Is perturbation theory misleading in general relativity?

Robert Geroch and Lee Lindblom
Enrico Fermi Institute, University of Chicago, Chicago, Illinois 60637

(Received 10 April 1985; accepted for publication 7 June 1985)

Two senses in which the perturbation equations of general relativity can be misleading are explored. (i) Under certain circumstances there exist solutions of the perturbation equations that appear to be gauge, in that the metric perturbation is the symmetrized derivative of a vector field, but which nonetheless are not true gauge. (ii) Under certain circumstances there exist solutions of the perturbation equations that cannot, even locally, be extended to higher order in perturbation theory. The latter is a local version of the well-known phenomenon of "linearization instability."

I. INTRODUCTION

Einstein's equation, a system of nonlinear second-order partial differential equations, is sufficiently complex that, for most situations of physical interest, there are available no corresponding exact solutions. Consequently, much of our insight into the physical implications of general relativity has come from the study of approximate solutions. The most common approximation method is perturbation theory: One introduces a background space-time—an exact solution of Einstein's equation—and then considers deviations, to first or higher orders, from this background.

To what extent do such approximate solutions correspond to exact solutions? The question has both a quantitative and a qualitative aspect. The quantitative question asks for some numerical measure of the extent to which an approximate solution corresponds to some exact solution. Consider, for example, use of the quadrupole formula to compute the amount of gravitational radiation emitted by a system. By how much does the result of this computation differ from the correct answer—that obtained from the full Einstein equation?[1] The qualitative question, on the other hand, asks whether the predictions of the approximate solutions agree, even in their broad, overall features, with those of exact solutions.

It was first noted by Brill[2] that, under certain circumstances, there is not even qualitative agreement between the linearized and full solutions of Einstein's vacuum equation. These circumstances are that the background space-time possess both a compact Cauchy surface and a Killing field.[3–6] One first writes down a certain integral over the Cauchy surface, where the integrand involves the Killing field and an arbitrary solution of the first-order perturbation equation. One then shows (i) that, by virture of the second-order perturbation equation, this integral must vanish, and (ii) that there exist solutions of the first-order perturbation equation for which the integral is nonzero. Then these first-order perturbations, since they cannot even be extended to second order, certainly cannot come from any family of solutions of the full Einstein equation.

We consider here a somewhat different class of circumstances under which perturbation theory is qualitatively incorrect. As in the result above, we require of the background space-time that it possess symmetries. But, in contrast to that result, (i) the perturbation fields are required to respect the symmetries, rather than being allowed to break those symmetries, and (ii) the arguments are purely local, rather than global.

An example will illustrate what we have in mind. Consider the external gravitational field of a static, plane-symmetric "sheet" of matter. We expect that the space-time appropriate to this situation will have three orthogonal, commuting Killing fields: t^a (timelike, giving the static character), and x^a and y^a (spacelike, giving the plane symmetry).

Consider first the limit in which the stress energy of the sheet is small. It should then be appropriate to treat the gravitational field as a linear perturbation off Minkowski space-time. Denote by g_{ab} the Minkowski metric, by ∇_a its associated derivative operator, and by t^a, x^a, y^a three orthonormal translations in this space-time. The first-order perturbation of the metric, denoted h_{ab}, must satisfy the linearized Einstein equation

$$\nabla^m\nabla_m h_{ab} - 2\nabla^m\nabla_{(a}h_{b)m} + \nabla_a\nabla_b h^m{}_m = 0, \tag{1}$$

and must respect the symmetries

$$\mathscr{L}_t h_{ab} = \mathscr{L}_x h_{ab} = \mathscr{L}_y h_{ab} = 0. \tag{2}$$

We now claim the following: *the most general solution of Eqs.* (1) *and* (2) *is given by*

$$h_{ab} = 2\nabla_{(a}\tau_{b)}\ , \tag{3}$$

where τ^a is any vector field such that each of $\mathscr{L}_t\tau^a$, $\mathscr{L}_x\tau^a$, and $\mathscr{L}_y\tau^a$ is a Killing field. Clearly, any such h_{ab} does indeed satisfy (1) and (2). To prove the converse, let h_{ab} satisfy (1) and (2). Equation (2) is the statement that $z_{[a}\nabla_{b]}h_{cd} = 0$, where z^a is the unit translation in the background orthogonal to the other three. But this in turn implies

$$z_{[a}R_{bc]de} = 0, \tag{4}$$

where

$$R_{abcd} = -2\nabla_{[a}\nabla_{|[c}h_{d]|b]} \tag{5}$$

is the linearized Riemann tensor. Equation (1) is the statement that all traces of R_{abcd} vanish. But this implies, taking the double dual of (4), that $z^aR_{abcd} = 0$. Contracting (4) with z^a and using this last equation, we obtain that $R_{abcd} = 0$, i.e., that h_{ab} is of the form (3) for *some* τ^a. Finally, for h_{ab} of this form, Eq. (2) is precisely that statement that each of $\mathscr{L}_t\tau^a$, $\mathscr{L}_x\tau^a$, and $\mathscr{L}_y\tau^a$ is a Killing field, completing the proof.

We thus conclude that every linear perturbation appro-

 0022-2488/85/102581-08$02.50

priate to this problem is "pure gauge." This conclusion is not surprising, for one might expect a static, plane-symmetric sheet of matter to result in a "uniform gravitational field," i.e., in flat space-time.

For this particular problem, however, we can compare the linearized approximation with the full theory, for all vacuum solutions of the full Einstein equation having three commuting Killing fields are known.[7,8] The most general such solution can be represented as

$$ds^2 = -(\lambda z+1)^{2p_1}\,dt^2 + (\lambda z+1)^{2p_2}\,dx^2 + (\lambda z+1)^{2p_3}\,dy^2 + dz^2, \tag{6}$$

where p_1, p_2, p_3 are any three numbers satisfying $p_1+p_2+p_3 = p_1{}^2+p_2{}^2+p_3{}^2 = 1$, and λ is any number. [The parameter λ, introduced here for later convenience, has the effect of merely rescaling the coordinates. So, the family of essentially different solutions forms a circle (intersection of a plane and a sphere in the space of the p_i).] For λ nonzero, these solutions (except those having one of the p_i one, the others zero) are nonflat; while, for λ zero, all the solutions are flat. Thus, there are exact, nonflat solutions to the problem arbitrarily close to Minkowski space-time. Yet all linear perturbations off Minkowski space-time are pure gauge! It would seem that first-order perturbation theory provides too few solutions to reflect adequately the full theory.

There is of course a direct way to see what is going on in this particular example: Take the family of metrics in (6), and linearize it in λ about $\lambda = 0$. There results a linearized metric h_{ab} of the form (3), but with a particular choice of τ^a, namely

$$\tau^a = -p_1 ztt^a + p_2 zxx^a + p_3 zyy^a + \tfrac{1}{2}(p_1 t^2 - p_2 x^2 - p_3 y^2)z^a. \tag{7}$$

This h_{ab} does indeed satisfy Eqs. (1) and (2), for each of

$$\mathscr{L}_t \tau^a = -p_1(zt^a - tz^a),\quad \mathscr{L}_x \tau^a = p_2(zx^a - xz^a),\quad \mathscr{L}_y \tau^a = p_3(zy^a - yz^a) \tag{8}$$

is a Killing field, as required. Thus, the situation is the following. First, it is indeed true that all linearized solutions in this plane-symmetric problem are symmetrized derivatives of vectors, as in Eq. (3). But this is the "right answer," for the linearization about Minkowski space-time of our family of exact solutions also yields a metric perturbation the symmerized derivative of a vector field. It is only at higher order in perturbation theory that the nonflat character of these exact solutions becomes apparent. What is happening, then, is that a first-order perturbation of the form $h_{ab} = 2\nabla_{(a}\tau_{b)}$ is necessary in order to provide access, at higher order in perturbation theory, to the nonflat exact solutions. Such first-order perturbations, then, should not in this example be regarded as "gauge."

This same example illustrates a second potential difficulty with perturbation theory. We saw above that the first-order theory appears to provide too few solutions to reflect the full theory, an appearance that resulted from a too-broad application of the term gauge. The second difficulty is that the first-order theory *does*—not only in appearance, but in fact—provide too *many* solutions to reflect the full theory. To see this, consider the linearized solution h_{ab} given by Eq. (3) with τ^a given by Eq. (7)—but with the constants p_i now chosen so that $(p_1+p_2+p_3)^2$ is not equal to $p_1{}^2+p_2{}^2+p_3{}^3$. This h_{ab} does not indeed satisfy the first-order linear equations (1) and (2). But, we claim, this particular linearized solution comes from no family of exact solutions. This claim (which will be discussed in more detail later) can be seen directly at this point by taking all linearizations about flat space-time of families of solutions taken from Eq. (6) (possibly allowing the p_i to depend on the parameter λ; possibly applying λ-dependent, Killing-field-preserving diffeomorphisms). This second difficulty is analogous to the "linearization instability" discussed earlier by Brill[2] and others.[3–6] But there are several significant differences. The earlier work required a compact Cauchy surface (while here there is none), dealt with arbitrary first-order perturbations (while here they must respect the symmetries), and involved a global argument (while here local).

This example illustrates the issues with which we shall be concerned in this paper. Under what circumstances can "apparent gauge" arise? What is the "correct" notion of gauge? Under what circumstances does perturbation theory yield solutions extraneous to the full theory? How can these extraneous solutions be identified?

II. PERTURBATION THEORY

In this section, we set up the general framework, consisting of a few definitions and their basic properties, for perturbation theory.

Fix a manifold $\mathscr{M}$. Specify a list of the types of fields to be considered on $\mathscr{M}$, as well as a list of the equations to be satisfied by those fields. In our earlier example, $\mathscr{M}$ was the manifold R^4; the fields were a symmetric, Lorentz-signature metric g_{ab} and three vectors fields, t^a, x^a, and y^a; the equations were Einstein's equation, Killing's equation for each of t^a, x^a, and y^a, and the equations asserting that all Lie brackets of these vector fields vanish.

Next, let the fields under consideration be divided into two classes: the *passive* fields and the *dynamic* fields. The passive fields will be fixed throughout, while the dynamic fields will be subject to perturbation. In our example, the three vector fields t^a, x^a, and y^a were passive, while the metric g_{ab} was dynamic.

Finally, fix a solution of the system. That is, fix actual fields, satisfying all the given equations, on $\mathscr{M}$. This solution will be called the *background*. In our example, the background consisted of a flat metric, together with three unit, orthogonal translations, on $\mathscr{M}$.

This, then, is the arrangement we contemplate for perturbation theory: a manifold $\mathscr{M}$, a list of fields and the equations they are to satisfy, a division of the fields into passive and dynamic, and a background solution of the system.

Now consider a one-parameter family of solutions of this system of equations, jointly smooth on $\mathscr{M}$ and in the parameter λ. Let this family be such that all passive fields are independent of λ, and such that the fields in the family reduce, for $\lambda = 0$, to the background fields. Taking the first derivatives, with respect to λ, of the dynamic fields in this family, and evaluating at $\lambda = 0$, we obtain the first-order perturbed fields. Taking the first derivatives, with respect to λ, of the equations applied to our family of fields, and evaluating at $\lambda = 0$, we obtain the *first-order perturbation equations*. This is a system of linear equations on the first-order perturbed fields. The first-order perturbation equations, once derived in this way, are then regarded as equations in their own right; one is free to consider solutions of this system of equations without reference to a one-parameter family—or even to whether such a family exists—giving rise to that solution. In our earlier example, the only first-order perturbed field was the perturbed metric h_{ab}, and the first-order perturbation equations were precisely Eqs. (1) and (2). (Nothing new results from the λ derivatives of the commutation relations, since these involve only the passive fields.) More generally, taking all derivatives up to the nth, with respect to λ, of the equations, and evaluating to $\lambda = 0$, we obtain the nth-*order perturbation equations*. These equations involve the background fields and the first n derivatives, with respect to λ, of the dynamic fields. The nth derivatives of the dynamic fields always appear linearly, but the other derivatives in general do not. Any solution of the nth-order perturbation equations yields immediately a solution of the mth-order equations (for $m < n$), by simply omitting all λ derivatives of order higher than the mth.

There is available a particularly simple class of solutions of the nth-order perturbation equations. Consider a one-parameter family $\mathscr{D}(\lambda)$ of diffeomorphisms on $\mathscr{M}$, jointly smooth on $\mathscr{M}$ and in λ. Let this family be such that all passive background fields are invariant under all $\mathscr{D}(\lambda)$, and such that $\mathscr{D}(0)$ is the identity diffeomorphism on $\mathscr{M}$. Applying these $\mathscr{D}(\lambda)$ to each of the dynamic background fields, we obtain a one-parameter family of solutions of the system. This family satisfies, by virtue of the conditions just imposed on $\mathscr{D}(\lambda)$, the conditions of the previous paragraph. Hence, taking all derivatives up to the nth with respect to λ, of the dynamic fields in this family, and evaluating at $\lambda = 0$, we obtain a solution of the nth-order perturbation equations. Solutions of the form so obtained will be called gauge solutions, reflecting the fact that the result of applying $\mathscr{D}(\lambda)$ to the background has exactly the same physical content as the background itself. Thus, the general gauge solution of the first-order perturbation equations is determined by a vector field τ^a on $\mathscr{M}$ [reflecting the first derivative of $\mathscr{D}(\lambda)$ at $\lambda = 0$], with respect to which the Lie derivatives of the passive background fields vanish [reflecting the condition that the passive background fields be invariant under the $\mathscr{D}(\lambda)$]. The first-order perturbed fields are those obtained by applying $\mathscr{L}_\tau$ to each of the dynamic background fields. The general gauge solution of the second-order perturbation equations is determined by two vector fields τ^a and σ^a on $\mathscr{M}$ [reflecting the first two derivatives $\mathscr{D}(\lambda)$ at $\lambda = 0$], with respect to which both of the Lie derivatives of the passive background fields vanish. The perturbed fields are those obtained by applying each of $\mathscr{L}_\tau$ and $\mathscr{L}_\tau \mathscr{L}_\tau + \mathscr{L}_\sigma$ to each of the dynamic background fields.

Fix a manifold $\mathscr{M}$, lists of the types of passive and dynamic fields to be considered on $\mathscr{M}$, a list of the equations to be satisfied by these fields, and a background solution of the system. We have seen above that certain one-parameter families of solutions of this system lead to solutions of the perturbation equations, and that certain one-parameter families of diffeomorphisms on $\mathscr{M}$ lead to gauge solutions of the perturbation equations. We now consider the extent to which these processes can be reversed. Does a given solution of the perturbation equations arise from some one-parameter family of solutions of the original system? Does a given gauge solution arise from some one-parameter family of diffeomorphisms?

The second question is easy to answer. We claim the following: Given any gauge solution of the nth-order perturbation equations, there exists a one-parameter family $\mathscr{D}(\lambda)$ of diffeomorphisms giving rise to that solution. To prove this, we must show that, given n vector fields $\tau^a, \sigma^a, ..., \kappa^a$ on $\mathscr{M}$, there exists a family $\mathscr{D}(\lambda)$ of diffeomorphisms whose first n derivatives with respect to λ, at $\lambda = 0$, are characterized by these vector fields. But for $n = 1$, a suitable family[9] is given by $\mathscr{D}_\tau(\lambda)$, the diffeomorphisms generated by the vector field τ^a itself; for $n = 2$, by $\mathscr{D}_\tau(\lambda) \circ \mathscr{D}_\sigma(\lambda^2/2)$; and similarly for other n.

The question of whether a given solution of the perturbation equations arises from some one-parameter family of exact solutions of the original system is more difficult. The source of the difficulty is that one does not in general have easy access to the exact solutions of the system. Indeed, it is the lack of such access that causes one to turn to perturbation theory in the first place. Fortunately, there is a notion closely related to "come from a family of exact solutions," but far easier to work with. We say that the nth-order perturbation equations are *reliable* if every solution of those equations can be extended [by some choice of the $(n+1)$st-order perturbed fields] to a solution of the $(n+1)$st-order perturbation equations. The condition, then, is that the solutions can be extended to one higher order in perturbation theory. The advantage of this definition is that it deals only with the perturbation equations and their solutions, with no reference to exact solutions of the full system. Should it happen that every solution of the nth-order perturbation equations comes from some family of exact solutions, then the perturbation equations must certainly be reliable; extend any perturbed solution from nth to $(n+1)$st order using the family of exact solutions. But there is no guarantee that, conversely, reliability implies that all perturbed solutions must come from families of exact solutions.

III. GAUGE

We now return to the example of Sec. I: the external gravitational field of a static, plane-symmetric sheet of matter. We found in that example that the general solution of the first-order perturbation equations (1) and (2) is

$$h_{ab} = 2\nabla_{(a}\tau_{b)}\,, \tag{9}$$

where τ^a is any vector field such that each of $\mathscr{L}_t\tau^a$, $\mathscr{L}_x\tau^a$, and $\mathscr{L}_y\tau^a$ is Killing field. We initially interpreted these solutions, in light of Eq. (9), as gauge; they give, for example, vanishing curvature tensor to first order. But this interpretation was found to be unacceptable, for there are in this example exact, nonflat solutions of the system arbitrarily close to the Minkowski background. A one-parameter family of such exact solutions does give rise to a first-order perturbed metric of the form (9), with τ^a given by Eq. (7).

In Sec. II, we introduced a general framework for perturbation theory, applicable to virtually any system of equations on fields. Within that framework, gauge solutions of the perturbation equations were defined quite generally as arising from certain families of diffeomorphisms preserving the passive fields. Applying the general definition to this example, we obtain the following: The gauge solutions of the first-order perturbation equations (1) and (2) are those of the form (9), but with τ^a now a vector field such that each of $\mathscr{L}_t\tau^a$, $\mathscr{L}_x\tau^a$, and $\mathscr{L}_y\tau^a$ vanishes. Thus, we correctly exclude from being gauge the solution with τ^a given by Eq. (7)—the solution which gives access to the nonflat exact solutions. While this solution arises from *some* diffeomorphisms, it does not arise from those preserving the passive fields. In short, gauge must be defined in this more restrictive way in the presence of passive fields.

This phenomenon—the existence of nongauge perturbed metrics that are nonetheless symmetrized derivatives of vector fields—is not just a special feature of the static, plane symmetric case. It is rather, as the following example shows, pervasive for Einstein's equation in the presence of symmetries. Fix a manifold $\mathscr{M}$. Let the dymanic field be a Lorentz-signature metric g_{ab}, and the passive fields n vector fields $\underset{i}{\xi}{}^a$. Let the equations be Einstein's equation, Killing's equation for each of the $\underset{i}{\xi}{}^a$, and a set of commutation relations

$$\mathscr{L}_{\underset{i}{\xi}}\underset{j}{\xi}{}^a = C^k{}_{ij}\underset{k}{\xi}{}^a, \tag{10}$$

where the $C^i{}_{jk}$ are fixed constants satisfying the Jacobi relation $C^m{}_{[ij}C^l{}_{k]m} = 0$. Fix a background solution $g_{ab}, \underset{i}{\xi}{}^a$. To simplify the discussion, we suppose that the background g_{ab} admits no Killing fields other than linear combinations, with constants coefficients, of the $\underset{i}{\xi}{}^a$. The first-order perturbation equations in this case are those analogous to Eqs. (1) and (2). One class of solutions of these equations is that with h_{ab} given by Eq. (9), where τ^a is any vector field such that each of the $\mathscr{L}_{\underset{i}{\xi}}\tau^a$ is a Killing field

$$\mathscr{L}_{\underset{i}{\xi}}\tau^a = U_i{}^k\underset{k}{\xi}{}^a, \tag{11}$$

where $U_i{}^k$ are constants. Applying $\mathscr{L}_{\underset{j}{\xi}}$ to Eq. (11) and antisymmetrizing over i and j, we find that $U_i{}^k$ must satisfy the further condition

$$U_i{}^mC^k{}_{mj} + U_j{}^mC^k{}_{im} - U_m{}^kC^m{}_{ij} = 0. \tag{12}$$

These solutions are "apparent gauge." The true gauge solutions, on the other hand, are those of the form above, but with the additional property that the $U_i{}^k$ in Eq. (11) vanish.

Under what condition are all apparent gauge solutions in fact true gauge? Note that we may, without altering the perturbed metric, add to the τ^a in Eq. (9) any Killing field, i.e., any field of the form $w^i\underset{i}{\xi}{}^a$ with the w^i constant. The condition, then, is that by adding such a field to τ^a we may achieve vanishing of the $U_i{}^k$ in Eq. (11), i.e., that

$$U_i{}^j = w^mC^j{}_{mi}, \tag{13}$$

for some w^i. We conclude the following: *Every solution that is apparent gauge is actually true gauge provided every $U_i{}^k$ satisfying Eq.* (12) *is of the form* (13). Note that this condition involves only the Lie-algebra structure on the Killing fields. It may, in fact, be restated thus: Every infinitesimal automorphism on the Lie algebra is inner. Note also that, by the Jacobi relation, every $U_i{}^k$ of the form (13) automatically satisfies Eq. (12).

It is generally easy to decide whether a given Lie algebra satisfies the condition above, i.e., whether apparent gauge, in the presence of passive symmetries with that Lie algebra, must be true gauge. For the zero-dimensional Lie algebra (no passive fields), the condition is of course satisfied. For all one- or two-dimensional Lie algebras, the condition is not. For three-dimensional Lie algebras, it depends on the algebra. It is satisfied, for example, for the Lie algebra of SO(3) (the rotations), and for SO(2,1), but not for the three-dimensional commutative Lie algebra. We remark that, when the condition above fails, then there normally do exist solutions that are apparent gauge but not true gauge.

The class of solutions considered here is restricted in that we allow no sources in Einstein's equation and no passive fields other than Killing fields, and yet broad enough to include curved background metrics and arbitrary Lie algebras of symmetries. It appears that apparent gauge is pervasive for Einstein's equation in the presence of passive symmetries.

IV. RELIABILITY

We now return again to the static, plane-symmetric example of Sec. I. We found in that example that there exist solutions of the first-order perturbation equations—namely, those of the form (3), with τ^a given by Eq. (7) with the constant p_i so chosen that $(p_1+p_2+p_3)^2 \neq p_1{}^2+p_2{}^2+p_3{}^2$—that do not come from any one-parameter families of exact solutions of the system. We interpreted this phenomenon as indicating that the first-order perturbation equations do not adequately reflect the full equations. In the general framework for perturbation theory in Sec. II, we introduced, for virtually any system of equations on fields, a closely related notion: The nth-order perturbation equations are reliable if every solution of those equations can be extended to a solution of the equations at next-highest order in perturbation theory. We now ask how this specific example fits into the general framework.

The first-order perturbation equations for this example are (1) and (2). The second-order perturbation equations are

$$\nabla^m\nabla_m i_{ab} - 2\nabla^m\nabla_{(a}i_{b)m} + \nabla_a\nabla_b i^m{}_m = t_{ab} - \tfrac{1}{2}g_{ab}t^m{}_m \tag{14}$$

and

$$\mathscr{L}_t i_{ab} = \mathscr{L}_x i_{ab} = \mathscr{L}_y i_{ab} = 0, \tag{15}$$

where i_{ab} denotes the second-order metric perturbation, and t_{ab} is given by

$$t_{ab} - \tfrac{1}{2} g_{ab} t^m{}_m = -4h^{mn}\nabla_{[a}\nabla_{|[b} h_{m]|n]} - \tfrac{1}{2}(\nabla_a h^{mn})(\nabla_b h_{mn}) - 2(\nabla^m h^n{}_a)(\nabla_{[m} h_{n]b}) + (\nabla_m h^{mn} - \tfrac{1}{2}\nabla^n h^m{}_m)(2\nabla_{(a} h_{b)n} - \nabla_n h_{ab}). \tag{16}$$

We now claim the following: *Given h_{ab} satisfying Eqs.* (1) *and* (2), *there exists i_{ab} satisfying Eqs.* (14) *and* (15) *if and only if*

$$u^{mn}u^{pq}s_{mp}s_{nq} - (u^{mn}s_{mn})^2 = 0, \tag{17}$$

where z^a is the unit translation orthogonal to the other three, $u^{ab} = g^{ab} - z^a z^b$ is the projection orthogonal to z^a, and s_{ab} is given by

$$\nabla_a h_{bc} = z_a s_{bc} \tag{18}$$

[noting that the existence of such an s_{ab} is guaranteed by (2)]. To prove this claim, let h_{ab} satisfy (1) and (2). Set, by Eq. (15), $\nabla_a \nabla_b i_{cd} = z_a z_b \kappa_{cd}$, for some κ_{cd}, and substitute into Eq. (14) to obtain the following: There exists i_{ab} satisfying Eqs. (14) and (15) if and only if $z^m t_{am} = 0$. Next, solve Eq. (16) [noting that, by (1) and (2), the first term on the right vanishes] for t_{ab}, contract with z^b, and substitute Eq. (18). The claim follows.

In particular, for the h_{ab} by given by Eq. (3) with τ^a given by (7), the condition, Eq. (17), for the existence of a second-order perturbed metric becomes $(p_1 + p_2 + p_3)^2 = p_1{}^2 + p_2{}^2 + p_3{}^2$. This is precisely the same as the condition, obtained in Sec. I, that our h_{ab} come from some one-parameter family of exact solutions. We conclude, then, that the first-order perturbation equations in this example are not reliable. That perturbation theory gives the "wrong answer" in this example is correctly detected by the notion of reliability. This is a local version of the linearization instability noted earlier.[2,6]

When, in general relativity, can local perturbation theory be trusted, and when can it not? To answer this question fully appears to be difficult. But the following discussion does suggest that reliability in general relativity is more the rule than the exception.

Let the dynamic field be a Lorentz-signature metric g_{ab}, the passive fields m vector fields $\underset{i}{\xi}{}^a$. Let the equations be Einstein's equation with vanishing sources, Killing's equation for each of the $\underset{i}{\xi}{}^a$, and set of commutation relations on the $\underset{i}{\xi}{}^a$. Fix a background solution. Then the nth-order perturbation equations are

$$\nabla^m \nabla_m n_{ab} - 2\nabla^m \nabla_{(a} n_{b)m} + \nabla_a \nabla_b n^m{}_m = t_{ab} - \tfrac{1}{2} g_{ab} t^m{}_m \tag{19}$$

and

$$\mathscr{L}_{\underset{i}{\xi}}\, n_{ab} = 0, \tag{20}$$

where n_{ab} is the nth-order perturbed metric and $t_{ab} = t_{(ab)}$ is some expression involving the perturbed metrics up to order $(n-1)$. It follows from the $(n-1)$st-order perturbation equations that this t_{ab} respects the symmetries $\mathscr{L}_{\underset{i}{\xi}}\, t_{ab} = 0$ and is conserved, $\nabla_m t^{am} = 0$. Hence, *perturbation theory is reliable at every order provided that, for any t_{ab} that respects the symmetries and is conserved there exists an n_{ab} satisfying Eqs.* (19) *and* (20).

We conjecture that one can always solve (19) and (20) locally, and so perturbation theory is locally reliable, whenever the number of passive Killing fields is two or less. Evidence for this conjecture comes from the important special case in which the Killing fields are linearly independent at each point and all Killing fields are spacelike. Choose, in this case, a spacelike slice S to which all Killing fields are tangent. The idea is to solve Eq. (19) for n_{ab} using an initial-value formulation on S. While this equation is not hyperbolic as it stands, it becomes such if there is imposed on n_{ab} the Lorentz-gauge condition

$$\nabla^m (n_{am} - \tfrac{1}{2} g_{am} n^s{}_s) = 0. \tag{21}$$

The initial data then consist of n_{ab}, together with its first normal derivative, evaluated on S. These must be so chosen that the gauge condition (21), together with its first normal derivative, are satisfied on S. To this end, choose as the data $\alpha t_a t_b + 2\alpha_{(a} t_{b)}$ for n_{ab} and $\beta t_a t_b + 2\beta_{(a} t_{b)}$ for its first normal derivative, where t^a is the unit normal to S and α^a and β^a are both orthogonal to t^a. Substituting, the gauge condition (21) gives expressions for β and β^a in terms of α and α^a, while its first normal derivative becomes

$$D^2 \alpha = \mu, \quad D^2 \alpha^a = \mu^a, \tag{22}$$

where D^2 is the Laplacian operator on S, and the source terms on the right involve α and α^a only through their values and first derivatives in S. We thus conclude the following: There exists a solution n_{ab} of Eqs. (19) and (20) provided there exists a solution α, α^a of the elliptic system (22) with α and α^a invariant under the Killing fields. With two or fewer Killing fields, linearly independent at each point, there does exist a solution of Eq. (22) with α, α^a invariant under the Killing fields, as one sees by passing to the manifold of trajectories[10] of the symmetries. It is false in general that there exists such a solution with three or more Killing fields. (The situation here is analogous to trying to solve the Newtonian gravitational equation $D^2\varphi = \rho$ such that the symmetries of ρ are also carried by φ. For ρ invariant under three translations, i.e., constant, there is no solution φ with the same symmetries.) It seems likely that one could prove the full conjecture by similar arguments.

So, the local unreliability apparently sets in only for space-times with a high degree of symmetry—three or more Killing fields. The situation may be contrasted with the global linearization instability, which sets in already with a single Killing field.

How pervasive is unreliability for space-times with high symmetry? A simple class of examples is that provided by certain spatially homogeneous space-times: those for which a symmetry group acts simply transitively on spacelike slices. Recall[11] that such a space-time is determined by a three-dimensional Lie algebra together with positive-definite metric q_{ab} and symmetric tensor p^{ab} over the vector space of the Lie algera, satisfying the constraint equations

$$(p^m{}_m)^2 - p^{mn}p_{mn} - \tfrac{3}{2}v^m v_m - s^{mn}s_{mn} + \tfrac{1}{2}(s^m{}_m)^2 = 0, \tag{23}$$

$$p^m{}_n s^{np}\epsilon_{pma} - \tfrac{1}{2}p^m{}_m v_a + \tfrac{3}{2}p_a{}^m v_m = 0. \tag{24}$$

Here, ϵ_{abc} is the alternating tensor for q_{ab}, and $s^{ab} = s^{(ab)}$ and

v_a result[12] from the decomposition of the structure-constant tensor of the Lie algebra

$$C^a{}_{bc} = s^{am}\epsilon_{mbc} - \delta^a{}_{[b}v_{c]}. \tag{25}$$

The Jacobi relation in the Lie algebra is just $s^{am}v_m = 0$.

Fix the Lie algebra, and consider the 12-dimensional manifold $\mathscr{S}$ of all pairs consisting of positive-definite q_{ab} and symmetric p^{ab}. Denote by $\mathscr{C}$ the subset of this manifold consisting of those points for which q_{ab}, p^{ab} satisfy the constraint equations (23) and (24). Thus, a point of $\mathscr{C}$ determines such a spatially homogeneous space-time. Fix a point q_{ab}, p^{ab} of $\mathscr{C}$, and suppose that $\mathscr{C}$ is a submanifold of $\mathscr{S}$ in a neighborhood of this point. Then a solution of the first-order perturbation equations with background determined by this point yields a tangent vector to $\mathscr{C}$ at this point. Since $\mathscr{C}$ is there a submanifold, this tangent vector is tangent to some curve in $\mathscr{C}$ through the point. Thus, every solution of the first-order perturbation equations with this background arises from some one-parameter family of exact solutions of the system. We conclude that the first-order perturbation equations (and, similarly, the higher-order equations) are reliable whenever the background is given by a point of $\mathscr{C}$ in a neighborhood of which $\mathscr{C}$ is a submanifold of $\mathscr{S}$. The issue of at what points $\mathscr{C}$ is a submanifold is analyzed in the Appendix, with the following result: *The perturbation equations are reliable to all orders for all such spatially homogeneous space-times, with the possible exception of those with $v_a = 0$ and s^{ab} and p_{ab} of the form*

$$s^{ab} = \alpha q^{ab} + \beta w^a w^b, \tag{26}$$

$$p_{ab} = \gamma q_{ab} + \delta w_a w_b, \tag{27}$$

where w^a is any q-unit vector in the Lie algebra, and α, β, γ, and δ are constants satisfying

$$4\gamma(3\gamma + 2\delta) + (\alpha + \beta)(3\alpha - \beta) = 0. \tag{28}$$

Thus, the perturbation equations are reliable to every order for the vast majority of spatially homogeneous background space-times, including in particular all those of Bianchi types III, IV, V, VI_h, and VII_h. The static, plane-symmetric example discussed earlier is analogous to the case $\alpha = \beta = \gamma = \delta = 0$ above. It is curious that all the "exceptional" backgrounds have a fourth Killing field.

Thus, the perturbation equations of general relativity have a pronounced tendency to be reliable.

V. DISCUSSION

We have been concerned here with two types of difficulties associated with perturbation theory in general relativity. The first is that there can exist solutions of the perturbation equations that appear to be gauge—in the sense that the perturbed metric is the symmetrized derivative of a vector field—but which nonetheless are not true gauge. The second is that there can exist solutions of the perturbation equations that cannot be extended to solutions at the next-higher order in perturbation theory. The first difficulty is more pervasive in general relativity than the second, but it also much easier to deal with.

One might think that, at least for situations similar to that of the static, plane-symmetric example, these difficulties are merely the result of a poor choice of variable. Recall that first-order perturbation theory for the vacuum Einstein equation with flat background is often formulated as follows. Let the field be, not the perturbed metric h_{ab}, but rather the perturbed Weyl tensor K_{abcd}, a field having all the symmetries and traces of a Weyl tensor. Let the first-order perturbation equation be, not Eq. (1), but rather the first-order Bianchi identity

$$\nabla_{[a}K_{bc]de} = 0. \tag{29}$$

There is no gauge freedom within K_{abcd} itself, only in its "potential," h_{ab}. This formulation, which does not fit within the general framework of Sec. II, has no second-order version, and is generally inapplicable with sources or curved backgrounds.

It is instructive to see what happens when this formulation is applied to the static, plane-symmetric example. Take as the first-order perturbation equations in this case Eq. (29) together with

$$\mathscr{L}_t K_{abcd} = \mathscr{L}_x K_{abcd} = \mathscr{L}_y K_{abcd} = 0. \tag{30}$$

The most general solution to this system is K_{abcd}, any constant field having the symmetries and traces of a Weyl tensor. By contrast, every one-parameter family of exact solutions of the system gives rise to a vanishing first-order perturbed Weyl tensor. Thus, on the one hand, the first-order perturbation equations in this formulation admit too many solutions; all constant K_{abcd} rather than just $K_{abcd} = 0$. But, on the other hand, these perturbation equations admit too few solutions to reflect adequately the full theory, for all families of exact solutions—even those that do not represent first-order gauge—collapse to $K_{abcd} = 0$ in the perturbation theory. Since a family of exact solutions of the system gives rise to a nonzero Weyl tensor only at second order, one might expect better agreement between the resulting second-order Weyl-tensor perturbations and the solutions K_{abcd} of Eqs. (29) and (30). While all such second-order perturbations of the Weyl tensor are in fact constant tensor fields, not all constants are allowed; there must be satisfied a condition analogous to $(p_1 + p_2 + p_3)^2 = p_1{}^2 + p_2{}^2 + p_3{}^2$ of Sec. I. Thus, there continue to be too many solutions of Eqs. (29) and (30). In short, use of the Weyl tensor as the perturbed field does not seem to alleviate these difficulties with perturbation theory.

One might think that, alternatively, these difficulties are merely the result of our introduction of passive fields. So, let the $\underset{i}{\xi}{}^a$ be active. Include, with the perturbed fields, those resulting from the $\underset{i}{\xi}{}^a$, and include, with the perturbation equations, those resulting from the commutation relations on the $\underset{i}{\xi}{}^a$. Nothing essential then changes. It turns out that there are still solutions of the first-order perturbation equations for which the perturbed metric is a symmetrized derivative but which are not gauge solutions. And there are still solutions that cannot be extended to second order in perturbation theory.

In the discussion of apparent gauge in Sec. III, we treated only the case of Einstein's equation with vanishing sources. The introduction of sources, constructed from additional dynamic fields, does not significantly change the situation. An apparent gauge solution is again determined by a vector field τ^a, where the perturbed fields are those that result from applying $\mathscr{L}_\tau$ to each of the dynamic background fields. Invariance of the active fields under the $\underset{i}{\xi}{}^a$ then requires invariance of these perturbed fields under the $\underset{i}{\xi}{}^a$, which is turn requires invariance of each of the dynamic background fields under the $\mathscr{L}_{\underset{i}{\xi}}\tau^a$. Thus, apparent gauge again results whenever each of the $\mathscr{L}_{\underset{i}{\xi}}\tau^a$ is some linear combination, with constant coefficients, of the $\underset{i}{\xi}{}^a$. For true gauge, on the other hand, we again require $\mathscr{L}_{\underset{i}{\xi}}\tau^a = 0$. So, there will again be solutions that are apparent gauge but not true gauge. A further restriction in Sec. III was that all passive fields be vector fields satisfying Killing's equation. What happens when other types of passive fields are included? Are there versions of apparent gauge at higher order in perturbation theory?

Is the conjecture of Sec. IV, that at least three Killing fields are required for unreliability, true? We showed in Sec. IV that the spatially homogeneous space-times, with certain possible exceptions, are reliable. Are all the exceptional cases actually unreliable? Does the presence of sources, or passive fields other than Killing fields, increase or decrease the chances of reliability? Is reliability less prevalent as one goes to higher orders in perturbation theory? A possible conjecture is that if the nth-order perturbation equations are not reliable, then neither are the $(n+1)$st. One might attack this by trying to show that, whenever Eqs. (19) and (20) admit no solution for some t_{ab}, then that t_{ab} can be reached through some choice of the lower-order perturbations. Even for the simple static, plane-symmetric example the situation is not immediately clear. Is it true that the nth-order perturbation equations in this example are unreliable for all n?

APPENDIX: RELIABILITY OF SPATIALLY HOMOGENEOUS SPACE-TIME

Fix a three-dimensional Lie algebra. As in Sec. IV, denote by $\mathscr{S}$ the 12-dimensional manifold of pairs (q_{ab}, p^{ab}) of tensors over the vector space of the Lie algebra, and by $\mathscr{C}$ the subset of $\mathscr{S}$ consisting of those pairs for which the constraint equations (23) and (24) are satisfied, where $s^{ab} = s^{(ab)}$ and v_a are given by the decomposition (25) of the structure-constant tensor.

Denote by $H(q,p)$ and $H_a(q,p)$ the respective left sides of the constraint equations (23) and (24). Thus, $\mathscr{C}$ consists of those points at which these functions on $\mathscr{S}$ vanish. The issue of at which of its points $\mathscr{C}$ is a submanifold turns on the issue of at which points of $\mathscr{C}$ the gradients in $\mathscr{S}$ of these functions are linearly independent. So, consider the linear combination $\mu H + \nu^a H_a$. Equating to zero the gradient in $\mathscr{S}$ of this combination, i.e., equating to zero the q_{ab}- and p^{ab}-partial derivatives, keeping the structure-constant tensor fixed, we obtain

$$\mu\left[2(p^{ab}p^m{}_m - p^a{}_m p^{bm}) + \tfrac{3}{2}v^a v^b + (s^{mn}s_{mn} - \tfrac{1}{2}(s^m{}_m)^2)q^{ab} - 2s^a{}_m s^{bm} + s^{ab}s^m{}_m\right] + U^{(a}{}_m p^{b)m} = 0, \tag{A1}$$

$$2\mu\left[q_{ab}p^m{}_m - p_{ab}\right] + U_{(ab)} = 0, \tag{A2}$$

where we have set

$$U^a{}_b = s^{am}\epsilon_{mbn}\nu^n + \tfrac{3}{2}\nu^a v_b - \tfrac{1}{2}\delta^a{}_b \nu^m v_m. \tag{A3}$$

Let, q_{ab}, p^{ab}, μ, and ν^a satisfy Eqs. (23), (24), (A1), and (A2). We first derive, as a consequence, the following three equations:

$$\mu\, p_{ab} = 0, \tag{A4}$$

$$\mu v_a = 0 \tag{A5}$$

$$\mu(s^a{}_m s^{bm} - \tfrac{1}{2}s^{ab}s^m{}_m) = 0. \tag{A6}$$

To derive Eq. (A4), first contract Eq. (A2) with q^{ab} and use Eq. (A3), to obtain $\mu p^m{}_m = 0$. Then contract Eq. (A2) with p^{ab}, using this and Eqs. (A3) and (24). To derive Eq. (A5), first multiply Eq. (23) by μ, noting that, by Eq. (A4), the first two terms drop out. Were μv_a nonzero, then the third term would be negative, while, by the Jacobi relation $s^{ab}v_b = 0$, s^{ab} would have rank at most two, whence the last two terms together would be nonpositive. This contradiction establishes $\mu v_a = 0$. To derive Eq. (A6), multiply Eq. (A1) by μ, using (A4), (A5), and the result of multiplying Eq. (23) by μ. We next derive, as a further consequence, the following four equations:

$$U_{(ab)} = 0, \tag{A7}$$

$$U^{(a}{}_m p^{b)m} = 0, \tag{A8}$$

$$U^{(a}{}_m s^{b)m} = 0, \tag{A9}$$

$$U_{ab} v_c = 0. \tag{A10}$$

To derive Eq. (A7), use Eq. (A2) with Eq. (A4). To derive Eq. (A8), use Eq. (A1) with Eqs. (A4)–(A6). To derive Eq. (A9), contract Eq. (A3) with s^{cb} and symmetrize over c and a. The first two terms give zero, while the last also vanishes as a consequence of contracting Eq. (A7) with $\nu^a\nu^b$. To derive Eq. (A10), note that, by antisymmetry of U_{ab}, it suffices to check this equation when contracted with q^{ac} and when the antisymmetrized over all indices. But these both follow from Eq. (A3).

To summarize, we have shown so far that Eqs. (23), (24), (A1), and (A2) together imply Eqs. (A4)–(A10).

Now let (q_{ab}, p^{ab}) be a point of $\mathscr{C}$ at which $\mathscr{C}$ is not a submanifold. Then there must be some linear combination $\mu H + \nu^a H_a$, other than the zero function, whose gradient vanishes at this point. Since $\mu H + \nu^a H_a$ is not the zero function, it follows from Eqs. (23) and (24) that either μ or $U^a{}_b$ is nonzero. But for μ nonzero it follows from Eqs. (A4)–(A6), and for $U^a{}_b$ nonzero from (A7)–(A10), that this is one of the exceptional cases of Sec. IV. We have shown, then, that any point of $\mathscr{C}$ at which $\mathscr{C}$ is not a submanifold must be of the form (26) and (27) with $v_a = 0$. Equation (28) is merely the result of substituting Eqs. (26) and (27) into Eq. (23).

[1]E. Balbinski, S. Detweiler, L. Lindblom, and B. Schutz, Mon. Not. R. Astron. Soc. **213**, 553 (1985).
[2]D. Brill, "Isolated solutions in general relativity," in *Gravitation: Problems and Prospects* (Naukova Dumka, Kiev, 1972), p. 17.

[3]D. Brill and S. Deser, Commun. Math. Phys. **32**, 291 (1973).
[4]A. Fischer and J. Marsden, Bull. Am. Math. Soc. **79**, 997 (1973); A. Fischer and J. Marsden, Proc. Symp. Pure Math. **27**, 219 (1975).
[5]V. Moncrief, J. Math. Phys. **16**, 493 (1975); **17**, 1893 (1976).
[6]A. Fischer, J. Marsden, and V. Moncrief, Ann. Inst. H. Poincaré **33**, 147 (1980); J. Arms, J. Marsden, and V. Moncrief, Ann. Phys. **144**, 81 (1982).
[7]E. Kasner, Am. J. Math. **43**, 217 (1921).
[8]A. Taub, Ann. Math. **53**, 472 (1951).
[9]In order that these diffeomorphisms be defined globally, one must require that the relevant vector fields on $\mathscr{M}$ be complete. But for $n = 1$, this global condition can be avoided by "cutting off" the diffeomorphisms in λ. A similar argument may work for other n.
[10]R. Geroch, J. Math. Phys. **13**, 394 (1972).
[11]R. Wald, *General Relativity* (Chicago U. P., Chicago, 1984), pp. 168–178.
[12]G. F. R. Ellis and M. A. H. MacCallum, Commun. Math. Phys. **12**, 108 (1969).