Spin-two fields and general covariance

Robert M. Wald
Enrico Fermi Institute and the Department of Physics, 5640 South Ellis Avenue, Chicago, Illinois 60637
(Received 10 December 1985)

It is widely believed that all "consistent" theories of a spin-two field coupled to matter or nonlinearly self-coupled must be generally covariant. The extent to which this statement is true is investigated here. We consider at the classical level nonlinear equations of motion for a field γ_{ab} in a flat background spacetime which are derived from a Lagrangian and which reduce, in linear order, to the equations of a spin-two field. In a perturbation expansion about $\gamma_{ab}=0$, we argue that in order for all the linearized solutions to give rise to a one-parameter family of exact solutions, the exact equations of motion must satisfy a certain type of divergence identity. (This is our "consistency" condition.) When the equations of motion arise from an action principle as we assume, this divergence identity implies an infinitesimal gauge invariance of the action. However, our main result is the demonstration that only a very restricted class of candidate infinitesimal gauge symmetries can actually arise from an exact (i.e., finite) gauge symmetry, as is necessary to realize the theory. Under some assumptions concerning the number of derivatives which occur in terms appearing in the divergence identity, we prove that only two types of gauge invariance are possible: (i) normal spin-two gauge invariance and (ii) general covariance. Explicit examples of nonlinear field theories of type (i) are constructed. When coupling to matter is considered, the requirement that in linear order γ_{ab} couple directly to the stress-energy tensor of matter may eliminate possibility (i), but I have shown this only in special cases. A similar analysis of nonlinear generalizations of the equations for a collection of spin-one fields is given, and it is shown that under analogous assumptions, the only possible type of gauge invariance for the nonlinear theory is Yang-Mills gauge invariance with respect to an arbitrary Lie algebra.

I. INTRODUCTION

It has been known for quite some time that severe consistency constraints arise when one attempts to couple a massless spin-two field γ_{ab} to matter or to couple it nonlinearly to itself. There is a widespread belief that these constraints force one to a theory which is generally covariant. (As will be discussed further at the end of Sec. III below, by "generally covariant" we mean that although the theory is initially formulated as the theory of a field γ_{ab} in a flat background spacetime $(\mathbb{R}^4,\eta_{ab})$, a change of dynamical field variable from γ_{ab} to $g_{ab}=g_{ab}[\eta_{cd},\gamma_{ef}]$ can be made so that the theory no longer depends on the flat background metric η_{ab}.) The main argument advanced that string theory is a theory of gravity is that, since it contains a massless, spin-two excitation, it must be a generally covariant theory.

Although there exists a rather extensive literature on spin-two fields and general covariance (see Refs. 1–4 and other references cited therein), it is not easy to extract from these references any mathematically precise claims (no less proofs) concerning the general class of consistent field theories. As will be discussed further in Sec. IV, the original Feynman[1] argument depends heavily on the particular form of the coupling of γ_{ab} to the (particle) matter which Feynman assumed. Other arguments rely heavily upon the assertion that for a consistent theory, the right side of the spin-two field equation must equal "the full stress-energy tensor of matter plus gravitation," which is then asserted to equal the functional derivative of the action with respect to the flat, background metric. Finally, other arguments concerning consistent nonlinear theories of a spin-two field have been formulated within the context of a (presumed to exist) quantum theory of gravity, although a translation to classical language presumably could be made. General relativity is then obtained[4] as the unique "low-energy limit" of such a theory, i.e., when only terms in the action containing no more than two derivatives of the field variable are considered. While the results of this approach are clearly related to the results we shall obtain below, a key difference is that we shall not impose any restrictions on the number of derivatives of the field variable appearing in the action, although as explained below we shall impose some restrictions on the number of derivatives in terms appearing in the gauge transformation symmetry of the theory.

The purpose of this paper is to give a systematic analysis at the classical level of the possible types of consistent, nonlinear generalizations of the theory of a spin-two field in flat spacetime. We shall confirm the above expectation that "consistency" of the theory (in the sense defined in Sec. II) together with the requirement that the equations of motion be derived from a Lagrangian does indeed impose very severe restrictions on the theory. However, we shall see that there do exist a class of theories which are not generally covariant; an explicit example is given by the Lagrangian of (3.18) below. Theories of this class have the "normal" gauge invariance of a spin-two field, $\gamma_{ab}\rightarrow\gamma_{ab}+\partial_{\langle a}\xi_{b\rangle}$. Nevertheless, the main result of this paper is the demonstration that, under some additional assumptions, consistent nonlinear generalizations of the theory of a spin-two field must either be of the above class or must be generally covariant. Thus, apart from the theories which have normal gauge invariance, general covariance appears to be the only possibility.

The key ideas behind the proof of our results are as follows. The linearized Einstein operator for the spin-two field satisfies a divergence identity: the linearized Bianchi identity. Consequently, in a nonlinear generalization of this theory, in a perturbation expansion, the divergence of the second-order equations will yield an equation involving only first-order quantities. In order that this equation not impose further restrictions on the linear solutions, the divergence of the second-order equation must vanish as a consequence of the first-order equation. More generally, in order for the nth-order equations not to impose restrictions on the lower-order quantities, we argue that the exact, nonlinear equations must satisfy a certain type of divergence identity. The assumption that the equations arise from an action principle is now imposed. The divergence identity implies a constraint on the action S, namely, that it be invariant under certain infinitesimal variations of the field γ_{ab}; in other words, it requires an "infinitesimal gauge invariance" of S. Suppose we denote by $\mathscr{V}_1$ and $\mathscr{V}_2$ two vector fields on the infinite-dimensional manifold of field configurations which represent these infinitesimal gauge directions. If the action is invariant under the motions generated by $\mathscr{V}_1$ and $\mathscr{V}_2$, then it must also be invariant under infinitesimal motion along their commutator $[\mathscr{V}_1,\mathscr{V}_2]$. The requirement that these commutators do not generate additional constraints on S—which we refer to as our *integrability condition*— places very severe restrictions on the possible types of infinitesimal gauge symmetries. It leads directly to our main result that only the above two classes of theories are possible.

We begin our investigation in Sec. II with an analysis of the simpler case of a massless spin-one field A_a. We seek "consistent" nonlinear generalizations of Maxwell's equations which arise from a Lagrangian. Our precise framework and assumptions (which will be carried over directly to the spin-two case) will be spelled out, and it then will be shown that all such theories must have the normal gauge invariance $A_a \rightarrow A_a + \partial_a \chi$. The analysis will then be generalized to a collection of spin-one fields, and it will be shown that all such theories must have Yang-Mills-type gauge invariance with respect to some Lie algebra. It should be noted that the Lie-algebra conditions arise solely from our integrability condition on gauge symmetries; no assumptions concerning group theory, connections on bundles, etc., are introduced.

In Sec. III we consider nonlinear generalizations of the spin-two equations in the simple case of no coupling to matter. We parallel the analysis of Sec. II and obtain our main result concerning the two classes of possible theories. The section concludes with a definition of "general covariance" and a demonstration that the second class of theories is generally covariant. Finally, in Sec. IV, we consider coupling of the spin-two field to matter. While noncovariant theories with coupling to matter certainly exist, some evidence (but not a general proof) is presented that there may not exist any theory that, in linear order, couples to the "usual stress-energy tensor of matter." The original Feynman result[1] is also briefly discussed from our viewpoint and some key assumptions implicit in his analysis are spelled out. Finally, it is remarked that to determine which of the two classes a given theory belongs to, one needs only examine the gravitational part of the Lagrangian to third order in γ_{ab}, or the matter part of the Lagrangian to second order in γ_{ab}.

While we attempt in this paper to give a careful and systematic analysis of the possible types of theories, it should be noted that a number of loopholes exist in our arguments. The most important of these is a strong assumption (for which, however, some motivation is given) concerning the number of derivatives that occur in terms appearing in the divergence identity satisfied by the field equations. (This restricts the possible forms of infinitesimal gauge transformations.) Furthermore, while our framework is quite general, it is not clear that it can encompass, for example, the equations arising from string theory.

Conversely, there are a number of reasonable requirements which we have not imposed. In particular, we have not required the equations to have a well-posed initial-value formulation, nor have we restricted the number of derivatives of γ_{ab} which appear in the equations (other than that this number be finite). Such additional requirements impose severe further limitations on the possible theories. Indeed, it appears likely that if one imposes the additional requirement that there appear no derivatives of γ_{ab} higher than second order, one would uniquely obtain Einstein's equation.

II. SPIN-ONE FIELDS

The purpose of this section is to set forth our framework, assumptions, and methods in the simpler context of massless spin-one fields (with no current sources). At the end of this section, we will see that for a collection of spin-one fields, the Yang-Mills-type of gauge invariance arises as a necessary requirement for a consistent nonlinear self-coupling. However, we shall begin in the simplest context of a single spin-one field, where we will show that the usual type of gauge invariance must hold for consistent nonlinear generalizations of Maxwell's equations.

The equations of motion for a spin-one field A_a in Minkowski spacetime—Maxwell's equations—are

$$\mathscr{F}^{(1)a}=0 , \tag{2.1}$$

where

$$\mathscr{F}^{(1)a} \equiv \partial_b F^{ab} \tag{2.2}$$

with

$$F_{ab} \equiv \partial_a A_b - \partial_b A_a . \tag{2.3}$$

[Here we use the abstract index conventions of Ref. 5. The superscript (1) on quantities such as $\mathscr{F}^{(1)a}$ is used here and throughout the paper to denote the number of powers of the field variable (in this case A_a) occurring in the expression.] These equations arise from variations with respect to A_a of the action

$$S^{(2)}=\int \mathscr{L}^{(2)}d^4\chi \ , \tag{2.4}$$

where

$$\mathscr{L}^{(2)}=\tfrac{1}{4}F_{ab}F^{ab} \ , \tag{2.5}$$

i.e., we have

$$\mathscr{F}^{(1)a}=\frac{\delta S^{(2)}}{\delta A_a} \ , \tag{2.6}$$

where the right-hand side of Eq. (2.6) denotes the functional derivative.

We seek, now, a nonlinear generalization of Maxwell's equations to be obtained by adding terms to $\mathscr{L}^{(2)}$ which are cubic or higher order in A_a. We denote the full Lagrangian density thus obtained by $\mathscr{L}$ and full action by S. We impose no restriction on the orders of A_a allowed (i.e., nonpolynomial Lagrangians are permissible) but we require that the total number of derivatives which occur in each order is bounded by a (fixed) integer N.

The exact equations of motion are

$$\mathscr{F}^a\equiv\frac{\delta S}{\delta A_a}=0 \ . \tag{2.7}$$

However, although any choice of $\mathscr{L}$ will, of course, yield equations of motion, a serious consistency problem will in general arise for these equations, as can be seen from the following. Suppose we imagine solving Eq. (2.7) by a perturbation expansion about the zero solution, $A_a=0$. The linearized perturbation $\dot{A}_a$ will, of course, satisfy Maxwell's equation:

$$\mathscr{F}^{(1)a}[\dot{A}_b]=0 \ . \tag{2.8}$$

The second-order equation will take the form

$$F^{(1)a}[\ddot{A}_b]+\mathscr{F}^{(2)a}[\dot{A}_b]=0 \ , \tag{2.9}$$

where $\ddot{A}_b$ denotes the second-order perturbation and $\mathscr{F}^{(2)a}$ denotes the quadratic part of $\mathscr{F}^a$. The potential problem is caused by the fact that $\mathscr{F}^{(1)a}$, Eq. (2.2), satisfies the identity

$$\partial_a\mathscr{F}^{(1)a}=0 \ . \tag{2.10}$$

Hence, if we take the divergence of Eq. (2.9), the term involving $\ddot{A}_a$ disappears and we obtain as an integrability condition for solving (2.9) an equation involving only the first-order perturbation:

$$\partial_a\mathscr{F}^{(2)a}[\dot{A}_b]=0 \ . \tag{2.11}$$

Thus, linearized perturbations must satisfy Eq. (2.11) in addition to Eq. (2.8); not all solutions of the original linear equation give rise to one-parameter families of exact solutions. Indeed, in general, Eq. (2.11) will be incompatible with Eq. (2.8) and few, if any, solutions will exist. Similar problems will also occur in higher orders in perturbation theory. This is the "consistency problem."

A sufficient condition to ensure that no such consistency problem arises is that the quantity $\mathscr{F}^a$ satisfy a divergence identity of the general form

$$\partial_a\mathscr{F}^a=\lambda_a\mathscr{F}^a+\rho_a{}^b\partial_b\mathscr{F}^a+\sigma_a{}^{bc}\partial_b\partial_c\mathscr{F}^a+\cdots \ , \tag{2.12}$$

where the tensors $\lambda_a,\rho_a{}^b,\sigma_a{}^{bc}$ are locally constructed out of η_{ab}, A_a, and derivatives of A_a and vanish when $A_a=0$. [The sum on the right-hand side of Eq. (2.12) is finite.] This identity will guarantee that the equation obtained from taking the divergence of the nth-order perturbation equation contains no new information beyond what is available from the lower-order equations; the integrability condition for the nth-order perturbation will be satisfied by virtue of the previous equations. For example, if we expand Eq. (2.12) to second order in A_a, we obtain

$$\partial_a\mathscr{F}^{(2)a}=\lambda_a^{(1)}\mathscr{F}^{(1)a}+\rho_a^{(1)b}\partial_b\mathscr{F}^{(1)a}+\sigma_a^{(1)bc}\partial_b\partial_c\mathscr{F}^{(1)a}+\cdots \ . \tag{2.13}$$

Thus, Eq. (2.8) now implies Eq. (2.11) and no consistency problem of the above sort arises to second order. The situation is the same for the higher orders.

Conversely, it appears plausible that an identity of the form (2.12) must be satisfied if the integrability conditions for the perturbation equations are to yield no additional equations on the lower-order quantities. Thus, I will adopt the existence of an identity of the form (2.12) as the fundamental requirement for a consistent nonlinear generalization of the theory of a spin-one field. It should be noted, however, that I cannot rule out the possible existence of theories which could be judged as consistent by some other reasonable criteria but which do not satisfy Eq. (2.12).

A further restrictive assumption will now be made. We wish Eq. (2.7) to be a local, partial differential equation, involving a bounded number of derivatives of A_a. This would be difficult—although not impossible—to achieve if the higher derivative terms beyond λ_a and $\rho_a{}^b$ are present in the identity (2.12). Namely, in a perturbation expansion, Eq. (2.12) then would relate a single derivative of $\mathscr{F}^{(n)a}$ to second and higher derivatives of the lower order $\mathscr{F}^{(j)a}$. Thus, unless the second and higher derivative terms acting on the highest derivative parts of the $\mathscr{F}^{(j)a}$ vanish identically or cancellations occur in the various terms, the successive $\mathscr{F}^{(n)a}$ will contain a larger and larger total number of derivatives. Similar problems will occur if λ_a contains more than one derivative of A_a or if $\rho_a{}^b$ contains any derivatives of A_a. Thus, I shall restrict attention to theories for which an identity of the form

$$\partial_a\mathscr{F}^a=\lambda_a\mathscr{F}^a+\rho_a{}^b\partial_b\mathscr{F}^a \tag{2.14}$$

holds, where $\rho_a{}^b$ is constructed locally out of A_a and η_{ab} (with no derivatives of A_a appearing) and λ_a is constructed locally out of A_a, η_{ab}, and ∂_aA_b, and is, at most, linear in ∂_aA_b. This restriction is probably the most significant loophole in our analysis.

Equation (2.14) may be rewritten more conveniently in the form

$$\partial_b(\beta_a{}^b\mathscr{F}^a)=\beta_a{}^b\alpha_b\mathscr{F}^a \ , \tag{2.15}$$

where α_a and $\beta_a{}^b$ have the same properties as λ_a and $\rho_a{}^b$

except that now when $A_a=0$ we have $\beta_a{}^b=\delta_a{}^b$ and $\alpha_a=0$. Note that there is some arbitrariness in α_a and $\beta_a{}^b$ in that the transformation

$$\beta_a{}^b \to f\beta_a{}^b \; , \tag{2.16a}$$

$$\alpha_a \to \alpha_a + f^{-1}\partial_a f \; , \tag{2.16b}$$

leaves the identity (2.15) unchanged, where f is an arbitrary function constructed locally out of A_a and η_{ab}, with $f=1$ when $A_a=0$.

Until now, we have not used the fact that our equations arise from an action via (2.7). We use this fact now to reexpress our identity (2.15) as an infinitesimal gauge symmetry of the action. Namely, for an arbitrary function χ, Eq. (2.15) implies

$$\begin{aligned} 0 &= \int \chi[\partial_b(\beta_a{}^b\mathcal{F}^a)-\beta_a{}^b\alpha_b\mathcal{F}^a]d^4x \\ &= -\int \beta_a{}^b(\partial_b\chi+\alpha_b\chi)\frac{\delta S}{\delta A_a}d^4x \end{aligned} \tag{2.17}$$

which states that S is unchanged under the infinitesimal variation

$$\delta A_a = \beta_a{}^b(\partial_b\chi+\alpha_b\chi) \; . \tag{2.18}$$

This infinitesimal gauge invariance can be viewed in the following manner. The collection of field configurations A_a on spacetime can be viewed as an infinite-dimensional manifold $\mathcal{M}$ (which has a natural vector-space structure). The action S is a scalar function on $\mathcal{M}$ and the infinitesimal variation (2.18) defines a vector field $\mathcal{V}_\chi$ on this manifold. Thus, Eq. (2.17) states that the directional derivative of S along $\mathcal{V}_\chi$ vanishes for all χ. In other words at each point $p\in\mathcal{M}$ (i.e., for each field configuration) there is an (infinite-dimensional) subspace W_p of the tangent space V_p at p such that the derivative of S vanishes for directions in this subspace. The key issue is whether these subspaces are integrable; that is, whether one can find submanifolds $\mathcal{S}_\alpha$ filling $\mathcal{M}$ which are everywhere tangent to the collection of subspaces $W=\{W_p \mid p\in\mathcal{M}\}$. If so, then by choosing the function S to be constant on the surfaces $\mathcal{S}_\alpha$, one can obtain a theory which realizes the infinitesimal gauge symmetry (2.18). In other words, the integrability of the subspaces implies that the infinitesimal gauge symmetry (2.18) corresponds to some exact gauge symmetry and one then need only choose a gauge-invariant action S to obtain a theory which satisfies the divergence identity (2.15). Conversely, if the subspace are not integrable, then there is no exact gauge symmetry corresponding to (2.18). It remains possible that there could exist a larger gauge invariance—i.e., that there still exist integrable subspaces which contain W and are not the entire tangent space—but this would imply a larger gauge invariance for the exact theory than occurs for the linear theory, and the behavior of the exact theory for small A_a presumably would not reduce to that of the linear theory. In any case, we shall not consider this possibility.

Thus, we adopt as our criterion for a consistent theory that the subspace W generated by the vector fields $\mathcal{V}_\chi$, Eq. (2.18), be integrable. By Frobenius's theorem this will occur if and only if the commutator of two vector fields in W lies in W. In other words, for each pair of functions ϕ and ψ on spacetime, there must exist a function χ such that

$$[\mathcal{V}_\phi,\mathcal{V}_\psi]=\mathcal{V}_\chi \; . \tag{2.19}$$

Thus, using our formula, Eq. (2.18), for these vector fields, this integrability condition (2.19) becomes

$$\begin{aligned} &(\delta_\phi\beta_a{}^b)(\partial_b\psi+\alpha_b\psi)-(\delta_\psi\beta_a{}^b)(\partial_b\phi+\alpha_b\phi) \\ &\quad +\beta_a{}^b(\delta_\phi\alpha_b)\psi-\beta_a{}^b(\delta_\psi\alpha_b)\phi=\beta_a{}^b(\partial_b\chi+\alpha_b\chi) \; , \end{aligned} \tag{2.20}$$

where δ_ϕ denotes the linear change in the quantity induced by the variation

$$\delta A_a = \beta_a{}^b(\partial_b\phi+\alpha_b\phi) \; .$$

Since $\beta_a{}^b$ is constructed locally out of only A_a and η_{ab}, we find

$$\begin{aligned} \delta_\phi\beta_a{}^b &= \frac{\partial\beta_a{}^b}{\partial A_c}\delta_\phi A_c \\ &= \frac{\partial\beta_a{}^b}{\partial A_c}\beta_c{}^d(\partial_d\phi+\alpha_d\phi) \; . \end{aligned} \tag{2.21}$$

Hence, Eq. (2.20) becomes

$$\begin{aligned} &2\frac{\partial\beta_a{}^{[b}}{\partial A_c}\beta_c{}^{d]}(\partial_d\phi+\alpha_d\phi)(\partial_b\psi+\alpha_b\psi) \\ &\quad +\beta_a{}^b[(\delta_\phi\alpha_b)\psi-(\delta_\psi\alpha_b)\phi]=\beta_a{}^b(\partial_b\chi+\alpha_b\chi) \; . \end{aligned} \tag{2.22}$$

Now, under an algebraic change of dynamical variables

$$A_a \to \tilde{A}_a(A_b,\eta_{cd}) \tag{2.23}$$

we have

$$\begin{aligned} \mathcal{F}^a \to \tilde{\mathcal{F}}^a &= \frac{\delta S}{\delta\tilde{A}_a} = \frac{\delta S}{\delta A_b}\frac{\partial A_b}{\partial\tilde{A}_a} \\ &= \mathcal{F}^b\frac{\partial A_b}{\partial\tilde{A}_a} \; . \end{aligned} \tag{2.24}$$

Comparing this with Eq. (2.15), we see that no change is induced in α_a and the change induced in $\beta_a{}^b$ is such that we can set $\beta_a{}^b=\delta_a{}^b$ by a change of variables provided that we can find an $\tilde{A}_a$ such that

$$(\beta^{-1})_a{}^b = \frac{\partial\tilde{A}_a}{\partial A_b} \; . \tag{2.25}$$

The integrability condition for this equation is that the second partial derivatives be symmetric:

$$\begin{aligned} 0 = 2\frac{\partial^2\tilde{A}_a}{\partial A_{[c}\partial A_{b]}} &= \frac{\partial(\beta^{-1})_a{}^b}{\partial A_c} - \frac{\partial(\beta^{-1})_a{}^c}{\partial A_b} \\ &= -(\beta^{-1})_d{}^b(\beta^{-1})_a{}^e\frac{\partial\beta_e{}^d}{\partial A_c} \\ &\quad +(\beta^{-1})_d{}^c(\beta^{-1})_a{}^e\frac{\partial\beta_e{}^d}{\partial A_b} \; . \end{aligned} \tag{2.26}$$

This is equivalent to the relation

$$\frac{\partial\beta_a{}^{[b}}{\partial A_c}\beta_c{}^{d]}=0 \; . \tag{2.27}$$

Hence, the condition that $\beta_a{}^b$ can be set equal to $\delta_a{}^b$ by a change of variables is equivalent to the vanishing of the first term in Eq. (2.22).

We shall now demonstrate that by a combination of the transformation (2.16) together with the change of variables (2.23), all solutions of Eq. (2.22) are equivalent to theories with $\beta_a{}^b=\delta_a{}^b$ and $\alpha_a=0$. In other words, with the caveats discussed above, any consistent nonlinear generalization of Maxwell's equations must have (possibly after redefinition of field variables) the normal gauge invariance under $A_a \to A_a + \partial_a \chi$.

In investigating Eq. (2.22), it should be noted that ϕ and ψ are (arbitrary) given functions and we are attempting to find a (ϕ- and ψ-independent) $\beta_a{}^b$, α_a, and (ϕ-and ψ-dependent) χ such that Eq. (2.22) holds for all ϕ and ψ. The complicated dependence of this equation on α_a and $\beta_a{}^b$ is such that a direct attack upon (2.22) as it stands does not appear promising. However, the situation improves dramatically if one expands the unknown quantities in a power series in A_a:

$$\alpha_a = \sum_n \alpha_a^{(n)} , \tag{2.28}$$

$$\beta_a{}^b = \sum_n \beta_a^{(n)b} , \tag{2.29}$$

$$\chi = \sum_n \chi^{(n)} , \tag{2.30}$$

where, as before, the superscript (n) denotes the terms containing precisely n powers of A_a (so that if $A_a \to \lambda A_a$ with λ constant, we have $\alpha_a^{(n)} \to \lambda^n \alpha_a$, etc.). Let us similarly expand Eq. (2.22) in powers of A_a. Consider, first, the zeroth-order terms in A_a in Eq. (2.22). Since the δ_ϕ and δ_ψ variations reduce the dependence on A_a by one order, there will be contributions from $\beta_a^{(1)b}$ and $\alpha_a^{(1)}$. We must have

$$\alpha_a^{(1)} = cA_a , \tag{2.31}$$

where c is a constant, for the simple reason that it is the only one-index tensor that can be constructed locally from η_{ab} and a single A_a or a single $\partial_a A_b$. On the other hand, we must have

$$\beta_a^{(1)b} = 0 \tag{2.32}$$

for the simple reason that no two-index tensor can be constructed algebraically from η_{ab} and a single A_a. Since $\beta_a^{(0)b}=\delta_a{}^b$ and $\alpha_a^{(0)}=0$, the zeroth-order terms in Eq. (2.22) are simply

$$c(\psi \partial_a \phi - \phi \partial_a \psi) = \partial_a \chi^{(0)} . \tag{2.33}$$

Taking the curl of this equation, we obtain

$$c\,\partial_{[b} \psi \partial_{a]} \phi = 0 \tag{2.34}$$

which can be satisfied for all ϕ and ψ only if $c=0$. Thus, we find

$$\alpha_a^{(1)} = 0 . \tag{2.35}$$

We now assume, inductively, that, by use of the transformation (2.16) as well as the change of variables (2.23), if necessary, we have obtained $\beta_a{}^b=\delta_a{}^b$ and $\alpha_a=0$ up to order n. (This inductive hypothesis has just been proven for $n=1$.) We wish to show that by further use of the transformation (2.16) and change of variables (2.23) we can obtain $\beta_a^{(n+1)b}=0$ and $\alpha_a^{(n+1)}=0$. Under the inductive hypothesis, the nth-order part of Eq. (2.22) is

$$\left[\frac{\partial \beta_a^{(n+1)b}}{\partial A_d} - \frac{\partial \beta_a^{(n+1)d}}{\partial A_b} \right] \partial_b \psi \partial_d \phi + (\delta_\phi^{(0)} \alpha_a^{(n+1)}) \psi - (\delta_\psi^{(0)} \alpha_a^{(n+1)}) \phi = \partial_a \chi^{(n)} , \tag{2.36}$$

where $\delta_\phi^{(0)}$ denotes the "zeroth-order variation," i.e., the linear change in the quantity induced by the variation

$$\delta_\phi^{(0)} A_a = \partial_a \phi . \tag{2.37}$$

Thus, explicitly, we have

$$\delta_\phi^{(0)} \alpha_a^{(n+1)} = \frac{\partial \alpha_a^{(n+1)}}{\partial A_b} \partial_b \phi + \frac{\partial \alpha_a^{(n+1)}}{\partial(\partial_c A_b)} \partial_c \partial_b \phi . \tag{2.38}$$

Now consider Eq. (2.36) in the case where $\psi=1$ but ϕ is an arbitrary function. Then $\delta_\psi^{(0)} \alpha_a^{(n+1)}=0$ and we find simply

$$\delta_\phi^{(0)} \alpha_a^{(n+1)} = \partial_a \chi^{(n)} \tag{2.39}$$

[where it should be remembered that $\chi^{(n)}$ depends on ϕ and ψ, so $\chi^{(n)}$ is not necessarily "the same function" in Eqs. (2.36) and (2.39)]. Thus, $\delta_\phi^{(0)} \alpha_a^{(n+1)}$ is a gradient. Now, in the linear case, Eq. (2.31), $\delta_\phi^{(0)} \alpha_a^{(1)}$ is a gradient even though $\alpha_a^{(1)}$ itself is not a gradient. However, this is not possible in the nonlinear case. To see this, we take the curl of Eq. (2.39) and obtain

$$\delta_\phi^{(0)} [\partial_{[b} \alpha_{a]}^{(n+1)}] = 0 . \tag{2.40}$$

But this equation states that $\partial_{[b} \alpha_{a]}^{(n+1)}$ is gauge invariant under the infinitesimal variation (2.37), which implies (by integrating these infinitesimal variations) that $\partial_{[b} \alpha_{a]}^{(n+1)}$ must be gauge invariant in the usual sense, i.e., under $A_a \to A_a + \partial_a \phi$. This means that, if nonvanishing, $\partial_{[b} \alpha_{a]}^{(n+1)}$ must be constructed entirely out of $F_{ab} = 2\partial_{[a} A_{b]}$ and η_{ab}. However, this is out of the question for $\partial_{[b} \alpha_{a]}^{(n+1)}$ with $n>1$ since that quantity contains $(n+1)$ A_a's but at most two derivatives. Examination of the limited possibilities available for $n=1$ also shows that $\partial_{[b} \alpha_{a]}^{(2)}$ cannot be nonvanishing and gauge invariant; only the case $\alpha_a^{(1)}=cA_a$ works. Thus, for $n \geq 1$, we have

$$\partial_{[b} \alpha_{a]}^{(n+1)} = 0 \tag{2.41}$$

and, thus,

$$\alpha_a^{(n+1)} = \partial_a f^{(n+1)}[A_b] , \tag{2.42}$$

where $f^{(n+1)}$ is a scalar function constructed locally out of A_a and η_{ab} (but no derivatives of A_a are permissible). But this is precisely the form which can be eliminated by the transformation (2.16). Thus, using this freedom we may set

$$\alpha_a^{(n+1)} = 0 . \tag{2.43}$$

Returning, now, to Eq. (2.36) for general ϕ and ψ, we find

$$\left[\frac{\partial \beta_a^{(n+1)b}}{\partial A_d} - \frac{\partial \beta_a^{(n+1)d}}{\partial A_b} \right] \partial_b \psi \partial_d \phi = \partial_a \chi^{(n)} \tag{2.44}$$

and hence, taking the curl of this equation, we obtain

$$\left[\partial_{[c}\frac{\partial\beta_{a]}^{(n+1)b}}{\partial A_d}-\partial_{[c}\frac{\partial\beta_{a]}^{(n+1)d}}{\partial A_b}\right]\partial_b\psi\partial_d\phi+\left[\frac{\partial\beta_{[a}^{(n+1)b}}{\partial A_d}-\frac{\partial\beta_{[a}^{(n+1)d}}{\partial A_b}\right]\partial_{c]}\partial_b\psi\partial_d\phi+\left[\frac{\partial\beta_{[a}^{(n+1)b}}{\partial A_d}-\frac{\partial\beta_{[a}{}^{d}}{\partial A_b}\right]\partial_{c]}\partial_d\phi\partial_b\psi=0\ . \tag{2.45}$$

By choosing special cases for ϕ and ψ (e.g., $\phi=x^\mu$, $\psi=x^\nu$ and $\phi=x^\mu x^\sigma,\psi=x^\nu$, where x^μ denotes a Cartesian coordinate), it is not difficult to see that Eq. (2.45) can hold for all ϕ and ψ only if

$$\frac{\partial\beta_a^{(n+1)b}}{\partial A_d}-\frac{\partial\beta_a^{(n+1)d}}{\partial A_b}=0\ . \tag{2.46}$$

But, given that $\beta_a^{(j)b}=0$ for $0<j\leq n$, this is precisely the $(n+1)$th-order part of the condition (2.27) that ensures that we can eliminate $\beta_a^{(n+1)b}$ by a redefinition of field variables. Thus, we may set $\beta_a^{(n+1)b}=0$, which completes our inductive proof.

Thus, we have proven that of all the possible infinitesimal gauge symmetries (2.18), the only one which corresponds to a finite gauge symmetry is equivalent (after a possible change of field variables) to the usual infinitesimal gauge invariance under $\delta A_a=\partial_a\chi$. Since this infinitesimal motion on $\mathcal{M}$ is constant, i.e., independent of A_a, the corresponding finite gauge symmetry is simply $A_a\to A_a+\partial_a\chi$. Thus, within the framework and assumptions discussed above, the only way to obtain a consistent nonlinear generalization of Maxwell's equations from an action principle is to add terms to Lagrangian which are gauge invariant in the usual sense of Maxwell theory. This can be achieved by adding arbitrary terms to $\mathcal{L}^{(2)}$ such that A_a appears only in the form F_{ab}, although since only the integral of $\mathcal{L}$ need be gauge invariant, it is not necessary that $\mathcal{L}$ be locally constructed out of F_{ab}. This completes our discussion of a single spin-one field.

However, an interesting change in this situation occurs if we start with a collection of k spin-one fields $A_a^1,\ldots,A_a^k$ or—to phrase it somewhat better—a spin-one field $A_a{}^\mu$ which takes values in a k-dimensional "internal vector space." We shall use greek letters for the internal abstract indices and will follow the convention of placing the internal indices to the right of the spacetime indices. The divergence identity required by our assumptions on $\mathcal{F}^a{}_\mu\equiv\delta S/\delta A_a{}^\mu$ now takes the form

$$\partial_b(\beta_a{}^b{}_\mu{}^\nu\mathcal{F}^a{}_\nu)=\beta_a{}^b{}_\lambda{}^\nu\alpha_{b\mu}{}^\lambda\mathcal{F}^a{}_\nu \tag{2.47}$$

[as compared with Eq. (2.15)] which implies the infinitesimal gauge symmetry of S under

$$\delta A_a{}^\mu=\beta_a{}^b{}_\nu{}^\mu(\partial_b\chi^\nu+\alpha_{b\lambda}{}^\nu\chi^\lambda) \tag{2.48}$$

[as compared with Eq. (2.18)]. The fundamental integrability condition analogous to (2.22) now takes the form

$$\left[\frac{\partial\beta_a{}^b{}_\mu{}^\nu}{\partial A_c{}^\sigma}\beta_c{}^d{}_\lambda{}^\sigma-\frac{\partial\beta_a{}^d{}_\lambda{}^\nu}{\partial A_c{}^\sigma}\beta_c{}^b{}_\mu{}^\sigma\right](\partial_d\phi^\lambda+\alpha_{d\rho}^\lambda\phi^\rho)(\partial_b\psi^\mu+\alpha_{b\rho}^\mu\psi^\rho)+\beta_a{}^b{}_\sigma{}^\nu(\delta_\phi\alpha_{b\mu}{}^\sigma\psi^\mu-\delta_\psi\alpha_{b\mu}{}^\sigma\phi^\mu)=\beta_a{}^b{}_\mu{}^\nu(\partial_b\chi^\mu+\alpha_{b\rho}^\mu\chi^\rho)\ . \tag{2.49}$$

The possible forms for the linear parts of $\alpha_{a\mu}{}^\nu$ and $\beta_a{}^b{}_\mu{}^\nu$ are now

$$\alpha_{a\mu}^{(1)\nu}=c^\nu{}_{\mu\lambda}A_a{}^\lambda\ , \tag{2.50}$$

$$\beta_a^{(1)b}{}_\mu{}^\nu=0\ . \tag{2.51}$$

The curl of the zeroth-order part of Eq. (2.49) now yields

$$c^\nu{}_{\mu\lambda}\partial_{[b}\psi^{(\mu}\partial_{a]}\phi^{\lambda)}=0\ . \tag{2.52}$$

In contrast with Eq. (2.34), Eq. (2.52) has nontrivial solutions (for $k>1$), namely, any $c^\nu{}_{\mu\lambda}$ which is antisymmetric in its lower indices:

$$c^\nu{}_{\mu\lambda}=-c^\nu{}_{\lambda\mu}\ . \tag{2.53}$$

The zeroth-order part of (2.49) then also implies that

$$\chi^{(0)\nu}=c^\nu{}_{\mu\lambda}\psi^\mu\phi^\lambda+b^\nu\ , \tag{2.54}$$

where b^ν is a constant.

We now substitute our solutions for $\alpha_{a\mu}^{(1)\nu}$ and $\chi^{(0)\nu}$ into the first-order part of Eq. (2.49), take the curl of this equation, and also set both ϕ^μ and ψ^μ equal to constants. By taking the curl, we eliminate the unknown function $\chi^{(1)\nu}$ and by making ϕ^μ and ψ^μ constant, all the terms involving the unknown quantities $\beta_a^{(2)b}{}_\mu{}^\nu$ and $\alpha_{a\mu}^{(2)\nu}$ drop out. We obtain

$$2c^\sigma{}_{\mu\lambda}c^\lambda{}_{\nu\rho}\phi^{[\nu}\psi^{\mu]}\partial_{[a}A_{b]}^\rho=c^\sigma{}_{\lambda\rho}(c^\lambda{}_{\mu\nu}\psi^\mu\phi^\nu+b^\lambda)\partial_{[a}A_{b]}^\rho\ . \tag{2.55}$$

The requirement that this equality holds for all A_a and all constant ϕ^μ and ψ^μ implies that $b^\nu=0$ and

$$2c^\sigma{}_{[\mu|\lambda|}c^\lambda{}_{\nu]\rho}=c^\sigma{}_{\lambda\rho}c^\lambda{}_{\mu\nu}\ , \tag{2.56}$$

i.e.,

$$c^\lambda{}_{[\mu\nu}c^\sigma{}_{\rho]\lambda}=0\ . \tag{2.57}$$

With Eq. (2.57) satisfied (and $b^\nu=0$) all of the terms involving $\alpha_{a\mu}^{(1)\nu}$ and $\chi^{(0)\nu}$ in the first-order part of Eq. (2.49) cancel, so the second-order quantities $\alpha_{a\mu}^{(2)\nu}$ and $\beta_a^{(2)b}{}_\mu{}^\nu$ satisfy the same equation as though $\alpha_{a\mu}^{(1)\nu}=0$. A repetition of the argument for the case of a single spin-one field then shows that by means of the analog of the transformation (2.16) and change of variables, we can set $\alpha_{a\mu}^{(2)\nu}=0$, $\beta_a^{(2)b}{}_\mu{}^\nu=0$. Similarly, by induction it follows that $\alpha_{a\mu}^{(n)\nu}=0$, $\beta_a^{(n)b}{}_\mu{}^\nu=0$ for all $n>2$. Thus, our general solution for $\alpha_{a\mu}{}^\nu$ and $\beta_a{}^b{}_\mu{}^\nu$ is

$$\beta_a{}^b{}_\mu{}^\nu = \delta_a{}^b \delta_\mu{}^\nu , \tag{2.58}$$

$$\alpha_{a\mu}{}^\nu = c^\nu{}_{\mu\lambda} A_a{}^\lambda , \tag{2.59}$$

where $c^\nu{}_{\mu\lambda}$ is a tensor over the "internal vector space" satisfying Eqs. (2.53) and (2.57).

Equations (2.53) and (2.57) are precisely the defining relations for a Lie algebra. [Equation (2.57) is the Jacobi identity.] Thus, the allowed infinitesimal gauge invariance of the general form (2.48), is precisely the Yang-Mills infinitesimal gauge invariance

$$\delta A_a{}^\mu = D_a \chi^\mu , \tag{2.60}$$

where

$$D_a \chi^\mu \equiv \partial_a \chi^\mu + [A_a, \chi]^\mu , \tag{2.61}$$

where [,] denotes the Lie-algebra bracket defined by $c^\nu{}_{\mu\lambda}$. Thus, we have shown that within our framework and assumptions, the only way of achieving a consistent nonlinear generalization of the theory of a collection of spin-one fields is to make the action gauge invariant in the usual Yang-Mills sense with respect to some Lie algebra. It should be emphasized that our analysis was in no way prejudiced in the direction of group theory, fiber bundles, etc.; the Lie-algebra requirement arose entirely from our fundamental integrability condition (2.49).

III. SPIN-TWO FIELDS

In this section we shall parallel the analysis of the preceding section to obtain the requirements on consistent nonlinear generalizations of the theory of a single, massless spin-two field. Only the "vacuum" case is considered in this section, i.e., we do not treat here the case of additional matter fields which couple to the spin-two field. Some additional results for this latter case we will be discussed in the next section.

The field equation for a spin-two field, $\gamma_{ab} = \gamma_{(ab)}$, in Minkowski spacetimes (R^4, η_{ab}) is

$$\mathscr{E}^{(1)ab} = 0 , \tag{3.1}$$

where $\mathscr{E}^{(1)ab}$ denotes the linearized Einstein tensor:

$$\mathscr{E}^{(1)ab} = -\tfrac{1}{2}\partial^c\partial_c \bar{\gamma}_{ab} + \partial^c \partial_{(b} \bar{\gamma}_{a)c} - \tfrac{1}{2}\eta_{ab}\partial^c\partial^d \bar{\gamma}_{cd} , \tag{3.2}$$

where

$$\bar{\gamma}_{ab} = \gamma_{ab} - \tfrac{1}{2}\eta_{ab}\gamma \tag{3.3}$$

with $\gamma = \eta^{ab}\gamma_{ab}$. These equations arise from the action

$$S^{(2)} = \int \mathscr{L}^{(2)} d^4x \tag{3.4}$$

with

$$\mathscr{L}^{(2)} = (\sqrt{-g}\,R)^{(2)} , \tag{3.5}$$

where R denotes the scalar curvature of the metric $\eta_{ab} + \gamma_{ab}$, g denotes the determinant of this metric as computed using the volume element ϵ_{abcd} associated with η_{ab}, and the superscript (2) on the right-hand side of Eq. (3.5) denotes the quadratic part (in γ_{ab}) of that expression. Note that in this section as in the preceding section we use the volume element associated with η_{ab} in all integrals such as (3.4). In addition, all indices in this section will be lowered and raised using the flat metric η_{ab} and its inverse η^{ab}.

As in the Maxwell case we seek to add cubic and higher-order terms to $\mathscr{L}^{(2)}$ to obtain a nonlinear theory with the field equation

$$\mathscr{E}^{ab} \equiv \frac{\delta S}{\delta \gamma_{ab}} = 0 . \tag{3.6}$$

Again, however, a serious potential consistency problem arises from the fact that $\mathscr{E}^{(1)ab}$ satisfies the linearized Bianchi identity:

$$\partial_a \mathscr{E}^{(1)ab} = 0 . \tag{3.7}$$

Hence, by exactly the same reasoning and assumptions which led to Eq. (2.15), we require $\mathscr{E}^{ab}$ to satisfy an identity of the form

$$\partial_a (B^{ab}{}_{cd} \mathscr{E}^{cd}) = B^{ef}{}_{cd} C^b{}_{ef} \mathscr{E}^{cd} , \tag{3.8}$$

where $B^{ab}{}_{cd} = B^{ab}{}_{(cd)}$ is constructed locally out of γ_{ab} and η_{ab} with $B^{ab}{}_{cd} = \delta^a{}_{(c}\delta^b{}_{d)}$ when $\gamma_{ab} = 0$, and $C^b{}_{ef}$ is constructed locally from γ_{ab}, η_{ab}, and, at most linearly, from $\partial_c \gamma_{ab}$, with $C^b{}_{ef} = 0$ when $\gamma_{ab} = 0$. In fact, since $C^b{}_{ef}$ has an odd number of indices, it cannot be constructed from γ_{ab} and η_{ab} alone, and thus it must contain $\partial_c \gamma_{ab}$ precisely once. The redefintion freedom corresponding to (2.16) is

$$B^{ab}{}_{cd} \to f^b{}_e B^{ae}{}_{cd} , \tag{3.9a}$$

$$C^b{}_{cd} \to f^b{}_g C^g{}_{ch} (f^{-1})^h{}_d + (f^{-1})^e{}_d \partial_c f^b{}_e , \tag{3.9b}$$

where $f^a{}_b$ is constructed locally from γ_{ab} and η_{ab}, with $f^a{}_b = \delta^a{}_b$ when $\gamma_{ab} = 0$.

The infinitesimal gauge invariance associated with (3.8) is

$$\delta \gamma_{cd} = B^{ab}{}_{cd} (\partial_a \psi_b + C^e{}_{ab} \psi_e) , \tag{3.10}$$

where ψ_e is an arbitrary one-form field.

Again, the key issue is whether this infinitesimal gauge invariance corresponds to an exact gauge invariance, i.e., whether the subspaces of the tangent space to the manifold of field configurations which are spanned by the vector fields (3.10) are integrable. Thus, for a spin-two field, the fundamental integrability condition becomes that for any two one-form fields ϕ_a and ψ_a, there must exist a one-form field χ_a such that

$$(\delta_\phi B^{ab}{}_{cd})(\partial_a \psi_b + C^e{}_{ab}\psi_e) - (\delta_\psi B^{ab}{}_{cd})(\partial_a \phi_b + C^e{}_{ab}\phi_e) + B^{ab}{}_{cd}[(\delta_\phi C^e{}_{ab})\psi_e - (\delta_\psi C^e{}_{ab})\phi_e] = B^{ab}{}_{cd}(\partial_a \chi_b + C^e{}_{ab}\chi_e) , \tag{3.11}$$

where now δ_ψ denotes the first-order variation resulting from (3.10).

Before proceeding to solve Eq. (3.11) in analogy with our solution of Eqs. (2.20) and (2.45), we note that there is one obvious solution to Eq. (3.11): namely,

$$B^{ab}{}_{cd} = \delta^a{}_{(c}\delta^b{}_{d)} , \tag{3.12}$$

$$C^c{}_{ab} = 0 , \tag{3.13}$$

$$\chi_a=0 . \quad (3.14)$$

Such theories have the infinitesimal gauge invariance

$$\delta\gamma_{ab}=\partial_{(a}\psi_{b)} . \quad (3.15)$$

Since this infinitesimal motion on the vector space of spin-two field configurations is constant (i.e., independent of γ_{ab}) the corresponding finite gauge transformations are simply

$$\gamma_{ab}\rightarrow\gamma_{ab}+\partial_{(a}\psi_{b)} , \quad (3.16)$$

which is the normal spin-two gauge invariance.

Can consistent nonlinear theories of a spin-two field of this type be derived from an action principle? Our general arguments say that it should be possible, and, indeed, it is quite easy to construct examples: we can make a normal gauge-invariant action by adding to $\mathcal{L}^{(2)}$ any scalar such that γ_{ab} appears only in the form of the linearized Riemann tensor:

$$R_{abc}^{(1)d}=\partial^d\partial_{[a}\gamma_{b]c}-\partial_c\partial_{[a}\gamma_{b]}{}^d . \quad (3.17)$$

Perhaps the simplest such term to add to the Lagrangian which gives nonlinear equations is

$$\mathcal{L}^{(2)}=(R^{(1)})^3=(\partial^a\partial^b\gamma_{ab}-\partial^a\partial_a\gamma)^3 . \quad (3.18)$$

Thus, the Lagrangian $\mathcal{L}=\mathcal{L}^{(2)}+\mathcal{L}^{(3)}$ where $\mathcal{L}^{(2)}$ is the usual spin-two Lagrangian (3.5), yields a nonlinear generalization of the spin-two equations which is completely consistent in the sense discussed in the preceding section. This theory, as well as all others with the gauge invariance (3.16), is *not* generally covariant, so it is demonstrably false that all consistent nonlinearly self-coupled spin-two theories must be generally covariant.

Returning to Eq. (3.11), we note that, in analogy with Eq. (2.21) we have

$$\delta_\phi B^{ab}{}_{cd}=\frac{\partial B^{ab}{}_{cd}}{\partial\gamma_{ef}}\delta\gamma_{ef}$$
$$=\frac{\partial B^{ab}{}_{cd}}{\partial\gamma_{ef}}B^{gh}{}_{ef}(\partial_g\phi_h+C^k{}_{gh}\phi_k) , \quad (3.19)$$

so Eq. (3.11) takes the form

$$\left[\frac{\partial B^{ab}{}_{cd}}{\partial\gamma_{ef}}B^{gh}{}_{ef}-\frac{\partial B^{gh}{}_{cd}}{\partial\gamma_{ef}}B^{ab}{}_{ef}\right](\partial_a\psi_b+C^k{}_{ab}\psi_k)(\partial_g\phi_h+C^l{}_{gh}\phi_l)+B^{ab}{}_{cd}[(\delta_\phi C^e{}_{ab})\psi_e-(\delta_\psi C^e{}_{ab})\phi^e]$$
$$=B^{ab}{}_{cd}(\partial_a\chi_b+C^e{}_{ab}\chi_e) . \quad (3.20)$$

Again, the condition that $B^{ab}{}_{cd}$ can be reduced to $\delta^a{}_{(c}\delta^b{}_{d)}$ by a change of variables $\gamma_{ab}\rightarrow\tilde{\gamma}_{ab}(\gamma_{cd},\eta_{ef})$ is the vanishing of the first term

$$\frac{\partial B^{ab}{}_{cd}}{\partial\gamma_{ef}}B^{gh}{}_{ef}-\frac{\partial B^{gh}{}_{cd}}{\partial\gamma_{ef}}B^{ab}{}_{ef}=0 . \quad (3.21)$$

We repeat, now, our previous procedure of expanding the unknown quantities in terms of powers of γ_{ab}:

$$C^c{}_{ab}=\sum_n C^{(n)c}{}_{ab} , \quad (3.22)$$

$$B^{ab}{}_{cd}=\sum_n B^{(n)ab}{}_{cd} , \quad (3.23)$$

$$\chi_a=\sum_n \chi_a^{(n)} . \quad (3.24)$$

The zeroth-order (in γ_{ab}) part of Eq. (3.20) yields

$$\left[\frac{\partial B^{(1)ab}{}_{cd}}{\partial\gamma_{gh}}-\frac{\partial B^{(1)gh}{}_{cd}}{\partial\gamma_{ab}}\right](\partial_a\psi_b)(\partial_g\phi_h)$$
$$+(\delta_\phi^{(0)}C^{(1)e}{}_{(cd)})\psi_e-(\delta_\psi^{(0)}C^{(1)e}{}_{(cd)})\phi_e=\partial_{(c}\chi_{d)}^{(0)} , \quad (3.25)$$

where $\delta_\psi^{(0)}$ denotes the linear change of a quantity caused by the variation

$$\delta_\psi^{(0)}\gamma_{ab}=\partial_{(a}\psi_{b)} . \quad (3.26)$$

Now, the most general tensor of type (1,2), symmetric in its lower indices, which can be constructed from η_{ab} and, linearly, from γ_{ab} (with no more than one derivative) is

$$C^{(1)e}{}_{(cd)}=a_1\partial^e\gamma_{cd}+a_2\partial_{(c}\gamma_{d)}{}^e+a_3\partial_{(c}\gamma\delta_{d)}{}^e$$
$$+a_4\partial_b\gamma_{(c}{}^b\delta_{d)}{}^e+a_5\partial^e\gamma\eta_{cd}$$
$$+a_6\partial_b\gamma^{eb}\eta_{cd} . \quad (3.27)$$

The most general tensor of type (2,2), symmetric in its lower indices, which can be constructed from η_{ab} and, linearly, from γ_{ab} (with no derivatives) is

$$B^{(1)ab}{}_{cd}=b_1\eta^{ab}\gamma_{cd}+b_2\gamma^{ab}\eta_{cd}+b_3\gamma\eta^{ab}\eta_{cd}$$
$$+b_4\gamma\delta^a{}_{(c}\delta^b{}_{d)}+b_5\delta^b{}_{(c}\gamma^a{}_{d)}+b_6\delta^a{}_{(c}\gamma^b{}_{d)} . \quad (3.28)$$

We use the redefinition freedom (3.9) with

$$f^b{}_e=\delta^b{}_e-a_3\gamma\delta^b{}_e-(b_6-b_5)\gamma^b{}_e \quad (3.29)$$

to set

$$a_3=0 , \quad (3.30)$$

$$b_5=b_6 . \quad (3.31)$$

Taking the variation of the quantities (3.27) and (3.28), we obtain

$$\delta_\phi^{(0)}C^{(1)e}{}_{(cd)}=(a_1+\tfrac{1}{2}a_2)\partial^e\partial_{(c}\phi_{d)}+\tfrac{1}{2}a_2\partial_c\partial_d\phi^e$$
$$+(a_3+\tfrac{1}{2}a_4)\delta^e{}_{(d}\partial_{c)}\partial^b\phi_b+\tfrac{1}{2}a_4\partial_b\partial^b\phi_{(c}\delta^e{}_{d)}+(a_5+\tfrac{1}{2}a_6)\partial^e\partial^b\phi_b\eta_{cd}+\tfrac{1}{2}a_6\partial_b\partial^b\phi^e\eta_{cd} \quad (3.32)$$

and

$$\frac{\partial B^{(1)ab}{}_{cd}}{\partial \gamma_{gh}} = b_1 \eta^{ab} \delta^g{}_{(c} \delta^h{}_{d)} + b_2 \eta_{cd} \eta^{a(g} \eta^{h)b} + b_3 \eta^{gh} \eta^{ab} \eta_{cd} + b_4 \eta^{gh} \delta^a{}_{(c} \delta^b{}_{d)} + 2b_5 \delta^{(a}{}_{(c} \eta^{b)(g} \delta^{h)}{}_{d)} \; . \tag{3.33}$$

We substitute these expressions back into Eq. (3.25), thereby obtaining an equation for the unknown coefficients analogous to (but noticeably more complicated than) Eq. (2.33) in the spin-one case:

$$\begin{aligned}(b_1 - b_4)(\partial^a \psi_a \partial_{(c} \phi_{d)} - \partial^a \phi_a \partial_{(c} \psi_{d)}) &+ (a_1 + \tfrac{1}{2} a_2)(\psi_e \partial^e \partial_{(c} \phi_{d)} - \phi_e \partial^e \partial_{(c} \psi_{d)}) \\ &+ \tfrac{1}{2} a_2 (\psi_e \partial_c \partial_d \phi^e - \phi_e \partial_c \partial_d \psi^e) + \tfrac{1}{2} a_4 [\psi_{(d} \partial_{c)} \partial^b \phi_b - \phi_{(d} \partial_{c)} \partial^b \psi_b + (\partial_b \partial^b \phi_{(c}) \psi_{d)} - (\partial_b \partial^b \psi_{(c}) \phi_{d)}] \\ &+ [(a_5 + \tfrac{1}{2} a_6)(\psi_e \partial^e \partial^b \phi_b - \phi_e \partial^e \partial^b \psi_b) + \tfrac{1}{2} a_6 (\psi_e \partial_b \partial^b \phi^e - \phi_e \partial_b \partial^b \psi^e)] \eta_{cd} = \partial_{(c} \chi^{(0)}{}_{d)} \; . \end{aligned} \tag{3.34}$$

The nonappearance of b_2, b_3, and b_5 in this equation reflects the fact that these terms satisfy the first-order part of Eq. (3.21) and hence can be eliminated by a redefinition of field variables $\gamma_{ab} \to \tilde{\gamma}_{ab}$. We assume this choice has been made so that

$$b_2 = b_3 = b_5 = 0 \; . \tag{3.35}$$

Similarly, the fact that b_1 and b_4 appear only in the combination $(b_1 - b_4)$ implies that, in addition, we can set any other combination of these coefficients, e.g., $(b_1 + b_4)$, to zero by redefinition of field variables:

$$b_1 + b_4 = 0 \; . \tag{3.36}$$

The analog here of taking the curl of Eq. (2.33) to obtain (2.34) is to perform the operation corresponding to calculating the linearized Riemann tensor: thus, we operate on Eq. (3.34) with $\partial_f \partial_g$ and then antisymmetrize over f and c and over g and d. The right-hand side then vanishes and (since $\chi^{(0)}_a$ was arbitrary) the full content of Eq. (3.34) is expressed by the vanishing of the left-hand side of this twice differentiated and antisymmetrized version of Eq. (3.34). Rather than write out all the terms occurring in this equation in general we consider several special cases.

First, if we take ψ_a to be a constant one-form field (i.e., a Minkowski translation) and let ϕ_a be arbitrary, the only surviving terms are

$$\begin{aligned}\tfrac{1}{4} a_4 (\psi_{[d} \partial_{g]} \partial_b \partial^b \partial_{[f} \phi_{c]} + \psi_{[c} \partial_{f]} \partial_b \partial^b \partial_{[g} \phi_{d]}) &\\ + (a_5 + \tfrac{1}{2} a_6) \psi_e \partial_{[f} (\eta_{c][d} \partial_{g]} \partial^e \partial^b \phi_b) &\\ + \tfrac{1}{2} a_6 \psi_e \partial_{[f} (\eta_{c][d} \partial_{g]} \partial_b \partial^b \phi^e) = 0 \; . \end{aligned} \tag{3.37}$$

However, this equation can hold for all ϕ_a only if

$$a_4 = a_5 = a_6 = 0 \; . \tag{3.38}$$

To see this explicitly, first choose ϕ_a such that $\phi_a \psi^a = 0$ and $\partial^b \phi_b = 0$, but let $\partial_{[a} \phi_{b]}$ and its derivatives be nonzero; then the last two terms vanish, but the first term does not, in general, unless $a_4 = 0$. Having established that $a_4 = 0$, we let $\psi_a \phi^a = 0$ but let $\partial^b \phi_b$ and its derivatives be nonzero; then the last term vanishes but the remaining term does not, in general, unless $a_5 + \frac{1}{2} a_6 = 0$. Finally, from the vanishing of the first two terms it follows immediately from Eq. (3.36) that $a_6 = 0$.

Next, we choose

$$\partial_a \psi_b = \eta_{ab} \tag{3.39}$$

(i.e., ψ_a is a dilation) and let ϕ_a be a gradient, $\phi_a = \partial_a \rho$. We obtain

$$-\tfrac{1}{2}(b_1 - b_4)(\eta_{c[d} \partial_{g]} \partial_f \partial^a \partial_a \rho - \eta_{f[d} \partial_{g]} \partial_c \partial^a \partial_a \rho) = 0 \tag{3.40}$$

which implies

$$b_1 - b_4 = 0 \; . \tag{3.41}$$

Finally, we consider the case where ψ_a is a general Killing field, $\partial_a \psi_b = L_{ab} = L_{[ab]}$ with L_{ab} constant, and now let ϕ_a be arbitrary. We obtain

$$-\tfrac{1}{2}(a_1 + \tfrac{1}{2} a_2)(L_{e[f} \partial_{c]} \partial^e \partial_{[g} \phi_{d]} + L_{e[g} \partial_{d]} \partial^e \partial_{[f} \phi_{c]}) = 0 \tag{3.42}$$

which implies

$$a_1 + \tfrac{1}{2} a_2 = 0 \; . \tag{3.43}$$

The only term still serving on the left-hand side of Eq. (3.34) can manifestly be expressed as a symmetrized derivative,

$$\tfrac{1}{2} a_2 (\psi_e \partial_c \partial_d \phi^e - \phi_e \partial_c \partial_d \psi^e) = \tfrac{1}{2} a_2 \partial_{(c} (\psi^e \partial_{d)} \phi_e - \phi^e \partial_{d)} \psi_e) \tag{3.44}$$

and hence cannot be eliminated. Thus, after using all our freedom (3.9) as well as the freedom to change variables, $\gamma_{ab} \to \tilde{\gamma}_{ab}$, the general solution to Eq. (3.25) can be expressed as

$$C^{(1)e}{}_{(cd)} = a_1 (\partial^e \gamma_{cd} - 2 \partial_{(c} \gamma_{d)}{}^e) \; , \tag{3.45}$$

$$B^{(1)ab}{}_{cd} = 0 \; . \tag{3.46}$$

Note that the expression for $C^{(1)e}{}_{(cd)}$ is just minus the standard formula for the linearized Christoffel tensor for the metric:

$$g_{ab} = \eta_{ab} + 2 a_1 \gamma_{ab} \; . \tag{3.47}$$

The antisymmetric part $C^{(1)e}{}_{[ab]}$ is unspecified, but since $C^e{}_{ab}$ enters Eq. (3.20) only contracted with $B^{ab}{}_{cd}$ and since $B^{(0)ab}{}_{cd} = B^{(0)(ab)}{}_{cd}$, the antisymmetric part of $C^e{}_{ab}$ will be irrelevant if all $B^{(n)ab}{}_{cd} = 0$ for $n > 0$, as we shall show.

Consider, first, the particular case $a_1 = 0$. We will now show that this uniquely corresponds to a theory with normal spin-two gauge invariance. We assume, as our inductive hypothesis, that by use of (3.9) and redefinition of field variables, we have $C^{(j)e}{}_{(ab)}=0$ for all $j\leq n$ and $B^{(j)ab}{}_{cd}=0$ for $0<j\leq n$. (Since $a_1=0$, this hypothesis holds for $n=1$.) The nth-order part of Eq. (3.20) then becomes

$$\left[\frac{\partial B^{(n+1)ab}{}_{cd}}{\partial\gamma_{gh}}-\frac{\partial B^{(n+1)gh}{}_{cd}}{\partial\gamma_{ab}}\right](\partial_a\psi_b)(\partial_g\phi_h)+(\delta_\phi^{(0)}C^{(n+1)e}{}_{(cd)})\psi_e-(\delta_\psi^{(0)}C^{(n+1)e}{}_{(cd)})\phi_e=\partial_{(c}\chi_{d)}^{(n)} . \quad (3.48)$$

In analogy with the derivation of Eq. (2.40), we let ψ_a be a constant one-form field. Then $\partial_a\psi_b=0$ and the $\delta_\psi^{(0)}$ variation also vanishes. We then operate on Eq. (3.48) with $\partial_f\partial_g$ and antisymmetric over f and c and over g and d. We obtain

$$\delta_\phi^{(0)}(\partial_f\partial_g C^{(n+1)e}{}_{(cd)}-\partial_c\partial_g C^{(n+1)e}{}_{(fd)}-\partial_f\partial_d C^{(n+1)e}{}_{(cg)}+\partial_c\partial_d C^{(n+1)e}{}_{(fg)})=0 . \quad (3.49)$$

But this equation states that the quantity in parentheses is gauge invariant in the usual spin-two sense (3.15). In order for this to be true, γ_{ab} can appear in this expression only in the form of $R_{abc}^{(1)d}$ in Eq. (3.17). However, each $R_{abc}^{(1)d}$ has two derivatives of γ_{ab}, whereas the term in parentheses has $(n+1)$ γ_{ab}'s and only three derivatives. Thus, for $n>1$ it cannot be gauge invariant in the sense of Eq. (3.15) unless it vanishes. However, the vanishing of the term in parentheses is equivalent to the statement that $C^{(n+1)e}{}_{(cd)}$ is a symmetrized derivative:

$$C^{(n+1)e}{}_{(cd)}=\partial_{(c}f^{(n+1)e}{}_{d)} . \quad (3.50)$$

But this is precisely the form of $C^{(n+1)e}{}_{(cd)}$ which can be eliminated by the redefinition freedom (3.9), with $f^e{}_b=\delta^e{}_b-f^{(n+1)e}{}_b$. Thus, we use this freedom to obtain

$$C^{(n+1)e}{}_{(cd)}=0 . \quad (3.51)$$

With the result established, it follows in analogy with (2.46) that

$$\frac{\partial B^{(n+1)ab}{}_{cd}}{\partial\gamma_{gh}}-\frac{\partial B^{(n+1)gh}{}_{cd}}{\partial\gamma_{ab}}=0 \quad (3.52)$$

which implies that we can set

$$B^{(n+1)ab}{}_{cd}=0 \quad (3.53)$$

by a redefinition of field variables. This completes our inductive proof that $C^c{}_{(ab)}=0$ and $B^{ab}{}_{cd}=\delta^a{}_{(c}\delta^b{}_{d)}$, so that we are in the case (3.15) discussed previously.

Consider, now, the case $a_1\neq 0$. We know of one candidate for the complete solution $B^{ab}{}_{cd}$ and $C^e{}_{(cd)}$, namely,

$$B^{ab}{}_{cd}=\delta^a{}_{(c}\delta^b{}_{d)} , \quad (3.54)$$

$$C^e{}_{(cd)}=-\Gamma^e{}_{cd} , \quad (3.55)$$

where $\Gamma^e{}_{cd}$ is the (full) Christoffel tensor constructed from the metric g_{ab}, Eq. (3.47), relative to η_{ab}. [This can be verified to be a solution of Eq. (3.20) by direct substitution.] It yields the infinitesimal gauge invariance

$$\delta\gamma_{ab}=\nabla_{(a}\psi_{b)} , \quad (3.56)$$

where ∇_a is the derivative operator associated with g_{ab}, and, as will be demonstrated explicitly below, it corresponds to a generally covariant theory.

We show, now, that by use of the allowed freedom (3.9) and redefinition of field variables, the solution (3.54) and (3.55) is, in fact, the only solution of Eq. (3.20) with the first-order part given by (3.45) and (3.46). To do so, we define $\Delta B^{(n)ab}{}_{cd}$ and $\Delta C^{(n)e}{}_{cd}$ to be the difference between the nth-order part of an arbitrary solution [with the first-order part given by (3.45) and (3.46)] and the nth-order part of our solution (3.54) and (3.55). Our inductive hypothesis now is that, by use of the available freedom, we can set $\Delta B^{(j)ab}{}_{cd}=0$ and $\Delta C^{(j)e}{}_{(cd)}=0$ for $j\leq n$. This hypothesis holds for $n=1$. Assuming our hypothesis, we take the difference between the nth-order parts of Eq. (3.20) for our arbitrary solution and for our known solutions (3.54) and (3.55). We find that $\Delta B^{(n+1)ab}{}_{cd}$ and $\Delta C^{(n+1)e}{}_{(cd)}$ satisfy Eq. (3.48). Hence a repetition of the argument given for the case $a_1=0$ establishes the desired result that these quantities can be made to vanish. Thus, the only types of infinitesimal gauge invariance that can be achieved are the normal spin-two gauge invariance (3.15) and the gauge invariance infinitesimally generated by (3.56).

We conclude this section with a demonstration that theories with the infinitesimal gauge invariance (3.56) are precisely the generally covariant theories. We define a theory of a field γ_{ab} in Minkowski spacetime $(\mathbb{R}^4,\eta_{ab})$ arising from an action $S[\gamma_{ab},\eta_{cd}]$ to be *generally covariant* if, by making a new choice of dynamical field variable $\gamma_{ab}\to g_{ab}$, if necessary, the action is independent of the choice of the flat background metric, i.e., for any two (complete) flat metrics η_{ab} and η'_{ab} we have

$$S[g_{ab},\eta_{cd}]=S[g_{ab},\eta'_{cd}] . \quad (3.57)$$

If Eq. (3.57) holds, then the dynamics of g_{ab} cannot depend upon η_{ab}. Furthermore if all other fields (including those describing the measuring apparatus) couple to g_{ab} in such a way that Eq. (3.57) continues to hold for the full action, then there will be no way to measure η_{ab}, i.e., the flat background metric will be physically irrelevant.

To see the connection between our infinitesimal gauge invariance (3.56) and general covariance, we replace γ_{ab} by g_{ab}, Eq. (3.47), as our dynamical variable. Let $\phi:\mathbb{R}^4\to\mathbb{R}^4$ be an arbitrary diffeomorphism and let ϕ^* denote the induced map on tensor fields. Since the action is constructed entirely from g_{ab} and η_{ab}, it cannot change when we apply a diffeomorphism to *both* g_{ab} and η_{ab}:

$$S[\phi^*g_{ab},\phi^*\eta_{cd}]=S[g_{ab},\eta_{cd}] . \quad (3.58)$$

In particular, this implies that for a one-parameter family of diffeomorphisms, ϕ_λ, we have

$$0=\frac{d}{d\lambda}S[\phi_\lambda^*g_{ab},\phi_\lambda^*\eta_{cd}] . \quad (3.59)$$

However, the right-hand side of Eq. (3.59) can be expressed in terms of functional derivatives as

$$\frac{d}{d\lambda}S[\phi_\lambda^*g_{ab},\phi_\lambda^*\eta_{cd}]=\int\left[\mathcal{L}_\xi g_{ab}\frac{\delta S}{\delta g_{ab}}+\mathcal{L}_\xi\eta_{cd}\frac{\delta S}{\delta\eta_{cd}}\right] , \quad (3.60)$$

where $\xi^a = \xi^a(\lambda)$ is the vector field generating the one-parameter family. (Note that since we have chosen to use the volume element associated with η_{ab} in our integrals, the functional derivative of S with respect to η_{ab} includes contributions from the variation of this volume element.) Since

$$\delta S/\delta g_{ab} = \frac{1}{2a_1}\delta S/\delta\gamma_{ab}$$

and

$$\mathscr{L}_\xi g_{ab} = 2\nabla_{(a}\xi_{b)} \tag{3.61}$$

the first term on the right-hand side of Eq. (3.60) vanishes for theories with the infinitesimal gauge invariance (3.56). Thus, we obtain

$$0 = \int \mathscr{L}_\xi \eta_{cd} \frac{\delta S}{\delta \eta_{cd}} = \frac{d}{d\lambda} S[g_{ab}, \phi_\lambda^* \eta_{cd}] \tag{3.62}$$

which implies that for any one-parameter family of diffeomorphisms starting at the identity, we have

$$S[g_{ab}, \phi_\lambda^* \eta_{cd}] = S[g_{ab}, \eta_{cd}] . \tag{3.63}$$

Now, any two complete, flat metrics η_{ab} and η'_{ab} on $\mathbb{R}^4$ can be related by a diffeomorphism: $\eta'_{ab} = \phi^* \eta_{ab}$. (Proof: choose global inertial coordinates for η'_{ab} and for η_{ab} and take ϕ to be the map which associates points with the same values of the coordinates in the two systems.) By appending an orientation reversing isometry to the diffeomorphism, if necessary, we can ensure that ϕ is orientation preserving. But any orientation preserving diffeomorphism can be smoothly deformed to the identity. [Proof: the one-parameter family ϕ_λ defined by $\phi_\lambda(x) = \phi(\lambda x)/\lambda$ deforms ϕ to a linear map in the limit $\lambda \to 0$, and any orientation preserving linear map can be deformed to the identity. However, it should be noted that we have been sloppy in not worrying about boundary conditions at infinity in our manifold of field configurations and when such boundary conditions are properly imposed and the allowed diffeomorphisms are correspondingly restricted, there is reason to believe that not all orientation preserving diffeomorphisms of $\mathbb{R}^4$ can be deformed to the identity.] Thus, any complete flat metric η'_{ab} can be written as $\eta'_{ab} = \phi_\lambda^* \eta_{ab}$ where ϕ_λ is a one-parameter family of diffeomorphisms starting at the identity. Thus, Eq. (3.63) becomes

$$S[g_{ab}, \eta'_{cd}] = S[g_{ab}, \eta_{cd}] , \tag{3.64}$$

i.e., we have proven general covariance. Conversely, by reversing the steps of this argument, it is easily seen that any generally covariant theory has the infinitesimal gauge invariance (3.56).

IV. COUPLING TO MATTER

In the preceding section we considered the simple case where the spin-two field γ_{ab} was coupled only to itself, i.e., the action depended only on γ_{ab} and η_{ab}. In this section, we shall make some additional remarks concerning the case where γ_{ab} is coupled to another field (or fields) which we shall denote as ϕ (although it is not assumed that ϕ is a scalar field or a single field). We consider an action of the form

$$S_\lambda[\gamma_{ab}, \eta_{cd}, \phi] = S_G[\gamma_{ab}, \eta_{cd}] - \lambda S_M[\gamma_{ab}, \eta_{cd}, \phi] , \tag{4.1}$$

where λ is a parameter. In cases usually considered, λ is the gravitational constant, and Eq. (4.1) gives the breakup of the total action into a "gravitational part" and a "matter part."

In the preceding section, we argued that $\mathscr{E}^{ab} = \delta S/\delta\gamma_{ab}$ must satisfy the identity (3.8). The only modification caused by the presence of an additional field ϕ is that the analog of the identity (3.8) for

$$\mathscr{E}_\lambda^{ab} \equiv \frac{\delta S_\lambda}{\delta \gamma_{ab}} \tag{4.2}$$

may now contain additional terms proportional to the matter equations of motion:

$$e_M \equiv \frac{\delta S}{\delta \phi} \tag{4.3}$$

and its derivatives. However, if we impose the matter equations of motion, $e_M = 0$ (but *not* the equations of motion for γ_{ab}), we again conclude that $\mathscr{E}_\lambda^{ab}$ must satisfy (3.8). But, if (3.8) holds for all λ (i.e., the theory is "consistent" for all values of the gravitational constant) then (3.8) must hold separately (with the same values of $B^{ab}{}_{cd}$ and $C^e{}_{cd}$) for

$$\mathscr{E}_G^{ab} \equiv \frac{\delta S_G}{\delta \gamma_{ab}} \tag{4.4}$$

and

$$\mathscr{E}_M^{ab} \equiv \frac{\delta S_M}{\delta \gamma_{ab}} . \tag{4.5}$$

As before, the integrability condition resulting from this identity establishes that the only possible types of theories are those with normal spin-two gauge invariance and those satisfying general covariance.

First, we remark that it certainly is possible to have consistent theories (in our sense of the term) with γ_{ab} coupled to matter which are not generally covariant. Perhaps the simplest type of example of such a theory is to take S_M of the form

$$S_M = S_M^{(0)} + \int \gamma_{ab} V^{ab} , \tag{4.6}$$

where $S_M^{(0)}$ is the zeroth order in γ_{ab} matter action and V^{ab} is any tensor field constructed from ϕ and η_{ab} which is identically conserved. In particular, such V^{ab}'s can be obtained by letting V^{abcd} be any tensor field locally constructed from η_{ab}, ϕ, and derivatives of ϕ, satisfying $V^{abcd} = V^{[ab][cd]} = V^{cdab}$, and setting

$$V^{ab} = \partial_c \partial_d V^{acbd} . \tag{4.7}$$

[A tensor field of the form (4.7) with a locally constructed V^{abcd} is referred to as a "Pauli term" by Weinberg.[2] In fact, any conserved V^{ab} can be written in the form (4.7) (see problem 5 of Chap. 4 of Ref. 5) but, in general, V^{abcd}

need not be local. I do not know if all conserved V^{ab} which are locally constructed out of η_{ab}, ϕ, and derivatives of ϕ must arise via (4.7) from a similarly locally constructed V^{abcd}.] Then S_M will satisfy the normal spin-two gauge invariance rather than general covariance, and when added to an S_G of this gauge type (such as discussed in the preceding section) will yield a consistent, nongenerally covariant theory. Perhaps the simplest explicit example in the case of a single scalar field ϕ is to choose $V^{abcd}=2\phi\eta^{a[c}\eta^{d]b}$, in which case Eq. (4.6) becomes

$$S_M=S_M^{(0)}-\int \gamma_{ab}(\partial^a\partial^b\phi-\eta^{ab}\partial^c\partial_c\phi)\ . \quad (4.8)$$

Note that for a general "Pauli term," Eq. (4.7), the coupling (4.6) can be reexpressed by integration by parts as $\int R^{(1)}_{abcd}V^{abcd}$ (where $R^{(1)}_{abcd}$, again, denotes the linearized Riemann tensor), thus making more manifest the normal spin-two gauge invariance of S_M. Numerous additional examples can be generated by nonlinear couplings of $R^{(1)}_{abcd}$ to the matter fields.

However, if additional assumptions are made concerning S_M, it may be possible to draw stronger conclusions. As an example, consider the argument for general covariance originally given by Feynman.[1] Feynman restricted consideration to particle matter, with zeroth-order action of the form

$$S_M^{(0)}=\int \eta_{ab}u^a u^b d\tau\ , \quad (4.9)$$

where u^a is the particle four-velocity and τ is the proper time (and we have made minor changes in notation and conventions). Feynman took the first-order coupling to be of the form

$$S_M^{(1)}=\int \gamma_{ab}u^a u^b d\tau\ . \quad (4.10)$$

He then assumed (implicity) that the *complete* matter Lagrangian is the sum of (4.9) and (4.10):

$$S_M=S_M^{(0)}+S_M^{(1)}\ ; \quad (4.11)$$

i.e., no quadratic or higher-order couplings in γ_{ab} to the matter are permitted. His analysis concluded that the only possibility for S_G which involves no derivatives of γ_{ab} higher than second is the Einstein Lagrangian for the metric $g_{ab}=\eta_{ab}+\gamma_{ab}$.

From our viewpoint, the Feynman result can be obtained as follows. The action (4.11) is manifestly generally covariant with respect to $g_{ab}=\eta_{ab}+\gamma_{ab}$. Consequently, the gauge invariance of S_M is of this type, and hence, as remarked above, S_G must also be generally covariant with respect to g_{ab}. Uniqueness of Einstein's equation then follows from the fact[6] that, aside from g^{ab}, the Einstein tensor G^{ab} is the only tensor constructable solely out of g_{ab} (i.e., which is "generally covariant") which involves no derivatives higher than second and satisfies $\nabla_a G^{ab}=0$ identically.

It should be noted, however, that the assumption that (4.11) describes the full matter Lagrangian is rather strong. Indeed, the possibility of obtaining a consistent theory with matter coupled only linearly to γ_{ab} is very special to the particle matter considered by Feynman. It is the (assumed) general covariance of S_M which leads directly to the general covariance of the theory. Indeed, one obtains the much stronger result that the theory is generally covariant with respect to $g_{ab}=\eta_{ab}+\gamma_{ab}$; the possibility of even a theory equivalent to this under a redefinition of field variables (e.g., a theory generally covariant with respect to $g_{ab}=\eta_{ab}+\gamma_{ab}+\eta^{cd}\gamma_{ac}\gamma_{db}$) is excluded by Feynman's assumptions.

However, at least in many cases, a much weaker and quite natural assumption concerning the coupling of γ_{ab} to matter may suffice to eliminate the possibility of a theory with normal spin-two gauge invariance, thus yielding the conclusion that the theory, if consistent, must be generally covariant: we consider the requirement that, in lowest order, γ_{ab} couple to the stress energy tensor, i.e., that the linear order in γ_{ab} part of the matter action be

$$S_M^{(1)}=\int \gamma_{ab}T^{ab}\ , \quad (4.12)$$

where T^{ab} is the "stress-energy tensor of ϕ in Minkowski spacetime." We make no assumption about the higher-order parts of S_M. The reason that I put quotes around "stress-energy tensor of ϕ in Minkowski spacetime" is that, in fact, this term is rather ambiguous. There are two basic procedures for defining T^{ab}. The first—which defines the "canonical stress-energy tensor"—uses Noether's theorem to obtain a conserved tensor, $\partial_a T_C^{ab}=0$, for an arbitrary Lagrangian field theory in Minkowski spacetime. However, T_C^{ab} depends on the choice of Lagrangian $\mathscr{L}$ (i.e., it changes if one adds a total divergence to $\mathscr{L}$, even though this does not affect the action S). Furthermore, T_C^{ab} is not, in general, symmetric and, hence, is not appropriate as a source term in the spin-two equations. Thus, we shall not employ this definition. The second approach to defining T^{ab} is applicable if the field theory is defined in a general, curved spacetime, with metric g_{ab}: one defines T^{ab} by functionally differentiating S_M with respect to g_{ab}. However, for a theory defined in flat spacetime, only the functional derivative of S_M in directions of variation toward other flat metrics are well defined, i.e., only

$$\int T^{ab}\mathscr{L}_\xi\eta_{ab}=2\int T^{ab}\partial_a\xi_b=-2\int(\partial_a T^{ab})\xi_b$$

is unambiguous. Thus, we obtain a well-defined expression for $\partial_a T^{ab}$ (which will be proportional to the matter equations of motion), but other aspects of T^{ab} defined by this procedure may depend upon how the theory is generalized to a curved spacetime.

Nevertheless, we shall now illustrate that, at least for conventional choices of T^{ab}, Eq. (4.12) already precludes the possibility of normal spin-two gauge invariance. As a specific example, we consider a massless Klein-Gordon scalar field ϕ with

$$S_M^{(0)}=\int \partial^a\phi\partial_a\phi \quad (4.13)$$

and we take its stress energy tensor in flat spacetime to be

$$T_{ab}=\partial_a\phi\partial_b\phi-\tfrac{1}{2}\eta_{ab}\partial^c\phi\partial_c\phi\ . \quad (4.14)$$

Now, if we are to obtain a theory with normal spin-two gauge invariance, S_M must have vanishing variation under the infinitesimal change $\delta\gamma_{ab}=\partial_{(a}\xi_{b)}$, provided that the matter equations of motion are satisfied. Thus, to obtain a normal spin-two gauge invariant theory, we must have

$$\int \partial_{(a}\xi_{b)}\frac{\delta S_M}{\delta\gamma_{ab}} = -\int \xi_b\partial_a\left[\frac{\delta S_M}{\delta\gamma_{ab}}\right]$$

$$=(\text{terms which vanish when } \delta S_M/\delta\phi=0) \ . \tag{4.15}$$

Plausibly, the only way Eq. (4.15) can be satisfied is if

$$\partial_a\left[\frac{\delta S_M}{\delta\gamma_{ab}}\right]=u^b\frac{\delta S_M}{\delta\phi}+v^{bc}\partial_c\left[\frac{\delta S_M}{\delta\phi}\right]+w^{bcd}\partial_c\partial_d\left[\frac{\delta S_M}{\delta\phi}\right]+\cdots \tag{4.16}$$

is satisfied identically, where $u^b, v^{bc}, w^{bcd},\ldots$ are constructed out of η_{ab}, ϕ, γ_{ab}, and derivatives of ϕ and γ_{ab}.

Now, we assume that $S_M^{(1)}$ is given by (4.12), with T_{ab} given by (4.14). The left-hand side of the zeroth order in the γ_{ab} part of Eq. (4.16) is then

$$\partial_a\left[\frac{\delta S_M^{(1)}}{\delta\gamma_{ab}}\right]=\partial_a T^{ab}=\partial^b\phi\partial^a\partial_a\phi=-2\partial^b\phi\frac{\delta S_M^{(0)}}{\delta\phi} \ . \tag{4.17}$$

(Note that T^{ab} appears in this equation only in the form $\partial_a T^{ab}$ which, as remarked above, is unambiguously defined.) Comparing the right-hand sides of Eqs. (4.16) and (4.17), we find

$$u^{(0)b}=-2\partial^b\phi \tag{4.18}$$

while the zeroth-order parts of $v^{bc}, w^{bcd},\ldots$ vanish. The first-order part of Eq. (4.16) then becomes

$$\partial_a\left[\frac{\delta S_M^{(2)}}{\delta\gamma_{ab}}\right]=-2\partial^b\phi\frac{\delta S_M^{(1)}}{\delta\phi}+u^{(1)b}\frac{\delta S_M^{(0)}}{\delta\phi}+v^{(1)bc}\partial_c\frac{\delta S_M^{(0)}}{\delta\phi}+\cdots \ . \tag{4.19}$$

Using the explicit form of T_{ab}, Eq. (4.14), we find

$$\frac{\delta S_M^{(1)}}{\delta\phi}=-2\partial^c[(\gamma_{cd}-\tfrac{1}{2}\eta_{cd}\gamma)\partial^d\phi] \tag{4.20}$$

and, hence, we obtain

$$\partial_a\left[\frac{\delta S_M^{(2)}}{\delta\gamma_{ab}}\right]=4\partial^b\phi\partial^c[(\gamma_{cd}-\tfrac{1}{2}\eta_{cd}\gamma)\partial^d\phi]-2u^{(1)b}\partial^c\partial_c\phi-2v^{(1)bc}\partial_c\partial^d\partial_d\phi+\cdots \ . \tag{4.21}$$

We argue, now, that Eq. (4.21) cannot be satisfied for the simple reason that the left-hand side is the divergence of a quantity that is locally constructed out of the fields, whereas the right-hand side cannot be expressed in that form. To verify this latter claim, we note that we can rewrite the first term as

$$\partial^b\phi\partial^c[(\gamma_{cd}-\tfrac{1}{2}\eta_{cd}\gamma)\partial^d\phi]=\partial^c[\partial^b\phi(\gamma_{cd}-\tfrac{1}{2}\eta_{cd}\gamma)\partial^d\phi]-(\partial^c\partial^b\phi)(\gamma_{cd}-\tfrac{1}{2}\eta_{cd}\gamma)\partial^d\phi \ . \tag{4.22}$$

The first term on the right-hand side of (4.22) is a divergence, but the second term cannot be since no derivatives of γ_{ab} appear, and the expansion of the divergence of any quantity involving γ_{ab} must contain at least one term with a derivative acting on γ_{ab}. Hence, the only possibility for making the right-hand side of Eq. (4.21) a divergence is to cancel the "bad" term in (4.22) with the remaining terms in (4.21). However, this is impossible, since (after adding further divergences, if necessary, to remove all derivatives from γ_{ab} in these terms) these remaining terms in (4.21) all contain the combination $\partial^c\partial_c\phi$, which does not occur in the bad term in (4.22). Thus, we conclude that Eq. (4.21), and, hence Eq. (4.16), cannot be satisfied, so no theory with normal spin-two gauge invariance can be constructed in which, in linear order, γ_{ab} couples directly to the stress-energy tensor (4.14). It seems clear that this conclusion is far more general than the particular example considered here, but I do not see how to give a general proof.

Finally, it is worth noting that in order to determine the gauge type of a consistent theory within the framework considered here, it suffices merely to know S_G to third order in γ_{ab} or (assuming $S_M^{(1)}\neq 0$) to know S_M to second order in γ_{ab}. The reason why this is so is that we can thereby calculate $\mathscr{E}_G^{ab}$ to second order or $\mathscr{E}_M^{ab}$ to first order. In either case, by substitution into Eq. (3.8), we can obtain $B^{(1)ab}{}_{cd}$ and $C^{(1)e}{}_{cd}$. If these quantities vanish or can be made to vanish by the allowed transformations discussed in the preceding section, then the theory must have normal spin-two gauge invariance. If not [in which case we can set $B^{(1)ab}{}_{cd}=0$ and have $C^{(1)e}{}_{cd}$ be given by Eq. (3.45)], then the theory must be generally covariant.

ACKNOWLEDGMENTS

I wish to thank Bill Unruh, Wayne Boucher, and, particularly, Bob Geroch for helpful discussions and comments. Special thanks are due to Curt Cutler for pointing out errors in the original manuscript. This research was supported, in part, by National Science Foundation Grant No. PHY84-16691 to the University of Chicago.

[1]R. P. Feynman, Lectures on Gravitation, California Institute of Technology Lecture Notes, 1962 (unpublished).
[2]S. Weinberg, Phys. Rev. **138**, 988 (1965).
[3]S. Deser, Gen. Relativ. Gravit. **1**, 9 (1970).
[4]D. G. Boulware and S. Deser, Ann. Phys. (N.Y.) **89**, 193 (1975).
[5]R. M. Wald, *General Relativity* (University of Chicago Press, Chicago, 1984).
[6]D. Lovelock, J. Math. Phys. **13**, 874 (1972).