Fixing Einstein's Equations

Arlen Anderson and James W. York, Jr.
Department of Physics and Astronomy, University of North Carolina, Chapel Hill, North Carolina 27599-3255
(Received 18 February 1999)

Einstein's equations are not a well-posed system of evolution equations for the spatial metric, except in special coordinates. A remarkable first-order symmetrizable hyperbolic formulation is found that is surprisingly close to Einstein's original equations yet does not require such coordinates. This system has only physical characteristic directions, the light cone and the normal to the spacelike foliation, and serves to unify all the physical hyperbolic formulations. [S0031-9007(99)09273-X]

PACS numbers: 04.20.Cv, 04.25.–g

Einstein's theory of general relativity has not only proven to be physically accurate [1], it also sets a standard for mathematical beauty and elegance—geometrically. When viewed as a dynamical system of equations for evolving initial data, however, these equations have a serious flaw: they cannot be proven to be well posed (except in special coordinates [2–4]). That is, they do not produce unique solutions that depend smoothly on the initial data. Because of this failing, there has been widespread interest recently in reformulating Einstein's theory as a hyperbolic system of differential equations [5–20]. The physical and geometrical content of the original theory remain unchanged, but dynamical evolution is made sound. Here we present a new hyperbolic formulation that is strikingly close to the space-plus-time ("3 + 1") form of Einstein's original equations. Indeed, the familiarity of its constituents make the existence of this formulation all the more unexpected. This is the most economical "physical" first-order symmetrizable hyperbolic formulation presently known to us. By physical, we mean that the only characteristic directions for all (nonmatter) variables are the light cone and the normal to the spacelike hypersurfaces. Our new system also serves as a foundation for unifying previous proposals.

The source of the imperfection in Einstein's theory lies in the fact that it is a constrained theory. Physical initial data cannot be freely specified, yet even infinitesimally perturbed data that violate the physical constraints can lead to results so wildly divergent that they spoil the desired smooth dependence on initial data. This is particularly troublesome in numerical evolution where such violations are unavoidable. The lack of well-posedness is also a serious problem when addressing such a basic question as the global nonlinear stability of flat Minkowski spacetime. In fact, the proof of such stability [21] employs the hyperbolic wave equation of [7] which we discuss below. A well-posed formulation of Einstein's equations would also seem to be an essential starting point for the conventional approach to quantum gravity in which one first quantizes the (unconstrained) classical theory and then imposes the constraints.

The desire to simulate numerically the full nonlinear evolution of Einstein's equations in three dimensions, such as in the collision of two black holes (cf., e.g., [22]), has motivated much of the recent effort on hyperbolic formulations: well-posed underlying equations make stable numerical evolution much more likely than otherwise would be the case, and formulations cast in first-order symmetrizable form are especially suited to numerical implementation. In addition, physical characteristic speeds make it easier to impose good boundary conditions, crucial to a successful numerical scheme. We amplify this point by noting that Einstein's equations contain many unphysical ("gauge") variables among its unknowns, and *a priori* they can travel at any speed. A formulation with only physical characteristic speeds has significant advantages because no explicit separation of physical and unphysical degrees of freedom is required. The physical and unphysical variables propagate at the same speeds and therefore satisfy boundary conditions on the same characteristic surfaces. This is particularly important, for example, at the horizon of a black hole, which is a characteristic boundary for physical variables but not for unphysical ones, unless the latter propagate at the speed of light.

The following system of thirty equations will be shown to be symmetrizable hyperbolic [23]

$$0 = \hat{\partial}_0 g_{ij} + 2NK_{ij}, \tag{1}$$

$$R_{ij} = -N^{-1}\hat{\partial}_0 K_{ij} + \bar{R}^{(e)}_{ij} - N^{-1}\bar{\nabla}_i\bar{\nabla}_j N + KK_{ij} - 2K_{ik}K^k{}_j, \tag{2}$$

$$2g_{ij}R_{k0} = \hat{\partial}_0\bar{\Gamma}_{kij} + \partial_j(NK_{ki}) + \partial_i(NK_{kj}) - \partial_k(NK_{ij}) - 2g_{ij}N\bar{\nabla}^m(K_{km} - g_{km}K) \tag{3}$$

(notation elaborated below). This form suggests the name "Einstein-Christoffel system." It is convenient to replace the third equation by the equivalent equation

$$4g_{k(i}R_{j)0} = \hat{\partial}_0\mathcal{G}_{kij} + \partial_k(2NK_{ij}) - 4g_{k(i}N\bar{\nabla}^m(K_{j)m} - g_{j)m}K), \tag{4}$$

where $A_{(i}B_{j)} = (1/2)(A_iB_j + A_jB_i)$ denotes symmetrization.

 0031-9007/99/82(22)/4384(4)$15.00

To establish notation, we assume that spacetime has topology $\Sigma \times R$ with metric given in the foliation-adapted cobasis,

$$ds^2 = -N^2(dt)^2 + g_{ij}(dx^i + \beta^i dt)(dx^j + \beta^j dt). \tag{5}$$

Here, $N(\alpha, g)$ is the lapse scalar, and $\beta^i(x,t)$ is the spatial shift vector, freely specifiable on the spacelike slices $t =$ constant. The lapse N is determined through $N = \alpha g^{1/2}$, where $\alpha(x,t)$ is a freely specified "slicing" density (of weight -1) and $g = \det g_{ij}$ is the determinant of the spatial metric g_{ij}. The spatial derivatives of the metric are denoted by

$$\mathcal{G}_{kij} = \partial_k g_{ij}. \tag{6}$$

This is a subtle element, as it will transpire that while this relation is imposed initially, it may not hold for the evolved quantities (see below). The spatial Christoffel symbols in this metric are, with $\bar{\Gamma}^m{}_{ij} = g^{mk}\bar{\Gamma}_{kij}$,

$$\bar{\Gamma}_{kij}(\mathcal{G}) \equiv (1/2)(\mathcal{G}_{jki} + \mathcal{G}_{ikj} - \mathcal{G}_{kij}). \tag{7}$$

To focus attention on $\mathcal{G}_{kij}$, we will not use the Christoffel symbols here as independent variables, though we could, but only as a compact notation for this expression in terms of $\mathcal{G}_{kij}$. Finally, K_{ij} denotes the extrinsic curvature of the slice Σ, and $K = K^k{}_k$ is its trace.

The derivative $\bar{\nabla}_k$ is the spatial covariant derivative operator in Σ. The derivative $\hat{\partial}_0 = \partial_t - \pounds_\beta$, where $\partial_t = \partial/\partial t$ and $\pounds_\beta$ is the Lie derivative along the shift vector $\boldsymbol{\beta}$ in a $t =$ constant slice, is the natural time derivative for evolving time-dependent spatial tensors. It is the extension to tensors of the (noncoordinate) basis vector $\partial_0 = \partial_t - \beta^k\partial_k$ ($\partial_k = \partial/\partial x^k$) that is normal to the slice Σ. Note that while $[\partial_0, \partial_j] = \partial_0\partial_j - \partial_j\partial_0 = (\partial_j\beta^k)\partial_k \neq 0$, we have the operator commutation rule

$$[\hat{\partial}_0, \partial_k] = 0. \tag{8}$$

On the left-hand sides of (2) and (3) or (4), R_{ij} and R_{j0} are spacetime Ricci curvature tensors and are to be replaced by their appropriate expressions in terms of matter stress tensors from Einstein's equations. $\bar{R}^{(e)}_{ij}$ is the spatial Ricci curvature tensor of the spacelike slice Σ. It is essential to manipulate the standard form $\bar{R}_{ij} = \partial_k\bar{\Gamma}^k{}_{ij} - \partial_j\bar{\Gamma}^k{}_{ik} + \bar{\Gamma}^k{}_{mk}\bar{\Gamma}^m{}_{ij} - \bar{\Gamma}^k{}_{mj}\bar{\Gamma}^m{}_{ik}$ into a distinct but (initially) equivalent form, indicated below, and the superscript "(e)" reflects this change.

Einsteinian initial data for the system (1), (2), (4) are g_{ij}, K_{ij}, and $\mathcal{G}_{kij}$ ($= \partial_k g_{ij}$), specified on an initial slice Σ_0, and presumed to satisfy the Einstein constraints, $G^0{}_0 = 8\pi T^0{}_0$ and $R^0{}_k = 8\pi T^0{}_k$. This system of initial constraints is well understood as a semilinear elliptic system [24–26]. A mathematically well-posed form of the twice-contracted Bianchi identities [27,28] shows that these initial-value constraints remain satisfied if the equations of motion are equivalent to $R_{ij} = 8\pi[T_{ij} - (1/2)g_{ij}T^\mu{}_\mu]$.

That the system (1), (2), (4) is hyperbolic is not obvious, but its content is easy to grasp. The first equation (1) is simply a definition of the extrinsic curvature K_{ij}. The second equation (2) is the 3 + 1 decomposition of the space-space components of the spacetime Ricci tensor. As such, these are the basic geometric ingredients of Einstein's original equations and of all 3 + 1 formulations of general relativity [24,25,29]. The remarkable fact is that Eq. (4) completes the first two into a symmetrizable hyperbolic system without altering the initial value problem. The content of Eq. (4) is also readily understood.

If one applies $\hat{\partial}_0$ to (6) and uses (8) and (1), one obtains the identity

$$\hat{\partial}_0\mathcal{G}_{kij} = -\partial_k(2NK_{ij}). \tag{9}$$

This is the right-hand side of (4) aside from the 3 + 1 decomposition of $4g_{k(i}R_{j)0}$. Ordinarily in Einstein's theory,

$$R_{0j} = -N\bar{\nabla}^m(K_{jm} - g_{jm}K) \tag{10}$$

is a constraint—the "momentum" constraint—because it involves no time derivatives. What is special about (4) is that it makes the momentum constraint dynamical by combining it with the identity (9) involving a time derivative. This defines a modified evolution of $\mathcal{G}_{kij}$ when the constraint is not satisfied, that is, when (10) does not hold after R_{0j} is replaced by its matter expression.

The identity (9) is closely related to metric compatibility of the connection. In a general spatial frame, a connection is metric compatible if and only if $\bar{\Gamma}_{ijk} + \bar{\Gamma}_{jik} = \partial_k g_{ij}$. Taking the time derivative of this condition and applying (4) shows that, if the momentum constraint is violated, metric compatibility of $\bar{\Gamma}$ is lost during evolution. While $2\bar{\Gamma}_{(ij)k} = \mathcal{G}_{kij}$ always holds, the evolved $\mathcal{G}_{kij}$ is no longer the spatial derivative ∂_k of the evolved g_{ij}.

To motivate the system (1), (2), (4) further, we consider two of its predecessors, the Einstein-Ricci formulation [7,12–14,18] and the Frittelli-Reula formulation [10,16]. The third-order Einstein-Ricci system consists of (1) and a wave equation built from (2) and (10) through the combination

$$\hat{\partial}_0 R_{ij} - \bar{\nabla}_i R_{j0} - \bar{\nabla}_j R_{i0} = N\hat{\Box}K_{ij} + J_{ij} + S_{ij}. \tag{11}$$

It is called third-order because of the effective number of derivatives of g_{ij} in (11). Here, $\hat{\Box} = -N^{-1}\hat{\partial}_0 N^{-1}\hat{\partial}_0 + \bar{\nabla}^k\bar{\nabla}_k$. J_{ij} is a nonlinear function of K_{ij}, N, their first derivatives, and the second derivatives of N. S_{ij} is a potentially troublesome term involving a second spatial derivative of K and a third derivative of N. The behavior of S_{ij} is tamed by using $N = \alpha(x,t)g^{1/2}$ [7,12] (or by imposing generalized harmonic slicing [18] with a gauge source [4]). Note that the use of α permits any time slicing to be employed.

This system can be put in first-order form by introducing new variables to represent the temporal and spatial derivatives of K_{ij} and of N. Together with (1) and the equations obtained by applying $\hat{\partial}_0$ to $\bar{\Gamma}^k{}_{ij}$, and using α

to eliminate N, one finds a system of 66 equations. This system is spatially covariant, is expressed in $3 + 1$ geometric variables, and has only physical characteristics.

One may wonder about the large number of equations and about a deeper meaning behind the combination in (11). Regarding the number of equations, the Einstein-Ricci system is equivalent (for Einsteinian initial data) to the Einstein-Bianchi system [19,20] which also has 66 equations. There, it is evident that this number of equations is precisely that needed to incorporate the full Bianchi identities and to propagate the Riemann curvature tensor explicitly in a system having only physical characteristics (otherwise, cf. [15]).

The Frittelli-Reula system [10,16], in contrast, has 30 equations, is expressed in noncovariant variables, and admits superluminal characteristic speeds for some (unphysical) degrees of freedom. Frittelli and Reula make their construction using a parametrized energy norm and find a family of hyperbolic systems with different characteristics, none wholly physical. Friedrich has observed that an equation for the metric constructed from (1) and (2), while not of known hyperbolic type, has only physical characteristics [15]. A natural question is whether there are further thirty-variable hyperbolic systems and if any have only physical characteristics. The Einstein-Christoffel system is such a system, and from it one sees how to extend the Frittelli-Reula construction.

The Einstein-Ricci system has only physical characteristics, but its equations number more than twice those of the Frittelli-Reula system. A natural question is whether the third-order form can be put in first-order form to achieve a thirty-variable system. To see why this might be possible, consider the wave equation

$$\partial_t^2 u - \partial_x^2 u = 0\,. \tag{12}$$

This can be put in first-order form in two ways. The easiest is to introduce the derivatives of the dependent variable as new variables. Introduce $U = \partial_t u$ and $V = \partial_x u$ to reach the system

$$\begin{gathered}\partial_t u = U\,,\\ \partial_t U - \partial_x V = 0, \qquad \partial_t V - \partial_x U = 0\,.\end{gathered} \tag{13}$$

The last equation is an integrability condition reflecting the commutativity of the partial derivatives. This parallels the way that the first-order form of the Einstein-Ricci system was obtained from (11).

The second way to get to first order form is to pull apart the wave equation to obtain first order pieces

$$\partial_t u - \partial_x v = 0, \qquad \partial_t v - \partial_x u = 0\,. \tag{14}$$

The first method is essentially one derivative higher. Note that the wave equation (12) is reconstructed from system (14) by taking a time derivative of the first equation and adding a spatial derivative of the second. This parallels the structure in (11) that leads to the third-order Einstein-Ricci system. This encourages the speculation that a "pulled-apart" system analogous to (14) is possible. The obstacle is that the momentum constraint as usually construed is not a dynamical equation, so the obvious pulled-apart system is not hyperbolic. The key idea is that adding a suitably chosen dynamical identity to the momentum constraint overcomes this obstacle and leads to a symmetrizable hyperbolic system.

To begin, we work with $\mathcal{G}_{kij}$ rather than $\bar\Gamma_{kij}$. Focus on the derivatives of the Christoffel symbols contained in $\bar R_{ij} - N^{-1}\bar\nabla_i\bar\nabla_j N$. These are the essential terms from the standpoint of hyperbolicity. [Recall that $N = \alpha g^{1/2}$, so $\bar\nabla_j N = g^{1/2}\partial_j\alpha + \bar\Gamma^k_{jk}(\mathcal{G})g^{1/2}\alpha$.] These terms can be reorganized as follows:

$$\begin{aligned}\partial_k\bar\Gamma^k_{ij}(\mathcal{G}) - \partial_j\bar\Gamma^k_{ik}(\mathcal{G}) - \partial_i\bar\Gamma^k_{jk}(\mathcal{G}) = {} & -\frac{1}{2}\,\partial_k(g^{km}\mathcal{G}_{mij}) + \partial_{(i}[g^{rs}(\mathcal{G}_{|rs|j)} - \mathcal{G}_{j)rs})]\\ & + g^{kr}g^{sm}[\mathcal{G}_{km(i}\mathcal{G}_{j)rs} - \mathcal{G}_{krs}\mathcal{G}_{(ij)m}]\end{aligned} \tag{15}$$

(where the indices between vertical bars are not symmetrized). Introducing

$$f_{kij} \equiv \frac{1}{2}\,\mathcal{G}_{kij} - g_{k(i}g^{rs}(\mathcal{G}_{|rs|j)} - \mathcal{G}_{j)rs}) = \bar\Gamma_{(ij)k} + g_{k(i}g^{rs}(\bar\Gamma_{|rs|j)} - \bar\Gamma_{j)rs}) \tag{16}$$

puts the leading derivatives of (2) in the familiar form

$$R_{ij} = -N^{-1}\hat\partial_0 K_{ij} - \partial^k f_{kij} + \text{l.o.}_{ij}\,, \tag{17}$$

where l.o._{ij} stands for lower order terms containing no derivatives of unknowns. They are

$$\begin{aligned}\text{l.o.}_{ij} = {} & KK_{ij} - 2K_{ik}K^k_j - \alpha^{-1}[\partial_i\partial_j - \bar\Gamma^k_{ij}(\mathcal{G})\partial_k]\alpha - [\bar\Gamma^k_{ki}(\mathcal{G}) + \alpha^{-1}\partial_i\alpha][\bar\Gamma^m_{mj}(\mathcal{G}) + \alpha^{-1}\partial_j\alpha]\\ & + 2\bar\Gamma^k_{mk}(\mathcal{G})\bar\Gamma^m_{ij}(\mathcal{G}) - \bar\Gamma^k_{mj}(\mathcal{G})\bar\Gamma^m_{ik}(\mathcal{G}) + g^{kr}g^{sm}[\mathcal{G}_{krs}f_{mij} + \mathcal{G}_{km(i}\mathcal{G}_{j)rs} - \mathcal{G}_{krs}\mathcal{G}_{(ij)m}]\,.\end{aligned} \tag{18}$$

Turn to consider (4). From (16), one computes

$$g_{ki}R_{j0} + g_{kj}R_{i0} = -\hat\partial_0 f_{kij} - \partial_k(NK_{ij}) + \text{l.o.}_{kij}\,. \tag{19}$$

The lower order terms are

$$\text{l.o.}_{kij} = 2NK_{k(i}g^{rs}(\mathcal{G}_{|rs|j)} - \mathcal{G}_{j)rs}) + 2g_{k(i}[K_{j)m}\partial^m N - K\partial_{j)}N + NK_{j)m}g^{rs}\bar\Gamma^m_{rs}(\mathcal{G}) + \frac{1}{2}\,N(\mathcal{G}_{j)rs} - 2\mathcal{G}_{|rs|j)})K^{rs}]\,. \tag{20}$$

The system (17) and (19) [completed by (1)] is obviously symmetrizable hyperbolic because it has the familiar structure of a wave equation in first-order form. It is also clear that to build a wave equation in K_{ij} from (17) and (19), one forms a combination similar to (11). This reveals the meaning behind this combination. The characteristic speed in the system (17), (19) is the speed of light, so the extrinsic curvature and the connection propagate at the speed of light. From (1), the metric propagates at speed zero. In the end, one need use only (say) f_{kij} to express all three-index symbols.

It should be emphasized that the wave equation obtained from (17), (19) is not exactly the same as the third-order Einstein-Ricci system. They differ by lower order terms proportional to constraints. This means that they will agree for Einsteinian initial data, but may disagree when the constraints are violated. Likewise, a third-order wave equation, equivalent for Einsteinian data to that above, can be constructed easily from the first-order Einstein-Bianchi system. One sees that "$\hat{\Box} K_{ij}$" is the central element of unification.

The energy norm for the system (1), (17), (19) is the integral over Σ of $K^{ij}K_{ij} + f^{kij}f_{kij}$, where $f^{kij} = g^{km}g^{ir}g^{js}f_{mrs}$. When the energy norm is expressed in terms of their variables, the result can be compared to the *ansatz* of Frittelli-Reula [16]. Additional terms are present beyond those that they considered. Their energy norm *ansatz* can be readily generalized. One finds a larger many-parameter family of symmetrizable hyperbolic systems equivalent to Einstein's equations. Some of these other systems also have only physical characteristics; for example, a multiple of the Hamiltonian constraint G^0_0 can be added to (2) if a multiple of the momentum constraint R_{0k} is added to (3) and (4).

The Einstein-Christoffel system discussed here is a well-posed system of 30 equations that has only physical characteristics and can be expressed in standard $3 + 1$ geometric variables. Spatial covariance of the system is not explicit, but present nonetheless. This formulation clarifies the relationships among Einstein's original equations and the Einstein-Ricci, Einstein-Bianchi, and Frittelli-Reula hyperbolic formulations. One can see further links to the Friedrich [5,6,8,15] and Bona-Massó [9,11] formulations. It will be interesting to implement this system numerically. Having only physical characteristics should prove useful when imposing boundary conditions on the horizon of a black hole since no information, physical or otherwise, can leave the black hole, and all information will enter the black hole in a physical way.

A. A. thanks Mark Hannam for discussions, and J. W. Y. thanks Y. Choquet-Bruhat. This work was supported in part by National Science Foundation Grant No. PHY-9413207.

[1] C. M. Will, *Theory and Experiment in Gravitational Physics* (Cambridge University Press, Cambridge, 1993), revised ed.

[2] Y. Choquet (Fourès)-Bruhat, Acta Math. **88**, 141 (1952).

[3] A. Fischer and J. Marsden, Commun. Math. Phys. **28**, 1 (1972).

[4] H. Friedrich, Commun. Math. Phys. **100**, 525 (1985).

[5] H. Friedrich, Proc. R. Soc. London A **375**, 169 (1981).

[6] H. Friedrich, Proc. R. Soc. London A **378**, 401 (1981).

[7] Y. Choquet-Bruhat and T. Ruggeri, Commun. Math. Phys. **89**, 269 (1983).

[8] H. Friedrich, J. Diff. Geom. **34**, 275 (1991).

[9] C. Bona and J. Massó, Phys. Rev. Lett. **68**, 1097 (1992).

[10] S. Frittelli and O. Reula, Commun. Math. Phys. **166**, 221 (1994).

[11] C. Bona, J. Massó, E. Seidel, and J. Stela, Phys. Rev. Lett. **75**, 600 (1995).

[12] Y. Choquet-Bruhat and J. W. York, C.R. Acad. Sci. Paris, **t. 321**, Série I, 1089 (1995).

[13] A. Abrahams, A. Anderson, Y. Choquet-Bruhat, and J. W. York, Phys. Rev. Lett. **75**, 3377 (1995).

[14] A. Abrahams, A. Anderson, Y. Choquet-Bruhat, and J. W. York, C.R. Acad. Sci. Paris, **t. 323**, Série II *b*, 835 (1996).

[15] H. Friedrich, Classical Quantum Gravity **13**, 1451 (1996).

[16] S. Frittelli and O. Reula, Phys. Rev. Lett. **76**, 4667 (1996).

[17] M. H. P. M. van Putten and D. M. Eardley, Phys. Rev. D **53**, 3056 (1996).

[18] A. Abrahams, A. Anderson, Y. Choquet-Bruhat, and J. W. York, Classical Quantum Gravity **14**, A9 (1997).

[19] Y. Choquet-Bruhat and J. W. York, *Mathematics of Gravitation*, Banach Center Publications (Polish Academy of Science, Warszawa, 1997), Vol. 41, Pt. 1, p. 119.

[20] A. Anderson, Y. Choquet-Bruhat, and J. W. York, Topol. Meth. Nonlinear Anal. **10**, 353 (1997).

[21] D. Christodoulou and S. Klainerman, *The Global Nonlinear Stability of the Minkowski Space* (Princeton University Press, Princeton, NJ, 1993).

[22] A. Abrahams *et al.*, Phys. Rev. Lett. **80**, 1812 (1998).

[23] A symmetrizable hyperbolic system is a system that can be put in the following form by an invertible transformation:

$$\mathbf{A}^t \partial_t \mathbf{u} + \mathbf{A}^k \partial_k \mathbf{u} = \mathbf{B},$$

where $\mathbf{u}$ is a vector of unknowns, $\mathbf{A}^t$ is a positive definite symmetric matrix, $\mathbf{A}^k$ are symmetric matrices, and $\mathbf{B}$ is a vector of sources. $\mathbf{A}^t$, $\mathbf{A}^k$, and $\mathbf{B}$ may depend on space and time and on $\mathbf{u}$ but not its derivatives. Such systems are well posed.

[24] J. W. York, in *Sources of Gravitational Radiation*, edited by L. Smarr (Cambridge University Press, Cambridge, 1979), p. 83.

[25] Y. Choquet-Bruhat and J. W. York, in *General Relativity and Gravitation, I*, edited by A. Held (Plenum, New York, 1980), p. 99.

[26] J. W. York, Phys. Rev. Lett. **82**, 1350 (1999).

[27] S. Frittelli, Phys. Rev. D **55**, 5992 (1997).

[28] A. Anderson and J. W. York, Phys. Rev. Lett. **81**, 1154 (1998).

[29] Y. Choquet (Fourés)-Bruhat, J. Rat. Mechanics Anal. **5**, 951 (1956).