psi)\mathbf{u}$ and $f(\psi)$, $g(\psi)$ are determined by the requirement $\langle P\xi,\mathbf{b}/\rho\rangle = \langle\xi,\mathbf{b}/\rho\rangle$ for all $\mathbf{b}$. f and g can be easily found if we determine them on a surface by surface basis, and observe that $d^3\mathbf{r} = d\psi\, dS/|\nabla\psi|$. We want

$$\oint P\xi\cdot\mathbf{b}\frac{dS}{|\nabla\psi|} = \oint \xi\cdot\mathbf{b}\frac{dS}{|\nabla\psi|}, \quad (30)$$

where dS is an area element on a flux surface. There are two independent $\mathbf{b}$ on each surface, and (30) completely determines $P\xi$.

Clearly,

$$P^2 = P = P^*, \quad (31)$$

which is a restatement of the fact the P is a projection onto S in the direction orthogonal to S. The property $F(\mathbf{b}/\rho) = 0$ implies $F'P = 0$. Taking the adjoint we have $PF' = 0$. To summarize,

$$PF' = F'P = 0. \quad (32)$$

Define the projection on the orthogonal complement of S, $Q = I - P$, where I is the identity operator. We have $Q = Q^2 = Q^*$ and $QP = PQ = 0$. Now multiply Eq. (29) by P and write $\xi = P\xi + Q\xi$. Because of (32) we have

$$\omega(2PA'P + 2PA'Q + \omega P)\xi = 0. \quad (33)$$

For stability purposes we are not interested in $\omega = 0$ and it can be divided out. We now claim

$$PA'P = 0. \quad (34)$$

To prove it we recall that $A' = i\mathbf{u}\cdot\nabla$ so that for any ξ, $A'P\xi = \mathbf{u}\cdot\nabla(\mathbf{b}/\rho)$ for some $\mathbf{b}$. Using (23) it can be directly shown that $\int\mathbf{b}'\cdot\mathbf{u}\cdot\nabla(\mathbf{b}/\rho)\, d^3\mathbf{r} = 0$ for any $\mathbf{b}$, $\mathbf{b}'$ as in (26). This shows that $A'P\xi$ is orthogonal to S and thus proves (34). From (33) we now have for an eigenvector with $\omega\neq 0$,

$$\omega P\xi = -2PA'Q\xi. \quad (35)$$

Now, multiplying (29) by Q we get [since $QF' = (Q+P)F' = F'$],

$$F'\xi + 2\omega QA'(Q+P)\xi + \omega^2 Q\xi = 0,$$

or, using (35),

$$[(F' - 4QA'PA'Q) + 2\omega QA'Q + \omega^2]Q\xi = 0. \quad (36)$$

This is the restriction of Eq. (29) to the subspace orthogonal to S. Notice that it is still quadratic in ω. Also, because of (34), and after denoting $\eta = Q\xi$, we have

$$[(F' - 4A'PA') + 2\omega QA'Q + \omega^2]\eta = 0, \quad (36')$$

and η automatically satisfies $\eta = Q\eta$ if $\omega\neq 0$. The eigenvector ξ of Eq. (29) is obtained using (35), and is $\xi = \eta - 2PA'\eta/\omega$. An immediate result of (36') is, as in Eq. (8),

Theorem 2: A sufficient condition for stability is $F + 4AP(1/\rho)A \leqslant 0$. [We returned to the original notations of the operators and use the inner product (6) with respect to which $P(1/\rho)$ is a symmetric operator.]

We now test our condition on the screw-pinch where we consider the mode which preserves both the axial and azimuthal symmetries, namely the $m = 0$, $k = 0$ mode.

Theorem 3: For the mode $m = 0$, $k = 0$ in the screw-pinch, the sufficient condition in Theorem 2 is also necessary for stability.

Proof: In this case, the effect of P is to eliminate the r component of ξ, while $Q\xi = \xi_r(r)\hat{r}$. $\mathbf{u}\cdot\nabla Q\xi$ is then a vector with a θ component only, so $QA'Q = 0$ and Eq. (36') reduces to an ordinary eigenvalue problem for a self-adjoint operator. The nonpositivity of this operator is necessary and sufficient for stability.

Remark: Even if the assumption of constant entropy on flux surfaces is violated, which is possible for nonergodic fluid trajectories, it is still true that $F(\mathbf{b}/\rho) = 0$, with $\mathbf{b} = f(\psi)\rho\mathbf{u}$. Thus Theorems 1 and 2 still hold, with P being the projection into this smaller subspace.

Our sufficient condition is somewhat reminiscent of the Frieman–Rotenberg condition (8). The two criteria actually complement each other. To see this, observe that A and F are real operators unless there is symmetry and one considers a particular Fourier mode denoted, say, by a wavenumber m. For $m = 0$, A and F are still real and considering a real test function ξ, $(A\xi,\xi) = 0$ with no imaginary part. For such vectors condition (8) reduces to $(F\xi,\xi) \leqslant 0$. Since F is also real the maximum of this quadratic form is attained for ξ real, so that condition (8) actually reduces to $(F\xi,\xi) \leqslant 0$ for *all* ξ, or $F \leqslant 0$. For $m \neq 0$, however, the A term in the condition offers an improvement. Considering our new condition, we notice that for $m \neq 0$ the right-hand side of Eq. (30) always vanishes [$\oint \exp(im\theta)\, d\theta = 0$ for $m \neq 0$] so $P = 0$ and the condition reduces to $F \leqslant 0$. Improvement then is possible only for $m = 0$. Notice that in the screw-pinch case, where there are two symmetries, only the mode $m = 0$, $k = 0$ could be considered. Theorem 3 shows that occassionally (but most likely not always) the improvement achieved is maximal. We conclude that our condition should be used for symmetry preserving modes while condition (8) should be used for all other modes.

IV. CIRCLE THEOREMS

In this section we concern ourselves with the location in the complex plane of eigenvalues of the Hamiltonian equation (7). ρ, iA, and F are assumed to be Hermitian operators in an inner product space with ρ invertible (not necessarily positive definite). Taking the inner product of (7) with ξ and solving for ω, we have

$$\omega = \{ -(iA\xi,\xi) + [(iA\xi,\xi)^2 - (\rho\xi,\xi)(F\xi,\xi)]^{1/2}\}/(\rho\xi,\xi). \tag{37}$$

For a complex ω one finds[24]

$$\mathrm{Re}\omega = -(iA\xi,\xi)/(\rho\xi,\xi), \tag{38}$$

$$|\omega|^2 = (F\xi,\xi)/(\rho\xi,\xi), \tag{39}$$

and the following estimates hold:

$$\operatorname*{Inf}_{\xi} -(iA\xi,\xi)/(\rho\xi,\xi) \leqslant \mathrm{Re}\omega \leqslant \operatorname*{Sup}_{\xi} -(iA\xi,\xi)/(\rho\xi,\xi), \tag{40}$$

$$\operatorname*{Inf}_{\xi}(F\xi,\xi)/(\rho\xi,\xi) \leqslant |\omega|^2 \leqslant \operatorname*{Sup}_{\xi}(F\xi,\xi)/(\rho\xi,\xi). \tag{41}$$

All complex eigenvalues ω are contained in the strip (40) and in a concentric annulus centered at $\omega = 0$. Notice, however, that the lower bound in (41) may be negative or the upper bound infinite in which case some of the bounds are not useful. This will certainly be the case if ρ is not definite. Improvement of the previous results is possible by observing that the bilinear transformation

$$\omega = (az + b)/(cz + d), \quad ad - bc \neq 0 \tag{42}$$

yields another quadratic equation in z which, moreover, preserves the symmetry properties of the operator coefficients if a, b, c, d are *real* numbers. This is assumed from now on. The coefficients transform according to the linear transformation rules of quadratic forms, that is,

$$\begin{pmatrix} \rho & iA' \\ iA' & F' \end{pmatrix} = \begin{pmatrix} a & c \\ b & d \end{pmatrix} \begin{pmatrix} \rho & iA \\ iA & F \end{pmatrix} \begin{pmatrix} a & b \\ c & d \end{pmatrix}, \tag{43}$$

where here primes indicate the new operators in (7) after ω is replaced by z. Estimates (40) and (41) can now be carried out for z. Since the transformation (42) maps and circles and lines into circles and lines, one gets (nonconcentric) annuli in the intersection of which all complex eigenvalues ω must be contained. Notice that $z = -\omega$ is a transformation which does not yield new bounds. Also it is possible to scale the coefficients in (42) without affecting anything. Hence we may restrict ourselves to real coefficients satisfying

$$ad - bc = 1. \tag{44}$$

To cover all possibilities, hopefully with some ρ' definite, involves a 3-parameter group of transformations. We remark that the transformation rule (43) applies also to the quadratic forms $(F\xi,\xi)$ and the like, and that the Frieman–Rotenberg condition (8) involves the determinant of the matrix in (43). The determinant is unchanged by the transformation considered so that the sufficient condition (8) is not improved.

The case of ρ positive definite is of particular interest. It is useful in this case to consider the special class of transformations

$$\omega = z + \alpha \tag{45}$$

with α real. Equation (7) takes a form quadratic in z,

$$[z^2\rho + 2z(iA + \alpha\rho) + L_\alpha]\xi = 0, \tag{46}$$

where L_α is defined in (7). The bounds on Rez yield (40) again, while an upper bound on $|z|$ is

$$|z|^2 \leqslant \operatorname*{Sup}_{\xi}(L_\alpha\xi,\xi)/(\rho\xi,\xi) \equiv R^2(\alpha), \tag{47}$$

We neglect here the lower bound which is sometimes useful.[10] One then obtains the following results[13,14]:

Theorem 4: All complex eigenvalues ω are contained in the intersection of all disks of center α and radius $R(\alpha)$ as defined in (47), for all real α.

Theorem 5: A sufficient condition for exponential stability is

$$\operatorname*{Inf}_{\omega\ \mathrm{real}} \operatorname*{Sup}_{\xi} \frac{(L_\omega\xi,\xi)}{(\rho\xi,\xi)} \leqslant 0. \tag{48}$$

(We changes our notation here from α to ω).

Condition (48) is not necessary for stability in general. To gain a better understanding of this condition we first point out that if L_ω is self-adjoint and ρ is positive definite, then $R^2(\omega)$ is the upper edge of the λ spectrum of L_ω defined by

$$L_\omega\xi = \lambda\rho\xi. \tag{49}$$

Consider now Eq. (7) in n-dimensional space with ρ, iA, F constant Hermitian matrices and ρ positive definite. Clearly there are $2n$ eigenvalues ω. It is possible to find all the real ones by considering problem (49) for all real ω. ω is an eigenvalue if and only if $\lambda_j(\omega) = 0$ for some j, $1 \leqslant j \leqslant n$. If the largest eigenvalue λ_1 becomes negative for some real ω, then all curves $\lambda_j(\omega)$ do so. Since for large $|\omega|$, $\lambda_j(\omega)$ behave like ω^2 they must have vanished at least twice, and therefore, there are exactly $2n$ real eigenvalues ω and stability follows (see Fig. 1(a)). The unnecessity of condition (48) is described in Figs. 1(b) and 1(c). Two eigenvalue curves $\lambda_j(\omega)$ may intersect and exchange their labels, or one of the lower eigenvalues λ_j, $j > 1$, may intersect the ω axis more than twice even though λ_1 does not intersect it at all. For the screw-pinch we will show, however, that the situation in Fig. 1(b) cannot occur.

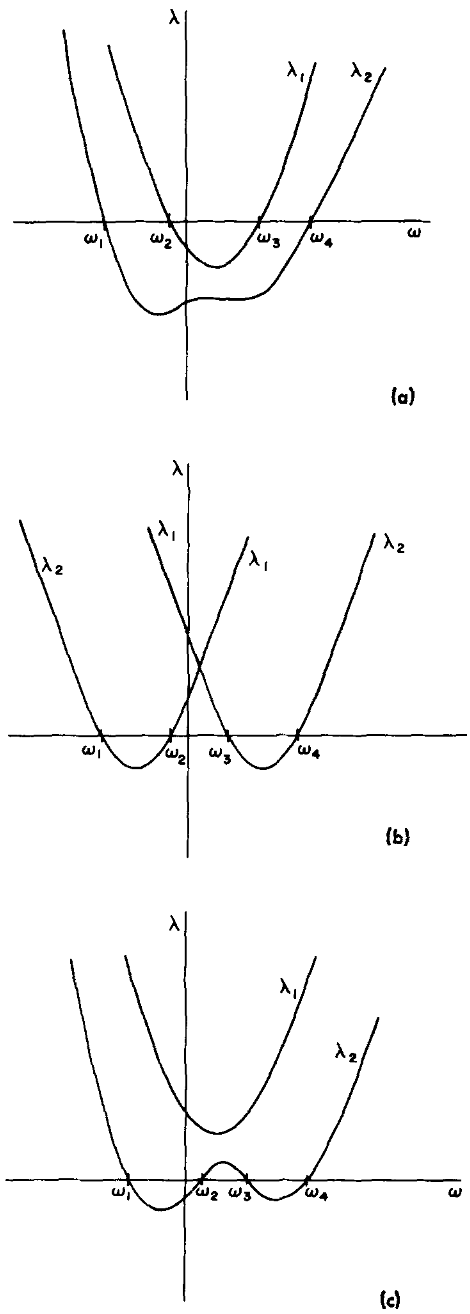


FIG. 1. Behavior of eigenvalues $\lambda_j(\omega)$ of Eq. (49) with 2×2 matrices in stable cases. (a) Sufficient condition (48) is satisfied. (b) and (c) The condition is violated.

V. SPECTRAL BOUNDS FOR THE SCREW-PINCH

The rotating screw-pinch with sufficiently smooth plasma profiles presents a workable model for the application of the theory described in the previous section, as well as a possibility to check the closeness of the spectral bounds to reality. This section is devoted to the derivation of results which will facilitate such an application. The λ-eigenvalue problem (49) is the main tool of our discussion. A computation of the λ 's can be made possible in the screw-pinch case by observing that a change $\omega^2 \to \omega^2 - \lambda$ in L_ω has a form similar to (7) and will likewise reduce to a system like (10). Notice that only ω^2 is changed but not ω. We can distinguish between the two by noting that ω in L_ω is mulitplied by $\mathbf{u}$ components while ω^2 is not. The system (10) then should be modified such that $\sigma^2 \to \sigma^2 - \lambda$ everywhere except for the last term in λ_{11} (and λ_{22}), while σ is not changed. In particular $A \to \hat{A} = H^2 - \rho(\sigma^2 - \lambda)$, $\rho^2\sigma^4 \to \rho^2(\sigma^2 - \lambda)^2$. We will denote modified expressions by adding a circumflex to the original notation, e. g., Eq. $(\hat{10})$. ω will be considered a real parameter. For convenience we cast some of the following results as claims.

Claim 1: For fixed wave numbers m, k and for fixed real ω, L_ω is essentially self-adjoint and bounded from above while the operator A is bounded.

Proof: L_ω is a second order Hermitian ordinary differential operator and $(L_\omega \boldsymbol{\xi}, \boldsymbol{\xi})$ involves derivatives of ξ only but not of ξ_θ, ξ_z. Since we imposed two boundary condition on ξ, self-adjointness can be shown in the same way that it is proved for more standard regular ordinary differential operators. (One shows that the Friedrichs self-adjoint extension of L_ω requires the same boundary conditions as we already imposed.[25]) The highest order term in ξ' in $(L_\omega \boldsymbol{\xi}, \boldsymbol{\xi})$ is $-2\pi \int r B_*^2 |\xi'|^2 \, dr \leqslant 0$. Thus it can be maximized with respect to ξ', considered as independent of ξ. What is left is an algebraic quadratic form in $\boldsymbol{\xi}$ which is bounded. The inefficient circle bounds based on this approximation appear in Ref. 13. The operator A reduces to a matrix mulitplication and therefore is bounded. Indeed $iA\,\boldsymbol{\xi} = -\rho \mathbf{k}\cdot\mathbf{u}\boldsymbol{\xi} + i\rho\Omega(-\xi_\theta \hat{r} + \xi\hat{\theta})$. Bounds (40) are given in terms of the λ spectrum for $-iA\,\boldsymbol{\xi} = \lambda\rho\boldsymbol{\xi}$. For each point in the plasma, $-iA$ has three λ eigenvalues, namely $\mathbf{k}\cdot\mathbf{u}$ and $\mathbf{k}\cdot\mathbf{u} \pm \Omega$. The spectrum of $-iA$ is the range of these values. It follows from (40) that

$$\operatorname*{Min}_r(\mathbf{k}\cdot\mathbf{u} - |\Omega|) \leqslant \operatorname{Re}\omega \leqslant \operatorname*{Max}_r(\mathbf{k}\cdot\mathbf{u} + |\Omega|). \tag{50}$$

Claim 2: L_ω has a λ-continuous spectrum consisting of all points $\lambda = \sigma^2 - \beta H^2/\rho$, $\lambda = \sigma^2 - H^2/\rho$ for any r in the plasma.

Proof: This is seen most directly from the analogy with the ω spectrum.[18] The continuous spectrum is visible in $(\hat{10})$ when letting $\hat{A}$ and $\hat{C}$ vanish. Note that since $\beta < 1$, the upper edge λ_c of the λ-continuous spectrum is given by

$$\lambda_c(\omega) = \operatorname*{Max}_r(\sigma^2 - \beta H^2/\rho). \tag{51}$$

Claim 3: Eigenvalue curves $\lambda(\omega)$ outside the λ-continuous spectrum do not intersect each other.

Proof: The λ eigenvalues can be obtained from the second order ordinary differential scalar equation ($\hat{15}$) when a solution satisfies the two boundary conditions. Intersection of two curves $\lambda(\omega)$ at $\omega = \omega_0$, $\lambda = \lambda_0$ means that at this point two independent solutions of ($\hat{15}$) vanish at the boundary. Hence every solution (their linear combination) should do so, which is impossible since one can prescribe an arbitrary value for a solution there.

Claim 4: The highest eigenvalue curve $\lambda_1(\omega)$ is convex as long as $\lambda_1(\omega) > \lambda_c(\omega)$.

Proof: $\lambda_1(\omega)$ is an analytic function of ω. Differentiate the relation $L_\omega \xi = \lambda_1 \rho \xi$ twice by ω and denote such derivatives by dots.

$$L_\omega \dot{\xi} + 2(\rho\omega + iA)\xi = \lambda_1 \rho \dot{\xi} + \dot{\lambda}_1 \rho \xi,$$

$$L_\omega \ddot{\xi} + 4(\rho\omega + iA)\dot{\xi} + 2\rho\xi = \lambda_1 \rho \ddot{\xi} + 2\dot{\lambda}_1 \rho \dot{\xi} + \ddot{\lambda}_1 \rho \xi.$$

Take the inner product of the second equation with ξ and of the first equation with $2\dot{\xi}$ and subtract. It follows that

$$\ddot{\lambda}_1(\rho\xi,\xi) = 2(\rho\xi,\xi) + 2[\lambda_1(\rho\dot{\xi},\dot{\xi}) - (L_\omega \dot{\xi},\dot{\xi})].$$

Since λ_1 is the uppermost point in the spectrum of L_ω, the bracket term is nonnegative and we have

$$d^2\lambda_1/d\omega^2 \geqslant 2 \quad \text{for } \lambda_1 > \lambda_c. \tag{52}$$

We have obtained the following result

Theorem 6: Let $R^2(\omega) = \max\{\lambda_1(\omega), \lambda_c(\omega)\}$ for ω real, where λ_1 is the largest isolated eigenvalue of L_ω. If $R^2(\omega) \leqslant 0$ for any real ω, the plasma is stable. If this condition is not satisfied then every complex ω eigenvalue of Eq. (7) is contained in the intersection of all circles with center ω and radius $R(\omega)$ for all real ω. If $\lambda_1(\omega) > \lambda_c(\omega)$ for all real ω, there is a *unique* circle of smallest radius. If $\lambda_c(\omega) > \lambda_1(\omega)$ for all ω, the circle (or circles) of smallest radius R is found by minimizing $\lambda_c(\omega)$ and

$$R^2 = \min_{\omega} \max_{r} (\sigma^2 - \beta H^2/\rho).$$

The following discussion deals with the question of how to decide the location of λ_1 relative to λ_c.

Claim 5: For fixed real ω and in the region $\lambda > \lambda_c(\omega)$, solutions of Eq. ($\hat{15}$) oscillate slower as λ increases.

Proof: By slower oscillation it is meant that all zeroes of a solution which vanish at a fixed point, move monotonically away from that point as λ increases. We prove the claim by noting that the quadratic form $\hat{Q}(\xi)$ corresponding to ($\hat{15}$) satisfies, similar to Eq. (17), $2\pi\hat{Q}(\xi) = (L_\omega \xi,\xi) - \lambda(\rho\xi,\xi)$, where ξ_θ, ξ_z take their stationary value for given ξ. For $\lambda > \lambda_c$ it can be shown directly that the stationary point is actually a maximum. Clearly $L_\omega - \lambda\rho$ decreases monotonically with λ, and thus so does $\hat{Q}(\xi)$. The oscillation result now follows similar to the usual one for ordinary differential equations.

In order to find whether $\lambda_1 < \lambda_c$ or not, it is sufficient to decide whether $L_\omega - \lambda_c(\omega)\rho$ is negative semidefinite. If it is negative then λ_c is the upper edge of the spectrum. Notice that for $\lambda = \lambda_c$, the coefficient $\hat{C}$ in ($\hat{10}$) vanishes somewhere in the plasma and the equation becomes singular. In fact, since λ_c has the maximum property (51), $\hat{C}$ will have in general a double zero. Yet, the singularity will still be a regular singularity as can be seen from Eq. (15) after we note that the matrix A can be written as $A = A_0 + CA_1$ and $\det A_0 \equiv 0$, so that $\det A$ is divisible by C. For the case of a regular singularity at r_0 we know that solutions behave like $(r - r_0)^{c_i}$ $(i = 1,2)$. If the c_i are complex, solutions oscillate rapidly near r_0. We can now state the following

Theorem 7: For a fixed real ω, λ_c as defined in (51) is the uppermost point in the spectrum of L_ω if and only if corresponding to all increasingly ordered points $r_1,...,r_s$ for which $\hat{C}$ vanishes in the plasma (with $\lambda = \lambda_c$), the indices $c_i(r_j)$ $(i = 1,2, j = 1,...,s)$ are real and no solution of Eq. ($\hat{15}$) vanishes more than once in any of the intervals $(0,r_1)$, (r_s, l), (r_j, r_{j+1}) for $j = 1,...,s-1$.

Proof: The theorem is an analog of the well-known Newcomb theorem[26] for the stationary pinch. The nonoscillation requirement of the theorem implies $\hat{Q}(\xi) \leqslant 0$, hence $L_\omega - \lambda_c\rho \leqslant 0$. Notice that the theorem applies only if $\hat{C}$ has at most a double zero at any r_j, but not higher. If $\hat{C}$ has a simple zero (the maximum in (51) is attained at the boundary) the singularity is still regular and the theorem holds.

From a practical point of view it may be more advisable to consider λ slightly larger than λ_c and then use claim 5 to track down λ_1 numerically. That is, if a solution of ($\hat{15}$) which satisfied the regularity condition at $r = 0$ (equivalent to having a zero there) vanishes before $r = l$, take λ larger until the first zero reaches $r = l$ and vice versa. We therefore do not pursue further the question of the value of the indices c_i in the last theorem or the question of optimizing the upper bound. We do discuss, however, our sufficient condition (48) for a particular point $\omega = \omega_0$ which is important in the derivation of Suydam's condition for rotating equilibria[13] and is also exceptional with respect to the previous discussion on λ_c. This is the point for which A and C both vanish at some point $r = r_0$. It follows that

$$H(r_0) \equiv \mathbf{k}\cdot\mathbf{B}(r_0) = 0, \quad \omega_0 = \mathbf{k}\cdot\mathbf{u}(r_0). \tag{53}$$

We assume that the wavenumbers are such that $\mathbf{k}\cdot\mathbf{B}$ does indeed vanish somewhere in the plasma. We also assume that h and α do not both vanish, where

$$h = H'(r_0), \quad \alpha = (\mathbf{k}\cdot\mathbf{u})'(r_0). \tag{54}$$

Equation (53) can be used to eliminate one of the wavenumbers from the definitions of h and α. Let μ be the pitch of the magnetic field,

$$\mu = B_\theta / rB_z. \tag{55}$$

We have

$$h = -kB_z\mu'/\mu, \quad \alpha = k(w' - \Omega'/\mu). \tag{56}$$

μ' is the magnetic shear while α is seen to be essentially the shear of the flow in a direction perpendicular to $\mathbf{B}$.

We would like to consider the possibility of $\lambda_c(\omega_0) = 0$. Notice that $\sigma^2 - \beta H^2/\rho$ has a stationary point at r_0 for $\omega = \omega_0$. It will be a local maximum if $\beta h^2 - \rho\alpha^2 > 0$, which we assume. This can be written as

$$\beta B_z^2 \mu'^2 > \rho(\Omega' - \mu w')^2, \tag{57}$$

and our assumption refers to the point $r = r_0$. In order to have $\lambda_c(\omega_0) = 0$, r_0 must be the global maximum in (51). If H vansihes in more than one point, say at $r = r_j$ $(j = 1,...,s)$, then corresponding to each point we have an ω_j as in (53). If not all the ω_j are equal we get $\lambda_c(\omega_j) \neq 0$ for all j, since at some point r_i, $\sigma \neq 0$ while $H = 0$. Equality of all the ω_j can be guaranteed if the flow is parallel to $\mathbf{B}$ up to a superimposed rigid rotation of the plasma column.

To apply the sufficient condition (48), consider Eq. $(\hat{15})$ (i. e., Eq. $(\hat{15})$ with $\lambda = 0$). Denoting $x = r - r_0$ we notice that near the singular point $\lambda_{11} = O(x^3)$, $\lambda_{12} = O(x^2)$, $\lambda_{21} = O(x^4) = \lambda_{22}$, and $A = O(x^2) = C$. Thus the coefficients of the equation have $p = O(x^2)$, $q = O(1)$ and the singularity is again regular. The indicial equation is of the form $c(c+1) + I = 0$ and the solutions do not oscillate rapidly if $I < \frac{1}{4}$. This is Suydam's condition[13] for local stability when (57) is satisfied. Since the explicit form of the condition is not readily available, we bring it here. Define

$$\phi = (\sqrt{\rho})(\Omega' - \mu w')/(B_z \mu'), \quad \psi = (\sqrt{\rho}) v / B_\theta.$$

The condition requires

$$\tfrac{1}{8} r B_z^2 (1 - \phi^2)\left(\frac{\mu'}{\mu}\right)^2 - \phi\psi B_z^2 \frac{\mu'}{\mu} + \tfrac{1}{2} r^2 \frac{\mathbf{B}^2}{B_\theta^2}(\rho\Omega^2)' + p' \quad (58)$$
$$> \frac{2}{\beta - \phi^2}\frac{B_\theta^2}{r}\{(1-\beta)[\tfrac{1}{4}\psi^4(1-\phi^2) - \psi^3\phi + \tfrac{3}{2}\phi^2\psi^2]$$
$$+ \beta[\tfrac{1}{2}\psi^2(1+\phi^2) - 2\phi\psi + \phi^2]\}.$$

When we let the flow vanish, (58) reduces to the classical Suydam condition[12]

$$\tfrac{1}{8} r B_z^2 \left(\frac{\mu'}{\mu}\right)^2 + p' > 0. \qquad (59)$$

If the flow is purely axial ($\psi = 0$) which is the expected case for toroidal devices,[1] (58) reduces to

$$\tfrac{1}{8} r B_z^2 \left(\frac{\mu'}{\mu}\right)^2 (1 - \phi^2) + p' > \frac{2\beta\phi^2}{\beta - \phi^2}\frac{B_\theta^2}{r} \qquad (60)$$

and is more difficult to satisfy than (59) since $\beta > \phi^2$ by (57). We can state the following result:

Theorem 8: Let $\mathbf{k}\cdot\mathbf{B}$ vanish in the plasma at $r = r_j$ $(j = 1,...,s)$ and $\omega_j = \mathbf{k}\cdot\mathbf{u}(r_j)$. Assume that all the ω_j $(j = 1,...,s)$ are equal (to ω_0) and that $(\omega_0 - \mathbf{k}\cdot\mathbf{u})^2 - \beta H^2/\rho$ is everywhere negative except at $r = r_j$. Then it is necessary for stability that condition (58) hold at $r = r_j$ $(j = 1,...,s)$ and it is sufficient for stability that no solution of Eq. (15) with $\omega = \omega_0$ vanishes more than once in any of the subintervals $(0, r_1)$, (r_s, l), (r_j, r_{j+1}), $j = 1,...,s-1$.

Proof: The necessary part was proven in Ref. 13 (without the maximality assumptions). The sufficiency part is a direct result of our sufficient condition (48) as described in the proof of Theorem 7. Notice that the violation of (58) implies that every solution vanishes infinitely many times in the appropriate subintervals. An interesting aspect of this theorem is that in the case of the static pinch, the sufficient condition is also *necessary* for stability.[26]

As a final topic of this section we discuss the dependence of our circle bounds on m and k. From the definition of λ_c we find

$$\lambda_c(s\omega;\ sm,\ sk) = s^2 \lambda_c(\omega;\ m,\ k) \qquad (61)$$

for any number s, which is assumed from now on to be positive. Notice that the matrix $\hat{\Lambda}$ in Eq. $(\hat{10})$ is homogeneous of degree four as a function of m, k, ω, $\sqrt{\lambda}$, except for the "exceptional" term $-\hat{C}\hat{A}^2$ in $\hat{\lambda}_{21}$, which is of degree six. As a result, $\hat{p}$ of Eq. $(\hat{15})$ is homogeneous of degree zero while $\hat{q} = \hat{q}_0 + r\hat{A}$, where $\hat{q}_0$ is also of degree zero while $\hat{A}$ is of course of degree two. Introducing s into the equation as a coefficient of m, k, ω, $\sqrt{\lambda}$ wherever these parameters appear and assuming $\lambda \geqslant \lambda_c$ (hence $\hat{A} > 0$), we find that $\hat{Q}(\xi;s)$ in $(\hat{16})$ is monotone decreasing in s. Denoting by $R(m,k)$ the radius of the circle of smallest radius (possibly zero) after minimizing over ω, we can now show

Theorem 9: $R(sm,sk)/s$ is a nonincreasing function of s.

Proof: Assume that for $m_0 = s_0 m$, $k_0 = s_0 k$, the smallest circle is centered at $\omega = \omega_0$ with radius R_0, $\lambda_0 = R_0^2$ and $\hat{Q}_0$ in $(\hat{16})$ (with obvious notation) is nonpositive. For $s > s_0$ and $m' = sm$, $k' = sk$, consider $\omega' = (s/s_0)\omega_0$ and $\lambda' = \lambda_0 (s/s_0)^2$. $\hat{Q}' \leqslant \hat{Q}_0$ and λ should be made smaller (unless $\lambda = \lambda_c$) to restore the upper bound of $\hat{Q}$ to zero. This proves the theorem.

Theorem 10: If a sufficient condition based on our circle estimates is satisfied for $m = 1$ and all k, and for $m = 0$, $k \to 0$, then the screw-pinch is stable to all modes.

Proof: This theorem is the counterpart of Theorem 1, Ref. 26 for the static pinch. The assertions of this theorem for $m > 1$ follow from Theorem 9 by taking $s = m$ and considering $R(m,k) = R(s, sk/m)$. If $m = 0$, it similarly follows that if $R(0,k_1) = 0$, it will also vanish for all $|k| > |k_1|$. Thus, the only mode to consider (except $k = 0$ which is a special case) is either $k_1 = 2\pi/L$ in a finite cylinder, or $k \to 0$ in the infinite cylinder. The limit $k \to 0$ should be taken by letting $s = k$, using $s\omega$ and $s^2\lambda$ instead of ω and λ, and then taking $s \to 0$. This amounts to using Eq. $(\hat{15})$ with $k = 1$; ω and λ still appear and the exceptional term $\hat{C}\hat{A}^2$ in q is dropped. If the circle estimates imply stability in this limit, it follows from Theorem 9 that all modes $m = 0$, $k \neq 0$ are stable. However, we can also show that the mode $m = 0$, $k = 0$ will be stable in this case. Indeed, stability of the $m = 0$, $k \to 0$ mode is shown by finding some ω and proper λ for which the limiting $\hat{Q}$ in $(\hat{16})$ is negative definite. Thus it will remain negative if we let $\lambda \to +\infty$. This last limit however is identical with what is obtained from (16) by setting $k = 0$, $m = 0$, and then $\omega = 0$. It is already known from Sec. III that the mode $k = 0$, $m = 0$ has a necessary and sufficient condition for stability, which is precisely the nonpositivity of the quadratic form we obtained. This completes the proof.

To conclude this section we summarize its practical consequences. A numerical calculation of ω eigenvalues and of our circle estimates will be facilitated by introducing two additional parameters to Eq. (10). The first is λ, as described at the beginning of the section, and the second is s^2 as a coefficient of CA^2 in λ_{21}. s will have the effect of enabling one to consider only modes with $m = 0$ or $m = 1$ while all other modes (m,k) can be replaced by $m \to 1$, $k \to k/m$, $s = m$. The benefit of this change is that the search for ω is done over a region smaller by a factor $1/s$ than in the actual problem. This is useful in particular if the actual complex eigenvalues ω are sought (in which case λ is set to 0). Of course, once an $\omega(s)$ eigenvalue is found in this way, it should be multiplied by s to correspond to its actual value. The computation should be first run for real ω in the range (50) (shrunk by a factor s). One should search for the first eigenvalue λ which will result in a circle bound on all complex eigenvalues. The λ parameter should then be set to 0 and the search for complex ω eigenvalue should be carried out inside that circle.

ACKNOWLEDGMENTS

The author wishes to thank Professor Harold Weitzner for introducing him to this area of plasma physics, and for many useful discussions. The author is also grateful to Dr. Abraham Kadish for his encouragement.

This work was supported by U. S. Department of Energy contract No. DE-AC02-76ER03077.

[1]A. B. Hassam and R. M. Kulsrud, Phys. Fluids **21**, 2271 (1978).
[2]J. P. Freidberg and J. A. Wesson, Phys. Fluids **13**, 1117 (1970).
[3]A. Kadish, Phys. Fluids **19**, 141 (1976).
[4]D. V. Anderson, D. C. Barnes, and C. E. Seyler, Bull. Am. Phys. Soc. **24**, 988 (1979).
[5]L. C. Steinhauer, Bull. Am. Phys. Soc. **24**, 927 (1979).
[6]I. B. Bernstein, E. A. Frieman, M. D. Kruskal, and R. M. Kulsrud, Proc. R. Soc. London Ser. A **224**, 1 (1958).
[7]E. Frieman and M. Rotenberg, Rev. Mod. Phys. **32**, 898 (1960).
[8]G. O. Spies, Phys. Fluids **21**, 580 (1978).
[9]J. P. Freidberg and D. S. Pearlstein, Phys. Fluids **21**, 1207 (1978).
[10]E. Hameiri, Phys. Fluids **22**, 89 (1979).
[11]G. O. Spies, Phys. Fluids **23**, 2017 (1980).
[12]B. R. Suydam, in *Proceedings of the Second International Conference on the Peaceful Uses of Atomic Energy* (United Nations, Geneva, 1958), Vol. 31, p. 157.
[13]E. Hameiri, Courant Institute Report MF-85 (COO-3077-123) (1976).
[14]E. M. Barston, J. Math. Phys. **18**, 750 (1977).
[15]L. N. Howard, J. Fluid Mech. **10**, 509 (1961).
[16]E. M. Barston, Int. J. Eng. Sci. **18**, 477 (1980).
[17]K. Hain and R. Lüst, Z. Naturforsch. **13a**, 936 (1958).
[18]H. Grad, Proc. Nat. Acad. Sci. USA **70**, 3377 (1973).
[19]J. P. Goedbloed and P. H. Sakanaka, Phys. Fluids **17**, 908 (1974).
[20]P. H. Sakanaka and J. P. Goedbloed, Phys. Fluids **17**, 919 (1974).
[21]E. Hameiri and J. H. Hammer, Phys. Fluids **22**, 1700 (1979).
[22]L. Woltjer, Astrophys. J. **130**, 405 (1959).
[23]E. Hameiri (to be published).
[24]E. M. Barston, J. Math. Phys. **8**, 523 (1967).
[25]M. Reed and B. Simon, *Methods of Modern Mathematical Physics* (Academic, New York, 1975), Vol. II, Chap. X. 3.
[26]W. A. Newcomb, Ann. Phys. **10**, 232 (1960).