Gauge, Diffeomorphisms, Initial-Value Formulation, Etc

Robert Geroch

ABSTRACT

We introduce a large class of systems of partial differential equations on a base manifold M, a class that, arguably, includes all systems of physical interest. We then give a general definition — applicable to any system in this class — of "having the diffeomorphisms on M as a gauge group", and "having an initial-value formulation, up to this gauge". This definition is algebraic in the coefficients of the differential equation. The Einstein system, of course, satisfies our definition. There do not, however, appear to be many other systems that do, suggesting that these properties are rather special to the Einstein system.

1 Introduction

Two features of Einstein's equation play a fundamental role in the structure of the general theory of relativity.

The first of these is that this equation manifests gauge freedom associated with diffeomorphisms on the underlying space-time manifold. This freedom is perhaps *the* striking feature of the relativity theory. It impacts the treatment, within the theory, of numerous topics, e.g., the character and structure of gravitational energy [2] [7], and of gravitational radiation [11] [5]. Indeed, while this gauge freedom was once regarded as a curious and novel feature of the theory, it is now sometimes argued (ref) that diffeomorphism gauge is a necessary feature of *any* viable physical theory based on partial differential equations on a manifold.

The second feature is that Einstein's equation manifests, up to this gauge freedom, a well-posed initial-value formulation [4]. Having an initial-value formulation is by now so much a part of our thinking that it is difficult to imagine doing physics without it. For example, in order even to discuss whether or not a physical system is stable, or whether or not it sends signals superluminally requires already such a formulation. In the case of the former, for instance, stability deals with the evolution of small perturbations of that system, while "evolution" refers to an initial-value formulation.

These two features are, clearly, closely intertwined with each other. In particular, Einstein's equation as it stands does *not* admit an initial-value formulation in the traditional sense, precisely because the gauge freedom prohibits this. Given the central role these two features play in the structure of the theory, then, it would be of some interest to understand better how they operate and how they interact. What is the mechanism by which general relativity manifests these features? For example, it is not obvious, merely by examining the Einstein system of partial differential equations, that it has an initial-value formulation up to gauge. One route to understanding these issues would be to formulate a general characterization of having "gauge freedom associated with diffeomorphisms, and, up to that freedom, a well-posed initial-value formulation". That is, one would like to formulate a precise definition of this notion, a definition that is applicable, say, to virtually any system of partial differential equations on a manifold. We here formulate such a definition.

Once in possession of this definition, we can ask what other systems of partial differential equations satisfy it; and whether any of those systems are likely to underlie viable physical theories. We shall find just two other classes of such systems. One is that for special relativity — regarding the Minkowski metric as dynamic. The other is a certain class of systems involving a preferred vector field, none of which seem to underlie physical theories. Are there any other systems of partial differential equations — either invented for this purpose or already available in the context of other, known, physical theories — that also satisfy our definition? Is there some theorem to the effect that the only systems satisfying our definition are those appearing in some short list?

In Sect 2, we review briefly systems of partial differential equations, and the conditions under which such a system admits an initial-value formulation. The key idea is to generate a universal framework, into which all the partial differential equations of physics fit. In Sect. 3, we introduce the notion of gauge for a system of partial differential equations; and, in particular, that of diffeomorphism gauge freedom. It turns out that "gauge" can be defined using only the partial differential equation itself, and not any physical interpretation of that equation. In Sect 4, we introduce our definition of a system's having an initial-value formulation up to gauge. This definition, while perhaps not as simple as one might have hoped, does have the important feature that it is essentially algebraic, i.e., it involves only testing the algebra of the coefficients in the equation. Some further examples and other issues are discussed briefly in the Conclusion, Sect 5. Appendix A is a plea for settlement of an important open question involving the existence of an initial-value formulation for general hyperbolic systems with constraints. Finally, in Appendix B we summarize a few facts about the linearized version of a system of partial differential equations and the relation of linearized solutions to gauge symmetries. While this material in these appendices has some relevance to the paper, it is not central thereto.

2 Partial Differential Equations

It might seem at first thought that the task of this section — to set up a universal framework for the physically interesting partial differential equations — is nearly an impossible one. There is clearly an enormous variety of possible systems of partial differential equations in this World: How, given this apparent diversity, will we ever get sufficient control over these equations to be able to analyze the character of the the "general" one? It turns out, however, that this problem is more apparent than real. There exists a general formulation of the subject of partial differential equations — a formulation that, on the one hand, is systematic, and, on the other, is sufficiently broad to include virtually all equations of physical interest. The key idea of this formulation is to restrict consideration to a certain class of systems of partial differential equations — namely to systems that are first-order (i.e., involve only first derivatives of the fields) and quasilinear (i.e., are linear in those first derivatives). This class is much broader than it appears at first. For example, higher-order systems are cast into this form by introducing new fields to represent the lower derivatives. This class, it turns out, admits a systematic treatment, and at the same time appears to be adequate for the description of physical phenomena. This general formulation of partial differential equations of physics (which is discussed in more detail, with many examples, in [6]) is summarized below.

Let there be given a fibre bundle, consisting of some base manifold M, some bundle manifold $\mathcal{B}$, and some smooth projection mapping $\mathcal{B} \xrightarrow{\pi} M$. Typically, M will be the 4-dimensional manifold of space-time events (but it could be any smooth manifold). By the *fibre* over a point x of M, we mean the set of all points b of $\mathcal{B}$ such that $\pi(b) = x$. Think of the fibre over $x \in M$ as "the set of possible field-values at x". Then $\mathcal{B}$ is interpreted as the set of "all possible choices of field-values at all points of M", and π as the mapping that assigns, to each such choice, the underlying point of M. Thus, point b of $\mathcal{B}$ could be written as $b = (x, \phi)$, with $x \in M$ and ϕ in the fibre over x. The action of the projection mapping would then be given by $\pi(x, \phi) = x$. Typically, the fibre over a point $x \in M$ will be some collection of tensors or other geometrical objects (such as derivative operators, spinors, etc.) at x, possibly subject to various symmetries or other algebraic conditions, whence $\mathcal{B}$ will be the manifold of all such collections of objects at all points of M. A tangent vector at a point of $\mathcal{B}$ is said to be *vertical* if it is tangent to the fibre at that point (or, what is the same thing, if its image under the projection π vanishes). Thus, a vertical vector represents an "infinitesimal change in the field-values" at a fixed point of M. We shall adopt the convention that all "algebraic constraints" on our fields have been incorporated already at the level of the construction of the bundle manifold $\mathcal{B}$, and thus that no such constraints are to be imposed as additional conditions on $\mathcal{B}$. While we shall think of $\mathcal{B}$ as representing "field values", it can in general be any smooth manifold, subject only to the local-product condition in the definition of a fibre bundle[1].

For Maxwell theory, for example, i) M is the four-dimensional space-time manifold; ii) $\mathcal{B}$ is the ten-manifold consisting of pairs (x, F_{ab}), where $x \in M$ and F_{ab} is a skew tensor at x; iii) the projection π sends (x, F_{ab}) to x; iv) the fibre over $x \in M$ is the six-manifold of skew F_{ab} at x; and v) a vertical vector at point (x, F_{ab}) of $\mathcal{B}$ may be represented by a skew tensor[2] δF_{ab} at

[1]Recall that this condition requires, essentially, that, locally in M, $\mathcal{B}$ can be written as a product, $M \times F$, of M with some other fixed manifold F, in such a way that the projection mapping π becomes the projection to the M-factor in this product. This condition guarantees, e.g., that, locally, all the fibres of the bundle are diffeomorphic with this fixed manifold F, and so with each other. We shall take the term "fibre bundle", to mean such a smooth mapping of manifolds, $\mathcal{B} \xrightarrow{\pi} M$, subject only to this local-product condition. That is, we shall not require (as is sometimes done [12]) that there also be given a group action on the bundle manifold.

[2]This δF_{ab}, which we may think of as a "small change in F_{ab}" is, more precisely, the tangent vector, at F_{ab}, to some curve in the manifold of skew tensors at x.

x. For general relativity, M is again the four-manifold of space-time events, while the fibre over $x \in M$ consists of pairs (g_{ab}, ∇_a), where g_{ab} is a Lorentz-signature metric and ∇_a a derivative operator[3], at x. Thus, the dimension of the fibres in this case is 50 (= 10 + 40); while the dimension of the manifold $\mathcal{B}$ is 54. A vertical vector at point (x, g_{ab}, ∇_a) of $\mathcal{B}$ may be represented by pair of tensors, $(\delta g_{ab}, s^m{}_{ab})$, with δg_{ab} (representing the g_{ab}-component of the vector) and $s^m{}_{ab}$ (representing the ∇_a-component of the vector[4]) both symmetric the indices "a" and "b".

Returning to the general case, by a *cross-section* of such a bundle we mean a smooth mapping $M \xrightarrow{\phi} \mathcal{B}$ such that $\pi \circ \phi$ is the identity map on M. In other words, a cross-section ϕ assigns, to each point x of M, some point, $\phi(x)$, of the fibre over x. Think of a given cross-section as representing a particular choice of a "field" (of the type represented by the bundle) over M. Thus, for Maxwell theory, a cross-section of the bundle is represented by a smooth skew tensor field F_{ab} on the space-time manifold M; for general relativity, by smooth fields g_{ab} and ∇_a on M.

Our partial differential equation will be an equation on such a cross-section map ϕ, linear in its first derivative. In order to write out this equation, we must introduce two new smooth fields, $k^{Aa}{}_\alpha$ and j^A, on $\mathcal{B}$. Being fields on $\mathcal{B}$, these depend of course on the point $b = (x, \phi)$ of $\mathcal{B}$, i.e., they depend on a choice of "point x of the base manifold, as well as field-value ϕ at that point". The index "α" on $k^{Aa}{}_\alpha$ is a tensor index in $\mathcal{B}$ at the point, $b \in \mathcal{B}$, at which this field is evaluated; the index "a" is a tensor index in M at the corresponding point, $\pi(b)$, of the base manifold. The index "A", of both $k^{Aa}{}_\alpha$ and j^A, lies in some new vector space (which will turn out, shortly, to be the vector space of equations). Finally, our partial differential equation, on a cross-section ϕ, is

$$k^{Aa}{}_\alpha (\nabla \phi)_a{}^\alpha = j^A. \tag{1}$$

This equation is to be imposed at each point $x \in M$, with the fields k and

[3] A (torsion-free) derivative operator at a point x of M could be defined, for example, as a map from smooth covector fields on M to second-rank covariant tensors at x, subject to additivity, the Leibnitz rule, and consistency with the exterior derivative.

[4] Recall that the difference of two (torsion-free) derivative operators, ∇_a and $\tilde{\nabla}_a$ on a manifold is represented by a tensor $C^m{}_{ab} = C^m{}_{(ab)}$, which is defined by the property that, for any smooth covector field k_a, $\nabla_a k_b - \tilde{\nabla}_a k_b = -C^m{}_{ab} k_m$. Thus, a tangent vector in the space of derivative operators at a point gives rise naturally, i.e., without any "reference" derivative oprator $\tilde{\nabla}_a$, to such a tensor $s^m{}_{ab}$.

j evaluated at $\phi(x) \in \mathcal{B}$, i.e., on the cross-section. Here, $(\nabla\phi)_a{}^\alpha$ denotes the derivative of the map ϕ (i.e., a map from tangent vectors in M at x to tangent vectors in $\mathcal{B}$ at $\phi(x)$). The index "A" in Eqn. (1) is free, i.e., Eqn. (1) represents a number of scalar equations equal to the dimension of the vector space in which "A" lies. We demand that $\nu_A k^{Aa}{}_\alpha = 0$ only when $\nu_A = 0$, i.e., that every the equation of the system (1) really is "differential".

For Maxwell theory, Eqn. (1) is Maxwell's equations,

$$\nabla_{[a}F_{bc]} = 0, \tag{2}$$

$$\nabla^a F_{ab} = 0. \tag{3}$$

The index "A" in this case stands for three antisymmetric M-tensor indices, together with a single M-tensor index (this being the index-structure of Eqns. (2)-(3)). Thus, "A" in this example lies in a vector space of dimension 8 (= 4 + 4). For general relativity (say, with vanishing sources), Eqn. (1) is

$$\nabla_a g_{bc} = 0, \tag{4}$$

$$R_{ab(c}{}^m g_{d)m} = 0, \tag{5}$$

$$R_{amb}{}^m = 0, \tag{6}$$

where we have defined the Riemann tensor $R_{abc}{}^d$ by the property that $\nabla_{[a}\nabla_{b]}k_c = 1/2 R_{abc}{}^d k_d$ for every covector field k_d on M. Note that these are indeed first-order, quasilinear equations in the fields g_{ab} and ∇_a (for $R_{abc}{}^d$ is the "derivative of the derivative operator"). The equation-index "A" in this example lies in a vector space of dimension 110 (= 40 + 60 + 10, these three terms corresponding, respectively, to the three equations (4)-(6)).

Returning to the general case, we are concerned with the issue of when the system (1) has an initial-value formulation. To this end, we consider a submanifold, T, of M of codimension one, together with a cross-section ϕ_0, of $\mathcal{B}$ over T[5]. Think of the cross section ϕ_0 as the "initial data" at the "time" represented by T. When do these data give rise to (i.e., are the restrictions to T of) some solution, ϕ, of (1), and when is that solution unique? In order to answer these questions, we require two further notions.

Fix a bundle, $\mathcal{B} \xrightarrow{\pi} M$, and a system (1) of partial differential equations on that bundle. By a *constraint* of this system, at a point of $\mathcal{B}$, we mean a

[5]That is, ϕ_0 is a smooth map from T to $\mathcal{B}$ such that $\pi \circ \phi_0$ is the identity on T.

tensor $c^a{}_A$ at that point satisfying $c^{(a}{}_A k^{|A|b)}{}_{\alpha'} = 0$ there. Here, and hereafter, a prime on a Greek subscript means "applied only to vertical vectors". [These primes will appear frequently, for it is usually only the "vertical parts" of things that are of interest.] Note that the constraints at each point of $\mathcal{B}$ form a vector space. Each constraint, it turns out, has two distinct facets. It is one of the beauties of this subject that these two, apparently quite dissimilar, facets coalesce into the simple, geometrical, definition above.

As to the first facet, each constraint gives rise to an integrability condition. Fix any constraint field, $c^a{}_A$, and any solution ϕ of Eqn. (1). Contract both sides of Eqn. (1) with $c^b{}_A$, and apply to both sides of the result some derivative operator, $\tilde{\nabla}_b$, on M. Then, by the defining equation for a constraint, terms involving second derivatives of ϕ disappear, leaving an algebraic equation (indeed, a polynomial of degree at most two) in the first derivative, $(\nabla\phi)_a{}^\alpha$, of ϕ. The given constraint field is said to be *integrable* if this equation is an algebraic consequence of Eqn. (1), i.e., if the difference of its two sides can be written as the product of two factors: one some expression (at most linear in field-derivatives), and the other the difference of the two sides of (1). Note that integrability of a fixed constraint field is independent of the choice, above, of the derivative operator $\tilde{\nabla}$, for a change in this choice is compensated for by a change in the first factor in the product above. Note also that any linear combination, with coefficients functions on $\mathcal{B}$, of integrable constraint fields is again an integrable constraint field. We say that the constraints of the system are *integrable* if every such constraint field is. Failure of integrability of the constraints may be interpreted as meaning that "not all of the partial differential equations appropriate to the given system have been included in (1)"[6].

As to the second facet, each constraint gives rise to a consistency condition on initial data. Fix a constraint field, $c^a{}_A$, and a solution ϕ of Eqn. (1); as

[6]On discovering that integrability fails for some given system of partial differential equations, one might contemplate constructing a new system, all of whose constraints *are* integrable, in the following manner. First augment the given system by additional quasilinear equations so chosen that they cause the original integrability conditions to be satisfied. This new system, so constructed, may now give rise to new constraints, and so to additional integrability conditions. If so, repeat the first step above, adding as necessary still more quasilinear equations; and continue in this way. In many cases, such a procedure terminates in an integrable system, although, as far as I am aware, there is no simple criterion for when this will happen.

well as some submanifold T of M of codimension one. Then, at each point of T, we have

$$n_m c^m{}_A k^{Aa}{}_\alpha (\nabla\phi)_a{}^\alpha = n_m c^m{}_A j^A, \tag{7}$$

where n_m is the normal to T at that point. But, by virtue of the definition of a constraint, the index "a" in the expression $n_m c^m{}_A k^{Aa}{}_\alpha$ is tangent to T (for the contraction of this expression with the normal, n_a, vanishes). So, Eqn. (7) takes the derivative of ϕ only in directions tangent to T, and so it refers only to the value of ϕ on T, i.e., only to the initial data induced on T from ϕ. In short, Eqn. (7) represents a consistency condition on initial data. If any consistency condition obtained in this way were not satisfied, by given data on T, then we would have have no hope of finding, for those data, a corresponding solution of Eqn. (1). We say that the constraints of the system (1) are *complete* if, in a certain sense, *every* consistency condition on initial data on T arises in the manner of Eqn. (7). The definition[7], in more detail, is the following. We demand that, for every point b of $\mathcal{B}$, there is an open set of covectors n_a at the corresponding base point, $\pi(b)$, such that the following holds: Given any n_a lying in that open set, and any ν_A, such that $n_a \nu_A k^{Aa}{}_{\alpha'} = 0$, then there exists a constraint $c^m{}_A$ of the system such that $\nu_A = c^m{}_A n_m$. Here, the n_a represents a normal to a possible initial surface (and the open set a restriction on the allowed surface-normals), while the ν_A represents the selection, from (1), of a particular consistency condition. The definition then requires that every such consistency condition arise, via (7), from some constraint $c^m{}_A$.

Integrability and completeness, as defined above, hold, as far as I am aware, for every system of partial differential equations of physical interest.

Consider, for example, the Maxwell system, with field F_{ab}, and equations (2)-(3). Then initial data consist of the specification of a skew field F_{ab} over the 3-submanifold T of space-time, where, of course, these tensor indices lie within the full 4-dimensional space-time M. A constraint, $c^m{}_A$ is represented by a pair of tensors, (c^{mabc}, cg^{ma}), where c^{mabc} is totally antisymmetric, and c is a number. Here, the equation-index "A" is represented by an antisymmetric triple of indices, "abc", together with a single index, "a". The vector

[7]This definition replaces a more awkward one, involving dimensions of various vector spaces, that was given earlier in [6]. While the two definitions are equivalent in the presence of a hyperbolization (defined below), the present condition is, in the general case, much more convenient than the earlier one. For example, the Einstein system satisfies the present definition of completeness, but not the earlier one.

space of constraints at each point, then, is 2-dimensional. The corresponding integrability conditions, obtained by taking the curl of (2) and the divergence of (3) are of course identities. So, the constraints are integrable. The corresponding consistency conditions (7) become, in this example, the familiar conditions $\nabla \cdot B = 0$ and $\nabla \cdot E = 0$ on the induced initial data on a spacelike T. This system, we claim, is complete. Indeed, let the open set of n_a be that consisting of the timelike covectors. Completeness then becomes the following assertion: Let, for fixed timelike n_a, $\nu^{abc} = \nu^{[abc]}$ and ν^a (i.e., let n_a in the open set and ν_A) be such that

$$\nu^{abc} n_a \delta F_{bc} + \nu^a n^b \delta F_{ab} = 0 \tag{8}$$

for every $\delta F_{ab} = \delta F_{[ab]}$ (i.e., such that $n_a \nu_A k^{aA}{}_\alpha = 0$). Then, for some $c^{abcd} = c^{[abcd]}$ and some number c, (i.e., for some $c^a{}_A$) we have $\nu^{bcd} = n_a c^{abcd}$ and $\nu^b = c g^{ab} n_a$ (i.e., we have $\nu_A = n_a c^a{}_A$). But, as is easily checked, this assertion is true. Note that the open set of n_a in this example depends only on the base point, and not the point of the fibre.

Consider, as a second example, the Einstein system, with fields (g_{ab}, ∇_a), and equations (4)-(6). The initial data in this case consist of the specification of fields g_{ab} and ∇_a on a 3-manifold T. Note that our initial data for the Einstein system consist of the *entire* spacetime metric g_{ab} at points of T (and not just its projection into T), as well as the *entire* derivative operator ∇_a at points of T (and not just that part of it associated with the extrinsic curvature). These choices are necessitated by our demand that the Einstein system fit into the general framework for systems of partial differential equations and the initial-value formulation of such systems. In my view, the present choices for the initial data, while not the standard ones, do serve to clarify the relationship between the surface T, the data on that surface, and the action of the gauge group.

For the Einstein system, the general constraint, $c^m{}_A$, at a point is represented by three tensors, $c^{mabc} = c^{[ma](bc)}$, $c^{mabcd} = c^{[mab](cd)}$, and $c^{mab} = g^{m(a}c^{b)}$. Here, the equation-index "A" is represented by the three index-combinations "abc", "$abcd$", and "ab", with appropriate symmetries, corresponding to the three equations, (4)-(6), of the Einstein system. The dimension of the vector space of constraints at a point is 104 (= 60 + 40 + 4). The corresponding integrability conditions correspond to taking a curl of Eqn. (4) and to applying the Bianchi identities to Eqns. (5)-(6). These constraints are integrable (the integrability condition for (4) being (5), and those for (5)-(6) being "identities"). Furthermore, we claim, this system is complete. Indeed, let the open set of n_a consist of the timelike covectors. [Note that this set, in contrast to the Maxwell case, *depends* on fibre-point.] Completeness then becomes the following assertion: Let, for fixed timelike n_a, $\nu^{abc} = \nu^{a(bc)}$, $\nu^{abcd} = \nu^{[ab](cd)}$ and $\nu^{ab} = \nu^{(ab)}$ (i.e., ν_A) be such that

$$\nu^{abc} n_a \delta g_{bc} + \nu^{abcd} n_a s^p{}_{bc} g_{dp} + \nu^{ab} n_{[a} s^m{}_{m]b} = 0 \tag{9}$$

for every $\delta g_{bc} = \delta g_{(bc)}$ and $s^m{}_{ab} = s^m{}_{(ab)}$ (i.e., such that $n_a \nu_A k^{aA}{}_\alpha = 0$). Then, for some $c^{mabc} = c^{[ma](bc)}$, $c^{mabcd} = c^{m[ab](cd)}$ and $c^{mab} = g^{m(a}c^{b)}$ (i.e., for some $c^a{}_A$) we have $\nu^{abc} = n_m c^{mabc}$, $\nu^{abcd} = n_m c^{mabcd}$ and $\nu^{ab} = n_m c^{mab}$ (i.e., we have $\nu_A = n_a c^a{}_A$). This assertion, again, is true.

We return now to the general case. The key to achieving an initial-value formulation for the system (1) is an object called a *hyperbolization*, a field $h_{\beta A}$ on the bundle manifold $\mathcal{B}$ having the properties described below. Fix any point, (x, ϕ), of $\mathcal{B}$, and consider, for any covector n_m at $x \in M$ and any two vertical vectors, $\delta\phi^\alpha, \delta'\phi^\alpha$, at $(x, \phi) \in \mathcal{B}$, the expression

$$n_m h_{\beta A} k^{Am}{}_\alpha \delta\phi^\alpha \delta'\phi^\beta. \tag{10}$$

This expression is a bilinear form in $\delta\phi^\alpha$ and $\delta'\phi^\alpha$. We demand, in order that this $h_{\beta A}$ be a hyperbolization, that this expression be symmetric under interchange of $\delta\phi^\alpha, \delta'\phi^\alpha$ for every n_m, and positive-definite (i.e., positive whenever $\delta'\phi^\beta = \delta\phi^\beta \neq 0$) for every n_m lying in some open set. Generally speaking, the most direct way to specify a hyperbolization for a system of partial differential equations is simply to give this bilinear expression. Such an expression indeed defines a hyperbolization provided it is symmetric and positive-definite, as described above; and furthermore, that it is of the form (10), i.e., that it is some multiple of the result of replacing, in the left side of Eqn. (1), "$(\nabla\phi)_a{}^\alpha$" by "$n_a \delta\phi^\alpha$". For example, the Maxwell system possess a hyperbolization. A corresponding quadratic form is given by

$$u_a n_b (\delta F^{am} \delta' F^b{}_m - 1/4 g^{ab} \delta F_{mn} \delta' F^{mn}). \tag{11}$$

where u^a is any fixed timelike vector. We note that this quadratic form is indeed symmetric under interchange of δF_{ab} and $\delta' F_{ab}$, and that it is indeed positive-definite for u^a and n_a timelike with $u^a n_a > 0$. Furthermore, this quadratic form does indeed arise as a linear combination of $n_{[a}\delta F_{bc]}$ and $n_a \delta F^{ab}$ (this fact being, e.g., the essence of the proof that conservation of the standard electromagnetic stress-energy follows from Maxwell's equations). Note that each choice of timelike u_a gives rise to a particular hyperbolization, i.e., that there are many hyperbolizations in this Maxwell example. More generally, consider any system of equations on fields, for which there is a symmetric stress-energy tensor that i) is a quadratic algebraic function of all the fields; ii) is, by virtue of the field equations, conserved; and ii) satisfies a suitable energy condition. Then that system of equations admits a hyperbolization via this stress-energy, just as for the Maxwell case. The Einstein system admits no hyperbolization, a feature that, as we shall see later, is closely related to the invariance of this system under diffeomorphisms.

A key assertion of this subject is to the effect that a system of partial differential equations, provided it satisfies certain conditions, must have an initial-value formulation. Consider the system (1), and let us suppose that i) the constraints of this system are integrable and complete, ii) this system admits some hyperbolization field, $h_{\alpha A}$, on $\mathcal{B}$, and iii) the open sets of covectors n_a for completeness and for the hyperbolization have, at each point of $\mathcal{B}$, nonempty intersection. Next, let there be given initial data, i.e., a cross section ϕ_0 over some submanifold T of M of codimension one. Let this initial-data set, (T, ϕ_o), i) satisfy the consistency conditions, (7), and ii) be such that, at each point of T, the normal, n_a, to T at that point lies within the two open sets above. Note that the second supposition, which is essentially the requirement that the data set (T, ϕ_o) be "non-characteristic", in general involves *both* T and the data ϕ_o thereon. Then we have the following assertion: There exists one and only one solution of the partial differential equation, (1), defined in a neighborhood of T, that manifests the given initial data. This assertion is discussed in more detail in Appendix A. It is, unfortunately, not theorem, because there is a gap in the proof. Closing this gap (possibly with the introduction of some benign further hypotheses) is, in my opinion, one of the important open questions in the subject of partial differential equations.

It is not hard to see, intuitively, that the assertion above is reasonable. Let the system (1) satisfy the conditions above, and fix n_a, lying in the open sets above. Now consider $n_a k^{Aa}{}_\alpha$, regarded as a linear mapping from the vector space of equations to the vector space of covectors in field-space. The domain of this mapping is the vector space of equations; the kernel of this mapping is (by completeness) the vector space of consistency conditions (7); and the range of this mapping is (by existence of a hyperbolization) the vector space of vertical vectors. Using an elementary fact about linear mappings, we now have the following relation: (dimension of space of equations) - (dimension of space of consistency conditions) = (dimension of space of fields). But this relation asserts that the number of equations in (1) involving "time-derivatives" (the left side of the relation) is equal to the number of field-components. In other words, this relation guarantees that Eqn. (1) can be solved for the "time-derivatives" of all fields, in terms of their values and space-derivatives. Then integrability guarantees, in a similar way, that the time-derivatives of the consistency conditions are expressed in terms of the values and space-derivatives of those conditions. In short, completeness, integrability and the existence of a hyperbolization, taken together, guarantee that the system (1) has a "naive" initial-value formulation.

3 Gauge

A key concept is the notion of a gauge transformation. We begin with the "infinitesimal" ones.

Fix a first-order, quasilinear system of partial differential equations, (1). By a *gauge vector field* for this system, we mean a smooth vector field, ξ^α, on the bundle manifold $\mathcal{B}$ that i) preserves the fibres, i.e., has the property that, for any two points b, b' of the bundle manifold lying in the same fibre, we have $(\nabla\pi)_\alpha{}^a(\xi^\alpha)|_b = (\nabla\pi)_\alpha{}^a(\xi^\alpha)|_{b'}$; and ii) preserves the system (1), i.e., satisfies the equations

$$\mathcal{L}_\xi k^{Aa}{}_\alpha = S^A{}_B k^{Ba}{}_\alpha + \gamma^{aA}{}_m (\nabla\pi)_\alpha{}^m, \tag{12}$$

$$\mathcal{L}_\xi j^A = S^A{}_B j^B + \gamma^{mA}{}_m, \tag{13}$$

for some fields $S^A{}_B$ and $\gamma^{aA}{}_m$ on $\mathcal{B}$.. The first condition, preservation of the fibres, is precisely the requirement that there exist some vector field in the base manifold (the *drop* of ξ) having ξ^α as its lift. So, for example, every vertical vector field on $\mathcal{B}$ is fibre-preserving (for it is a lift of the zero vector field on the base manifold). In the second condition, the Lie-derivative operator, $\mathcal{L}_\xi$ is defined as follows: Consider the one-parameter family of diffeomorphisms on the bundle manifold generated by ξ^α. Take the image of the field ($k^{Aa}{}_\alpha$ or j^A) under that family of diffeomorphisms (noting that this is well-defined, in the case of $k^{Aa}{}_\alpha$, by condition i)). Finally, take the parameter-derivative of this family at parameter-value zero. The arbitrary fields $S^A{}_B$ and $\gamma^{aA}{}_m$ in Eqns. (12)-(13) reflect a certain freedom in how our basic differential equation (1) is represented in terms of (k, j). Indeed, replacement of $(k^{Aa}{}_\alpha, j^A)$ by $(W^A{}_B k^{Ba}{}_\alpha, W^A{}_B j^B)$, for $W^A{}_B$ an arbitrary invertible tensor field, results in an identical system of equations. This is the origin of the $S^A{}_B$ in (12)-(13). Furthermore, adding to $k^{Aa}{}_\alpha$ and j^A the fields $\gamma^{aA}{}_m(\nabla\pi)_\alpha{}^m$ and $\gamma^{mA}{}_m$, respectively, where $\gamma^{aA}{}_m$ is an arbitrary field on $\mathcal{B}$, also results in an identical system of equations[8]. This is the origin of the $\gamma^{aA}{}_m$ in (12)-(13).

These gauge vector fields represent, of course, "infinitesimal gauge transformations". We may also define, similarly, a (full) *gauge transformation* as a diffeomorphism, Ψ, on $\mathcal{B}$ that sends fibres to fibres, and the fields $k^{Aa}{}_\alpha, j^A$ to an equivalent pair. It is immediate that, for Ψ a gauge transformation, there exists a unique diffeomorphism, ψ, (the *drop* of Ψ) on M such that $\pi \circ \Psi = \psi \circ \pi$. These two versions of gauge are indeed related as we would expect: A vector field on $\mathcal{B}$ generates (locally) gauge transformations if and only if it is a gauge vector field. The gauge vector fields are generally simpler to work with computationally, while the full gauge transformations are easier to think about.

Every gauge transformation sends every solution cross-section of (1) to another solution cross-section (clearly, since a diffeomorphism on $\mathcal{B}$, in order to be a gauge transformation, must preserve everything involved in (1))[9]. The Lie bracket of two gauge vector fields is, of course, a gauge vector field, and so the gauge vector fields form a Lie algebra. Furthermore, the drop of the bracket is the bracket of the drops. That is, "drop" is a homomorphism from the Lie algebra of gauge vector fields to the Lie algebra of smooth vector fields on M. Similarly, the gauge transformations form, under composition,

[8]To see this, use the relation $(\nabla\phi)_a{}^\alpha(\nabla\pi)_\alpha{}^b = \delta^b{}_a$, which is precisely the derivative of the equation $\pi \circ \phi =$ (identity on M).

[9]The converse of this assertion — that a fibre-preserving diffeomorphism on $\mathcal{B}$ that sends solution cross-sections to solution cross-sections must be a gauge transformation — also holds, under the additional condition that there are "sufficiently many" solutions of (1). What is required, in more detail, is that, given any tensor $\mu_a{}^\alpha$ at any point of $\mathcal{B}$ such that $k^{aA}{}_\alpha \mu_a{}^\alpha = j^A$ at that point, then there exists some solution cross-section through that point, with $(\nabla\phi)_a{}^\alpha = \mu_a{}^\alpha$ there.

a group; and "drop" is a homomorphism from this group to the group of diffeomorphisms on the manifold M.

The traditional picture of a gauge transformation is of a change in the mathematical objects used to describe a physical system, which, however, reflects no change in the physical system itself. According to this picture, then, the notion of a gauge transformation is intimately connected with the physical interpretation that is attached to the mathematical objects. But here, by contrast, the definition of "gauge transformation" refers only to mathematical objects — specifically, only to the differential equation (1) — with no direct reference to any physical interpretation. The present definition, in my view, represents a useful observation about the character of gauge transformations. The idea is that the "physical interpretation" is already inherent, in some sense, in the partial differential equation itself. This interpretation, after all, is merely a compilation of physical effects the fields produce; and these effects, in turn, are described entirely by the equations those fields satisfy (possibly involving additional fields that describe the various measuring instruments). This remark will be illustrated by the examples below.

In the Maxwell case, with zero sources in flat space-time, the gauge transformations consist precisely of duality rotations and scaling transformations on the Maxwell field (replacement F_{ab} by any linear combination, with constant coefficients, of itself and its dual), and, the replacement of F_{ab} by the result of applying to that field a Poincare transformation. That is, the gauge group is a Lie group, of dimension twelve. The corresponding gauge vector fields are vertical for the duality rotations and scaling, but not for the Poincare transformations. If we replace this system by one in a curved background space-time (say, having no symmetries), then the Poincare group of gauge transformations disappears. If we further replace this Maxwell system by one with a fixed charge-current source, then we no longer have duality or scaling as gauge transformations. Quite generally, gauge transformations lose their character when there is turned on an interaction that breaks the corresponding gauge-symmetry. Suppose that the charge-current source was not fixed, but instead was expressed in terms of additional fields, where these were included within the system, (1), of partial differential equations? In this case, the scalings would normally reappear as gauge transformations. Note that in order to recover the "usual" gauge transformations in the Maxwell case, it is necessary to introduce an additional field, A_a, (thus enlarging $\mathcal{B}$

to a fourteen-manifold), and an additional equation,

$$\nabla_{[a}A_{b]} = F_{ab}. \tag{14}$$

For this enlarged system, the gauge transformations include adding to A_a the gradient of any smooth scalar field on M. The corresponding gauge vector fields are all vertical. Note that it not possible, within the present framework, to admit only the field A_a, and not F_{ab}, for we are requiring right from the beginning that our system be first-order.

For the Einstein system with zero sources, the gauge transformations consist [1] of scaling transformations (multiplication of g_{ab} by a constant factor, keeping ∇_a fixed), and the diffeomorphisms on M (which we shall consider in more detail shortly). If sources are included in Einstein's equation — and if those sources are represented by fields, which are included within the bundle manifold $\mathcal{B}$ and on which equations are imposed — then these gauge transformations above will generally remain such. Suppose that we wished to restrict consideration to a particular class of solutions of Einstein's equation — say, those having a Killing field. What are the gauge transformations in this case? We are not permitted simply to impose "having a Killing field" on top of Eqns. (4)-(6), for the general rule is that *all* information is to be encoded, once and for all, into the single system, (1), of partial differential equations. So, we might proceed as follows. First, introduce a new bundle, with fibre over $x \in M$ consisting of quadruples, $(g_{ab}, \nabla_a, \zeta^a, \zeta_a{}^b)$, with g_{ab} Lorentz-signature and the combination $g_{ma}\zeta_b{}^m$ symmetric. [Thus, the fibres in this case have dimension 64 (= 10 + 40 + 4 + 10).] Let the equations on these fields consist of all those already given for the Einstein system, (4)-(6), together with two new equations: $\nabla_a\zeta^b = \zeta_a{}^b$ and $\nabla_a\zeta_b{}^m + R_{sab}{}^m\zeta^s = 0$. Thus, the new field ζ^a represents the Killing field, and $\zeta_a{}^b$ its derivative (the latter being necessary to retain the first-order character). Note that this new system has, for its gauge transformations, not only the g-scalings and diffeomorphisms above, but also ζ-scalings. An alternative treatment of the Einstein system with Killing field is the following. Fix, once and for all, a (say, nowhere vanishing) vector field ζ^a on M. Let the bundle $\mathcal{B}$ consist of all (g_{ab}, ∇_a), with g_{ab} Lorentz signature, such that $g_{m(a}\nabla_{b)}\zeta^m = 0$. Note that the latter is an *algebraic* condition on the fields. The fibres in this case have dimension forty. Let the equations be the usual ones for the Einstein system, (4)-(6). For this system, the gauge transformations include, not all M-diffeomorphisms, but rather only those that are ζ^a-preserving.

The examples above — of various Maxwell and Einstein systems — are sufficiently familiar that it was possible to determine their gauge transformations relatively easily. But what about more complicated systems? Is there some simple, general procedure that, applied to any system, (1), of partial differential equations, will yield the complete class of gauge vector fields for that system? None, apparently, is known. But here is a possible line that might yield such a procedure. Fix the system, (1). First, find all fields ξ^α and $S^A{}_B$ such that the equation

$$\mathcal{L}_\xi k^{Aa}{}_{\alpha'} = S^A{}_B k^{Ba}{}_{\alpha'} \tag{15}$$

holds. Eqn. (15) is precisely Eqn. (12), restricted to vertical vectors. But this equation has the great advantage over (12) that it acts within each fibre separately: It is virtually "algebraic". Next, find the $\gamma^{aA}{}_m$ (which, by (15), must exist; and which must be unique) such that Eqn. (12) holds. And finally, demand (as a condition on the original choices of ξ^α and $S^A{}_B$) that this $\gamma^{aA}{}_\alpha$ also satisfy Eqn. (13). It might be of interest to see if there could be generated, along this line, some simple procedure for finding the gauge vector fields.

Up to this point, we have been dealing with "gauge" in a very general context. We now wish to consider a special case of particular interest: The gauge of diffeomorphisms. The system (1) of partial differential equations will be said to admit *diffeomorphism gauge* if every smooth vector field ξ^a on M admits a lifting to some gauge vector field, ξ^α, on $\mathcal{B}$. In terms of the full gauge transformations, the condition is that every diffeomorphism on M lifts to a gauge transformation on $\mathcal{B}$. In the presence of diffeomorphism gauge, every gauge vector field can be written as the sum of one of the ξ^α above and a vertical gauge vector field. It is apparently not known whether, for a system with diffeomorphism gauge, one can always make a specific choice of lifting, for each vector field ξ^a on M, such that the resulting gauge vector fields themselves form a Lie algebra. Note that, in the presence of diffeomorphism gauge, the Lie algebra of gauge vector fields is always infinite-dmensional.

The Einstein system, of course, admits diffeomorphism gauge. The Maxwell system in Minkowski space-time does not. However, if we modify the latter by including the metric and derivative operator among the fields, and including Eqn. (4) and the vanishing of the Riemann tensor among the equations, then we again recover a system having diffeomorphism gauge.

Diffeomorphism gauge is particularly tractable in the case in which the gauge vector field ξ^α on $\mathcal{B}$ arises from the vector field ξ^a on M by means of a differential operator. Fix any derivative operator $\tilde{\nabla}_a$ on M, and consider the equation

$$\xi^\alpha = \delta^{\alpha a \cdots c}{}_m \tilde{\nabla}_a \cdots \tilde{\nabla}_c \xi^m + \cdots . \tag{16}$$

Here, the right side represents a general linear combination of derivatives of ξ^a up to the n-th (with $n \geq 1$). We have written out only the highest-order term, with the remaining terms (with orders $n-1$ down to zero) indicated by dots. The coefficient, $\delta^{\alpha a \cdots c}{}_m$, of this highest-order term is a natural tensor, independent of the choice of derivative operator $\tilde{\nabla}_a$ used on the right in (16). The coefficients of the lower-order terms do, of course, depend on this choice. In (16), the coefficients depend on point of the bundle manifold (and so in particular $\delta^{\alpha a \cdots c}{}_m$ is a field on $\mathcal{B}$). The right side of Eqn. (16) is thus a function on $\mathcal{B}$, where, for $b \in \mathcal{B}$, the expressions $\tilde{\nabla}_a \cdots \tilde{\nabla}_c \xi^m$ are to be evaluated at $\pi(b) \in M$. Thus, this right side indeed defines a vector field, ξ^α, on $\mathcal{B}$. We next demand that, for every smooth field ξ^a on M, the field ξ^α on $\mathcal{B}$ given by (16) be a lift of ξ^a. This implies that the index "α" of $\delta^{\alpha a \cdots c}{}_m$ is vertical. [To see this, apply $(\nabla \pi)_\alpha{}^s$ to both sides of Eqn. (16), using $n \geq 1$ and noting that the left side then involves no derivatives of ξ^a.] Finally, we impose on the ξ^α of (16) the condition that it be a gauge vector field. We claim: If, for every smooth field ξ^a on M, the field ξ^α on $\mathcal{B}$ given by (16) is a gauge vector field, then the coefficient $\delta^{\alpha a \cdots b}{}_m$ in (16) must satisfy the following two conditions: i) For every tensor field $L_{a \cdots c}{}^m$ on M, the vertical vector field given by $\kappa^\alpha = \delta^{\alpha a \cdots c}{}_m L_{a \cdots c}{}^m$ satisfies

$$\mathcal{L}_\kappa k^{Aa}{}_{\alpha'} - [k^{Ac}{}_{\alpha'} L_c{}^m] = S^A{}_B k^{Ba}{}_{\alpha'}, \tag{17}$$

for some $S^A{}_B$; and ii)

$$k^{A(d}{}_\alpha \delta^{|\alpha| a \cdots c)}{}_m = 0. \tag{18}$$

In condition i), the term in square brackets in (17) is to be included *only* for the case $n = 1$. [In fact, this term only makes sense for $n = 1$.] Condition i) follows from Eqn. (15), retaining in that equation only the part of highest order (namely, n) in ξ^a-derivatives. Since the vertical part is being taken in (15), i.e., since the index α is primed there, only the vertical derivative of ξ^α is taken in that equation, and so no further derivatives of ξ^a are introduced there. The square-bracketed term appears in (17) because the Lie derivative on the left in (15) involves a term $k^{Am}{}_{\alpha}\tilde{\nabla}_m\xi^a$, which must be retained when and only when $n = 1$. Note that the left side of (17) is effectively the Lie derivative within the fibre, since κ^α is vertical. For condition ii), first take that part of Eqn. (12) of highest-order (namely, $(n+1)$) in ξ^a-derivatives. There results $\gamma^{dA}{}_m = k^{Ad}{}_\alpha\delta^{\alpha a\cdots c}{}_s\tilde{\nabla}_m\tilde{\nabla}_a\cdots\tilde{\nabla}_c\xi^s$. [The highest-order term on the left of (12) is that arising from $k^{Aa}{}_\alpha\tilde{\nabla}_\beta\xi^\alpha$.] Now substitute this $\gamma^{aA}{}_m$ into Eqn. (13), and again take the part highest-order (namely, $(n+1)$) in ξ^m-derivatives. [Neither the left side of (13), nor the first term on the right, contribute at all, for they both involve terms of order at most n. Thus, it is only the $\gamma^{aA}{}_m$-term that contributes at this order.]

I do not believe that there are any further simple conditions on $\delta^{\alpha a\cdots c}{}_m$ that follow from the demand that ξ^α be a gauge vector field[10]. On the other hand, it appears that any sort of converse of the above (i.e., any result to the effect that every $\delta^{\alpha a\cdots c}{}_m$ satisfying the two conditions above, together, possibly, with some further conditions, leads, via (16) to a gauge vector field) is likely to be extremely complicated. The problem, of course, is that the demand that the ξ^α given by (16) be a gauge vector field imposes conditions on all the lower-order terms on the right. These terms will already be awkward (since they will be derivative-operator dependent), and the conditions that must be imposed on them will surely be complicated.

I am not aware of any example of a system, (1), of first-order, quasilinear partial differential equations, having diffeomorphism gauge freedom, for which the corresponding gauge vector fields, ξ^α cannot be expressed as the result of applying a suitable differential operator to ξ^a, in the manner of (16).

Let us now return to the example of general relativity. Recall that the fields in this case consist of a Lorentz-signature metric, g_{ab}, together with a (torsion-free) derivative operator, ∇_a; and that the equations are (4)-(6). Fix a smooth vector field, ξ^a, on M. Then the corresponding gauge vector field is given by

$$\xi^\alpha = (2g_{m(b}\nabla_{a)}\xi^m, \nabla_{(a}\nabla_{b)}\xi^c) + \cdots. \tag{19}$$

The two components of the first term on the right reflect the behavior of

[10] A further condition we might reasonably impose (but won't, becuse it is not needed for what follows) is that the differential operator in (16) commute with taking the Lie bracket, i.e., that $[\xi^\alpha(\xi^a), \xi^\beta(\xi'^b)] = \xi^\gamma([\xi^a, \xi'^b])$. Note that the drop of this equation is automatic. In the case $n > 1$, this equation implies: For any two tensor fields, $L_{a\cdots c}{}^m$ and $L'_{a\cdots c}{}^m$ on M, the vertical vector fields that result from contracting these with $\delta^{\alpha a\cdots c}{}_m$ commute.

the space-time metric and of the derivative operator, respectively, under diffeomorphisms. This term is vertical (as, as we have seen, it must be). The remaining terms on the right are lower-order[11] in derivatives of ξ^a, and are not all vertical. Thus, the order of the differential operator relating ξ^α and ξ^a in (16), is, in this example, $n = 2$; and the tensor $\delta^{\alpha b \cdots c}{}_m$ in (16) is given by

$$\delta^{\alpha a \cdots b}{}_m = (0, \delta^{(a}{}_{(r}\delta^{b)}{}_{s)}\delta^p{}_m). \tag{20}$$

The two expressions on the right in this equation correspond, respectively, to the g-component and the ∇-component inherent in the index "α". One can check (with a little algebra) that this $\delta^{\alpha ab}{}_m$ does indeed satisfy Eqn. (18).

It appears to be difficult to find a tractable criterion that decides, given a general system (1) of partial differential equations, whether or not that system has diffeomorphism gauge freedom. For common examples (such as Einstein system, above), all the fields are "geometrical objects", i.e., are fields on which the action of diffeomorphisms has been pre-specified. In this case, there is a simple intuitive criterion: Diffeomorphisms act as gauge when and only when all fields are "dynamical", i.e., all are included within the fibres of the bundle $\mathcal{B}$. But in a general case — with the bundle $\mathcal{B}$ and the system (1) specified in some less concrete way — the situation may not be so clear. In this connection, it would be useful, at least as a first step, to be able to find, given the general system (1), those tensors $\delta^{\alpha b \cdots c}{}_m$ satisfying the two conditions given above. From simple examples, these conditions appear to be rather stringent, i.e., there doesn't appear to be an excessive number of solutions. Eqn. (18) is purely algebraic, and so it may be possible to "solve" it. Note that every solution, $\delta^{\alpha b \cdots c}{}_m$, of Eqn. (18) gives rise to a whole class of solutions namely, those given by $\delta'^{\alpha a \cdots e}{}_m = \delta^{\alpha (a \cdots c}{}_n \lambda^{d \cdots e) n}{}_m$, where $\lambda^{d \cdots en}{}_m = \lambda^{(d \cdots e) n}{}_m$, but is otherwise arbitrary. In this way, one can easily raise (or, at minimum, keep the same) the order of δ. It is typically the lowest-order solutions of Eqn. (18) that are of interest. With respect to condition i), fix any point of the base manifold M; and restrict to order $n > 1$. Then the vertical vector fields κ^α on the fibre over this point, satisfying

[11] There is, unfortunately, a confusing technical issue regarding (19). The term given explicitly on the right is *not* a vector field on $\mathcal{B}$, for it involves not only a point, (x, g_{ab}, ∇_a), of $\mathcal{B}$, but also the "derivative of ∇_a" there. For this term to make sense requires, if you like, a cross-section. However, this dependence on the derivative of ∇_a is lower-order in ξ^a, and is compensated for by the remaining terms in (19) (which aren't vector fields either). The final result is, indeed, a vector field on $\mathcal{B}$.

$\mathcal{L}_\kappa k^{Aa}{}_{\alpha'} = S^A{}_B k^{aA}{}_{\alpha'}$ for some $S^A{}_B$, form a vector space. [Essentially, this equation asserts that the Lie derivative, by κ^α, of each of a certain collection of covector fields in the fibre, yields again a certain linear combination of those covectors.] This vector space is "usually" finite-dimensional. It is possible that finite-dimensionality follows already from the conditions of the following section. In any case, fixing now a point of M, $\delta^{\alpha a \cdots c}{}_m$ in the fibre over this point can now be regarded as a linear map from the vector space of tensors $S_{a \cdots c}{}^m$ in M at this point to the vector space such vector fields κ^α on the fibre.

As an example of the usefulness of the differential operator, (16), we now establish the following result: Given any the system (1) of equations that admits diffeomorphism gauge freedom via (16), then that system cannot have a hyperbolization. Intuitively, we might have expected such a result, for the diffeomorphism gauge freedom could be invoked to change any given solution in a region away from some initial surface T while leaving that solution unchanged on T itself. This possibility would seem to be inconsistent with the presence of an initial-value formulation. The proof of this result is quite simple. Let, $\delta^{\alpha a \cdots c}{}_m$ yield the diffeomorphism gauge freedom, via (16); and suppose, for contradiction, that there also exists a hyperbolization, $h_{A\alpha}$. Let, at a point, n_a be such that the quadratic form (10) is positive-definite. Consider the expression

$$n_r \cdots n_s \delta^{\alpha r \cdots s}{}_m [n_a h_{A\alpha} k^{Aa}{}_\beta] n_p \cdots n_q \delta^{\beta p \cdots q}{}_n. \tag{21}$$

On the one hand, this expression must vanish, by (18), since the n's enforce symmetrization over the contravariant indices $a, p, \cdots q$. But, on the other hand, the tensor in square brackets is positive-definite. We conclude that $n_p \cdots n_q \delta^{\beta p \cdots q}{}_n = 0$. But this must hold for every n_a in an open set, and so we have that $\delta^{\beta p \cdots q}{}_n = 0$.

So, diffeomorphism gauge freedom indeed precludes an initial-value formulation. But other types of gauge freedom need not. For example, the system of Maxwell's equations with zero sources, (2)-(3), admits gauge transformations associated with duality and scaling, and yet this system has an initial-value formulation.

4 Initial-Value Formulation for Systems with Gauge

Consider a first-order, quasilinear system, (1), of partial differential equations. Let the constraints of this system be integrable and complete. Suppose further that this system manifests some nonzero group of gauge transformations. Then this system as it stands may have no initial-value formulation, for, while we have required integrability and completeness, we have not required the existence of a hyperbolization. Indeed, as we have just seen, a large degree of gauge freedom (such as that for diffeomorphism gauge) is typically incompatible with a hyperbolization. Nevertheless, it may be possible, for certain such systems, to recover an "effective" initial-value formulation. Consider initial data for such a system, consisting of a submanifold T of M of codimension one and a cross-section, ϕ_o, of our bundle over T, subject to the consistency conditions, (7), on those data that flow from the constraints. Roughly speaking, this system has an initial-value formulation "up to gauge" provided the following holds: Given such initial data, subject to some further inequality to the effect that (T, ϕ_o) is "non-characteristic", then there exists a solution of the system (1) manifesting that initial data, and that solution is unique up to gauge transformations[12]. Clearly, this is the closest one could reasonably expect to an initial-value formulation, in the presence of gauge freedom.

Gauge Conditions

There is a standard technique for demonstrating that certain classes of systems of partial differential equations have an initial-value formulation up to gauge, in the sense described above. This technique involves introducing what are called "gauge conditions" — certain additional equations imposed on the fields of the system. These additional equations can be purely algebraic (i.e., requiring passage to a subbundle of the original bundle $\mathcal{B}$),

[12] In more detail, we require, for existence, that there exist a solution in some neighborhood of T. For uniqueness, we require that, given two such solutions in neighborhoods, there exists a gauge transformation that leaves the initial data (T, ϕ_o) invariant (and so, in particular, that leaves the submanifold T pointwise invariant), and that sends the first solution to one that coincides, in some neighborhood of T, with the second.

purely differential (i.e., requiring a simple enlargement of the system (1)) or some combination of the two. The idea is to choose these equations such that, at least locally, they have the following two properties: i) Given any solution of the original system, (1), these additional equations can always be achieved by some suitable gauge transformation; and ii) the system that results from combining the original system, (1), with these additional equations *does* have an initial-value formulation in the traditional sense, i.e., that described in Sect. 2. Note that these additional equations cannot themselves be gauge-invariant: Indeed, if they were, then *neither* of the two conditions above could hold. The gauge transformation whose existence is demanded by property i) need not be unique. The initial-value formulation demanded by property ii) can be achieved in a variety of ways. For example, the additional equations can, among other things, create or destroy constraints, turn some of the original differential equations in (1) into identities, or cause the appearance of a hyperbolization when there was none before. Clearly, if we can manage to find gauge conditions having the two properties above, then our system does have, in the sense described above, an initial-value formulation up to gauge. We should emphasize, however, that such gauge conditions are by no means necessary. There could well be a system of partial differential equations that has an initial-value formulation up to gauge, and yet for which there exist no suitable gauge conditions whatever. Here are two examples of gauge conditions.

Consider the Maxwell system, with fields F_{ab}, A_a, and equations (2), (3), and (14). The gauge transformations of interest here are given by addition to A_a of the gradient of a smooth scalar field on M. Consider the gauge condition given by

$$\nabla^a A_a = 0. \tag{22}$$

This equation satisfies the two properties given above. Indeed, given any solution of (2), (3), and (14), the gauge condition (22) can always by achieved by some transformation. [The gauge scalar field must be chosen to satisfy the wave equation with a suitable source.] Furthermore, the system that results from combining Eqns. (2), (3), and (14) with (22) does indeed have an initial-value formulation. [The vector potential, A_a, now satisfies the wave equation.] So, Eqn. (22) is a suitable gauge condition: It yields, for this Maxwell system, an initial-value formulation up to gauge. This gauge condition is purely differential. The gauge transformation to achieve (22) is never unique. Inclusion of Eqn. (22) with Eqns. (2), (3), and (14) neither creates nor destroys constraints; nor does it render any of the original equations identities. However, (22) does give rise to a hyperbolization where there was none before. Eqn. (22) is, of course, the familiar Lorentz gauge condition in Maxwell theory.

As a second example, consider the Einstein system, given by (4)-(6). The gauge transformations of interest in this case are the M-diffeomorphisms. Fix any totally symmetric tensor field $W^{abc} = W^{(abc)}$ on M such that at each point of M there exists a covector n_a (and, hence, an open set of n_a) such that the combination $n_a W^{abc}$ is positive-definite. Now consider the following gauge conditions:

$$W^{abc} g_{bc} = 0, \tag{23}$$

$$(\nabla_m W^{abc}) g_{bc} = 0. \tag{24}$$

Note that these equations are purely *algebraic* in the fields: Eqn. (23) is algebraic in the metric g_{ab} alone; while (24) is algebraic in (g_{ab}, ∇_a). Eqn. (24) is, of course, merely the result of combining Eqns. (23) and (4). To understand what Eqn (23) means, consider a particular choice of W^{abc}, given by $W^{abc} = u^{(a} h^{bc)}$, where u^a is any nowhere-vanishing vector field, and h^{bc} any positive-definite metric. This choice indeed satisfies the positive-definiteness condition above, e.g., for $n_a = u^b h_{ab}$, where h_{ab} denotes the inverse of h^{ab}. For this particular W^{abc}, Eqn. (23) requires that $g_{ab} u^b$ be a certain multiple of the fixed vector $u^b h_{ab}$. Think of u^b as a "time-direction". Then Eqn. (23) fixes the "time-time" and "time-space" components of the metric g_{ab}. Thus, in this special case, the gauge condition (23) is a version of the familiar lapse-shift gauge in the traditional treatment (ref) of the initial-value formulation for general relativity.

There are many other possible choices for the W^{abc} above. Another choice, for instance, is $W^{abc} = u^{(a} h^{bc)}$, where h^{ab} is Lorentz-signature and u^a is h-timelike. But note that we cannot choose $h^{bc} = g^{bc}$, for we did not permit W^{abc} to depend on the space-time metric. In fact, what follows is actually true for a class of gauge conditions more general than that given above. This class is described as follows. Fix any vector-valued function, v^a, which depends, at each point of M, algebraically on the value there of the space-time metric g_{ab} and of some additional tensor fields. Let this function be such that $\partial(v^a)/\partial(g_{bc}) = W^{abc}$ is totally symmetric, and satisfies the positive-definiteness condition above. The new gauge conditions are now those that result from replacing Eqn. (23) by the equation $v^a = 0$, and retaining Eqn. (24) as given. This is indeed a generalization, for the original formulation (23)-(24), arises as the special case $v^a = W^{abc} g_{bc}$.

We turn now to the issue of whether, for the Einstein system, Eqns. (23)-(24) qualify as a set of gauge conditions, in the sense above.

Can Eqns. (23)-(24) be achieved, at least locally, via a gauge transformation? Fix a solution, (g_{ab}, ∇_a), of the Einstein system. Then, by virtue of Eqn. (4) of that system, it suffices to achieve only (23), for Eqn. (24) then follows. The statement that a diffeomorphism, ψ, on M achieve (23) is

$$W^{abc} (\nabla\psi)_b{}^{d'} (\nabla\psi)_c{}^{e'} g_{d'e'} = 0, \tag{25}$$

where primed indices refer to the image point, $x' = \psi(x)$, and $g_{d'e'}$ denotes the metric evaluated at this point. We may understand this equation in the following manner. Consider the fibre bundle over M whose fibre, over each point $x \in M$, is a copy of M itself. Then the diffeomorphism ψ can be regarded as a cross-section of this bundle; whence (25) becomes a first-order equation on this cross-section. But this equation is not even quasilinear! Consider, however, its linearized version. [See Appendix B.] In this version, the diffeomorphism ψ is replaced by its generator, vector field ξ^a on M; and Eqn. (25) is replaced by an equation on this ξ^a, with principal part $W^{abc} \nabla_b (g_{cd} \xi^d)$. But this linearized equation *does* admit an initial-value formulation: Its only constraint is zero, and it admits a hyperbolization, by virtue of precisely the conditions imposed above on W^{abc}. Thus, while we cannot guarantee solutions of the full system (25), we can guarantee solutions of its linearized version. This behavior — a first-order system of partial differential equations that is not even quasilinear, whose corresponding linearized system is not only quasilinear but actually has an initial-value formulation — seems surprising. It seems likely that, at least in the present case, the initial-value formulation for the linearized system will imply appropriate solutions also of Eqn. (25). It would be of interest to try to prove this — either for the present case, or more generally. A possible method might be first to choose a diffeomorphism ψ that makes the right side of Eqn. (25) "small", and then perform a sequence of corrections via generators ξ^a of infinitesimal diffeomorphisms.

Note that for the Einsein system, in contrast to the Maxwell example above, we only expect to be able to satisfy the gauge conditions via a gauge transformation when the original cross-section ϕ satisfies the field equations, (4)-(6).

Does the system whose fields consist of (g_{ab}, ∇_a) subject to the algebraic conditions (23)-(24), and whose equations consist of (4)-(6), have an initial-value formulation? The constraints of this system, as is not hard to check, are integrable and complete. But does this system admit a hyperbolization?

In order to answer this question, let us return briefly to the original Einstein system, i.e., that without (23)-(24). For this system, the dimension of the space of equations (4)-(6) is 110 (= 40 + 60 + 10), while the dimension of the space of consistency conditions (7) is 64 (= 30 + 30 + 4). This leaves 46 (= 110 - 64) dynamical equations. But the space of independent variables has dimension 50 (= 10 + 40). This excess — four more independent variables than equations — reflects the diffeomorphism gauge freedom inherent in the Einstein system. Let us now see how this arithmetic is affected by the imposition of the gauge equations (23)-(24). The dimension of the space of equations is reduced by the gauge equations by 40 (= 16 (contraction of (4) with W^{bcd}) + 24 (contraction of (5) with W^{cdn})). The dimension of the space of consistency conditions is reduced by 24 (= 12 (from (4)) + 12 (from (5))). Hence the dynamical equations are reduced in dimension by 16 (= 40 - 24). But the independent variables are reduced in dimension by 20 (= 4 (from (23)) + 16 (from (24))). Thus, imposition of the gauge equations (23)-(24) reduces dimension of the dependent variables by four more than it reduces that of the dynamical equations. In other words, in the Einstein system, supplemented with the algebraic equations (23)-(24), the number of dependent variables precisely matches the number of dynamical equations. What this means is that, for this system, there are equations that give the "time-derivatives" of all the fields, in terms of their values and space-derivatives. Naively, then, one would expect an initial-value formulation for this system.

But is there really such a formulation, i.e., does there actually exist a hyperbolization for this system? This is, apparently, an open question. But it seems extremely likely that there exists none. Here is a possible line for a proof. Suppose for a moment that we replace Eqns. (5)-(6) by the single equation ${R_{abc}}^d = 0$ That is, we replace the Einstein system by that of "special relativity". This replacement merely adds to the Einstein system some additional constraints, from which it follows that every hyperbolization (if any) of the Einstein system must also be a hyperbolization of this new system. In this new system, the constraints continue to be integrable and complete, and, just as above, there is a matching between dimensions of the space of dynamical equations and dependent variables. However, in this case, there *is* a hyperbolization. The corresponding quadratic form is

$$n_a[W^{abc} g_{bm} g_{cn} s^m{}_{rs} s'^n{}_{pq} + u^a \delta g_{rs} \delta' g_{pq}] \; V^{rspq} \tag{26}$$

where $V^{rspq} = V^{(pq)(rs)}$ is any tensor positive-definite in its two symmetric index-pairs and u^a is any timelike vector. It now suffices to prove two assertions: i) that the expression (26) represents the *only* hyperbolization for this new system, and ii) that the expression (26) is not a hyperbolization of the constrained Einstein system. Both of these assertions seem plausible. It would be of interest to prove, by this method or otherwise, that this constrained Einstein system, (4)-(6), (23)-(24), has no initial-value formulation.

These two examples illustrate the point that imposing gauge conditions on systems of partial differential equations with gauge freedom is potentially a complicated business. The basic problem is that there is so much variety: The gauge conditions themselves can be algebraic or differential; these conditions can be achieved by a gauge transformation in a variety of ways; and these conditions can restore to the system an initial-value formulation through a variety of mechanisms. It appears to be necessary to deal with systems of equations on a case by case basis.

We shall now introduce an alternative method for determining that a given system of first-order, quasilinear partial differential equations has an initial-value formulation up to gauge. Our method is systematic, relatively simple (although not as simple as one might have hoped), and manifestly gauge-invariant.

Fix a first-order, quasilinear system, (1) of partial differential equations, and let this system have some gauge group. The idea is to introduce certain additional geometrical structure on the bundle manifold $\mathcal{B}$, this structure consisting, at each point of $\mathcal{B}$, of a vertical flat σ, i.e., of a subspace of the vector space of vertical vectors at that point of $\mathcal{B}$. We demand that this field of flats be smooth in its dependence on point of $\mathcal{B}$. We further demand that this field be integrable, so the integral surfaces of σ give a foliation of each fibre by submanifolds; and each of these submanifolds has, for its tangent space at each of its points, precisely the flat σ at that point. In all the examples of which I am aware, these flats are also invariant under the action of gauge transformations on the system (1), although this invariance is not actually needed in what follows.

Think of the flat σ, at a fixed point of $\mathcal{B}$, as representing "physical directions" in field-space. Then the vertical vectors not lying in this flat are to be thought of as having "unphysical" (i.e., gauge) components. A crucial point here is that we do not attempt to single out any particular complementary subspace, i.e. any space of specifically "unphysical directions". In the case in which the gauge group consists only of the identity (as well as in certain other cases with small gauge group), the flat σ, at each point of $\mathcal{B}$, will consist of *all* vertical vectors at that point, whence the integral surfaces of the flats will be precisely the fibres of the fibre bundle $\mathcal{B}$.

Existence

Recall that a first-order, quasilinear system of partial differential equations has (subject to some resolution of the gap discussed in Appendix A) an initial-value formulation provided the following four conditions are satisfied: i) its constraints are integrable, ii) its constraints are complete, iii) it admits a hyperbolization, and iv) some single open set of covectors n_a suffices for both completeness and the hyperbolization. The idea is to demonstrate existence of an initial-value formulation up to gauge (subject to the same gap) by suitably modifying, taking into account the flats σ, these four conditions.

The first condition, integrability of the constraints, remains unchanged. For the second condition, recall, from Sect. 2, that completeness means that, for every n_a in some open set of covectors, whenever ν_A is such that $\nu_A n_a k^{Aa}{}_\alpha$ annihilates all vertical vectors, then $\nu_A = c_A{}^a n_a$ for some constraint $c_A{}^a$. We now replace this second condition with what we call *σ-completeness*: For every such n_a, whenever $\nu_A n_a k^{Aa}{}_\alpha$ annihilates all vectors in the flat σ, then $\nu_A = c_A{}^a n_a$ for some constraint $c_A{}^a$. Note that σ-complete implies complete; and, furthermore, that σ-completeness follows from completeness, together with the property: If $\nu_A n_a k^{Aa}{}_\alpha$ annihilates all vectors in σ, then it annihilates all vertical vectors. The condition of σ-completeness means physically that none of the equations serves to restrict the "time-derivatives" *only* of unphysical degrees of freedom. The third condition is modified to the following. We demand that there exist, at each point of the bundle manifold $\mathcal{B}$, a *σ-hyperbolization*, i.e., a tensor $h_{A\beta}$ at that point such that the combination $n_a h_{A\beta} k^{Aa}{}_\alpha$, applied only to vectors in σ, is symmetric, and, for all n_a lying in some open set, is positive-definite. This is precisely the same as the original definition of a hyperbolization, except that now the quadratic form is restricted to σ, i.e., to "physical degrees of freedom". In particular, every hyperbolization is a σ-hyperbolization. Finally, the fourth condition remains unchanged. Thus, what we have done here is expand the notion of completeness, and contract that of a hyperbolization, to take account of the flats σ.

We shall show shortly that, under these four conditions, as broadened in the paragraph above, existence holds for the system (1) of partial differential equations. But first, we give two examples.

Consider first the Maxwell system with vector potential — the system with equations (2), (3), and (14) and gauge the addition to A_a of the gradient of a scalar field — in a general curved spacetime. Now fix any smooth, nowhere vanishing timelike vector field u^a on the manifold M. Let, at each point of $\mathcal{B}$, the flat σ at that point consist of all vertical tangent vectors, $(\delta F_{ab}, \delta A_a)$, at that point satisfying $u^a \delta A_a = 0$. Thus, the flats in this case are nine-dimensional (in ten-dimensional fibres). These flats are smooth and integrable, the integral surfaces being those of constant $u^a A_a$. Furthermore, these flats are gauge-invariant, for the action of a gauge transformation, within a fixed fibre, is to send the point (F_{ab}, A_a) to the point $(F_{ab}, A_a + w_a)$, where w_a is some fixed vector (which, however, is different for different gauge transformations). But clearly these leave the integral surfaces surfaces above, and so the flats σ, invariant.

We now claim that this Maxwell system, with these flats σ, satisfies the four conditions given above. Indeed, the first condition, integrability, was already shown in Sect. 2; as was completeness of the system (2)-(3). Hence, the second condition, σ-completeness, (with the open set of covectors n_a consisting of the timelike ones): Given any tensor $\nu^{ab} = \nu^{[ab]}$ and any timelike n_a, such that $\nu^{ab} n_a \delta A_b = 0$ for all δA_b satisfying $u^b \delta A_b = 0$, then we have $\nu^{ab} n_a = 0$. But this assertion is true, for the vanishing of $\nu^{ab} n_a \delta A_b$ for all δA_b with $u^b \delta A_b = 0$ implies that $\nu^{ab} n_a$ is a multiple of u^b; whence, contracting with n_b and using the timelike character, that $\nu^{ab} n_a = 0$. For the third condition, existence of a σ-hyperbolization, consider the quadratic form given by the usual Maxwell one, (11), on the Maxwell field, plus the additional term $n_{[a} \delta A_{b]} u^a g^{bc} \delta' A_c$ This expression i) arises from (14) (for it is the contraction of $n_{[a} \delta A_{b]}$ with something); and, with δA_b and $\delta' A_c$ restricted by $u^b \delta A_b = u^c \delta' A_c = 0$, is ii) symmetric; and iii) positive-definite, whenever $u^a n_a > 0$. The fourth condition is immediate (the appropriate open set of n_a consisting, say, of the timelike ones with $u^a n_a > 0$).

Consider, as a second example, the Einstein system, (4)-(6) with diffeomorphism gauge. Recall that, at a general point, (g_{ab}, ∇_a), of the fibre over base point $x \in M$, vertical tangent vectors are represented by $(\delta g_{ab}, s^m{}_{ab})$, where $s^m{}_{ab} = s^m{}_{(ab)}$ represents the "change in the derivative operator". Let the flat σ at that point consist of those vertical vectors satisfying $g^{ab}s^m{}_{ab} = 0$. Thus, the flats in this case are forty-six-dimensional (in fifty-dimensional fibres). These flats are smooth and integrable. To see what the integral surfaces are, fix any derivative operator, $\tilde{\nabla}_a$, at x, and represent ∇_a there by the tensor, $C^m{}_{ab}$, expressing the difference between ∇_a and $\tilde{\nabla})_a$. The integral surfaces of the flats σ within this fibre are now given by the surfaces of constant $g^{ab}C^m{}_{ab}$. [Note that these surfaces are independent of the choice of the auxiliary operator $\tilde{\nabla}_a$. However, which constant vector represents which surface will, of course, depend on this choice.] That these flats are gauge-invariant follows from the fact that they are defined without reference to any "external objects".

We now claim that this Einstein system, with these flats σ, satisfies the four conditions given above. The first condition, integrability, was already shown in Sect. 2; as was completeness of the system (4)-(6). Hence, the second condition, σ-completeness, requires (with the open set of covectors n_a consisting of the timelike ones): Given any tensors $\nu^{abcd} = \nu^{[ab](cd)}$ and $\nu^{ab} = \nu^{(ab)}$ and any timelike n_a such that

$$\nu^{abcd} n_a s^m{}_{bc} g_{md} + 1/2 \nu^{ab}(n_m s^m{}_{ab} - n_b s^m{}_{am}) = 0 \tag{27}$$

for all $s^m{}_{ab} = s^m{}_{(ab)}$ with $g^{ab}s^m{}_{ab} = 0$, then Eqn. (27) holds for *all* $s^m{}_{ab} = s^m{}_{(ab)}$. [Here, the left side of Eqn. (27) is the equation that results from replacing the derivative of ∇ by $n_a s^b{}_{cd}$ in Eqns. (5) and (6), contracting with ν^{abcd} and ν^{ab}, respectively, and adding.] To see that this assertion is true, note that the hypothesis implies that $n_r[\nu^{r(ab)s}g_{sm} + 1/2\delta^r{}_m \nu^{ab} - 1/2\delta^{(a}{}_m \nu^{b)r}]$ is a multiple of g^{ab}. Contracting with $n_a n_b$ and using the timelike character, we conclude that that multiple is zero. But this is precisely the statement that Eqn. (27) holds for all $s^m{}_{ab} = s^m{}_{(ab)}$. For the third condition, the existence of a σ-hyperbolization, we proceed as follows. Let $s^m{}_{ab} = s^m{}_{(ab)}$ and $s'^m{}_{ab} = s'^m{}_{(ab)}$, be any two tensors, and n_a and u^a any two vectors. We then have the following identity, whose proof is straightforward but a little tedious:

$$s'_{(cd)n} u^n \Sigma(ab)[-2n_{[m} s^m{}_{a]b} + 4n_{[q} s^m{}_{b](a} g_{p)m} g^{pq}] - s'_{(cd)}{}^m u^n [s_{(ab)[m} n_{n]}] \tag{28}$$

$$= s_{(ab)m}s'_{(cd)n}[-(u^p n_p)g^{mn} + 2u^{(m}n^{n)}] - n_{(a}s_{b)m}{}^m s'_{(cd)n}u^n, \tag{29}$$

where we have freely raised and lowered indices with the metric g_{ab}, and where the symbol "$\Sigma(ab)$" is the instruction to symmetrize over the indices a and b. Now, the expressions in square brackets on the left involve only the combinations that result from the equations, (5)-(6), of the Einstein system by replacing the derivative of ∇ by $n_a s^b{}_{cd}$. The last term on the right vanishes for $s^m{}_{ab}$ lying in the flat σ, for this implies $s_{bm}{}^m = 0$. Now fix timelike u^a, and let n_a also be timelike, with $u^a n_a > 0$. Then the right side, contracted with $(g^{ac} + 2u^a u^c/(u^m u_m))(g^{bd} + 2u^b u^d/(u^m u_m))$, is the required quadratic form for the σ-hyperbolization. Finally, the fourth condition is again immediate, with, again, the open set of n_a consisting of the timelike ones with $u^a n_a > 0$. Thus, we have verified that these flats σ for the Einstein system satisfy our four conditions above. It is perhaps a reasonable conjecture that these flats are the unique ones for the Einstein system satisfying our conditions. It would be most interesting to prove this.

We now return to the general case. Why do we not not further require that the hyperbolization $h_{A\alpha}$ of the third condition also be invariant under gauge transformations? We would if we could, but, unfortunately, such invariance cannot be achieved in examples (e.g., the Einstein system). Note that the conditions above refer *only* to the flats σ, and not at all to gauge transformations. The gauge transformations enter at this stage only indirectly, through our inability to find a hyperbolization that is symmetric when applied to *all* vertical vectors. Note, incidentally, σ-completeness favors larger flats σ, while existence of a σ-hyperbolization favors smaller. The final flats represent a compromise between these competing demands.

The condition of σ-completeness has several immediate, and very useful, consequences. Let us adopt the convention, for purposes of this paragraph that a σ, appended to any covariant Greek index, means "applied only to vectors in the flat σ". Then σ-completeness implies in particular: $n_a \nu_A k^{Aa}{}_{\alpha^\sigma} = 0$ implies $n_a \nu_A k^{Aa}{}_{\alpha'} = 0$. [Recall that a prime on a covariant Greek index means "applied only to vertical vectors".] The first consequence is this: Whenever $\nu_A k^{Aa}{}_{\alpha^\sigma} = 0$, then $\nu_A = 0$. [To see this, note that the hypothesis implies $\nu_A n_a k^{Aa}{}_{\alpha^\sigma} = 0$ for every n_a; and so, by σ-completeness, that $\nu_A n_a k^{Aa}{}_{\alpha'} = 0$ for every n_a in some open set; and so, that $\nu_A k^{Aa}{}_{\alpha'} = 0$; which in turn implies $\nu_A = 0$.] What this means is that no equation of the system has the property that, when restricted to physical degrees of freedom, it becomes purely algebraic. We turn next to the constraints. First, note that, quite generally, every constraint of the system (1), i.e., every $c^a{}_A$ satisfying $c^{(a}{}_A k^{|A|b)}{}_{\alpha'} = 0$, is also automatically a σ-*constraint*, i.e., also satisfies $c^{(a}{}_A k^{|A|b)}{}_{\alpha^\sigma} = 0$. But σ-completeness implies the reverse, i.e., that *every* σ-constraint arises in this manner from some constraint of the original system. [Indeed, let $\hat{c}^{(a}{}_A k^{|A|b)}{}_{\alpha^\sigma} = 0$. Contract with $n_a n_b$ to obtain $(n_a \hat{c}^a{}_A) n_b k^{Ab}{}_{\alpha^\sigma} = 0$. Choosing n_a to lie in the open set, this condition now implies, by σ-completeness, that $(n_a \hat{c}^a{}_A) n_b k^{Ab}{}_{\alpha'} = 0$; which, since this now holds for all n_a in an open set, implies $(\hat{c}^{(a}{}_A) k^{|A|b)}{}_{\alpha'} = 0$.] In addition, again from σ-completeness, no nonzero constraint gives rise to the zero σ-constraint. To summarize, σ-completeness implies that the equations of the original system (1), as well as its constraints, go over, essentially unchanged in number and character, when restricted to the physical degrees of freedom.

Fix a first-order, quasilinear system of partial differential equations, together with a smooth, integrable field of flats σ, satisfying the four conditions (integrability, σ-completeness and existence of a σ-hyperbolization, for a common open set of n_a) above. There is a key result to the effect that this system must, for suitable initial data, manifest existence of solutions. This result, in more detail, is the following. Let (T, ϕ_o) be initial data for this system, so T is a submanifold of the base manifold M of codimension one, and ϕ_o is a cross-section of the bundle $\mathcal{B}$ over T. Let these data satisfy all the consistency conditions (7) arising from the constraints of the system. Let, furthermore, these data be non-characteristic, in the sense that the normal n_a to T at each point lies within the open set specified in the fourth condition. Then there exists a solution ϕ of the system (1), defined in a neighborhood of T, that reduces to the given data, ϕ_o, on T. The proof of this result will emerge from the discussion that follows.

Consider, in this system, a submanifold, $\tilde{\mathcal{B}}$, of $\mathcal{B}$ having the following two properties: i) $\tilde{\mathcal{B}}$ meets each fibre of $\mathcal{B}$ in a single integral surface of σ, and ii) the projection mapping $\tilde{\mathcal{B}} \xrightarrow{\pi} M$ is a surjection onto M, i.e., every tangent vector at every point of M is the image, under the derivative of this map, of some tangent vector at some point of $\tilde{\mathcal{B}}$. We remark that there are many such submanifolds. Indeed, fixing any cross-section, ϕ, of $\mathcal{B}$, the union of all the σ-integral surfaces that meet $\phi[M]$ is such a submanifold (and, in fact, the unique one containing the image of this cross-section). Furthermore, every such submanifold $\tilde{\mathcal{B}}$ arises in this way. Next, note that every such submanifold $\tilde{\mathcal{B}}$ is itself also a bundle over the same base manifold M. In this bundle, the fibre over $x \in M$ is the corresponding integral surface of σ in the $\mathcal{B}$-fibre over x. Note further that every cross-section of $\tilde{\mathcal{B}}$ is automatically also a cross-section of $\mathcal{B}$; and that every cross-section of $\mathcal{B}$ — provided it lies within the submanifold $\tilde{\mathcal{B}}$ — is also a cross-section of the bundle $\tilde{\mathcal{B}}$.

We now introduce a certain partial differential equation based on this subbundle $\tilde{\mathcal{B}} \xrightarrow{\pi} M$, namely, that which results from simply restricting our original equation, (1), to $\tilde{\mathcal{B}}$. That is, we regard a cross-section $\tilde{\phi}$, of $\tilde{\mathcal{B}}$ as a cross-section also of the original bundle $\mathcal{B}$, and, as such, impose on it Eqn. (1). The result is a first-order, quaslilinear system of partial differential equations, based on the bundle $\tilde{\mathcal{B}}$. The "k" for this new system is simply the restriction of the original $k^{Aa}{}_{\alpha}$ to σ, and the "j" the original j^A; where both of these are now evaluted on $\tilde{\mathcal{B}}$ (since that is where the cross-section $\tilde{\phi}$ lives). We note that, by the discussion of σ-completeness above, none of the partial differential equations, (1), of the original system are "lost" (i.e., become algebraic equations) on restriction to $\tilde{\mathcal{B}}$. Hence, every solution of the restricted equation, in $\tilde{\mathcal{B}}$, is also, when regarded as a cross-section in $\mathcal{B}$, a solution of (1). We now claim that this new system of partial differential equations satisfies the conditions of Sect. 2 for having an initial-value formulation. Indeed, this is immediate. Integrability passes from the original system, (1), to the new one because the two systems have the same constraints; σ-completeness and the σ-hyperbolization yield completeness and a hyperbolization for the new system.

Now consider our given initial data (T, ϕ_o), satisfying the consistency conditions (7), and such that the normal, n_a, to T, at each of its points, lies in the open set of given above. Extend this cross-section ϕ_o over T to *any* cross-section over the entire base manifold M; and then expand this cross-section to a submanifold $\tilde{\mathcal{B}}$ of $\mathcal{B}$, in the manner described above. There results a subbundle, $\tilde{\mathcal{B}} \xrightarrow{\pi} M$, of our original bundle, $\mathcal{B} \xrightarrow{\pi} M$, and, as we have seen above, this subbundle inherits, from (1), its own first-order, quasilinear partial differential equation. Now, the original initial data, (T, ϕ_o), for (1) is (since $\phi_0[T] \subset \tilde{\mathcal{B}}$, by construction) also initial data for this new system. Furthermore, these initial data on $\tilde{\mathcal{B}}$ also satisfy the consistency conditions there (since the original initial data satisfied the consistency conditions for (1), and since the two systems have precisely the same constraints). Hence, since the equation in $\tilde{\mathcal{B}}$ has an initial-value formulation, there exists a unique solution, $\tilde{\phi}$, of the $\tilde{\mathcal{B}}$-equation manifesting these initial data. But, as we remarked above, this $\tilde{\phi}$ is also a solution of the original system, (1), in $\mathcal{B}$. In short, the original initial data, (T, ϕ_o), in $\mathcal{B}$ admit a solution of (1).

This completes the proof of existence of solutions for systems of partial differential equations satisfying the four conditions above. Note that we do not in general have uniqueness, because we had the freedom, in the argument above, in the choice of the subbundle $\tilde{\mathcal{B}}$. Note also that the gauge transformations themselves nowhere entered the discussion above (at least not directly: They did enter indirectly through the flats σ).

Uniqueness

Under certain further conditions, the solution obtained above will be unique, up to gauge. To make things more explicit, let us restrict consideration, for the moment, to the case in which the gauge is that of diffeomorphisms. What follows can be generalized to other gauge groups. In any case, we have, under this assumption, the tensor $\delta^{\alpha a \cdots c}{}_m$ (with n indices $a \cdots c$) of Eqn. (16), which describes the lifting of any smooth vector field ξ^a on M to a gauge vector field ξ^α on $\mathcal{B}$.

The idea is to consider two extensions of the given cross-section ϕ_o over T to subbundles $\tilde{\mathcal{B}}$ of $\mathcal{B}$. We must show that there exists a gauge transformation sends one such subbundle to the other. In more detail, we shall rewrite "sends one such subbundle to the other" as a system of partial differential equations on the gauge transformation itself. This system will turn out to be n^{th} order, with principal part $\delta^{\alpha a \cdots c}{}_m$. We will then demand that this system have an initial-value formulation. This system, when rendered first-order, will automatically have its constraints integrable (as systems usually do!). But we shall have to impose on it (as appropriate conditions on $\delta^{\alpha a \cdots c}{}_m$) completeness and existence of a hyperbolization. These conditions, in more detail, are the following.

First, we shall demand that, for every covector n_a in our open set, we have: Given any ν_α that annihilates all vectors in σ and satisfies $\nu_\alpha \delta^{\alpha a \cdots c}{}_m n_a \cdots n_c = 0$, then ν_α annihilates all vertical vectors. This is a kind of completeness, but now referring to the action of the gauge group[13]. It not only has a form similar to σ-completeness, but also has similar consequences. The following,

[13] In fact, it appears that, in practice, this condition may provide the simplest route to "finding", appropriate flats σ, given only some partial differential equation (1), and the action on the bundle $\mathcal{B}$ of the diffeomorphisms.

for example, are direct consequences of this condition (with proofs completely analogous to those for σ-completeness): i) For any ν_α that annihilates all vectors in σ and satisfies $\nu_\alpha \delta^{\alpha a \cdots c}{}_m = 0$, we have that this ν_α annihilates all vertical vectors; ii) For $c^a{}_\alpha$ annihilating all vectors in σ, and satisfying $c^{(a}{}_\alpha \delta^{|\alpha|c \cdots d)}{}_m = 0$, then $c^a{}_\alpha$ annihilates all vertical vectors.

Second, we shall demand that there exist a tensor $H_n{}^{d \cdots e}{}_\alpha = H_n{}^{(d \cdots e)}{}_\alpha$, vanishing when contracted with any $\mu^\alpha \in \sigma$, and a tensor $S_n{}^{d \cdots e a b \cdots c}{}_m = S_n{}^{(d \cdots e) a (b \cdots c)}{}_m$, with symmetries $S_n{}^{d \cdots e (a b \cdots c)}{}_m = 0$, such that the combination

$$H_n{}^{d \cdots e}{}_\alpha \delta^{\alpha a b \cdots c}{}_m + S_n{}^{d \cdots e a b \cdots c}{}_m \tag{30}$$

is symmetric under interchange of the index-pairs "$n\ d \cdots e$" and "$m\ b \cdots c$", and, contracted with any covector n_a lying in our open set, is positive-definite in these index-pairs. This H will turn out to be an effective hyperbolization for our system of equations on the gauge transformation.

As we shall see shortly, these two conditions taken together guarantee, in an appropriate sense, uniqueness of solutions up to gauge.

Let us return to the example of the Einstein system. Recall that the tensor $\delta^{\alpha a \cdots b}{}_m$ in this case has rank two, and is given by Eqn. (20). The flats are given by tangent vectors, $(\delta g_{ab}, s^m{}_{ab})$, satisfying $g^{ab} s^m{}_{ab} = 0$; and the open set of n_a is that consisting of the timelike covectors. This system, we claim, satisfies the two conditions above. The first asserts, in this example, the following. For every timelike n_a, if tensor $\nu^{ab}{}_m$ (i.e., ν_α) is such that $\nu^{ab}{}_m$ is a multiple of g^{ab} (i.e., $\nu_{\alpha^\sigma} = 0$) and $\nu^{ab}{}_m n_a n_b = 0$ (i.e., $\nu_\alpha \delta^{\alpha ab}{}_m n_a n_b = 0$), then $\nu^{ab}{}_m = 0$. But this assertion, clearly, is true. For the second condition, consider the choices

$$H_n{}^{drs}{}_m = -(g_{mn} + u_m u_n/(u^q u_q)) g^{rs} u^d, \tag{31}$$

$$S_n{}^{dab}{}_m = (g_{mn} + u_m u_n/(u^q u_a)) u^{[a} g^{b]d}, \tag{32}$$

where the index combination "mrs" in $H_n{}^{drs}{}_m$ stands for "α". These two clearly satisfy the symmetries given above. Furthermore, the quadratic form given by (30) becomes $(g_{mn} + u_m u_n/(u^q u_q))[u^a g^{db} - 2g^{a(d} u^{b)}]$, which has the requisite symmetry and positive-definite properties. Thus, the Einstein system, (4)-(6), satisfies our two conditions.

The key result is that the two conditions above imply uniqueness, up to gauge, of our solution of (1) with the given initial data (T, ϕ_o). The method, as discussed earlier, is to show that given any two extensions of the given cross-section, ϕ_o, over T to a subbundles, $\tilde{\mathcal{B}}$ and $\tilde{\mathcal{B}}'$, of $\mathcal{B}$, then these two subbundles are gauge-related.

It is convenient to begin with the "infinitesimal argument", which is somewhat easier conceptually. Fix any cross-section, ϕ, of the original bundle $\mathcal{B}$, and any vertical field, μ^α, defined at points of $\phi[M]$. Consider now the following partial differential equation

$$[\delta^{\alpha a\cdots c}{}_m \tilde{\nabla}_a \cdots \tilde{\nabla}_c \xi^m + \cdots] - (\nabla\phi)_a{}^\alpha \xi^a - \mu^\alpha \in \sigma \tag{33}$$

The term in square brackets on the left in Eqn. (33) is precisely the expression, (16) for the gauge vector field defined by ξ^a. [Here, $\tilde{\nabla}_a$ denote some arbitrary, fixed derivative operator on the manifold M.] The second term on the left in (33), which is just the lift of ξ^a to the cross-section, serves to make the sum of the first two terms vertical. That is, Eqn. (33) is the assertion that the difference between the vertical vector given by the first two terms on the left and the given vertical vector μ^α lies in the flat σ. This is a first-order, quasilinear system of partial differential equations on the vector field ξ^a on M (as is seen, e.g., by contracting (33) with covectors in the fibre orthogonal to σ). The geometrical meaning of this equation is the following. Think of the given cross-section, ϕ, as defining some subbundle, $\tilde{\mathcal{B}}$, of the original bundle $\mathcal{B}$. Think of the vertical vector μ^α as representing the vertical "connecting vector" from $\tilde{\mathcal{B}}$ to some nearby subbundle $\tilde{\mathcal{B}}'$. Then this μ^α is meaningful only up to addition of vertical vectors tangent to $\tilde{\mathcal{B}}'$, i.e., only up to vectors lying in the flat σ. [In other words, if we change μ^α by addition to it of a vector in σ, then this new μ^α connects to the same nearby subbundle $\tilde{\mathcal{B}}'$.] Eqn. (33), then, is the assertion that the vertical part of the gauge vector field generated by the vector field ξ^a on M send $\tilde{\mathcal{B}}$ to $\tilde{\mathcal{B}}'$.

Eqn. (33) is a system of n-th order, quasilinear partial differential equation. [The "vector space of equations" is the quotient space of the vector space of vertical vectors by the vector subspace σ.] We now claim: By virtue of the two conditions above, this system has an initial-value formulation. The first step is to convert (33) to a first-order system. To this end, we introduce, in addition to the original vector field ξ^m, tensor fields $\xi_a{}^m, \xi_{ab}{}^m, \cdots \xi_{a\cdots c}{}^m$, each totally symmetric in its covariant indices, and having covariant indices ranging in number from from one to $(n-1)$. On these fields, we now impose the following three sets of equations:

$$\tilde{\nabla}_a \xi^m = \xi_a{}^m, \cdots, \tilde{\nabla}_{(a} \xi_{b \cdots c)}{}^m = \xi_{a \cdots c}{}^m, \tag{34}$$

$$\tilde{\nabla}_{[a} \xi_{b]}{}^m = \cdots, \quad \cdots, \quad \tilde{\nabla}_{[a} \xi_{b] \cdots c} = \cdots, \tag{35}$$

$$[\delta^{\alpha a \cdots c}{}_m \tilde{\nabla}_a \xi_{b \cdots c}{}^m + \cdots] - (\nabla \phi)_a{}^\alpha \xi^a - \mu^\alpha \in \sigma. \tag{36}$$

Eqn. (34) consists of $(n-1)$ equations, which allow us to "interpret" each of $\xi_a{}^m, \cdots, \xi_{a \cdots c}{}^m$ as the symmetrized derivative of its predecessor, and so, ultimately, as a symmetrized derivative of ξ^m. Eqn. (35), consists of the $(n-1)$ equations resulting from taking the curl of each equation in (34). Thus, Eqns. (35) are the integrability conditions for Eqns. (34). Finally, Eqn. (36) is precisely Eqn. (33), rewritten in first-order form.

Note that (34)-(36) is indeed a first-order, quasilinear system of partial differential equations, on the n tensor fields $\xi^m, \xi_a{}^m, \cdots, \xi_{a \cdots c}{}^m$. In the case of Eqn. (36), the equation lies in the vector space given by the quotient of the space of vertical vectors by the subspace consisting of those vectors lying in flat σ. This system, we now claim, satisfies the conditions, of Sect 2, for having an initial-value formulation. To see this, consider first the equations consisting of all $(n-1)$ in (34) and all but the last in (35). These $(2n-3)$ equations express the derivatives of each of the the first $(n-1)$ of our fields (i.e., all but the last one, the $\xi_{a \cdots c}{}^m$ appearing in (36)) algebraically in terms of these fields. This system has an obvious initial-value formulation. [The constraints correspond to taking the curls of (34) and of (35), the corresponding integrability conditions being (35) and identities, respectively. A hyperbolization is immediate.] The critical field is the final one, $\xi_{a \cdots c}{}^m$, and the equations on it are the last equation in (35), and Eqn. (36). The first of these equations has a constraint corresponding to taking its curl. The corresponding integrability condition is an identity (by the choice of right-hand side of the last equation in (35)). A constraint for (36) is represented by a tensor $c^a{}_\alpha$ such that $c^a{}_{\alpha^\sigma} = 0$, and such that $c^{(a}{}_\alpha \delta^{|\alpha| b \cdots c)}{}_m = 0$. But our first condition above, completeness, implies that the only such $c^a{}_\alpha$ is zero. Thus, the equations on $\xi_{a \cdots c}{}^m$ are integrable and complete. But these equations also have a hyperbolization, namely the H of (30), for our second condition is precisely the statement that this H has the requisite properties. Thus, we have shown that the system (34)-(36) satisfies the conditions of Sect. 2 for an initial-value formulation.

Next, let there be given initial data, (T, ϕ_o), satisfying the consistency conditions (7) and with the normal n_a to T lying, at each of its points, in our open set. Let the given vertical field μ^α be chosen to vanish on (T, ϕ_o). Now choose, on T, initial data for the system (34)-(36) with all the fields $\xi^m, \cdots, \xi_{a\cdots c}{}^m$ vanishing there. These data clearly satisfy all the consistency conditions for the system (34)-(36). So, there exists a solution of this system. That is, there exists a vector field ξ^m on M whose corresponding gauge vector field, ξ^α, coincides with the given vertical μ^α on the subbundle $\tilde{\mathcal{B}}$, and leaves invariant the given initial data, (T, ϕ_o). We have shown, then, uniqueness of solutions up to "infinitesimal" gauge.

There follows immediately a corresponding result for the full gauge group. Consider the fibre bundle over M, whose fibre, over each point x of M, is a copy of M itself. Then a cross-section of this bundle is precisely a smooth map ψ from M to M. Now demand that the gauge transformation $\mathcal{B} \xrightarrow{\Psi} \mathcal{B}$ arising from this ψ send the submanifold $\tilde{\mathcal{B}}$ to $\tilde{\mathcal{B}}'$. This is an n-th order, quasilinear differential equation on the cross-section ψ, an equation having precisely the principal part of Eqn. (33). Just as with (33), we convert this to a first-order partial differential equation (so the fields will now be "point of M", together with certain tensors that can be interpreted as the first $(n-1)$ derivatives of the smooth map ψ). These equations will have precisely the character of the system (34)-(36), and for the same reason will have an initial-value formulation. We conclude that there exists an M-diffeomorphism whose corresponding gauge transformation sends the submanifold $\tilde{\mathcal{B}}$ to $\tilde{\mathcal{B}}'$ — and, therefore, a diffeomorphism that sends the one solution ϕ of (1) to the other solution ϕ'.

This completes the demonstration of uniqueness of solutions of (1) up to diffeomorphism gauge, under the two conditions listed above. In particular solutions of the Einstein system are unique up to the diffeomorphism gauge. Note that even though the diffeomorphism gauge group is infinite-dimensional, all manifolds in the present treatment are finite-dimensional. We also remark that the gauge transformation generated by the argument above is unique.

Finally, we note that similar considerations apply to certain other gauge groups. The crucial property we needed, above, is that the gauge transformations be represented by cross-sections of a suitable bundle, and that the requirement that a gauge transformation send one submanifold $\tilde{\mathcal{B}}$ to another be a system of quasilinear partial differential equations on cross-sections of that bundle. An example is the Maxwell system considered earlier, with flats σ given by $u^a \delta A_a = 0$. In this case, the corresponding bundle has fibre, over each point x of M, consisting of the reals (the possible values of the gauge-potential). The conditions analogous to those above again hold; and again we conclude uniqueness of solutions, up to gauge transformations.

5 Conclusion

Fix any first-order, quasilinear system, (1), of partial differential equations, on which there acts gauge transformations. We have shown that, under certain conditions, this system manifests an initial-value formulation up to gauge, i.e., that, given suitable initial data, solutions for that data exist, and are unique, up to gauge transformations. These "certain conditions", while more complicated than one might wish, are at least algebraic in the coefficients of the equation. The key to the proof is to introduce flats σ representing "physical directions" in the fibres of $\mathcal{B}$. To prove existence, we effectively restrict the original equation, (1), to "physical degrees of freedom". Then the conditions in this case require that the equation, so restricted, manifest completeness of its constraints, and a hyperbolization. To prove uniqueness, we write out a partial differential on a gauge transformation, which guarantees that that that gauge transformation send one solution to another. The conditions in this case guarantee that this system has an initial-value formulation, i.e., that it have completeness of its constraints, and a hyperbolization. Of course, this scheme is applicable, in particular, to the Einstein system.

In the example of the Einstein system, the flats σ that satisfy our conditions are those usually associated with what is called "harmonic gauge". For the Maxwell system, with "Coulomb gauge". But note that we do not directly impose any gauge conditions on the fields. In particular, we do not impose, on initial data for either system, any equations beyond the consistency conditions (7) that flow directly from (1). Note that, in the present scheme, the harmonic gauge for the Einstein system is placed on the same footing with the Coulomb gauge for the Maxwell system, for both these gauge conditions are algebraic in the fields. Yet, one might have expected the former to be more analogous to the Lorentz gauge for the Maxwell system.

Are there, besides the Einstein system, other systems of partial differential equations with diffeomorphism gauge freedom that satisfy the conditions of the previous section? I am aware of just two.

One such system, it turns out, is that of special relativity. The fields are the same as those for the Einstein system: a Lorentz-signature metric g_{ab} and a derivative operator ∇_a. However, the equations for this system consist of (4) together with $R_{abc}{}^d = 0$ (the latter replacing (5)-(6) for the Einstein system). That is, this system is merely the Einstein system, augmented with some additional "constraint equations" Note that here the Minkowski metric is taken to be "dynamical", resulting in a system subject to the gauge freedom of diffeomorphisms. The M-diffeomorphisms act on (g_{ab}, ∇_a) in the same manner as in the Einstein system, and so in particular we have the same tensor $\delta^{\alpha ab}{}_m$ of Eqn. (20). For the flats σ, choose precisely the same ones as for the Einstein system. We now claim that this system — special relativity — satisfies all the conditions given in Sect. 4. For the σ-hyperbolization $h_{A\alpha}$ and the $H_n{}^{dab}{}_m$ choose precisely the same objects as for the Einstein system. The demonstrations that the conditions are satisfied is virtually identical to the corresponding demonstrations for the Einstein system.

Here is a second example. Let the fields, on the base manifold M, consist of a nowhere vanishing vector field u^a together with a tensor field $\alpha^b{}_{cd}$. Let the equation be $\mathcal{L}_u \alpha^b{}_{cd} = 0$, i.e., the requirement that the Lie derivative of $\alpha^b{}_{cd}$ by u^a vanish. This system has an action of M-diffeomorphisms as gauge: These diffeomorphisms act on $(u^a, \alpha^b{}_{cd})$ in the usual way, and this action clearly preserves the equation of the system. Note that the order of this action is one, in contrast to the order, two, for the Einstein system. We now claim that this system satisfies all the conditions of Sect. 4. First note that the only constraint of this system is zero. Let, at each point of the bundle manifold $\mathcal{B}$, the flat σ consist of those tangent vectors, $(\delta u^a, \delta\alpha^b{}_{cd})$, with $\delta u^a = 0$. This field of flats is clearly smooth and integrable. Let the open set of covectors n_a, at each point of $\mathcal{B}$ consist of those with $n_a u^a > 0$. Then σ-completeness becomes: Let n_a satisfy $n_a u^a > 0$, and let $\nu_b{}^{cd}$ be any tensor such that the expression

$$\nu_b{}^{cd}[u^m n_m \delta\alpha^b{}_{cd} + \alpha^m{}_{cd}\delta u^b n_m - \alpha^b{}_{md}\delta u^m n_c - \alpha^b{}_{cm}\delta u^m n_d] \qquad (37)$$

vanishes for all $(\delta u^a, \delta\alpha^b{}_{cd})$ with $\delta u^a = 0$. Then this expression vanishes for *all* $(\delta u^a, \delta\alpha^b{}_{cd})$. But this assertion is true (as follows immediately from the fact that the hypothesis implies, using $u^m n_m > 0$, that $\nu_b{}^{cd} = 0$). For the σ-hyperbolization, let $h_{A\alpha}$ be represented by the tensor $h_m{}^{rs}{}_b{}^{cd}$ given by $h_m{}^{rs}{}_b{}^{cd} = p_{mb} p^{rc} p^{sd}$, where p^{ab} is any positive-definite metric, and p_{ab} its inverse. The completeness condition for the gauge transformation reads: If $(\nu_a, \nu_b{}^{cd})$ (i.e., ν_α) is such that $\nu_b{}^{cd} = 0$ (i.e., $\nu_{\alpha^\sigma} = 0$) and such that $(n_s u^s)\nu_m = 0$ (i.e., $n_s \nu_\alpha \delta^{\alpha s}{}_m = 0$), then $(\nu_a, \nu_b{}^{cd}) = 0$. This is immediate. Finally, for the hyperbolization for the gauge transformation, choose $H_{n\alpha}$ such that $H_{n\alpha}\delta\phi^\alpha = p_{na}\delta u^a$, and $S_n{}^a{}_m = 0$. Then the corresponding quadratic form, (30), is given by $u^a p_{nm}$; and this form indeed has the required symmetry and positive-definiteness.

Thus, the system above has an initial-value formulation, up to the gauge freedom of diffeomorphisms. This result, by the way, is already clear geometrically: Use the diffeomorphisms to "fix" the field u^a, whence the equation, which simply requires that $\alpha_b{}^{cd}$ be invariant under u-motions, determines $\alpha_b{}^{cd}$. There are similar examples in which the field $\alpha_b{}^{cd}$ is replaced by a tensor field with general index structure. Furthermore, the single field $\alpha_b{}^{cd}$ could be replaced by any number of tensor fields, each with some index structure, at the same time replacing the single equation $\mathcal{L}_u \alpha^b{}_{cd} = 0$ by the equations that specify that each of these fields has vanishing Lie-derivative by u^a. Still more generally, there could be imposed various algebraic or differential equations (such as, e.g., the vanishing of a contraction, or of the exterior derivative of a form) on the fields, and furthermore the Lie derivative of each field could be equated, not to zero, but rather to some expression algebraic in the fields.

Further examples can be obtained from these by combining systems. Thus, for example, the Einstein-perfect fluid system, as well as the special-relativity-Maxwell system, also represent systems having an initial-value formulation up to diffeomorphism gauge. Also, presumably, further examples can be constructed by taking the derivative system [8] of some known example.

The examples above are the only ones of which I am aware that have an initial value formulation up to the gauge freedom of diffeomorphisms. Is there any hope of finding *all* systems of first-order, quasilinear systems of partial differential equations having this property? This may be feasible, for the conditions of Sect. 4 do not look terribly complicated. A good starting point might be to look for systems whose diffeomorphism-order exceeds two. There may exist no such systems. Such a classification — particularly if relatively few systems were permitted — would be of interest. After all, one could argue that the system of partial differential equations for any viable physical theory shoud manifest an initial-value formulation and diffeomorphism gauge. Thus, such a classification would represent the beginnings of a classification of allowed physical theories — at least, physical theories described by partial differential equations on a manifold.

Appendix A — Initial-Value Formulation

Fix a of first-order, quasilinear system, (1), of partial differential equations. Let this system have constraints that are integrable and complete; and let it also admit a hyperbolization, $h_{A\alpha}$. Let T be a submanifold of M of codimension one, and ϕ_0 a cross-section over T, such that i) this (T, ϕ_0) satisfies all the consistency conditions, (7), arising from (1), and ii) the normal n_a of T at each of its points lies in the open sets of covectors associated with completeness and with the hyperbolization. [That is, let (T, ϕ_0) be "non-characteristic".] We sketch a partial proof of the following assertion: Under the conditions above, there exists, in some neighborhood of T, one and only one smooth solution ϕ of our system, (1), such that $\phi|_T = \phi_0$. What follows is only a "partial proof" because, as we shall see, it contains a gap. This appendix is an invitation to fill that gap.

Consider first the system of equations given by

$$h_{A\beta} k^{Aa}{}_{\alpha} (\nabla \phi)_a{}^{\alpha} = h_{A\beta} j^A, \tag{38}$$

i.e., the subsystem of (1) that results from contracting it with $h_{A\beta}$. This system also has a hyperbolization (and, indeed, it is already in "symmetric-hyperbolic form"), and its only constraint is zero. There is a standard theorem [3] [10] [9] [6], whose proof uses an energy argument, to the effect that *this* system, by virtue of its having a hyperbolization and zero constraints, admits one and only one smooth solution, ϕ, in some neighborhood of T, manifesting the given initial data (T, ϕ_o). Thus (since every solution of (1) is certainly a solution of (38)), we have shown uniqueness of solutions of (1). There remains only existence, and the plan for this is to show that the solution ϕ, just obtained, of the subsystem (38) in fact satisfies the *full* system, (1) of partial differential equations. To this end, consider the field I^A, defined in a neighborhood of T by

$$I^A = k^{Aa}{}_{\alpha} (\nabla \phi)_a{}^{\alpha} - j^A, \tag{39}$$

where, on the right, we have substituted our solution ϕ of (38). What we must show, to prove existence, is $I^A = 0$. To this end, we note that this field I^A has the following three properties. First, this I^A satisfies the algebraic (linear) system

$$h_{A\alpha} I^A = 0 \tag{40}$$

everywhere in our neighborhood of T. This is exactly (38). Second, this I^A vanishes on the submanifold T itself. To see this, first note that, by the consistency conditions (7), we have that $n_a c^a{}_A I^A = 0$ on T, for every constraint $c^a{}_A$ of our system. But, by completeness, the $h_{A\alpha}$ and $n_a c^a{}_A$ (as $c^a{}_A$ runs through the constraints) together span the equation covector space. The result now follows from (40). The third property is that this I^A satisfies a system of (in fact, linear) partial differential equations of the form

$$c^a{}_A \nabla_a I^A = W^A{}_B I^B. \tag{41}$$

Here, $c^a{}_A$ denotes any constraint of the original system, (1), and there is one equation in (41) for each such constraint. Indeed, fixing the constraint $c^a{}_A$, that this equation hold, for some field $W^A{}_B$ on M, follows immediately from integrability of that constraint.

Now consider the following first-order, quasilinear (in fact, linear) system of partial differential equations: The field is I^A subject to (40), and the system of equations is (41). One solution of this system is the I^A given by (39), and this particular solution has $I^A = 0$ on T. So, in order to prove that the I^A of (39) vanishes in our neighborhood of T, we need only show uniqueness for this system, (40)-(41), with given initial values on T.

We first remark that, by completeness of (1), the system above takes "evolution form", i.e., it expresses the derivatives, off T, of every component of I^A in terms of I^A and its derivatives within T. But this property alone is not, apparently, sufficient to establish uniqueness. Uniqueness would follow, from the standard existence/uniqueness theorem discussed earlier, if we could find a hyperbolization for the system (40)-(41) (for it has only the zero constraint). In fact, for every physical example of which I am aware, this system does admit a hyperbolization. Thus, for all these physical examples, the system (1) does indeed manifest existence and uniqueness. But, unfortunately, there has not been given, as far as I am aware, any general proof of the existence of a hyperbolization for (40)-(41). Such a proof could, of course, make use of the assumed integrability and completeness of the constraints, as well as the existence of a hyperbolization, for (1). It would also be acceptable to impose, on the original system (1), suitable additional, mild hypotheses. Of course, it would also suffice to prove uniqueness of the system (40)-(41)) by some other means, i.e., without invoking the standard existence/uniqueness theorem. This, then, is the gap. It would, I believe, be of some interest to fill it.

Consider, as an example, the Maxwell system, (2)-(3). A hyperbolization for this system is characterized by a timelike vector field u^a on space-time. Eqn. (38) reads in this case

$$u^c \nabla_{[a} F_{bc]} = 0, \quad u_{[a} \nabla^m F_{b]m} = 0. \tag{42}$$

The I^A of (39) is given by

$$I_{abc} = \nabla_{[a} F_{bc]}, \quad I_b = \nabla^a F_{ab}. \tag{43}$$

The algebraic conditions, (40), on the I^A become

$$u^c I_{abc} = 0, \quad u_{[a} I_{b]} = 0. \tag{44}$$

Finally, the equations, (41) on the I^A become

$$\nabla_{[d} I_{abc]} = 0, \quad \nabla^b I_b = 0. \tag{45}$$

So, our question becomes in this case: Does uniqueness hold for the system (45) of partial differential equations, on fields consisting of (I_{abc}, I_b) subject to (44)? The answer to this question is yes. Consider, e.g., the field I_b. Then the first equation in (44) is precisely the statement that $I_b = \gamma u_b$ for some function γ, whence (45) becomes $u^b \nabla_b \gamma = -\gamma \nabla_b u^b$ on gamma. But, clearly, this partial differential equation on function γ on M satisfies uniqueness. [In fact, it not only has a hyperbolization and so an initial-value formulation; but it is actually an ordinary differential equation.] Dually for I_{abc}.

So, by the argument above, the Maxwell system admits an initial-value formulation. A similar argument (checking directly and explicitly that the system (40), (41) satisfies uniqueness) works for other systems (1) of partial differential equations of physical interest. In particular, such an argument works for the Einstein system, reduced, as in Sect 4, to a hyperbolic system. What is now needed, in place of this piecemeal approach, is a general theorm.

Appendix B — Linearization

In this appendix, we introduce, for any first-order, quasilinear system of partial differential equations and any solution cross-section for that system, the notion of the corresponding linearized system off that background solution. As an application of this notion, we show that, the gauge vector fields, as we would expect, generate linearized solutions.

Fix a first-order, quasilinear system (1), of partial differential equations; and fix also any solution, ϕ, of this system, the "background field". A *linearized solution* off this background is a vertical vector field, μ^α, defined at points of the image of ϕ, satisfying

$$[\mathcal{L}_\mu k^{Aa}{}_\alpha](\nabla\phi)_a{}^\alpha = \mathcal{L}_\mu j^A \tag{46}$$

there. We must first check that the left side of this equation makes sense. To this end, introduce any derivative operator ∇_α on the bundle manifold $\mathcal{B}$, and write

$$[\mathcal{L}_\mu k^{Aa}{}_\alpha](\nabla\phi)_a{}^\alpha = (\nabla\phi)_a{}^\alpha \mu^\beta \nabla_\beta k^{Aa}{}_\alpha + (\nabla\phi)_a{}^\alpha k^{Aa}{}_\beta \nabla_\alpha \mu^\beta. \tag{47}$$

The derivative in the first term on the right makes sense, despite the Latin index "a" on $k^{Aa}{}_\alpha$, because, by virtue of the factor μ^β, that derivative is taken only in a vertical direction. The derivative in the second term on the right makes sense, despite the fact that μ^α is defined only on the cross-section, because, by virtue of the factor $(\nabla\phi)_a{}^\alpha$, that derivative is taken only in directions tangent to the cross-section. Note that Eqn. (46) is linear: The linearized solutions, for a given background cross-section, form a vector space.

The motivation for this definition is the following. Fix a one-parameter family, ϕ_λ, with $\lambda \in \mathbf{R}$, of cross-sections, each satisfying the partial differential equation (1). Then $d\phi_\lambda/d\lambda|_{\lambda=0}$ defines, at each point of the $\lambda = 0$ cross-section, a vertical vector field, μ^α. This vector, we claim, indeed satisfies (46). To see this, take $d/d\lambda$, at $\lambda = 0$, of each of the equations $k^{Aa}{}_\alpha(\nabla\phi_\lambda)_a{}^\alpha = j^A$ and $\pi \circ \phi_\lambda =$ identity. We may think of the field μ^α as the "connecting vector" from a point of the background cross-section to a nearby cross-section; and of Eqn. (46) as the condition that this nearby cross-section also satisfy our original differential equation (1), to first order.

The linearized system, (46), has exactly the same coefficients as the original system, (1), where the "unknown", ϕ, in the coefficients in the latter is replaced by the fixed background field in the former. It follows that the linearized system inherits all the properties of the original system. Thus, each constraint of the original system gives rise to a constraint of the linearized system; and the integrability and completeness of these constraints (should these properties obtain in the original system) pass to the linearized system. Furthermore, each hyperbolization of the original system gives rise to a corresponding hyperbolization of the linearized system. We remark that all these inheritances go also in the reverse direction, provided only that there exists a "sufficient number" of solutions of Eqn. (1). What we require, in more detail, is that some solution cross-section of (1) passes through each point of $\mathcal{B}$. In short, the linearized system is a good mirror of the original system of partial differential equations.

Consider, for example, the Einstein system. Fix a solution, (g_{ab}, ∇_a), of this system, (4)-(6). Then a vertical vector field on this cross-section is described by a pair of tensor fields, $(\mu_{bc}, \mu^a{}_{bc})$, on the space-time manifold M, each symmetric in indices "b, c". The linearized Einstein equations (i.e., the linearized versions of (4)-(6)) are $\nabla_a \mu_{bc} - 2g_{m(b}\mu^m{}_{c)a} = 0$, $\nabla_{[a}\mu^m{}_{b](c}g_{d)m} = 0$, and $\nabla_{[a}\mu^m{}_{m]b} = 0$, respectively.

We now return to the general case. As we have seen, every gauge transformation sends every solution cross-section of Eqn. (1) to another solution cross-section. We now establish, as an example of the notion of the linearized system, an an "infinitesimal version" of this statement. Let ξ^α be any gauge vector field. Fix any cross-section, ϕ, satisfying Eqn (1). Now consider, at each point of this cross-section, the vector given by $\mu^\alpha = \xi^\alpha - \xi^\mu(\nabla\pi)_\mu{}^a(\nabla\phi)_a{}^\alpha$. It follows immediately, applying $(\nabla\pi)_\alpha{}^a$, that this μ^α is vertical. Thus, μ^α is just the gauge vector field ξ^α "projected vertically via the cross-section ϕ". Furthermore, it follows from Eqns. (12)-(13) that this μ^α satisfies Eqn. (46), i.e., that it defines a linearized solution off the solution ϕ. [The $\mathcal{L}_\xi$-part of $\mathcal{L}_\mu$ in Eqn. (46) satisfies this equation, by (12)-(13). For the $(\nabla\phi)_a{}^\alpha\xi^a$-part, use that this vector field is tangent to the cross-section, and so automatically Lie-derives $k^{Aa}{}_\alpha(\nabla\phi)_a{}^\alpha - j^A$.] Thus, each gauge vector field assigns, to each cross-section satisfying the equation, a linearized solution. Is there a converse to this? There is, subject to a caveat. Let ξ^α be any vector field on $\mathcal{B}$ with the following property: Given any cross-section, ϕ, satisfying Eqn. (1), then the field μ^α, defined at points of the cross-section by $\mu^\alpha = \xi^\alpha - \xi^a(\nabla\phi)_a{}^\alpha$, is a linearized solution. It then follows that this ξ^α satisfies the result of contracting (12) with $(\nabla\phi)_a{}^\alpha$ and subtracting (13), for every cross-section ϕ satisfying our equation. So, if there is a "sufficient number" of solutions — e.g., if every $\nu_a{}^\alpha$ at a bundle-point satisfying $k^{Aa}{}_\alpha \nu_a{}^\alpha = j^A$ and $\nu_a{}^\alpha (\nabla\pi)_\alpha{}^b = \delta_a{}^b$ is the derivative of some solution cross-section ϕ through that point — then we are guaranteed to have (12)and (13) separately, and so are guaranteed that ξ^α is a gauge vector field.

Consider, for example, the Einstein system, (4)-(6). Fix any smooth vector field, ξ^a, on M. Then this ξ^a gives rise to a one-parameter family of diffeomorphisms on the manifold M, and so, since we have the rule for how such diffeomorphisms act on g_{ab} and ∇_a, to a corresponding one-parameter family of diffeomorphisms on the bundle manifold $\mathcal{B}$. The generator of this family is a certain gauge vector field ξ^α, for the Einstein system. This ξ^α is, of course, a lift of the original field ξ^a on M. Now fix a solution cross-section of this system, i.e., fix (g_{ab}, ∇_a) satisfying (4)-(6). Then this gauge vector field gives rise, in the presence of this solution cross-section as background, to a linearized solution, given by

$$\mu_{ab} = 2\nabla_{(a}\xi_{b)}, \tag{48}$$

$$\mu^m{}_{ab} = \nabla_{(a}\nabla_{b)}\xi^m + R_{d(ab)}{}^m \xi^d. \tag{49}$$

Note that the vertical vector μ^α given by (48)-(49) *depends* on the cross-section, and not merely on the point (g_{ab}, ∇_a) of $\mathcal{B}$. This is seen by noting that the right side of (49) involves not only the derivative operator ∇_a, but also its derivative.

Acknowledgement. I would like to thank Oscar Reula, Gabriel Nagy, David Garfinkle, and Maciej Dunajski for discussions.

References

[1] I. M. Anderson, C. G. Torre, *Classification of Local Generalized Symmetries for the Vacuum Einstein Equations*, Comm. Math. Phys 176, 479 (1996).

[2] R. Arnowitt, S. Deser, C. W. Misner, in *Gravitation: An Introduction to Current Research*, L. Witten, ed, (Wiley, NY, 1962).

[3] K. O. Fridrichs, *Comm Pure and Appl Math 7*, 345 (1954).

[4] H. Friedrich, A. Rendall, *The Cauchy Problem for the Einstein Equations*, in *Einstein's Field Equations and their Physical Interpretation*, B. G. Schmidt, ed, (Springer, Berlin, 2000). Available as gr-qc/0002074.

[5] R. Geroch, in *Asymptotic Structure of Space-Time*, F. P. Esposito, L. Witten, eds (Plenum, NY, 1977).

[6] R. Geroch, *Partial Differential Equations of Physics*, in General Relativity (Proceedings of the Forty-Sixth Scottish Summer School in Physics), G. S. Hall, J. R. Pulham, Ed. (SUSSP Publ, Edinburgh). Available as gr-qc/9602055

[7] R. Geroch, S. M. Perng, *Total Mass-Momentum of Arbitrary Initial-Data Sets in General Relativity*, J Math Phys 35, 4157 (1994).

[8] R. Geroch, G. Nagy, O. Reula, *Relativistic Lagrange Formulation*, J. Math. Phys 42, 3789 (2001). Available as gr-qc/0102071

[9] F. Johns. *Partial Differential Equations*, Springer-Verlag (New York), 1982.

[10] P.D. Lax, *Comm Pure and Appl Math 8*, 615,(1955).

[11] R. Penrose, *Asymptotically Flat Spacetimes*, Proc R. Soc A284, 159 (1965).

[12] N. Steenrod, *The Topology of Fibre Bundles*, (Princeton University Press, Princeton (1954)).