Constraint and gauge shocks in one-dimensional numerical relativity

Bernd Reimann,[1,2] Miguel Alcubierre,[1] José A. González,[3] and Darío Núñez[1]

[1]*Instituto de Ciencias Nucleares, Universidad Nacional Autónoma de México, A.P. 70-543, México D.F. 04510, México*
[2]*Max Planck Institut für Gravitationsphysik, Albert Einstein Institut, Am Mühlenberg 1, 14476 Golm, Germany*
[3]*Institute for Gravitational Physics and Geometry, Penn State University, University Park, Pennsylvania 16802, USA*

We study how different types of blowups can occur in systems of hyperbolic evolution equations of the type found in general relativity. In particular, we discuss two independent criteria that can be used to determine when such blowups can be expected. One criteria is related to the so-called geometric blowup leading to gradient catastrophes, while the other is based upon the ODE-mechanism leading to blowups within finite time. We show how both mechanisms work in the case of a simple one-dimensional wave equation with a dynamic wave speed and sources, and later explore how those blowups can appear in one-dimensional numerical relativity. In the latter case we recover the well known "gauge shocks" associated with Bona-Massó–type slicing conditions. However, a crucial result of this study has been the identification of a second family of blowups associated with the way in which the constraints have been used to construct a hyperbolic formulation. We call these blowups "constraint shocks" and show that they are formulation specific, and that choices can be made to eliminate them or at least make them less severe.

I. INTRODUCTION

When studying the Cauchy problem of field theories in physics, one has to worry about the existence and uniqueness of solutions to the system of evolution equations considered. In mathematical terms, one is interested in working with problems that are well-posed, by which one understands that a unique solution exists (at least locally), and solutions are stable in the sense that small changes in the initial data produce small changes in the solution. In this respect, one usually looks for either symmetric or strongly hyperbolic systems of equations since Cauchy problems for such systems are known to be well-posed under very general conditions [1].

In the case of general relativity, the Cauchy problem was studied since the 1950s with the pioneering work of Choquet-Bruhat [2], and by the mid 1980s a number of hyperbolic reductions were known (see for example [3–6], and more recent reviews in [7,8]). Still, those reductions played a minor role in numerical relativity, where practically all work using the Cauchy approach was based on the Arnowitt-Deser-Misner (ADM) system of equations [9,10]. Interest in hyperbolic formulations in numerical relativity started in the early 1990s, with the work of Bona and Massó [11–13], but continued as a small side branch for a number of years. This situation remained until Baumgarte and Shapiro showed in [14] that a reformulation of the ADM equations originally proposed by Nakamura, Oohara and Kojima [15], and Shibata and Nakamura [16], had far superior numerical stability properties than ADM. Baumgarte and Shapiro attributed this to the fact that the new formulation, which has since become known as BSSN, had a "more hyperbolic flavor." This rather informal statement was later put on firmer ground in [17–19], and today it is understood that ADM is only

weakly hyperbolic (and thus not well-posed) [20], whereas BSSN is strongly hyperbolic [18,19].

The recognition by the numerical relativity community of the fact that well-posedness is a crucial ingredient for having long term stable and well behaved numerical simulations (see [21]) has led, in recent years, to an explosion in the search for ever more general hyperbolic reductions of the Einstein evolution equations. At this time many such hyperbolic formulations exist, several of which have dozens of free parameters (see for example [22,23]). The large number of ways in which one can construct strongly or even symmetric hyperbolic formulations has taken us to a situation where there are now many more proposed formulations than numerical groups capable of testing them. At the same time, there is a growing realization that in some respects well-posedness is not enough, as empirically some hyperbolic formulations have proven to be far more robust than others. Some work has been done on the analytic side trying to understand what makes some hyperbolic formulations better suited for numerical work. In particular, one can mention the work of Shinkai and Yoneda [24], where the propagation of constraints for different formulations is studied by linearizing the evolution system around the Schwarzschild background and looking for the eigenvalues of the evolution matrix in Fourier space, and the work of Lindblom and Scheel [25], where the rate of growth of the constraint violation is analyzed for a family of symmetric hyperbolic formulations using the fact that for such systems one can construct an "energy norm." The consensus is that one should look beyond the principal part of the system, and study the effect of the source terms on the stability.

In this paper, we want to focus on a different aspect that can differentiate between hyperbolic formulations and that has been so far overlooked. Well-posed formulations are

known to have well behaved solutions locally, but there is no guarantee that these solutions can exist beyond a certain finite time. In fact, on physical grounds we expect solutions to fail after a finite time in some circumstances, due, for example, to the formation of singularities in gravitational collapse. But there is another way in which solutions can become singular after a finite time, the best example of which is the formation of shock waves in hydrodynamics. In general relativity, and particularly in vacuum, we do not expect this type of shock waves to develop. Nevertheless, one must remember that the evolution equations evolve more than just the physical degrees of freedom. In particular, there are also gauge degrees of freedom that can cause coordinate singularities to arise during the evolution. In [26] one of us showed that coordinate singularities caused by the crossing of the characteristic lines associated with the propagation of the gauge can in fact easily form. These so-called ''gauge shocks'' are now known to be responsible for the fact that some gauge choices are much better behaved than others. In particular, in Ref. [27] it was shown that the well known ''$1 + \log$'' slicing condition, which empirically has been found to be very robust, is in fact the only member of its family that avoids gauge shocks approximately. More recently it has been found that, for the evolution of Brill wave spacetimes that are very close to the critical threshold for black hole formation, the use of shock avoiding slicing conditions is crucial [28].

But there are degrees of freedom other than the physical and gauge modes that also appear in the evolution equations of general relativity. These extra degrees of freedom have to do with the violation of the constraints and, even though for physical initial data they should vanish identically, truncation errors make their presence unavoidable in numerical simulations. It is therefore very important to understand how these constraint violating modes behave analytically. A main result of this paper is the recognition that constraint modes can also give rise to the development of blowups in a finite time that are very similar to the gauge shocks studied before. These blowups are a property of the specific form of the evolution equations and their effects can be significantly reduced if one chooses carefully how the constraints are used when constructing a hyperbolic system. In this paper we show how blowups can arise in spherically symmetric relativity, and how they can best be avoided by modifying the evolution system. We believe that the study of such constraint shocks might help us understand why otherwise well-posed and ''nice'' formulations might behave poorly in numerical simulations.

A comment about our terminology is in order. Borrowing the language from hydrodynamics, throughout the paper we will refer in a somewhat loose way to blowups as ''shocks,'' though this term strictly only refers to blowups caused by the crossing of characteristic lines.

This paper is organized as follows. In Sec. II we introduce the concept of hyperbolicity and describe two differ-

ent criteria that can be used to determine when blowups in the solutions to hyperbolic systems of equations can be expected. Section III introduces the simple one-dimensional wave equation with sources and a dynamic wave speed as an example of how these blowups develop. In Sec. IV we apply the blowup criteria to the evolution equations of general relativity. We start with the simple case of $1 + 1$ dimensions, where we recover the well known gauge shocks, and later study the case of spherical symmetry where we find that constraint shocks can also arise. We conclude in Sec. V.

II. HYPERBOLICITY AND SHOCKS

The evolution equations we are interested in analyzing are the evolution equations for the Cauchy problem of general relativity. In particular we are interested in studying the appearance of singular nonphysical solutions. Such an analysis can be best made using the characteristic structure of hyperbolic systems, so we will start from the definition of hyperbolicity. We will also concentrate on systems with only one spatial dimension as this makes the analysis so much simpler. The important point of what happens in the multidimensional case is a matter for future research. Notice that one-dimensional systems are in fact relevant in general relativity, as they can represent the evolution of systems with, for example, spherical symmetry.

There is one important point that should be mentioned. Throughout this section, and in the rest of the paper, we manipulate differential equations by assuming that partial derivatives in time and space commute. This type of manipulation leaves smooth (''classical'') solutions unchanged, but can easily change the speed of propagation of shock waves [29]. Still, in this paper we will only be interested in smooth solutions, and we will consider the development of a shock as a pathology. Our whole emphasis is in finding ways to avoid shocks.

A. Hyperbolic systems

We will consider quasilinear systems of evolution equations that can be split into two subsystems with the following structure:

$$\partial_t \vec{u} = -\mathbf{M}_u^1 \partial_x \vec{u} - \mathbf{M}_u^2 \vec{K}, \qquad (2.1)$$

$$\partial_t \vec{K} + \mathbf{M}_K^1 \partial_x^2 \vec{u} + \mathbf{M}_K^2 \partial_x \vec{K} = \vec{p}_K(\vec{u}, \partial_x \vec{u}, \vec{K}). \qquad (2.2)$$

Here $\vec{u}$ and $\vec{K}$ are n and m dimensional vectors, respectively, and $\mathbf{M}_u^{1,2}$ and $\mathbf{M}_K^{1,2}$ are matrices whose coefficients may depend on the u's, but not on the K's.

In order to have a first order system we will introduce the spatial derivatives $D_i := \partial_x u_i$ as extra independent variables, whose evolution equations are obtained directly from

those of the u's,

$$\partial_t \vec{D} + \partial_x(\mathbf{M}_u^1 \vec{D} + \mathbf{M}_u^2 \vec{K}) = 0. \qquad (2.3)$$

In the following we will always assume that the initial data satisfies the constraints $D_i = \partial_x u_i$. This implies that spatial derivatives of the u's can always be substituted for D's and treated as source terms.

Let us now define the $(n + m)$ dimensional vector $\vec{v} := (\vec{D}, \vec{K})$. We can then rewrite the system of evolution equations as

$$\partial_t \vec{u} = -\mathbf{M}(\vec{u})\vec{v}, \qquad (2.4)$$

$$\partial_t \vec{v} + \mathbf{A}(\vec{u})\partial_x \vec{v} = \vec{q}_v(\vec{u}, \vec{v}), \qquad (2.5)$$

where $\mathbf{M}$ and $\mathbf{A}$ are $n \times (n + m)$ and $(n + m) \times (n + m)$ matrices,

$$\mathbf{M} = (\mathbf{M}_u^1 \quad \mathbf{M}_u^2), \qquad \mathbf{A} = \begin{pmatrix} \mathbf{M}_u^1 & \mathbf{M}_u^2 \\ \mathbf{M}_K^1 & \mathbf{M}_K^2 \end{pmatrix}, \qquad (2.6)$$

and where the source vector $\vec{q}_v$ is given by

$$\vec{q}_v = \left(-\sum_{i=1}^{n} D_i[(\partial_{u_i}\mathbf{M}_u^1)\vec{D} + (\partial_{u_i}\mathbf{M}_u^2)\vec{K}], \vec{p}_K \right). \quad (2.7)$$

In our primary example, the Einstein equations, the vector $\vec{u}$ consists of gauge variables and components of the 3-metric, whereas $\vec{v}$ contains both variables associated with the spatial derivatives of the gauge variables and metric components (the D's) and also variables arising from the time derivatives of the metric components (the K's). Note, furthermore, that the source terms $\vec{q}_v$ appearing on the right-hand side of (2.5) are in general functions of both the u's and v's (typically quadratic on the v's).

The system of equations above will be hyperbolic if the matrix $\mathbf{A}$ has $(n + m)$ real eigenvalues λ_i. Furthermore, it will be strongly hyperbolic if it has a complete set of eigenvectors $\vec{\xi}_i$,

$$\mathbf{A}\vec{\xi}_i = \lambda_i\vec{\xi}_i. \qquad (2.8)$$

If we denote the matrix of column eigenvectors by $\mathbf{R}$,

$$\mathbf{R} = (\vec{\xi}_1 \cdots \vec{\xi}_{(n+m)}), \qquad (2.9)$$

then the matrix $\mathbf{A}$ can be diagonalized as

$$\mathbf{R}^{-1}\mathbf{A}\mathbf{R} = \mathbf{diag}[\lambda_1, \ldots, \lambda_{(n+m)}] := \mathbf{\Lambda}. \qquad (2.10)$$

Notice that for systems with only one spatial dimension the otherwise important distinction between strongly and symmetric hyperbolic systems does not arise.

For a strongly hyperbolic system we define the eigenfields as

$$\vec{w} = \mathbf{R}^{-1}\vec{v}. \qquad (2.11)$$

By analyzing the time evolution of the eigenfields we now want to study by which mechanisms blowups (i.e., singu-

larities) in the solutions can arise. As pointed out in [30], there are basically two different blowup mechanisms which are, somewhat misleadingly, referred to as "geometric blowup" and the "ODE-mechanism." Since in the first case the derivative of an evolution variable, and in the second case an evolution variable itself, becomes infinite within a finite time, the names "gradient catastrophe" [31] and "blowup within finite time" are probably more appropriate. In the following sections we will explain the basic idea behind these two mechanisms using as a prototype a simple scalar equation, and we will show how both mechanisms are in fact closely related. We will also point out how these mechanisms generalize from a single ordinary or partial differential equation (ODE or PDE) to systems of PDE's.

B. Geometric blowup and linear degeneracy

1. Scalar conservation laws

The first mechanism responsible for blowups involves only quasilinear systems of equations. Here the solution u under consideration has a well-defined limit at a given point and only the derivatives of u become infinite there. Typical examples of this situation are obtained when solving scalar conservation equations of the form

$$\partial_t u + \partial_x F(u) = \partial_t u + a(u)\partial_x u = 0, \qquad (2.12)$$

where $a(u) := \partial F(u)/\partial u$. The blowup is then due to the focusing of characteristics at a point, and the mechanism is referred to as geometric blowup. Taking the simplest nonlinear function $F(u) = u^2/2$, we obtain Burgers' equation

$$\partial_t u + u\partial_x u = 0, \qquad (2.13)$$

which is frequently discussed in the literature as an example of a genuinely nonlinear PDE leading to shock formation (see for example [31,32]). The solution of Eq. (2.13) can be interpreted as a time-dependent one-dimensional velocity field u. The equation then states that the characteristics (i.e., the "flow lines") have zero acceleration, that is $du/dt = \partial_t u + (dx/dt)\partial_x u = \partial_t u + u\partial_x u = 0$, which means that particles following those trajectories move with constant velocity $u = dx/dt$. However, unless the initial velocity distribution $u_0(x)$ is a nondecreasing function of x (so that the particles "spread out"), eventually a particle with higher velocity will collide with one ahead of it having a lower velocity. In particular, as particles initially at rest are not moving at all, u is forced to become singular at a finite time if its initial velocity distribution has compact support [except in the trivial case when $u_0(x)$ vanishes everywhere].

2. Linear degeneracy

The previous argument coming from Burgers' Eq. (2.13) can be easily generalized to the case of Eq. (2.12). Consider two locations x_1 and x_2 with $x_1 < x_2$ and corresponding

initial values $u_1 = u_0(x_1)$ and $u_2 = u_0(x_2)$. Just as before, the values of u are conserved along characteristics, so unless $a[u(x)]$ is purely increasing in x, it is always possible to find locations such that $a(u_1) > a(u_2)$, so the characteristic lines on the left go faster than those on the right [as $a(u_i)$ represents their velocity]. One may readily verify that the lines will intersect at a time given by

$$t^* = -\left(\frac{x_1 - x_2}{a(u_1) - a(u_2)}\right). \qquad (2.14)$$

At the point of intersection u has to take both values u_1 and u_2, so a unique solution ceases to exist. When this happens the spatial derivative of u becomes infinite and the differential equation breaks down, in other words no smooth solution of (2.12) exists after $t = t^*$.

For smooth initial data, taking the limit $|x_1 - x_2| \to 0$ in the expression for t^*, we see that this gradient catastrophe will occur at a finite time

$$t^* = -\frac{1}{\min[\partial_x a(u_0)]} = -\frac{1}{\min[a'(u_0)\partial_x u_0]} \qquad (2.15)$$

in the "future" if initially we have $\partial_x a[u_0(x)] < 0$, whereas the problem arises in the "past" if $\partial_x a[u_0(x)] > 0$ holds initially. Since in general one cannot guarantee that every initial data set one would like to use satisfies such a condition, the criteria for not forming a shock demands that the function $a(u)$ should be linear, that is $a'(u) = 0$.

The above argument can also be generalized to (strongly) hyperbolic systems of equations (see [33]). For such systems, the condition one needs in order to avoid the formation of shocks associated with the propagation of a given eigenfield w_i is for this eigenfield to be "linearly degenerate," which means that the eigenvalue λ_i associated with w_i must be constant along integral curves of the corresponding eigenvector $\vec{\xi}_i$:

$$\nabla_v \lambda_i \cdot \vec{\xi}_i = \frac{\partial \lambda_i}{\partial w_i} = \sum_{j=1}^{(n+m)} \frac{\partial \lambda_i}{\partial v_j} \frac{\partial v_j}{\partial w_i} = 0. \qquad (2.16)$$

C. ODE-mechanism and the source criteria

1. ODE's with quadratic sources

In ODE's, PDE's and systems of PDE's, an evolution variable itself can become infinite at a point by a process of "self-increase" in the domain of influence leading to this point. In a somewhat misleading way, the underlying mechanism has been given the name ODE-mechanism (see [30]), since prototype examples are based on simple ODE's such as

$$\frac{du}{dt} = cu^2, \qquad c = \text{const} \neq 0. \qquad (2.17)$$

For nontrivial initial data the solution of (2.17) is

$$u(t) = \frac{u_0}{1 - u_0 ct}, \qquad u_0 \neq 0. \qquad (2.18)$$

This solution clearly blows up at a finite time given by

$$t^* = \frac{1}{u_0 c}, \qquad (2.19)$$

either in the past or in the future, depending on the sign of $u_0 c$. One can also expect such blowups to happen in the case when c is not a constant but instead a function of time. If $c(t)$ is bounded, one can apply theorems 1 and 2 of [34]: Supposing that the function $c(t)$ satisfies the inequality $0 < C < c(t)$ for $0 \le t \le T$, and that u_0 is positive, then $T < 1/(u_0 C)$ since $u(t)$ for all positive t is bounded from below by $u_0/(1 - u_0 Ct)$. Similarly, supposing that $c(t)$ satisfies the inequality $|c(t)| < \tilde{C}$, then the initial value problem has a solution for at least $|t| < 1/|u_0 \tilde{C}|$.

2. The source criteria for avoiding shocks

Let us now go back to our original system of Eqs. (2.4) and (2.5). Multiplying Eq. (2.5) from the left by $\mathbf{R}^{-1}$ we find

$$\partial_t(\mathbf{R}^{-1}\vec{v}) + (\mathbf{R}^{-1}\mathbf{A}\mathbf{R})\partial_x(\mathbf{R}^{-1}\vec{v})$$
$$= \mathbf{R}^{-1}\vec{q}_v + [\partial_t\mathbf{R}^{-1} + (\mathbf{R}^{-1}\mathbf{A}\mathbf{R})\partial_x\mathbf{R}^{-1}]\vec{v}, \qquad (2.20)$$

which, by making use of (2.10) and (2.11), yields

$$\partial_t\vec{w} + \mathbf{\Lambda}\partial_x\vec{w} = \vec{q}_w, \qquad (2.21)$$

where

$$\vec{q}_w := \mathbf{R}^{-1}\vec{q}_v + [\partial_t\mathbf{R}^{-1} + \mathbf{\Lambda}\partial_x\mathbf{R}^{-1}]\vec{v}. \qquad (2.22)$$

In this way we have obtained an evolution system where on the left-hand side of (2.21) the different eigenfields w_i are decoupled. However, in general the equations are still coupled through the source terms q_{w_i}. In particular, if the original sources were quadratic in the v's, we will have

$$\frac{dw_i}{dt} = \partial_t w_i + \lambda_i \partial_x w_i = \sum_{j,k=1}^{(n+m)} c_{ijk} w_j w_k + \mathcal{O}(w), \quad (2.23)$$

where $d/dt := \partial_t + \lambda_i \partial_x$ is the derivative along the corresponding characteristic. As pointed out for a similar system in [35,36], here the $c_{iii}w_i^2$ component of the source term can be expected to dominate, so mixed and lower order terms can be neglected. Though we have no proof of this statement in the general case, one can expect it to be true at least for systems with distinct eigenspeeds, as mixed terms will then be suppressed when pulses moving at different speeds separate from each other. The effect of the term $c_{iii}w_i^2$, on the other hand, will remain even as the pulse moves. The numerical simulations presented in the following sections show empirical evidence that reinforces this argument.

We can then rewrite the above equation as

$$\frac{dw_i}{dt} \approx c_{iii} w_i^2, \qquad (2.24)$$

which has precisely the form of the ODE studied in the previous section. In order to avoid a blowup, one would then have to demand that the coefficients c_{iii} vanish. We call this the "source criteria" for avoiding blowups.

There is a very important property of our system of equations regarding the coefficients c_{iii} that come from the source terms $\vec{q}_w$. From (2.22) one could expect contributions to these coefficients coming both from the original sources $\vec{q}_v$ and from the term in brackets involving derivatives of $\mathbf{R}^{-1}$. However, one can show that this is not the case and for the systems under study the contributions to c_{iii} coming from the term in brackets cancel out, that is, all contributions to c_{iii} come only from the original sources $\vec{q}_v$.

In order to see this we start by rewriting the term inside brackets on the right-hand side of (2.22) as

$$\partial_t \mathbf{R}^{-1} + \mathbf{\Lambda} \partial_x \mathbf{R}^{-1} = \sum_{l=1}^{n} (\partial_t u_l + \mathbf{\Lambda} \partial_x u_l) \partial_{u_l} \mathbf{R}^{-1}. \quad (2.25)$$

Since both the time and space derivatives of u_l can be written in terms of v's, and the above term multiplies the vector $\vec{v}$ in Eq. (2.22), it is clear that this term will give rise to quadratic terms in the v's, and hence in the w's. The question is whether these quadratic terms will produce a contribution to the coefficients c_{iii}. From the last equation it is clear that no such contribution will exist if the following condition holds:

$$\frac{\partial}{\partial w_i} (\partial_t u_l + \lambda_i \partial_x u_l) = 0, \quad \forall \ i \leq (n+m), \quad l \leq n. \quad (2.26)$$

We will now show that this condition is indeed always satisfied. Notice first that, from the definition of the eigenfields $\vec{w}$ and the matrix $\mathbf{R}$, one can easily see that

$$\frac{\partial}{\partial w_i} = \vec{\xi}_i \cdot \nabla_v, \qquad (2.27)$$

with the eigenvector $\vec{\xi}_i$ corresponding to the eigenvalue λ_i. Now, from Eq. (2.4) and the definition of the D's we have

$$\partial_t u_l + \lambda_i \partial_x u_l = - \sum_{j=1}^{(n+m)} M_{lj} v_j + \lambda_i D_l, \qquad (2.28)$$

which implies that

$$\vec{\xi}_i \cdot \nabla_v (\partial_t u_l + \lambda_i \partial_x u_l) = \lambda_i \xi_{il} - \sum_{j=1}^{(n+m)} M_{lj} \xi_{ij}, \quad (2.29)$$

where ξ_{ij} is the j component of the vector $\vec{\xi}_i$, and where we have used the fact that the first n components of $\vec{v}$ are precisely the D's (remember that by construction $l \leq n$).

To finish the proof we now use the fact that the first n rows of the matrix $\mathbf{A}$ are given by the matrix $\mathbf{M}$, and also the fact that $\vec{\xi}_i$ is an eigenvector of $\mathbf{A}$ with eigenvalue λ_i, which implies

$$\sum_{j=1}^{(n+m)} M_{lj} \xi_{ij} = \sum_{j=1}^{(n+m)} A_{lj} \xi_{ij} = \lambda_i \xi_{il}, \qquad (2.30)$$

from which we finally find

$$\vec{\xi}_i \cdot \nabla_v (\partial_t u_l + \lambda_i \partial_x u_l) = 0. \qquad (2.31)$$

This completes the proof that condition (2.26) always holds (as long as the constraints $D_i = \partial_x u_i$ are satisfied), which in turn means that the term in square brackets in (2.22) does not contribute to the coefficients c_{iii}.

We want to make another important comment here: As the eigenvectors $\vec{\xi}_i$ diagonalizing the matrix $\mathbf{A}$ are obtained only up to an arbitrary rescaling, also the eigenfields w_i are not unique. In particular, any w_i can be multiplied by an arbitrary function of the u's to obtain $\tilde{w}_i = \Omega_i(\vec{u}) w_i$. However, since the v's are related to derivatives of the u's, such a rescaling will introduce new quadratic source terms, so one would in general not expect the coefficients c_{iii} to be invariant under rescalings of the eigenfields.

Remarkably, for the systems of the type (2.4) and (2.5) that we are interested in, it turns out that such rescalings of the eigenfields have no effect on the coefficients c_{iii}. The proof of this is again related to condition (2.26). In general, if we rescale the eigenfunctions as $\tilde{w}_i = \Omega_i(\vec{u}) w_i$, we will find that

$$\partial_t \tilde{w}_i + \lambda_i \partial_x \tilde{w}_i = \Omega(\partial_t w_i + \lambda_i \partial_x w_i)$$
$$+ w_i \sum_{l=1}^{n} \partial_{u_l} \Omega(\partial_t u_l + \lambda_i \partial_x u_l)$$
$$= \Omega q_{w_i} + w_i \sum_{l=1}^{n} \partial_{u_l} \Omega(\partial_t u_l + \lambda_i \partial_x u_l).$$
$$(2.32)$$

From this we see that, although the rescaling does introduce new quadratic terms, condition (2.26) guarantees that no new contributions to the coefficient c_{iii} will arise, i.e., the source criteria for avoiding blowups is invariant with respect to rescalings of the eigenfunctions (again, as long as the constraints $D_i = \partial_x u_i$ are satisfied).

3. Is the source criteria necessary and sufficient in order to avoid blowups?

A question one might immediately ask is whether the source criteria introduced above is necessary and sufficient to avoid blowups. Although we have no proof at this time, numerical experiments (such as those shown in later sections) indicate that whenever the source criteria is not satisfied blowups do develop, which would support our

conjecture that the criteria is indeed necessary in order to avoid blowups.

About it being sufficient, it clearly is not. This can be seen from the following example. Consider the system of two equations,

$$\partial_t v_1 + \partial_x v_2 = v_1^2 + v_2^2, \qquad (2.33)$$

$$\partial_t v_2 + \partial_x v_1 = -2v_1 v_2, \qquad (2.34)$$

which can easily be diagonalized to find

$$\partial_t w_+ + \partial_x w_+ = w_-^2, \qquad (2.35)$$

$$\partial_t w_- - \partial_x w_- = w_+^2, \qquad (2.36)$$

with $w_\pm := v_1 \pm v_2$. The diagonalized system clearly satisfies the source criteria. Now, take initial data such that $v_1 = k = \text{const}$ and $v_2 = 0$ in a large spatial region. This implies that in that region $w_1 = w_2 = k$. Since the spatial derivatives vanish and both fields are in fact equal, the equations above are of the form (2.17) and the fields will blow up in finite time (provided the region where the fields were initially equal is large enough). This initial data is clearly very special, but it does show that the source criteria is not sufficient in order to avoid a blowup. Nevertheless, for more generic data this situation will be very rare.

We have in fact performed numerical experiments with systems of the above form, but generalizing the source terms to

$$\partial_t w_\pm \pm \partial_x w_\pm = a_\pm w_\pm^2 + b_\pm w_+ w_- + c_\pm w_\mp^2. \quad (2.37)$$

When using Gaussians with a small amplitude of order $\mathcal{O}(\epsilon)$ as initial data for $w_\pm$, then whenever the coefficients $a_\pm$ are nonzero one typically finds that blowups occur on a time scale of order $\mathcal{O}(1/\epsilon)$. If, on the other hand, $a_\pm = 0$ and one has only mixed terms and/or terms quadratic in the other eigenfield in the sources, blowups again eventually develop, but now on a time scale of order $\mathcal{O}(1/\epsilon^2)$. Hence, if one is interested in propagating small perturbations, then satisfying the source criteria should allow one to obtain longer evolutions.

D. Relationship between the different blowup mechanisms

In order to understand the relationship between the geometric blowup and the ODE-mechanism, we will study a system of two variables constructed from the simple scalar conservation law (2.12) by introducing either the time or space derivative of the function u as an extra independent variable.

We start by introducing $D := \partial_x u$ as a new variable. One then obtains the system

$$\partial_t u = -a(u)D, \qquad (2.38)$$

$$\partial_t D + a(u)\partial_x D = -a'(u)D^2, \qquad (2.39)$$

where the evolution equation for D has been found by differentiating (2.12) with respect to x and exchanging the order of ∂_t and ∂_x.

As we are interested in studying solutions of the original scalar conservation law, but seen from a different perspective, will only consider initial data such that the constraint $D := \partial_x u$ is satisfied. Remembering that along a characteristic line of (2.12) u is constant [and hence also $a(u)$ and $a'(u)$], the equation

$$\frac{dD}{dt} = \partial_t D + a(u)\partial_x D = -a'(u)D^2 \qquad (2.40)$$

arising from (2.39) can be easily integrated. We find that, along the characteristic, the following relation holds:

$$D(t) = \frac{D_0}{1 + D_0 a'(u)t}. \qquad (2.41)$$

This clearly becomes infinite at a time t^* given by

$$t^* = -\frac{1}{D_0 a'(u)}. \qquad (2.42)$$

Let us now introduce $K := \partial_t u$ instead of D as an extra variable. We then obtain the system

$$\partial_t u = K, \qquad (2.43)$$

$$\partial_t K + a(u)\partial_x K = \frac{a'(u)}{a(u)} K^2. \qquad (2.44)$$

Here the evolution equation for K has been derived by taking a partial derivative with respect to t of (2.12). As before, by integrating the equation

$$\frac{dK}{dt} = \partial_t K + a(u)\partial_x K = \frac{a'(u)}{a(u)} K^2, \qquad (2.45)$$

along the characteristic one finds

$$K(t) = \frac{K_0}{1 - [K_0 a'(u)/a(u)]t}, \qquad (2.46)$$

which diverges at a time given by

$$t^* = \frac{a(u)}{K_0 a'(u)}. \qquad (2.47)$$

These two examples are nothing more than our original scalar equation (2.12) in disguise. However, they are in fact linearly degenerate by the definition given above as the only eigenvalue $a(u)$ is independent of D and K, respectively. They still give rise to a blowup, as they should, but this time the blowup appears through the ODE-mechanism instead of the original geometric blowup mechanism. Notice that, from (2.42) and (2.47), one can conclude that a condition for not having a blowup in finite time is $a'(u) = 0$, which is the same condition we found in Sec. II B 2 above. This shows clearly that what can be considered a

geometric blowup of a given variable u can always be reinterpreted as an ODE-type blowup of its derivatives, so both blowup mechanisms are closely related.

E. Indirect linear degeneracy

Linear degeneracy turns out to be insufficient for avoiding blowups in the particular case of system (2.4) and (2.5) for two reasons. The first reason has to do with the presence of nonvanishing source terms $\vec{q}_v$ and has been discussed in Sec. II C above. The other reason is simply the fact that for this type of systems the eigenvalues of the characteristic matrix $\mathbf{A}$ depend only the u's and not on the v's, which means that all eigenfields are linearly degenerate in a trivial way. We have already seen an example of this in the previous section when we considered the simple scalar conservation law and introduced derivatives as extra independent variables.

For this reason the concept of "indirect linear degeneracy" was introduced in [26]. This simply replaces the eigenvalue λ_i in Eq. (2.16) by its time derivative:

$$\nabla_v \dot{\lambda}_i \cdot \vec{\xi}_i = \frac{\partial \dot{\lambda}_i}{\partial w_i} = \sum_{j=1}^{(n+m)} \frac{\partial \dot{\lambda}_i}{\partial v_j} \frac{\partial v_j}{\partial w_i} = 0. \qquad (2.48)$$

This new condition yields nontrivial results for the system (2.4) and (2.5) if the time derivatives of the u's, appearing when differentiating λ_i with respect to time, depend on the corresponding w_i.

It is in fact not difficult to see where the indirect linear degeneracy condition comes from. Consider the system of two equations

$$\partial_t u = p(u, v, \partial_x u), \qquad (2.49)$$

$$\partial_t v + \lambda(u)\partial_x v = q(u, v, \partial_x u), \qquad (2.50)$$

with p linear in v and $\partial_x u$. We now extend the above system by introducing the variable $D := \partial_x u$. This means that the sources p and q are now functions of (u, v, D). The full system will then be

$$\partial_t u = p, \qquad (2.51)$$

$$\partial_t D - \partial_x p = 0, \qquad (2.52)$$

$$\partial_t v + \lambda(u)\partial_x v = q, \qquad (2.53)$$

which is exactly of the form (2.4) and (2.5). Let us for a moment assume that $q = 0$. In that case it is clear that v will be constant along the characteristics lines $x = x_0 + \lambda(u)t$. The simplest example is obtained when $\lambda(u) = u$ and $p = v - uD$, since then we find that along the characteristics $du/dt = v$ (provided that the constraint $D = \partial_x u$ remains satisfied). This means that along those lines we have $u = u_0 + vt$, so the characteristics have constant acceleration given by v (since u is the characteristic speed). If initially $v_0(x) = v(t = 0, x)$ has negative slope

in a given region, the characteristics are then guaranteed to cross (as lines behind accelerate faster than those in front). At the point where this happens the gradient of v will become infinite and we will have a blowup. For cases when p is a different function one cannot integrate the equations exactly, but the same general idea will hold. Of course, when the source term q is not zero one could imagine that q can be chosen in such a way as to avoid the crossing of characteristics, but such a choice would clearly not be generic. The only way to be sure that there will be no blowup is to ask for $\partial p / \partial v = 0$. Indirect linear degeneracy is simply the generalization of this condition to the case of a system with more equations.

The argument given above, however, is clearly not rigorous. Indirect linear degeneracy is therefore still a more or less *ad hoc* condition. Part of the reason for discussing it here is precisely to study its relevance in different cases by numerical experiments. As our results in the following sections show, indirect linear degeneracy and the source criteria often yield the same conditions for avoiding blowups. When they do not, the source criteria seems to be more important. Exploring the link between indirect linear degeneracy and the source criteria is something that should be further investigated, and we are currently working in that direction.

III. THE WAVE EQUATION WITH SOURCES AND A DYNAMIC WAVE SPEED

A. Blowup formation

As an example for the type of evolution systems studied in the previous sections we will consider the simple scalar wave equation with sources,

$$\partial_t^2 u - c^2(u)\partial_x^2 u = q(u, \partial_t u, \partial_x u). \qquad (3.1)$$

Here we allow the wave speed c to be a function of the wave function u. The source term q, on the other hand, can depend both on u and its first derivatives. Introducing $D = \partial_x u$ and $K = \partial_t u$, we can rewrite the wave equation as

$$\partial_t u = K, \qquad (3.2)$$

$$\partial_t D - \partial_x K = 0, \qquad (3.3)$$

$$\partial_t K - c^2 \partial_x D = q, \qquad (3.4)$$

which is of the form (2.4) and (2.5). One can readily verify that the eigenvalues of the characteristic matrix,

$$\mathbf{A} = \begin{pmatrix} 0 & -1 \\ -c^2 & 0 \end{pmatrix}, \qquad (3.5)$$

are $\lambda_\pm = \pm c$, with corresponding eigenfields (the normalization is chosen for convenience)

$$w_\pm = \tfrac{1}{2}(K \mp cD). \qquad (3.6)$$

The linear degeneracy criteria then is trivially satisfied since the eigenvalues depend only on u. However, when we calculate the time derivative of the eigenvalues we find

$$\dot{\lambda}_{\pm} = \pm c' \partial_t u = \pm c' K = \pm c'(w_+ + w_-). \quad (3.7)$$

The indirect linear degeneracy condition (2.48) asks for the derivatives $\partial \dot{\lambda}_{\pm}/\partial w_{\pm}$ to vanish, which implies

$$c' = 0. \quad (3.8)$$

This is no longer trivial and corresponds to what one would in fact expect: The wave speed should be independent of the wave function if we want no shocks to develop.

Let us now turn to the source criteria. We find

$$\frac{dw_{\pm}}{dt} = \frac{1}{2} q(u, w_{\pm}) + \frac{c'}{c}(w_+ w_- - w_{\mp}^2). \quad (3.9)$$

Notice that the evolution equation for a given eigenfield $w_{\pm}$ contains no quadratic terms on itself other than those that might come from q (it has only mixed terms and terms quadratic in the other eigenfield). The source criteria then demands that the source term q should be free of the quadratic terms w_+^2 and w_-^2. If we assume that q is of the form

$$q = \mathcal{A} c^2 D^2 + \mathcal{B} c D K + \mathcal{C} K^2$$
$$= (\mathcal{A} - \mathcal{B} + \mathcal{C}) w_+^2 - 2(\mathcal{A} - \mathcal{C}) w_+ w_-$$
$$+ (\mathcal{A} + \mathcal{B} + \mathcal{C}) w_-^2, \quad (3.10)$$

with $\mathcal{A}$, $\mathcal{B}$ and $\mathcal{C}$ arbitrary functions of u, it follows that in order to avoid blowups these functions have to satisfy

$$\mathcal{B} = 0, \qquad \mathcal{A} + \mathcal{C} = 0. \quad (3.11)$$

B. Numerical results

We have performed a series of numerical simulations for wave equations which satisfy or violate indirect linear degeneracy and/or the source criteria. The results from these simulations are summarized in Table I. All simulations have been performed using a method of lines with fourth order Runge-Kutta integration in time, and standard second order centered differences in space (with no artificial dissipation added[1]).

As initial data we have taken $u(t = 0) = 1$ and hence $D(t = 0) = 0$, together with the derivative of a Gaussian for the time derivative of u, i.e.,

$$K(t = 0) = -(2\kappa x/\sigma^2) \exp(-x^2/\sigma^2). \quad (3.12)$$

---

[1]Artificial dissipation (or some other more advanced technique like shock capturing) is crucial to follow shock waves, but here we do not want to go beyond shock formation, so we do not actually need it. In hydrodynamics there are physical principles that allow one to follow the solution past the shock formation (we know that physically there is no true discontinuity), but in this case there is no analog of that.

TABLE I. Summary of results from simulations of a wave equation that satisfies (Yes) or violates (No) both indirect linear degeneracy (i.l.d.) and the shock criteria (s.c.).

i.l.d.	s.c.	(c, q)	Result
No	No	$c = u,\ q = 2uD^2$	Blowup (Fig. 1)
No	Yes	$c = u,\ q = 2(uD^2 - K^2/u)$	Blowup? (Fig. 2)
Yes	No	$c = 1,\ q = 2D^2$	Blowup (Fig. 3)
Yes	Yes	$c = 1,\ q = 2(D^2 - K^2)$	No blowup (Fig. 4)

Here we used the derivative of a Gaussian and not a simple Gaussian in order to excite perturbations where u is both smaller and larger than its initial value.

For all the runs shown here we have used the particular values $\kappa = 0.1$ and $\sigma = 0.3$. These rather strong and localized perturbations are motivated by the fact that we wanted to see shock formation early, in particular, before the variable u changes sign (since the first two examples with eigenspeeds given by $\pm u$ are not strongly hyperbolic for $u = 0$). However, also for other values of κ and σ we have seen qualitatively very similar behavior (and whether or not u crosses zero does not seem to play a role in the formation of shocks).

For the runs with highest resolution we have used 80 000 grid points and a resolution of $\Delta x = 5 \times 10^{-4}$, which places the boundaries at ± 20, together with time steps of $\Delta t = \Delta x/2$. In addition, for each evolution variable we have computed the convergence factor η which, using three runs with high (u^h), medium (u^m) and low (u^l) resolutions differing in each case by a factor of 2, can be calculated as, e.g.,

$$\eta_u = \frac{\frac{1}{N_i} \sum_{i=1}^{N_i} |u_i^m - u_i^l|}{\frac{1}{N_j} \sum_{j=1}^{N_j} |u_j^h - u_j^m|}. \quad (3.13)$$

In the plots we show four different convergence factors: We denote with a triangle the convergence factors obtained when comparing runs with 80 000, 40 000 and 20 000 grid points and a spatial resolution of 5×10^{-4}, 10^{-3} and 2×10^{-3}. We use boxes, diamonds and stars to denote the convergence factors when gradually lowering all three resolutions by a factor of 2. For second order convergence we expect $\eta \simeq 4$.

The first test we have done corresponds to a case that violates both blowup criteria. We obtain such a system by simply taking a time derivative of Burgers' equation. We find $\partial_t^2 u - u^2 \partial_x^2 u = 2u(\partial_x u)^2$ and hence identify

$$c = u, \qquad q = 2uD^2. \quad (3.14)$$

Results for this simulation can be found in Fig. 1. In the left panel we show snapshots of the evolution of the variables u, D and K in steps of $\Delta t = 1$, and in the right panel we show convergence factors for the previously mentioned series of different resolutions. From the figure we clearly
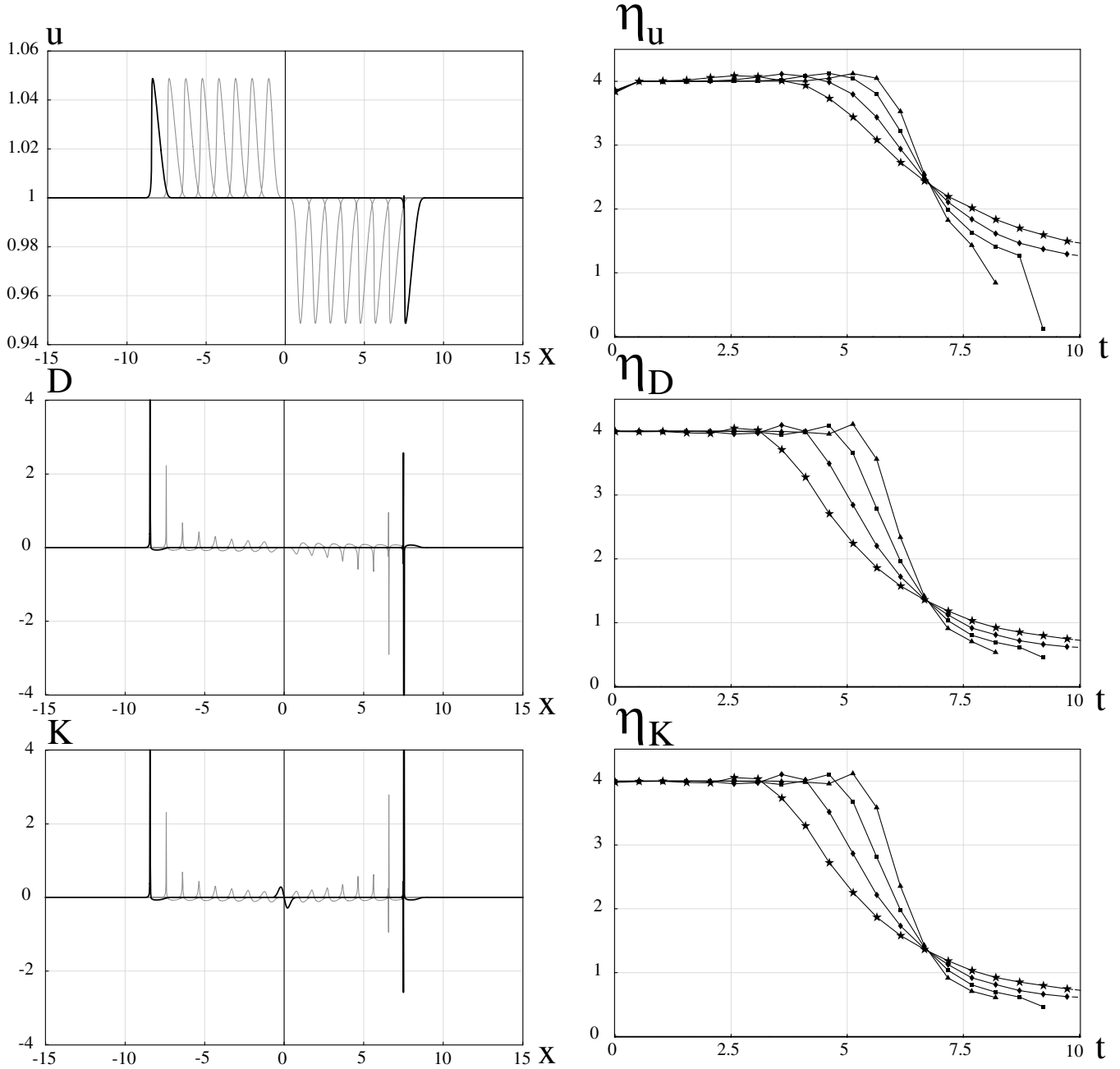
FIG. 1.   Results for the case when both criteria are violated. This simulation corresponds to the time derivative of Burgers' equation, for which we have $c = u$ and $q = 2uD^2$. In the left panel a sharp gradient is clearly visible on u, and large peaks can also be seen on D and K. In the right panel we see a clear loss of convergence that becomes sharper as resolution is increased in the order "star, diamond, box and triangle," indicating a blowup at a time $t \simeq 7$.

see that, as expected, shocks do form, with large gradients developing on u and large peaks on D and K. Moreover, from the convergence plots we see that there is a clear loss of convergence, and as the resolution is increased, this loss of convergence becomes more sharply centered around a specific time $t \simeq 7$, indicating that the blowup happens at this time.

For the second example we have chosen a situation where indirect linear degeneracy is violated but the source

criteria holds. Numerical results for the case

$$c = u, \qquad q = 2(uD^2 - K^2/u) \qquad (3.15)$$

are shown in Fig. 2 (a simpler case arising for a vanishing source term, $q = 0$, yields very similar results). The figure shows large peaks developing in both D and K, and a sharp gradient developing in u. The convergence plots show some loss of convergence for the lower resolutions but, in contrast with the previous example, convergence seems
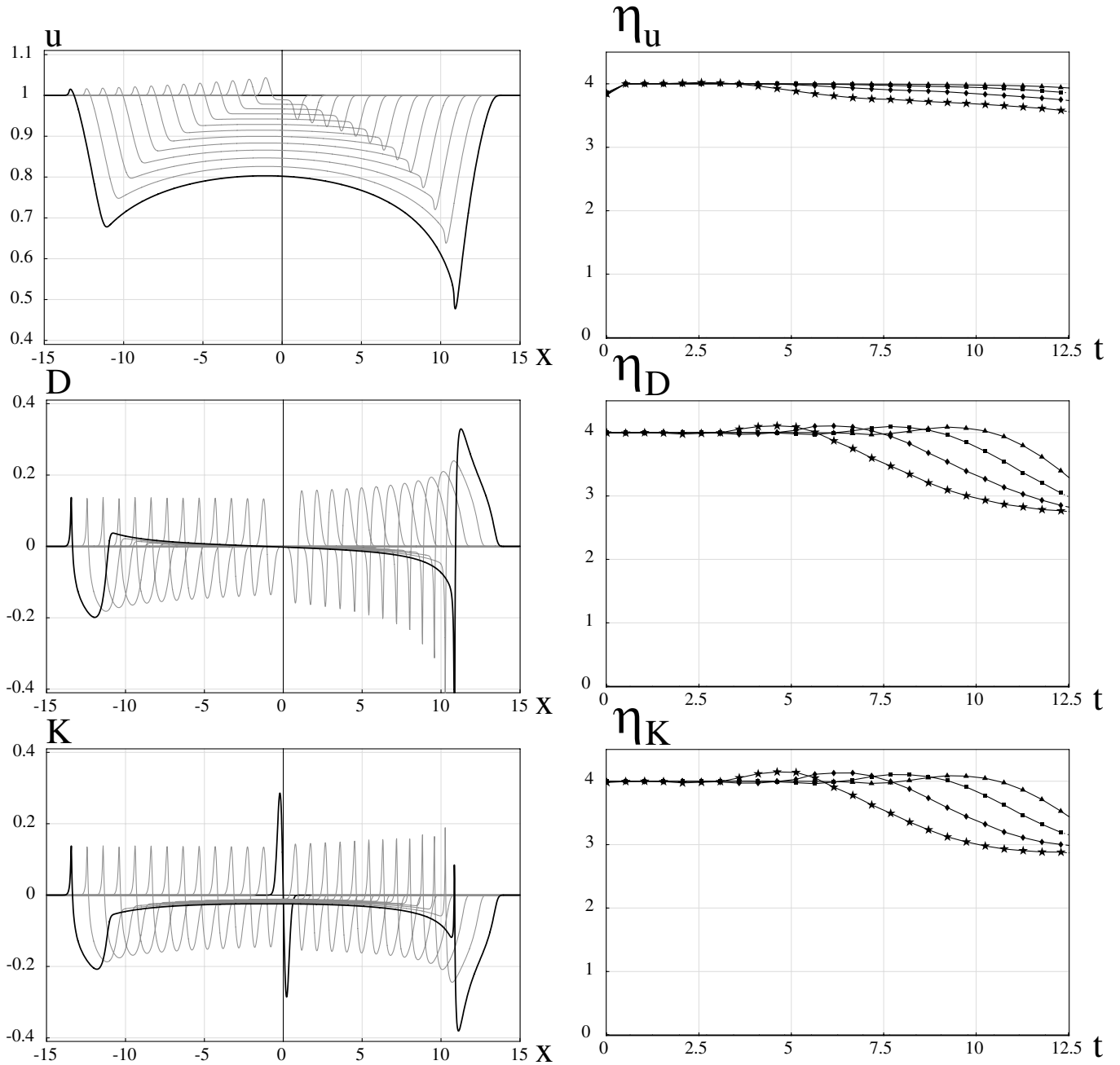
FIG. 2. Simulation for a case that violates indirect linear degeneracy, but satisfies the source criteria ($c = u$, $q = 2(uD^2 - K^2/u)$). The left panel shows that D and K develop large peaks, while u develops a sharp gradient but no peak. The convergence plots on the right panel indicate also loss of convergence in D and K, however convergence seems to improve with resolution.

to improve with resolution. This would seem to indicate that, although sharp gradients do develop, a real blowup has not occurred. Nevertheless, such sharp gradients are difficult to resolve numerically, so their presence is undesirable.

The third simulation corresponds to a case that satisfies indirect linear degeneracy, but violates the source criteria, with wave speed and source term given by

$$c = 1, \qquad q = 2D^2. \qquad (3.16)$$

Results for this run are shown in Fig. 3. As before, we see that both D and K are developing large peaks. The evolution variable u is developing both a large peak and a large gradient. The convergence plots show a loss of convergence at lower resolutions that improves as the resolution is increased. In this case, however, all runs crash at $t \simeq 7$, indicating that a blowup has indeed occurred at that time.

Finally, our last test corresponds to the case when both criteria are satisfied, with the wave speed c and source term q given by
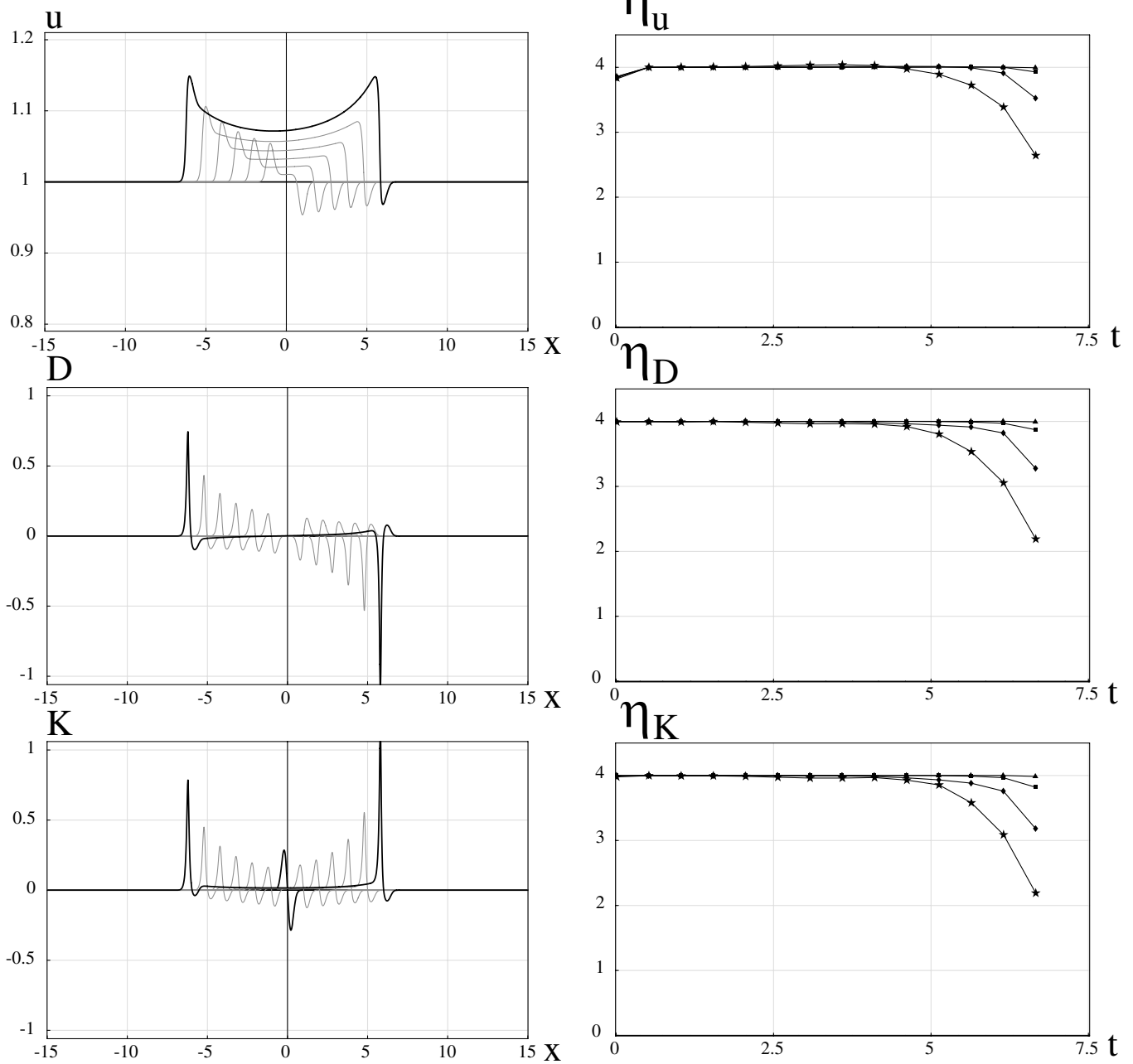
FIG. 3.   Results of a simulation for the case that satisfies indirect linear degeneracy, but violates the source criteria ($c = 1$, $q = 2D^2$). The left panel shows that D and K develop large peaks, while u develops both a peak and a large gradient. The right panel shows that convergence is lost at low resolutions but improves at higher resolutions, until a final time of $t \simeq 7$ when runs at all resolutions crash, indicating a true blowup.

$$c = 1, \qquad q = 2(D^2 - K^2). \qquad (3.17)$$

Results for this simulation are shown in Fig. 4. We see that the solution behaves in a wavelike manner, with no evidence of a blowup. This result is reinforced by the convergence plots indicating that we have close to second order convergence during the whole run for all resolutions considered, with no evidence of loss of convergence at any

time (notice the change in scale with respect to previous plots).

From the previous simulations it is clear that, for the scalar wave equation with a dynamic wave speed and sources, sharp gradients and blowups are only avoided when both indirect linear degeneracy and the source criteria are satisfied. In particular, the case with $c = 1$ and $q = 2(D^2 - K^2)$ behaves very similar to what one would expect
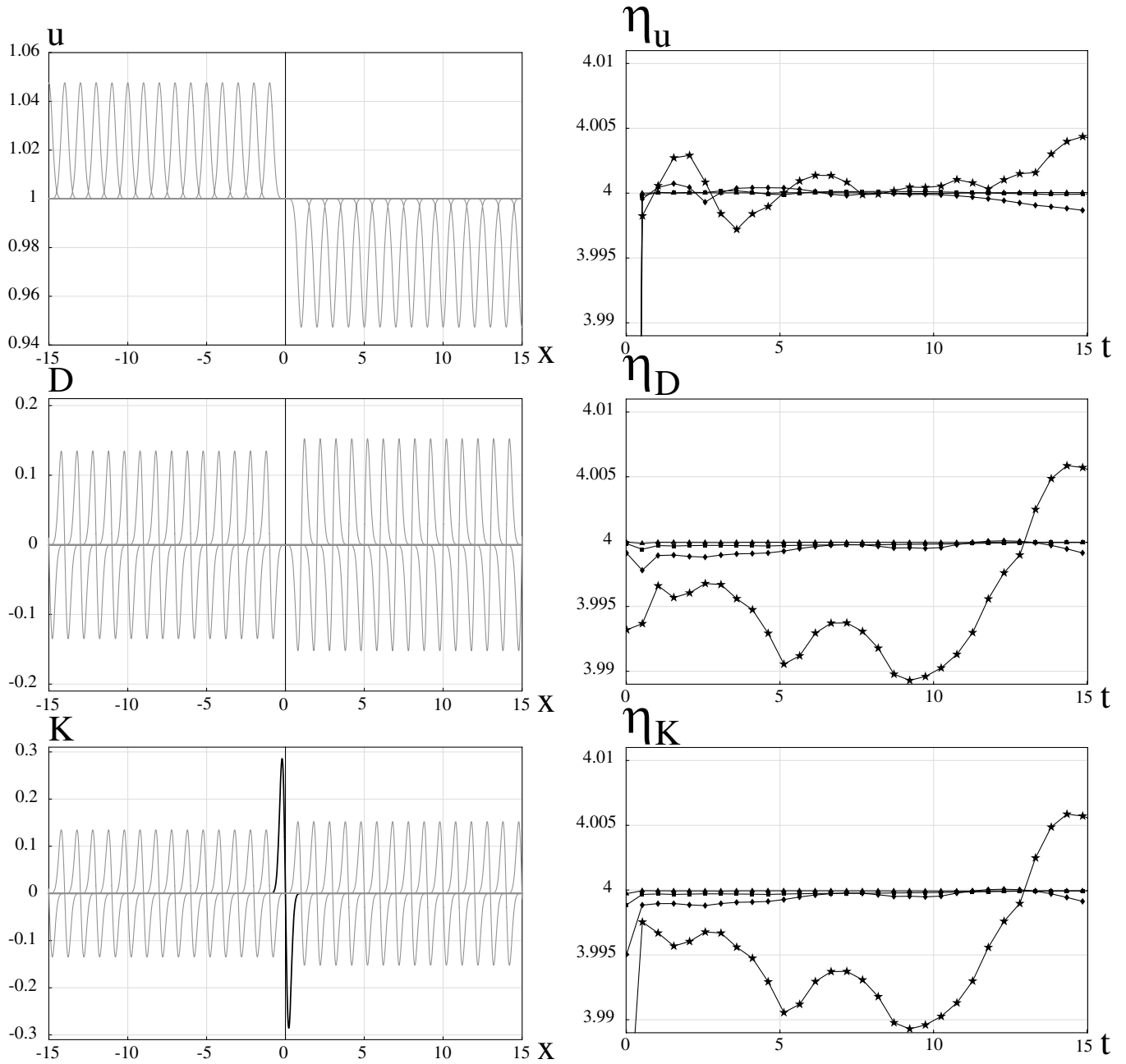
FIG. 4. Simulation for the case that satisfies both blowup avoiding criteria. In this case we have taken $c = 1$ and $q = 2(D^2 - K^2)$, and we find wavelike behavior with no evidence of a blowup.

from the standard wave equation with unit wave speed and a vanishing source term.

This observation can be easily understood by generalizing an example suggested by Nirenberg in [37]: Smooth solutions that extend globally in time will always exist if one can find a smooth transformation of the form $\tilde{u} = \tilde{u}(u)$, for which $\tilde{u}$ satisfies the standard wave equation, $\partial_t^2 \tilde{u} - \partial_x^2 \tilde{u} = 0$ (see [38,39]). One may readily verify that, for our particular example with wave speed and source term given by (3.17), this is indeed the case for the variable $\tilde{u} = \exp(2u)$.

IV. THE EINSTEIN EQUATIONS

In the previous sections we have described how blowups can be produced in systems of hyperbolic equations, and what conditions need to be satisfied in order for these to be avoided. We have also considered one simple example, the wave equation with sources and a dynamic wave speed. We will now turn our attention to the system we are most interested in, namely, the evolution equations of general relativity. In this paper, we will restrict ourselves to two cases, "toy" $1 + 1$ relativity, and spherically symmetric

relativity, and leave the important three-dimensional case for a future work.

We believe that it is important to mention here the main results which will be presented in this section. In the first place, we will recover the results regarding "gauge shocks" discussed already in [26,27,40]. But the most important result that we will show is the fact that, for the spherically symmetric case, one can also identify a second family of blowups that are not associated with the gauge but rather with the violation of the constraints. We will refer to such blowups as "constraint shocks," since they are clearly associated with the way in which the constraints have been added to the evolution equations. These constraint shocks will correspond to blowups in the Hamiltonian and momentum constraints at a finite time as the numerical example at the end of Sec. IV C 2 shows.[2]

Since the blowup analysis assumes that we have a strongly hyperbolic system, we will in each case begin by constructing such a hyperbolic system for the Einstein equations. Notice that there is no unique way to obtain hyperbolic evolution systems from the Einstein equations and we will use this fact to explicitly construct formulations that avoid constraint shocks.

A. Einstein equations in $1 + 1$ dimensions

Let us assume that we have standard general relativity in one spatial dimension. It is well known that in such a case the gravitational field is trivial and there are no true dynamics. However, one can still have nontrivial gauge dynamics that can be used as a simple example of the type of behavior one can expect in the higher dimensional case. We will start from the "standard" Arnowitt-Deser-Misner (ADM) equations for one spatial dimension [10], where by standard we mean the version of York [9]. Since we want a hyperbolic system of evolution equations that includes the gauge, we will use the Bona-Massó family of slicing conditions [11]:

$$\partial_t \alpha = -\alpha^2 f(\alpha) \, \mathrm{tr} \, K, \tag{4.1}$$

where $f = f(\alpha) > 0$ identifies the member of the Bona-Massó family being used (for example, $f = 1$ corresponds to harmonic slicing, and $f = 2/\alpha$ to the so-called $1 + \log$ slicing). For simplicity we will also restrict ourselves to the case of vanishing shift vector.

Following [26], the two-dimensional vector $\vec{u}$ will consist of the lapse function α and the spatial metric function $g := g_{xx}$ as components. The vector $\vec{v}$, on the other hand, is

[2]A question arises as to whether one could predict the constraint shocks by applying the source criteria directly to the constraint evolution system. The answer is that one cannot, since in the constraint evolution system the sources are linear in the constraints, so if we keep all coefficients constant no blowup would be expected (at worst one would have exponential growth). But this is inconsistent, as in fact the coefficients are not constant and depend on the metric and its derivatives.

a three-dimensional vector with components given by the logarithmic spatial derivatives of α and g, and the unique component of the extrinsic curvature (with mixed indices). That is,

$$\vec{u} = (\alpha, g), \qquad \vec{v} = (D_\alpha, D_g, K), \tag{4.2}$$

where

$$D_\alpha := \partial_x \ln\alpha, \qquad D_g := \partial_x \ln g, \qquad K := K^x_x. \tag{4.3}$$

(Note that in [26] the variable $D = \partial_x g/2$ is used instead of D_g and K_{xx} is used instead of K^x_x.) The fact that we define the D's as logarithmic derivatives instead of simple derivatives is in order to simplify the resulting equations, and makes no significant difference in the analysis of Sec. II.

The ADM evolution equations for the vectors $\vec{u}$ and $\vec{v}$ turn out to be

$$\partial_t \alpha = -\alpha^2 f K, \tag{4.4}$$

$$\partial_t g = -2\alpha g K, \tag{4.5}$$

and

$$\partial_t D_\alpha + \partial_x(\alpha f K) = 0, \tag{4.6}$$

$$\partial_t D_g + \partial_x(2\alpha K) = 0, \tag{4.7}$$

$$\partial_t K + \partial_x(\alpha D_\alpha/g) = \alpha(K^2 - D_\alpha D_g/2g). \tag{4.8}$$

This system has again the form (2.4) and (2.5). In particular, the last three equations can be written as $\partial_t \vec{v} + \mathbf{A}(\vec{u})\partial_x \vec{v} = \vec{q}_v$, where

$$\mathbf{A} = \begin{pmatrix} 0 & 0 & \alpha f \\ 0 & 0 & 2\alpha \\ \alpha/g & 0 & 0 \end{pmatrix}, \tag{4.9}$$

and

$$\vec{q}_v = \begin{pmatrix} -\alpha(f + \alpha f')D_\alpha K \\ -2\alpha D_\alpha K \\ \alpha(K^2 - D_\alpha^2/g + D_\alpha D_g/2g) \end{pmatrix}. \tag{4.10}$$

When studying the characteristic structure of the system of equations above we find the following eigenvalues

$$\lambda_0 = 0, \qquad \lambda_\pm^f = \pm\alpha\sqrt{\frac{f}{g}}, \tag{4.11}$$

with corresponding eigenfunctions

$$w_0 = D_\alpha - \frac{f}{2}D_g, \tag{4.12}$$

$$w_\pm^f = \sqrt{fg}K \pm D_\alpha. \tag{4.13}$$

The system is therefore strongly hyperbolic as long as $f > 0$, with one eigenfield propagating along the time lines and the other two propagating with the "gauge speeds"

$\lambda_\pm^f = \pm\alpha\sqrt{f/g}.$

In order to study the possible formation of shocks for the propagating eigenfields, one can immediately see that the direct linear degeneracy criteria as formulated in (2.16) cannot be used, since $\lambda_\pm^f$ does not depend on either D_α, D_g or K. The indirect linear degeneracy condition, however, yields

$$\frac{\partial \dot\lambda_\pm^f}{\partial w_\pm^f} = \sum_{j=1}^{(n+m)} \frac{\partial \dot\lambda_\pm^f}{\partial v_j} \frac{\partial v_j}{\partial w_\pm^f} = \pm\frac{\alpha^2}{2g}\left(1 - f - \frac{\alpha f'}{2}\right) = 0,$$

(4.14)

where the last step comes from expressing the time derivatives of α and g contained in $\dot\lambda_\pm^f$ in terms of K using (4.4) and (4.5).

For the source criteria, on the other hand, we need to determine the term quadratic in $w_\pm^f$ in the source terms associated to the evolution equation for $w_\pm^f$ itself. We find

$$\frac{dw_\pm^f}{dt} = \frac{\alpha}{2\sqrt{fg}}\left(1 - f - \frac{\alpha f'}{2}\right)w_\pm^{f2} + \mathcal{O}(w_0 w_\pm^f, w_\pm^f w_\mp^f).$$

(4.15)

Asking for the coefficient of the quadratic term to be zero, one finds

$$c_{\pm\pm\pm}^{fff} = \frac{\alpha}{2\sqrt{fg}}\left(1 - f - \frac{\alpha f'}{2}\right) = 0.$$

(4.16)

It is interesting to note that here both indirect linear degeneracy and the source criteria yield the same condition for avoiding blowups, namely

$$1 - f - \frac{\alpha f'}{2} = 0.$$

(4.17)

The reason why this is so is not completely clear, but it probably implies that in this case the sources and characteristic speeds are not independent of each other.

The shock avoiding condition (4.17) has been studied before in [26,27,40]. Its general solution is

$$f(\alpha) = 1 + \frac{\text{const}}{\alpha^2}.$$

(4.18)

Reference [27] also considers some approximate solutions that are more useful for numerical simulations. Notice also that, since in this simple case we only have gauge dynamics, these shocks are directly associated with the foliation itself, and for this reason they are known as "gauge shocks."

B. Einstein equations in spherical symmetry

1. Standard ADM equations

In order to generalize the previous system to spherical symmetry, we start with the spatial line element

$$dl^2 = A(t, r)dr^2 + r^2 B(t, r)d\Omega^2,$$

(4.19)

where $d\Omega^2 = d\theta^2 + \sin^2\theta d\phi^2$ denotes the usual solid angle element. Note that in this case the vector $\vec{u}$ will consist of the lapse α and the metric components A and B. Furthermore, the v's are given by the spatial derivatives of these quantities

$$D_\alpha = \partial_r \ln\alpha, \qquad D_A = \partial_r \ln A, \qquad D_B = \partial_r \ln B,$$

(4.20)

together with the extrinsic curvature variables (again with mixed indices)

$$K_A = K_r^r, \qquad K_B = K_\theta^\theta = K_\phi^\phi.$$

(4.21)

That is,

$$\vec{u} = (\alpha, A, B), \qquad \vec{v} = (D_\alpha, D_A, D_B, K_A, K_B).$$

(4.22)

In the following we will assume for simplicity that we are in vacuum and that the shift vector vanishes. Using the Bona-Massó slicing condition (4.1), the evolution equations for the u's become

$$\partial_t \alpha = -\alpha^2 f(K_A + 2K_B),$$

(4.23)

$$\partial_t A = -2\alpha A K_A,$$

(4.24)

$$\partial_t B = -2\alpha B K_B.$$

(4.25)

The evolution equation for the v's can be obtained directly from the ADM equations and the definition of the D's. These equations can again be written in the form $\partial_t \vec{v} + \mathbf{A}(\vec{u})\partial_r \vec{v} = \vec{q}_v$, where the characteristic matrix is

$$\mathbf{A} = \begin{pmatrix} 0 & 0 & 0 & \alpha f & 2\alpha f \\ 0 & 0 & 0 & 2\alpha & 0 \\ 0 & 0 & 0 & 0 & 2\alpha \\ \alpha/A & 0 & \alpha/A & 0 & 0 \\ 0 & 0 & \alpha/2A & 0 & 0 \end{pmatrix},$$

(4.26)

and the source terms are given by

$$q_{D_\alpha} = -\alpha(f + \alpha f')D_\alpha(K_A + 2K_B),$$

(4.27)

$$q_{D_A} = -2\alpha D_\alpha K_A,$$

(4.28)

$$q_{D_B} = -2\alpha D_\alpha K_B,$$

(4.29)

$$q_{K_A} = -\frac{\alpha}{A}\left[D_\alpha\left(D_\alpha - \frac{D_A}{2}\right) - \frac{D_B}{2}(D_A - D_B)\right.$$
$$\left. - AK_A(K_A + 2K_B) - \frac{1}{r}(D_A - 2D_B)\right],$$

(4.30)

$$q_{K_B} = -\frac{\alpha}{2A}\left[D_B\left(D_\alpha - \frac{D_A}{2} + D_B\right) - 2AK_B(K_A + 2K_B)\right.$$
$$\left. + \frac{1}{r}(2D_\alpha - D_A + 4D_B) - \frac{2}{r^2 B}(A - B)\right].$$

(4.31)

Furthermore, the Hamiltonian and momentum constraints

take the form

$$C_h := -\partial_r D_B + \frac{D_B}{2}\left(D_A - \frac{3D_B}{2}\right) + AK_B(2K_A + K_B)$$
$$+ \frac{D_A - 3D_B}{r} + \frac{A - B}{r^2 B} = 0, \tag{4.32}$$

$$C_m := -\partial_r K_B + \left(\frac{D_B}{2} + \frac{1}{r}\right)(K_A - K_B) = 0. \tag{4.33}$$

2. ADM equations in new variables

Rather than working with the standard ADM equations described in the last section, following [41] we will introduce the "antitrace" of the spatial derivatives of the metric components, $D = D_A - 2D_B$, and the trace of the extrinsic curvature, $K = K_A + 2K_B$, as fundamental variables instead of D_A and K_A. This choice of variables makes the hyperbolicity analysis more transparent. The vector $\vec{v}$ will then be

$$\vec{v} = (D_\alpha, D, D_B, K, K_B), \tag{4.34}$$

and the evolution of the u's will be given by

$$\partial_t \alpha = -\alpha^2 f K, \tag{4.35}$$

$$\partial_t A = -2\alpha A(K - 2K_B), \tag{4.36}$$

$$\partial_t B = -2\alpha B K_B. \tag{4.37}$$

For the v's, one again obtains a system of the form $\partial_t \vec{v} + \mathbf{A}(\vec{u})\partial_r \vec{v} = \vec{q}_v$, but this time with characteristic matrix

$$\mathbf{A} = \begin{pmatrix} 0 & 0 & 0 & \alpha f & 0 \\ 0 & 0 & 0 & 2\alpha & -8\alpha \\ 0 & 0 & 0 & 0 & 2\alpha \\ \alpha/A & 0 & 2\alpha/A & 0 & 0 \\ 0 & 0 & \alpha/2A & 0 & 0 \end{pmatrix}, \tag{4.38}$$

and source terms

$$q_{D_\alpha} = -\alpha(f + \alpha f')D_\alpha K, \tag{4.39}$$

$$q_D = -2\alpha D_\alpha(K - 4K_B), \tag{4.40}$$

$$q_{D_B} = -2\alpha D_\alpha K_B, \tag{4.41}$$

$$q_K = -\frac{\alpha}{A}\left[D_\alpha\left(D_\alpha - \frac{D}{2}\right) - D_B\left(D + \frac{D_B}{2}\right) - AK^2 \right.$$
$$\left. + \frac{2}{r}(D_\alpha - D + D_B) - \frac{2}{r^2 B}(A - B)\right], \tag{4.42}$$

$$q_{K_B} = -\frac{\alpha}{2A}\left[D_B\left(D_\alpha - \frac{D}{2}\right) - 2AKK_B \right.$$
$$\left. + \frac{1}{r}(2D_\alpha - D + 2D_B) - \frac{2}{r^2 B}(A - B)\right]. \tag{4.43}$$

Finally, the Hamiltonian and momentum constraints take the form

$$C_h := -\partial_r D_B + \frac{D_B}{2}\left(D + \frac{D_B}{2}\right) + AK_B(2K - 3K_B)$$
$$+ \frac{D - D_B}{r} + \frac{A - B}{r^2 B} = 0, \tag{4.44}$$

$$C_m := -\partial_r K_B + \left(\frac{D_B}{2} + \frac{1}{r}\right)(K - 3K_B) = 0. \tag{4.45}$$

3. Modifying the equations by using the constraints

It turns out that, as they stand, neither the matrix $\mathbf{A}$ of the original ADM system, (4.26), nor the one of the rewritten system, (4.38), has a complete set of eigenvectors for all $f > 0$, so the systems of evolution equations are not strongly hyperbolic. Interestingly, strong hyperbolicity only fails for $f = 1$, which corresponds to harmonic slicing (this only happens in spherical symmetry, in the full 3-dimensional case strong hyperbolicity fails for ADM much more severely). Since harmonic slicing is such an important condition for both theoretical and practical reasons, the systems described above are not very useful.

Let us concentrate on the second system of evolution equations. By making use of the constraint equations, (4.44) and (4.45), its principal part can be modified to construct a strongly hyperbolic system for all $f > 0$. In particular, adding multiples of the constraints will modify the third and fifth columns of the matrix (4.38). We will consider adjustments to the evolution equations for the v's of the form

$$\partial_t v_i + \sum_{j=1}^{(n+m)} A_{ij}\partial_r v_j + \alpha(h_i A^{\mathcal{H}_i} C_h + m_i A^{\mathcal{M}_i} C_m) = q_i. \tag{4.46}$$

Here the terms $\{h_i, m_i\}$ are allowed to depend on $f(\alpha)$, such that for harmonic slicing these coefficients reduce to constants. Furthermore, the exponents $\{\mathcal{H}_i, \mathcal{M}_i\}$ are fixed by looking at the characteristic matrix and demanding that for $f = 1$ its entries have homogeneous powers of A. Doing this we find

$$\mathcal{H}_{D_\alpha} = \mathcal{H}_D = \mathcal{H}_{D_B} = -1/2, \tag{4.47}$$

$$\mathcal{H}_K = \mathcal{H}_{K_B} = -1, \tag{4.48}$$

$$\mathcal{M}_{D_\alpha} = \mathcal{M}_D = \mathcal{M}_{D_B} = 0, \tag{4.49}$$

$$\mathcal{M}_K = \mathcal{M}_{K_B} = -1/2, \tag{4.50}$$

and the characteristic matrix then takes the general form

$$\mathbf{A} = \begin{pmatrix} 0 & 0 & \alpha h_{D_\alpha}/A^{1/2} & \alpha f & \alpha m_{D_\alpha} \\ 0 & 0 & \alpha h_D/A^{1/2} & 2\alpha & \alpha(m_D - 8) \\ 0 & 0 & \alpha h_{D_B}/A^{1/2} & 0 & \alpha(2 + m_{D_B}) \\ \alpha/A & 0 & \alpha(2 + h_K)/A & 0 & \alpha m_K/A^{1/2} \\ 0 & 0 & \alpha(1/2 + h_{K_B})/A & 0 & \alpha m_{K_B}/A^{1/2} \end{pmatrix}.$$

$$(4.51)$$

We now need to determine the coefficients h_i and m_i in order to obtain a well behaved system of evolution equations.

4. Integrability and hyperbolicity

Since the D's arise as spatial derivatives of the lapse and metric components, their evolution equations are obtained by taking a time derivative of their definition and then changing the order of the partial derivatives. If one later adds multiples of the Hamiltonian and momentum constraints to the evolution equations for the D's, one finds that whenever these constraints are violated the D's in fact cease to be derivatives of metric functions. One consequence of this is the fact that, in the sources of the evolution equations for the eigenfields w_i, the coefficients c_{iii} will no longer be invariant under rescalings of the form $\tilde{w}_i = \Omega(\alpha, A, B)w_i$ (the proof presented at the end of Sec. II C 2 of the invariance of these coefficients under such rescalings relied on the derivative constraints being satisfied). Such a property of our system of equations is undesirable, as it makes the source criteria for avoiding blowups impossible to apply in practice.

This leads us to the "integrability criteria," which states that the D's should remain derivatives of the metric functions independently of the constraints, and implies that we must set the h_D's and m_D's to zero and consider only adjustments in the evolution equations of the extrinsic curvature variables.

Doing this we obtain for the characteristic matrix (4.51) the following eigenvalues

$$\lambda_0 = 0, \tag{4.52}$$

$$\lambda_\pm^f = \pm\alpha\sqrt{\frac{f}{A}}, \tag{4.53}$$

$$\lambda_\pm^c = \frac{\alpha}{2\sqrt{A}}(m_{K_B} \pm \sqrt{4 + 8h_{K_B} + m_{K_B}^2}). \tag{4.54}$$

The first three eigenvalues (λ_0, $\lambda_\pm^f$) are precisely the ones we found for the 1 + 1 dimensional case, so clearly $\lambda_\pm^f$ are again gauge speeds (they depend on the gauge function f). The last two eigenvalues depend on the choices we have made to add constraints to the evolution equations, so we will call them "constraint speeds." In fact, it is not surprising that we find only characteristic speeds associated with the gauge and the constraints, since in spherical symmetry

it is well known that there are no gravitational waves, i.e., no physical waves propagating at the speed of light.

From the last expressions we see that if we want the constraint speeds to be centered along the time lines we must ask for

$$m_{K_B} = 0. \tag{4.55}$$

If we now rewrite the parameter h_{K_B} as

$$h_{K_B} = -\frac{1}{2} + \frac{\mu^2}{2}f, \tag{4.56}$$

with μ a constant, then the eigenvalues $\lambda_\pm^c$ take the following simple form:

$$\lambda_\pm^c = \pm\mu\alpha\sqrt{\frac{f}{A}}. \tag{4.57}$$

At this point, only the adjustments to the evolution equation of the trace of the extrinsic curvature, h_K and m_K, and the constant μ remain as free parameters. It is now not difficult to show that the system of evolution equations will be strongly hyperbolic, i.e., the matrix (4.51) can be diagonalized and a complete set of eigenvectors exists, as long as $f > 0$ and $\mu \notin \{0, \pm1\}$. We want to point out here that the latter condition implies that all characteristic speeds have to be distinct. However, there is one important exception, since for $|\mu| = 1$ the adjustments $h_K = -2$ together with $m_K = 0$ also yield a strongly hyperbolic system (which will be of some importance later). Furthermore we observe that no particular problems arise for the case of harmonic slicing ($f = 1$). For completeness we explicitly state here the general form of the eigenfields:

$$w_0 = D_\alpha - \frac{f}{2}D - 2fD_B, \tag{4.58}$$

$$\begin{aligned} w_\pm^f = &\sqrt{f}(1 - \mu^2)[D_\alpha \pm \sqrt{fA}K] \\ &+ \left[\sqrt{f}(2 + h_K) \pm \frac{m_K f \mu^2}{2}\right]D_B \\ &\pm [2(2 + h_K) \pm m_K\sqrt{f}]\sqrt{A}K_B, \end{aligned} \tag{4.59}$$

$$w_\pm^c = \mu\sqrt{f}D_B \pm 2\sqrt{A}K_B. \tag{4.60}$$

In the case when $|\mu| = 1$, $h_K = -2$ and $m_K = 0$, the eigenfields $w_\pm^f$ and $w_\pm^c$ take a different form given by

$$w_\pm^f = \sqrt{fA}K \pm D_\alpha, \tag{4.61}$$

$$w_\pm^c = \sqrt{f}D_B \pm 2\sqrt{A}K_B. \tag{4.62}$$

5. Indirect linear degeneracy

Just as before, the linear degeneracy criteria is trivially satisfied. Applying the indirect linear degeneracy criteria, $\partial \lambda_i/\partial w_i = 0$, to the eigenvalues $\lambda_\pm^f$, we find

$$\frac{\partial \dot{\lambda}_{\pm}^f}{\partial w_{\pm}^f} = \frac{\alpha^2}{2(1 - \mu^2)\sqrt{f}A}\left(1 - f - \frac{\alpha f'}{2}\right). \qquad (4.63)$$

This gives us the same condition on f for avoiding a blowup as before, namely, condition (4.17), which is precisely the result we obtained for the $1 + 1$ dimensional case. In addition, however, we note that blowups can also arise from the second pair of eigenvalues, for which we find

$$\frac{\partial \dot{\lambda}_{\pm}^c}{\partial w_{\pm}^c} = -\frac{\mu \alpha^2}{4\sqrt{f}A}\left[\left(1 - f - \frac{\alpha f'}{2}\right)\right.$$
$$\left. \times \left(\frac{4 + 2h_K \pm \mu m_K \sqrt{f}}{1 - \mu^2}\right) + 2f\right]. \qquad (4.64)$$

The condition for avoiding these blowups is then

$$\mu\left[\left(1 - f - \frac{\alpha f'}{2}\right)\left(\frac{4 + 2h_K \pm \mu m_K \sqrt{f}}{1 - \mu^2}\right) + 2f\right] = 0. \qquad (4.65)$$

The first thing to notice is that we clearly must take

$$m_K = 0, \qquad (4.66)$$

in which case the condition reduces to

$$\mu\left[\left(1 - f - \frac{\alpha f'}{2}\right)\left(\frac{4 + 2h_K}{1 - \mu^2}\right) + 2f\right] = 0. \qquad (4.67)$$

Now, if we insert a member of the gauge shock avoiding family into this condition, we find

$$\mu f = 0, \qquad (4.68)$$

which, remembering that for strong hyperbolicity one must have $f > 0$ and $\mu \neq 0$, brings us to the rather discouraging result that—for the adjustments considered here—we cannot avoid at the same time both gauge shocks and constraint shocks coming from the indirect linear degeneracy criteria.

Instead of using a full blown member of the gauge shock avoiding family, we could be less ambitious and use a solution that avoids gauge shocks only approximately. For example, in Ref. [27] it was shown that the standard $1 + \log$ slicing, corresponding to the choice $f = 2/\alpha$, avoids gauge shocks to first order (which explains why it is so robust in practice). Taking such a form of f we find that the condition for avoiding constraint shocks is

$$h_K = -2\frac{\alpha - \mu^2}{\alpha - 1}. \qquad (4.69)$$

For $|\mu| \approx 1$ this condition implies $h_K \approx -2$ which, as we have seen in the previous section, is the only value of h_K that yields a strongly hyperbolic system when $|\mu| = 1$. However, as mentioned before, for the special case $\{|\mu| = 1, h_K = -2\}$ the eigenfields are different and the analysis should be performed separately. When we apply the indirect linear degeneracy criteria to the eigenfields $w_{\pm}^c$

given by (4.62), we find that $\partial \dot{\lambda}_{\pm}^f/\partial w_{\pm}^f$ is nonzero, and since in this case we have no free parameters, it is hard to say if the condition is "closely satisfied" or not. Still, Eq. (4.69) does seem to indicate that the combination $\{|\mu| = 1, h_K = -2\}$ is preferred by indirect linear degeneracy. We will return to this case when we consider some numerical examples below.

6. Source criteria

The source criteria for the gauge eigenfields $w_{\pm}^f$ yields the same condition on f as indirect linear degeneracy since the quadratic coefficient turns out to be

$$c_{\pm\pm\pm}^{fff} = \pm\frac{\alpha}{2(1 - \mu^2)f\sqrt{A}}\left(1 - f - \frac{\alpha f'}{2}\right). \qquad (4.70)$$

On the other hand, for the constraint eigenfields $w_{\pm}^c$ we find the following quadratic coefficient:

$$c_{\pm\pm\pm}^{ccc} = \pm\frac{(1 + \mu^2 f)\alpha}{16\mu^2 f\sqrt{A}}(7 + 4h_K - 3\mu^2 f \pm 2\mu m_K \sqrt{f}). \qquad (4.71)$$

If we want to avoid a blowup this coefficient must vanish,

$$7 + 4h_K - 3\mu^2 f \pm 2\mu m_K \sqrt{f} = 0. \qquad (4.72)$$

This can be accomplished for any f if we choose

$$m_K = 0, \qquad (4.73)$$

and

$$h_K = -2 + \frac{1 + 3\mu^2 f}{4}. \qquad (4.74)$$

We first notice that again we obtain the condition $m_K = 0$, just as we found in the previous section. It is also interesting to notice that for a given choice of f we have a one-parameter family of solutions that avoids this type of shocks. From (4.56) and (4.74) we see that the parameter μ relates h_K and h_{K_B} according to

$$h_K = -1 + \frac{3h_{K_B}}{2}. \qquad (4.75)$$

Considering the restrictions on μ imposed by strong hyperbolicity, $\mu \notin \{0, \pm 1\}$, we see that in the (h_K, h_{K_B}) plane this shock avoiding family must be such that

$$h_{K_B} > -\frac{1}{2}, \qquad h_{K_B} \neq \frac{1}{2}(f - 1). \qquad (4.76)$$

We see hence that by appropriate choices of $f(\alpha)$ and suitable adjustments to the evolution equations, it is possible to avoid at the same time the gauge and constraint shocks identified by the source criteria.

As a final comment it is important to point out that, in contrast to what we found for the gauge shocks, in this case indirect linear degeneracy and the source criteria yield different statements for avoiding constraint shocks.

C. Numerical tests in spherical symmetry

We will now describe some numerical experiments designed to test the shock avoiding conditions found in the previous sections. We will concentrate on two different types of tests: the robust stability test [42], and a test of Minkowski initial data with a violation of the constraints added by hand.

1. Robust stability test

As a first numerical experiment we have performed the robust stability test as described in [42]. For this test, one takes Minkowski initial data and adds random noise with a small amplitude to all dynamical variables. For the evolution we have used both harmonic slicing with $f = 1$ and standard $1 + \log$ slicing with $f = 2/\alpha$.

For the simulations discussed below we used 1000 grid points, a grid spacing of $\Delta r = 0.001$ and a time step $\Delta t = \Delta r/2$. Furthermore, we have demanded periodic boundary conditions and have set all $1/r$ and $1/r^2$ terms to zero by hand (i.e., we are performing the run "at infinity"). Setting these terms to zero should have no important effect on the

presence of blowups since one can readily verify that such terms are only linear in the eigenfields. Moreover, one could easily regularize the equations at the origin by following the procedure described in [41], but doing so would only obscure the study of blowups by mixing them with purely geometric effects.

We start the simulations with Minkowski initial data plus random noise of order 10^{-6} on all evolution variables. At latter times we compute the error δ as the average absolute value over the grid of the quantity $(\sum_{i=1}^{3} |u_i - 1| + \sum_{i=1}^{5} |v_i|)/8$.

We first performed runs for the ADM system without adding constraints, and observed the well known growth in the average error caused by the fact that ADM is only weakly hyperbolic. Next, we implemented the μ-family given by (4.56) and (4.74), which according to the source criteria is shock avoiding. Figure 5 shows the behavior of the average error for this family. In the top panel we plot the growth of the error for the case of harmonic slicing as a function of time (measured in units of the light-crossing time of our computational grid), for several different values of $|\mu|$. We see that for ADM (which is not a member of the
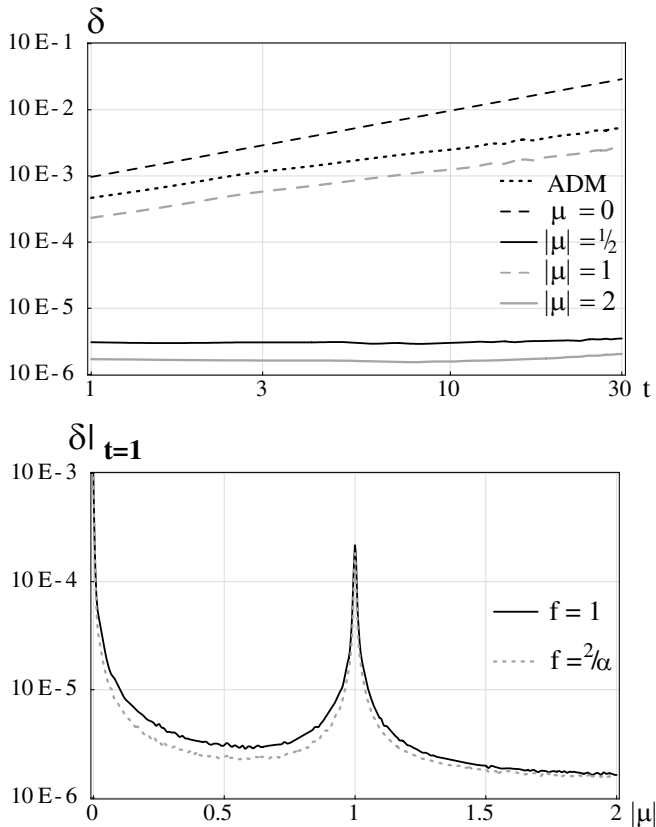

FIG. 5. *Top.* Using harmonic slicing, the average error for the robust stability test is shown for ADM and different members of the μ-family as a function of time (measured in units of the light-crossing time of our grid). *Bottom.* The value of the average error after one light-crossing time is plotted as a function of the parameter μ.
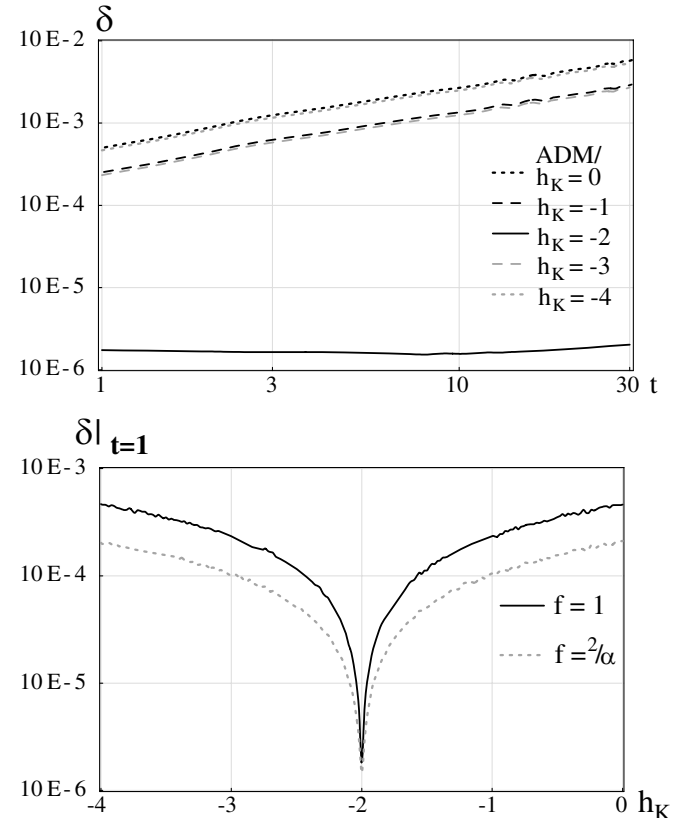
FIG. 6. *Top:* Setting $|\mu| = 1$, the average error for harmonic slicing is found to grow only moderately for $h_K = -2$, whereas for other values of h_K (and, in particular, for ADM corresponding to $h_K = 0$) a rapid growth is observed. *Bottom:* The average error after one crossing time is plotted as a function of h_K.

μ-family), and for the cases with $\mu = 0$ and $|\mu| = 1$, the error grows linearly with time, while for $|\mu| = 1/2$ and $|\mu| = 2$ the error initially does not grow at all (at later times, however, also in these cases a linear growth with a very small gradient develops). The lower panel in this figure shows the value of the average error after one light-crossing time as a function of $|\mu|$, as obtained both for harmonic and $1 + \log$ slicings. We see that for values of $|\mu|$ close to 0 or 1, the error is already very large after one light-crossing time, while away from these values the error remains small. The poor behavior of the simulations with $\mu = 0$ and $|\mu| = 1$ is caused by the fact that for such cases the evolution system is not strongly hyperbolic. Also, from (4.57) we see that for values of $|\mu|$ close to but different from either 0 or 1, the eigenspeeds $\lambda_{\pm}^c$ associated with the constraint modes become very similar to either λ_0 or $\lambda_{\pm}^f$. This means that even if the system is still strongly hyperbolic, the argument used for ignoring mixed terms in the sources will not apply.

In particular, for the case $|\mu| = 1$ the adjustment $h_K = -2$ suggested by the indirect linear degeneracy criteria turned out to be helpful. Because of this, in Fig. 6 we also show similar plots testing different values of the parameter h_K in the case of $|\mu| = 1$. Here one can observe that, for values of h_K other than -2, and, in particular, for ADM corresponding to $h_K = 0$, a linear growth in the average error is present. For the adjustment suggested by the indirect linear degeneracy criteria (which is the only value here which yields a strongly hyperbolic system), initially no error growth is found.

2. Minkowski initial data plus constraint violation

As we saw in the previous section, the robust stability test is very good at distinguishing between strongly and weakly hyperbolic systems, but does not show any clear indication that, among strongly hyperbolic systems, those that avoid shocks are better behaved. This should not be surprising as the robust stability test uses random, uncorrelated and nonsmooth initial data, while shock formation is the result of the coherent evolution of smooth initial data.

For this reason, using harmonic slicing we have performed evolutions of Minkowski initial data with a smooth violation of the constraints added by hand. Here we have chosen a perturbation in K_B similar to the one we used for the scalar wave equation, namely

$$K_B(t = 0) = -(2\kappa r/\sigma^2) \exp(-r^2/\sigma^2), \qquad (4.77)$$

with parameters $\kappa = 0.05$ and $\sigma = 1$.

For the simulations shown below we used 8000 grid points and a grid spacing of $\Delta r = 0.01$ (which places the boundaries at ± 40) together with a time step $\Delta t = \Delta r/2$. Furthermore, we have again neglected $1/r$ and $1/r^2$ terms as the simulation can be ''shifted'' to large values of r.

In Fig. 7 we show contour plots of the root mean square (rms) of the Hamiltonian constraint as a function of the
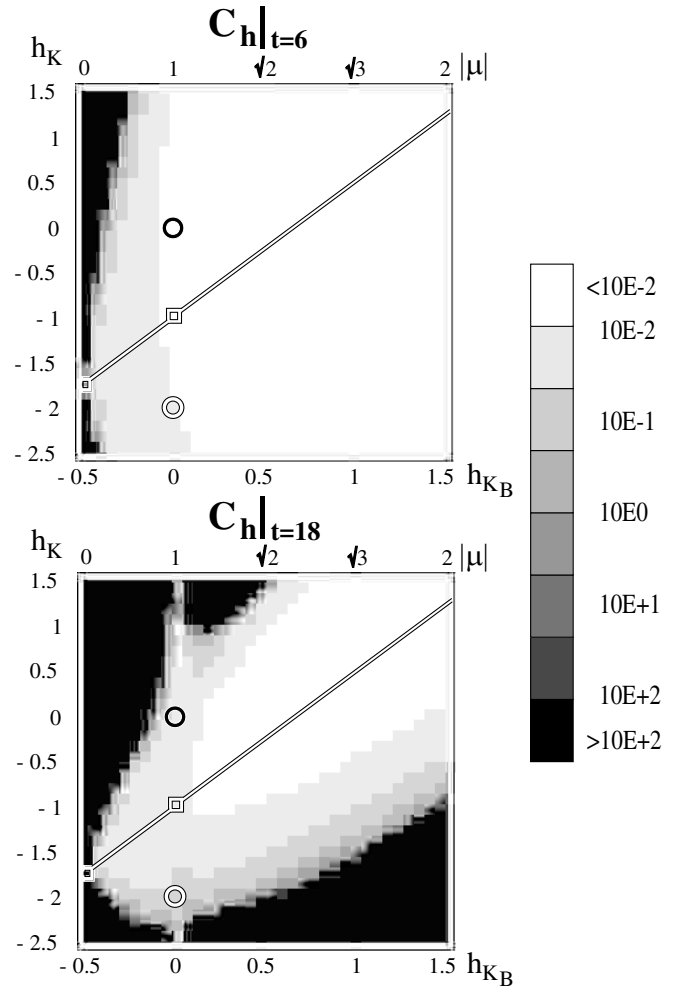

FIG. 7. Contour plots of the rms of the Hamiltonian constraint are shown at times $t = 6$ and $t = 18$.

adjustment parameters h_K and h_{K_B} at two different times during the evolution, using 40 equidistant parameter choices in each direction. The momentum constraint, not shown here, has a very similar behavior. Note that black and dark gray colors correspond to regions where the rms of the Hamiltonian constraint grows rapidly, and light gray denotes regions where it grows very slowly or not at all. For ADM (corresponding to $h_K = h_{K_B} = 0$ and denoted by a black circle) a rapid growth of the constraints is found. We also observe rapid growth close to the white circle, representing the special case with $h_K = -2$ and $h_{K_B} = 0$ which corresponds to the only strongly hyperbolic system along the line $|\mu| = 1$ and is preferred by indirect linear degeneracy.

The white diagonal line corresponds to the shock-avoiding μ-family obtained from the source criteria (not defined for the points $\mu = 0$ and $|\mu| = 1$ represented as boxes). We clearly see that this one-parameter family falls in the middle of the region where the rms of the Hamiltonian constraint does not grow, indicating that it does correspond to a preferred region of parameter space.

There is a final point related to Fig. 7. The figure shows that the line $|\mu| = 1$ also seems to produce slow growth of the constraints. However, as mentioned before, in that case the system is only weakly hyperbolic and our whole analysis breaks down, so we have no explanation as to why this line represents a preferred region.
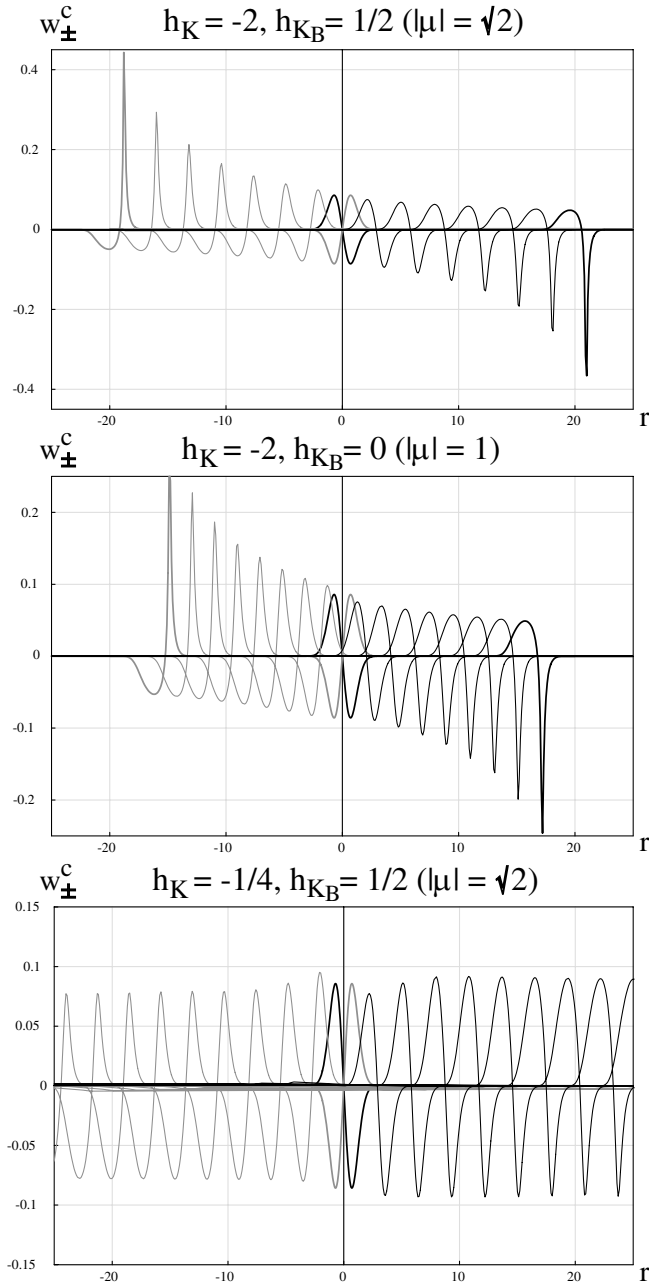

$w^c_\pm$ $h_K = -2$, $h_{K_B} = 1/2$ $(|\mu| = \sqrt{2})$

$w^c_\pm$ $h_K = -2$, $h_{K_B} = 0$ $(|\mu| = 1)$

$w^c_\pm$ $h_K = -1/4$, $h_{K_B} = 1/2$ $(|\mu| = \sqrt{2})$

FIG. 8.   Evolution of the eigenfields w^c_+ (black) and w^c_- (gray) associated with constraint shocks. *Top.* Blowups occur for a strongly hyperbolic case that is not shock avoiding. *Middle.* Shocks also form for the adjustments suggested by the indirect linear degeneracy criteria. *Bottom.* No shocks are found for a member of the μ-family obtained by the source criteria.

In order to see the formation of shocks more clearly, in Fig. 8 we show the time evolution (shown every $\Delta t = 2$) of the eigenfields $w^c_\pm$ which are associated with the formation of constraint shocks. The upper panel shows the evolution for the parameter choice $h_K = -2$ and $h_{K_B} = 1/2$ (which for harmonic slicing implies $|\mu| = \sqrt{2}$), corresponding to a strongly hyperbolic case that does not avoid constraint shocks. From the figure we see how a shock is clearly forming, as expected. The middle panel corresponds to the parameters $h_K = -2$ and $h_{K_B} = 0$ (and hence $|\mu| = 1$) preferred by the indirect linear degeneracy criteria. Again we observe the formation of shocks. Finally, the lower panel shows the evolution for a member of the shock avoiding μ-family with parameters $h_K = -1/4$ and $h_{K_B} = 1/2$ (i.e., $|\mu| = \sqrt{2}$). In this case the evolution shows a wavelike character, and no shocks form on the time scale considered here.

V. CONCLUSIONS

We have presented an analysis of two different blowup mechanisms for systems of hyperbolic equations of the type found in general relativity, namely, the geometric blowup or gradient catastrophe, and the ODE-mechanism. As an example of how these mechanisms operate we have used the simple one-dimensional wave equation with dynamic wave speed and nontrivial source terms.

We have later performed a blowup analysis of one-dimensional systems in general relativity, concentrating on toy $1 + 1$ relativity and spherically symmetric relativity, and using the hyperbolic Bona-Massó family of slicing conditions. In the first case, we have recovered the well known gauge shocks and found, somewhat surprisingly, that both blowup criteria give precisely the same condition for shock avoidance. When studying the spherically symmetric case, we have also found the same type of gauge shocks but, much more importantly, we have found a second family of modes that can produce blowups. The characteristic speed of these extra modes is controlled by the way in which the constraint equations are used to modify the evolution equations, so we have called such blowups ''constraint shocks.'' We have also studied how one can adjust the way in which constraints are added to the evolution equations to avoid as much as possible the formation of such constraint shocks. Finally, we have presented numerical simulations that confirm that constraint shocks can and do indeed form, and that they can be avoided by using the results of our previous analysis.

We would like to mention here that although in the present study we did not work with adjustments that can avoid constraint shocks according to both blowup criteria, this is simply because we have considered only a restricted form of adjustments to the evolution equations. More elaborate adjustments can give rise to completely shock

free formulations.[3] Moreover, empirically we have found that the source criteria seems to be far more important than indirect linear degeneracy (though it is still not clear why in the case of gauge shocks both criteria give rise to the same condition for avoiding blowups).

As a final note, we want to point out that there is a very important difference between gauge and constraint shocks. When a gauge shock develops, it means that our coordinate system has broken down and there is no way to continue the evolution past that point other than choosing a different gauge. Gauge shocks are certainly nonphysical, as the geometry of spacetime remains smooth, but they do represent very real pathologies in the spatial foliation. Constraint shocks, on the other hand, are not only non-

physical, but they should not arise at all if the constraints remain satisfied exactly. In a numerical simulation the constraints are of course violated, but as the resolution is increased this violation should become smaller and smaller, so the possible formation of constraint shocks becomes less of an issue. Nevertheless, at any fixed resolution constraint shocks can form at a finite time unless care is taken to use a form of the equations that avoids them.

Because of this we recommend that out of the many possible ways of constructing strongly hyperbolic evolution systems in general relativity, for numerical evolutions one should concentrate on those that have better shock avoiding properties, both in the gauge and constraint violating sectors, as these should prove to be more robust in practice.

ACKNOWLEDGMENTS

---

[3]In fact, it is not very difficult to find adjustments that satisfy both criteria. By making use of an arbitrary function $c(\alpha, A, B)$, the adjustment (4.56) can by generalized to obtain $h_{K_B} = (c - 1)/2$. The source criteria then implies that $h_K = -2 + (1 + 3c)/4$, which generalizes the previous expression (4.74) and again yields the relation (4.75). Furthermore, since the eigenvalues $\lambda^c_\pm$ are now given by $\lambda^c_\pm = \pm\alpha\sqrt{c/A}$, by choosing $c = \mu^2 A/\bar{\alpha}^2$ one obtains $\lambda^c_\pm = \pm\mu = $ const, which trivially satisfies the indirect linear degeneracy criteria. Numerical experiments show that systems with these adjustments behave in a very similar way to those described in the text.

[1] H. O. Kreiss and J. Lorenz, *Initial-Boundary Value Problems and the Navier-Stokes Equations* (Academic, San Diego, 1989).

[2] Y. Bruhat, Acta Mathematica **88**, 141 (1952).

[3] Y. Choquet-Bruhat and Y. York, in *General Relativity and Gravitation*, edited by A. Held (Plenum, New York, 1980), Vol. I, pp. 99–172.

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

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

[6] Y. Choquet-Bruhat, in *Gravitation: An Introduction to Current Research*, edited by L. Witten (Wiley, New York, 1962).

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

[8] H. Friedrich and A. D. Rendall, Lect. Notes Phys. **540**, 127 (2000).

[9] J. York, in *Sources of Gravitational Radiation*, edited by L. Smarr (Cambridge University Press, Cambridge, England, 1979).

[10] R. Arnowitt, S. Deser, and C. W. Misner, in *Gravitation: An Introduction to Current Research*, edited by L. Witten (Wiley, New York, 1962), pp. 227–265.

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

[12] C. Bona and J. Massó, Phys. Rev. D **40**, 1022 (1989).

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

[14] T. Baumgarte and S. Shapiro, Phys. Rev. D **59**, 024007 (1999).

[15] T. Nakamura, K. Oohara, and Y. Kojima, Prog. Theor. Phys. Suppl. **90**, 1 (1987).

[16] M. Shibata and T. Nakamura, Phys. Rev. D **52**, 5428 (1995).

[17] M. Alcubierre, G. Allen, B. Brügmann, E. Seidel, and W.-M. Suen, Phys. Rev. D **62**, 124011 (2000).

[18] O. Sarbach, G. Calabrese, J. Pullin, and M. Tiglio, Phys. Rev. D **66**, 064002 (2002).

[19] H. Beyer and O. Sarbach (2004), Phys. Rev. D **70**, 104004 (2004).

[20] S. Frittelli and R. Gómez, J. Math. Phys. (N.Y.) **41**, 5535 (2000).

[21] B. Gustafsson, H. Kreiss, and J. Oliger, *Time Dependent Problems and Difference Methods* (Wiley, New York, 1995).

[22] L. E. Kidder, M. A. Scheel, and S. A. Teukolsky, Phys. Rev. D **64**, 064017 (2001).

[23] O. Sarbach and M. Tiglio, Phys. Rev. D **66**, 064023 (2002).

[24] H.-a. Shinkai and G. Yoneda, Classical Quantum Gravity **19**, 1027 (2002).

[25] L. Lindblom and M. A. Scheel, Phys. Rev. D **66**, 084014 (2002).

[26] M. Alcubierre, Phys. Rev. D **55**, 5981 (1997).

[27] M. Alcubierre, Classical Quantum Gravity **20**, 607 (2003).

[28] P. Diener (private communication).

[29] R. J. Leveque, *Numerical Methods for Conservation Laws* (Birkhauser Verlag, Basel, 1992).

[30] S. Alinhac, *Blowup for Nonlinear Hyperbolic Equations* (Birkhäuser, Boston, 1995).

[31] F. John, *Partial Differential Equations* (Springer-Verlag, New York, 1986).

[32] E. Hopf, Commun. Pure Appl. Math. **3**, 201 (1950).

[33] P. Lax, *Development of Singularities of Solutions of Nonlinear Hyperbolic Partial Differential Equations* (SIAM, Philadelphia, 1973).

[34] P. Lax, J. Math. Phys. (N.Y.) **5**, 611 (1964).

[35] F. John, in *Nonlinear Wave Equations, Formation of Singularities*, University Lecture Series (American Mathematical Society, Providence, Rhode Island, 1990),

Vol. 2.

[36] F. John, Commun. Pure Appl. Math. **27**, 377 (1974).

[37] S. Klainerman, Commun. Pure Appl. Math. **33**, 43 (1980).

[38] F. John, Commun. Pure Appl. Math. **34**, 29 (1981).

[39] W. Strauss, in *Nonlinear Wave Equations*, Regional Conference Series in Mathematics No. 73 (American Mathematical Society, Providence, Rhode Island, 1989).

[40] M. Alcubierre and J. Massó, Phys. Rev. D **57**, R4511 (1998).

[41] M. Alcubierre and J. González, gr-qc/0401113 [Comput. Phys. Commun. (to be published)].

[42] M. Alcubierre *et al.*, Classical Quantum Gravity **21**, 589 (2004).