PHYSICAL REVIEW D VOLUME 57, NUMBER 8 15 APRIL 1998

Pathologies of hyperbolic gauges in general relativity and other field theories

Miguel Alcubierre and Joan Massó
Max-Planck-Institut für Gravitationsphysik (Albert-Einstein-Institut), Schlaatzweg 1, D-14473 Potsdam, Germany
(Received 12 September 1997; published 16 March 1998)

We present a mathematical characterization of hyperbolic gauge pathologies in electrodynamics and general relativity. We show analytically how non-linear gauge terms can produce a blow-up of some fields along characteristics. We expect similar phenomena to appear in any other gauge field theory. We also present numerical simulations where such blow-ups develop and show how they can be properly identified by performing a convergence analysis. We stress the importance of these results for the particular case of numerical relativity, where we offer some cures based on the use of non-hyperbolic gauges. [S0556-2821(98)50208-1]

PACS number(s): 04.25.Dm, 04.20.Ex

During the last decades, gauge field theories have become the paradigm of fundamental physics. In such theories, gauge invariance implies that the ''physics'' is independent of our choice of gauge. Yet in some cases, of which general relativity (GR) is a very good example, it can be difficult to separate the gauge from the physical degrees of freedom. This is of particular importance in numerical simulations of non-linear field theories, where one needs to distinguish physical singularities from pure gauge pathologies. In such cases, exact solutions are not available, and the simple idea of a change of gauge can become a daunting task. Because of this, understanding the properties of particular gauge choices becomes an issue of great relevance.

In a recent paper [1] it was shown, to the surprise of many in the numerical relativity community, that certain gauge choices (throughout we use the term ''gauge'' applied to relativity to mean only slicing conditions) in relatively simple scenarios could lead to the development of what were there called ''coordinate shocks.'' Here we will show that the term ''shock'' used in that paper is misleading and will refer to them instead as gauge pathologies. The recent Bona-Massó hyperbolic formulation of the Einstein equations [2] allowed the author of Ref. [1] to study the structure of the gauge condition and how its non-linearity could cause gauge pathologies. It was shown numerically that these pathologies seemed to occur, but no proof that they were real discontinuities was given. This work left some basic questions unanswered: How can we characterize these pathologies mathematically and numerically? How generic are they? Do similar phenomena occur in other gauge theories?

Here we give for the first time a mathematical characterization of hyperbolic gauge pathologies (where by hyperbolic gauge we mean choices where the gauge is evolved using a hyperbolic equation) based on the theory of nonlinear waves [3]. We focus our attention on two cases: electrodynamics (ED) and spherically symmetric general relativity (GR). We analyze the structure of the equations and show how non-linear gauge terms can produce a blow-up of some fields along characteristics [3]. Such a blow-up indicates a gauge singularity and corresponds to the ''coordinate shocks'' of Ref. [1]. The blow-ups described here are much stronger singularities than shocks, since some of the propagating fields become in fact infinite. We also show how the blow-ups can be identified numerically by analyzing the convergence properties of computational simulations. Finally, we propose some ''cures'' for these pathologies based on changing the hyperbolic nature of the gauge condition. Though we only study the cases of ED and GR, we expect similar phenomena to appear in any other gauge field theory.

Electrodynamics—Surprisingly, gauge effects can produce blow-ups along characteristics even in simple systems such as ED. We are not aware of any previous analysis of gauge pathologies in ED, probably because well behaved gauge choices are intuitively clear there. In the following, we will follow closely the notation of Refs. [1,2].

We will write the equations of ED as a first order initial value problem in the following way:

$$\partial_t A_i = -(E_i + \psi_i), \tag{1a}$$

$$\partial_t D_{ij} = -\partial_i (E_j + \psi_j), \tag{1b}$$

$$\partial_t E_i = \partial_i \ \mathrm{tr}\ D - \sum_j \partial_j D_{ji}, \tag{1c}$$

where ϕ and $\vec{A}$ are the scalar and vector potentials, $\vec{E}$ the electric field, and where we have introduced the quantities $D_{ij} \equiv \partial_i A_j$ and $\psi_i \equiv \partial_i \phi$. We also have one constraint which in vacuum takes the form $\nabla \cdot \vec{E} = 0$.

We are now free to decide how $\{\phi, \psi_i\}$ will evolve, *i.e.* we are free to choose the gauge. Here we make a choice similar to the one we will make later in relativity

$$\partial_t \phi = -f(\phi)\,\mathrm{tr}\ D, \tag{2}$$

with $f(\phi) > 0$, but otherwise arbitrary ($f = 1$ corresponds to the familiar Lorentz gauge). It is through the gauge function f that we introduce a non-linearity into ED, which is otherwise a completely linear theory.

The resulting system of equations can be shown to be hyperbolic (in the sense of having a complete family of eigenfields). For a given direction x we find the characteristic structure: four fields propagate along the physical light cones (speed ± 1)

$$\omega_{p_\pm} \equiv (E_p + \psi_p) \pm D_{xp}, \quad p \neq x, \tag{3}$$

and 2 fields propagate with the ''gauge speeds'' $\pm\sqrt{f}$

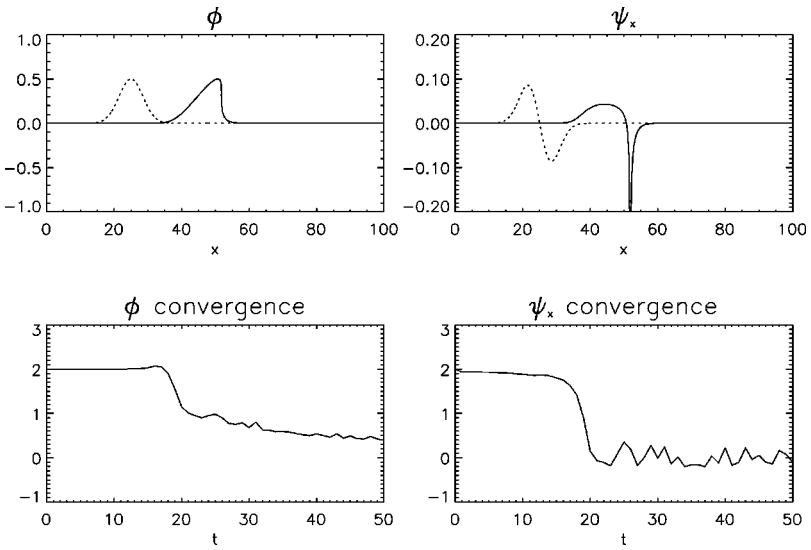

FIG. 1. Numerical simulation of a gauge wave in electrodynamics. The first two plots show the initial (dotted line) and final (solid line) values at $t=20$ of the scalar potential ϕ and its derivative ψ_x. The lower plots show the convergence rates as a function of time.

$$\omega_{g_\pm} := \psi_x \pm \sqrt{f}\ \mathrm{tr}\ D. \tag{4}$$

The remaining 13 fields move along the time lines.

Let us now define $\Omega_{g_\pm} := \sqrt{f}\omega_{g_\pm}$. We find

$$\partial_t \Omega_{g_\pm} \pm \sqrt{f}\partial_x \Omega_{g_\pm} = \pm \frac{f'}{4f}(\Omega_{g_\pm}^2 + \Omega_{g_+}\Omega_{g_-}). \tag{5}$$

For $f' \neq 0$, the quadratic term in $\Omega_{g_\pm}$ can produce a blow-up along a characteristic in a finite time. We can predict the time when this will happen, if we restrict ourselves to the case $\Omega_{g_-}=0$, and take $f=e^{a\phi}$. Along a characteristic we will have

$$\frac{d\Omega_{g_+}}{dt} = \frac{a}{4}\Omega_{g_+}^2, \tag{6}$$

which can be easily integrated to find

$$\Omega_{g_+} = \frac{\Omega_0}{1 - a\Omega_0 t/4}. \tag{7}$$

For $a\Omega_0 > 0$, a gauge pathology (a blow-up of Ω_{g_+}) will appear at a finite time T given by

$$T = 4/a\Omega_0. \tag{8}$$

We have constructed a numerical code to evolve this system of equations and reproduce this blow-up. Figure 1 shows the results of a one-dimensional simulation for which a shock is expected at $T=19.92$. The first two plots show the initial (dotted line) and final (solid line) values at $t=20$ of ϕ and ψ_x. A sharp gradient has developed in ϕ, while a large spike has appeared in ψ_x.

Now, if we had not performed the mathematical analysis of the system, we could ask ourselves how can we know that these features correspond to a real blow-up and not to the development of large, but smooth gradients. After all, nonlinear systems may develop features that are difficult to resolve numerically. As an example, one of the key successes of numerical relativity has been the discovery of critical phenomena [4]. Confidence in these results required advanced adaptive mesh refinement and the careful use of convergence tests [5]. Here we have performed similar convergence studies: we consider a series of different resolutions and find the rate at which a global measure of their relative errors converges to zero (the ''self-consistent'' convergence rate). Our code uses a second order accurate scheme, so we expect the convergence rate of our solutions to be 2. The lower two plots of Fig. 1 show the convergence rates for ϕ and ψ_x obtained from runs at three resolutions (5000, 10000 and 20000 points). As expected, for $t<20$ the convergence rate is close to 2, while after that it drops dramatically indicating loss of convergence. Moreover, the spikes in ψ_x become larger and larger with increased resolution. We then conclude that they correspond to a real infinity and not just a large gradient. The lesson is clear: we can characterize numerically the appearance of a blow-up using convergence analysis.

It is important to consider the effect that the choice of finite difference method has on the results described above. When we used standard methods such as a leapfrog scheme, the simulations crashed very soon after the spike in ψ_x started to develop. Using instead shock capturing [6] techniques, we were able to follow the evolution much further, which allowed us to determine more precisely the time of the blow-up. Notice that this further evolution is non-physical: the shock capturing techniques help to maintain the simulation stable, but this is of no physical relevance, since the true solution is not really a shock wave. A real infinity has developed and no numerical method will converge after that.

Spherically symmetric general relativity—Let us now consider the case of GR in spherical symmetry. The basic variables are the lapse function α, the spatial metric components $\{g_{rr}, g_{\theta\theta}\}$ and the extrinsic curvature components $\{K_{rr}, K_{\theta\theta}\}$. Since we are interested in hyperbolic gauges, we will use the following gauge condition [2]:

$$\partial_t \alpha = -\alpha^2 f(\alpha)\mathrm{tr}\ K, \tag{9}$$

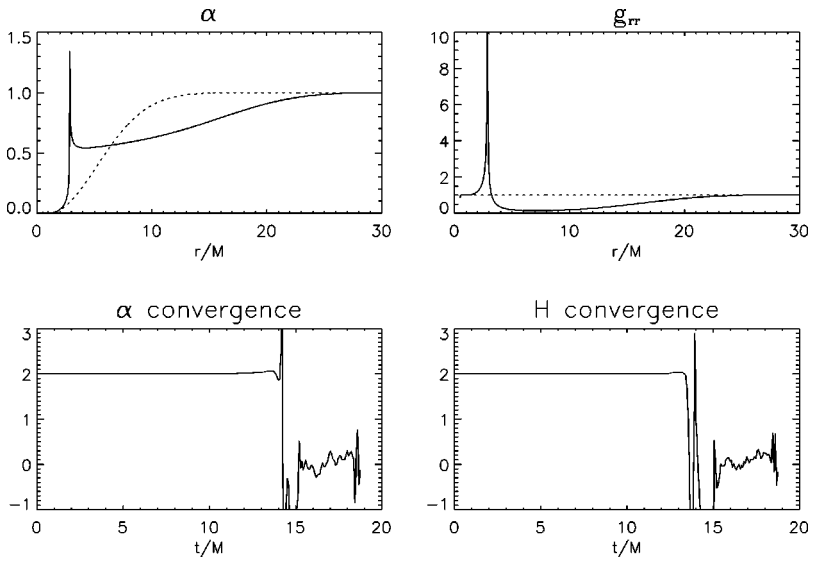

FIG. 2. Numerical simulation of a spherically symmetric black hole spacetime. The first two plots show the initial (dotted line) and final (solid line) values at $t=15M$ of the lapse α and the conformal metric function g_{rr}. The lower plots show the convergence rates for the lapse and the Hamiltonian constraint.

with $f(\alpha)>0$, but otherwise arbitrary ($f=1$ now corresponds to harmonic slicing). It can be shown that with this gauge condition, the evolution equations of GR can be written in first order hyperbolic form [2].

The particular form of the equations for spherical symmetry and its characteristic structure can be found in [1]. There it was shown that there are two families of travelling modes, one that moves with the speed of light, and one that moves with a ''gauge speed'' that depends on the value of the function f and that reduces to the speed of light for harmonic slicing. In general, both types of travelling modes can develop gauge pathologies. Here we will concentrate in the particular case of harmonic slicing ($f=1$), where only one type of pathology appears. Note that this gauge is the preferred choice for many of the new hyperbolic formulations of the Einstein equations [7]. The analysis of the system for other forms of f will be considered elsewhere [12].

We will start by defining

$$\Omega_{\pm} \equiv \alpha/g_{\theta\theta}(K_{\theta\theta} \pm D_{r\theta\theta}/g_{rr}^{1/2}), \tag{10}$$

where $D_{r\theta\theta} \equiv 1/2\partial_r g_{\theta\theta}$. From the form of the evolution equations given in [1], it is easy to show that, for $f=1$, we will have

$$\partial_t \Omega_{\pm} \pm \alpha/g_{rr}^{1/2}\partial_r \Omega_{\pm} = \Omega_{+}\Omega_{-} - \Omega_{\pm}^2 + \alpha^2/g_{\theta\theta}. \tag{11}$$

We see that the $\Omega_{\pm}$ represent outgoing (+) and ingoing (−) modes travelling with the speed of light. Unfortunately, the last term on the right hand side of the above equations makes it impossible to separate them, and so prevents one from predicting when a blow-up caused by the quadratic source terms might occur. Nevertheless, one can see that the quadratic term in $\Omega_{\pm}$ appearing in Eq. (11) is only dangerous for negative values of $\Omega_{\pm}$. Moreover, it is easy to see that if $\Omega_{\pm}$ is initially positive, it will not change sign. Now, if our initial data is time-symmetric ($K_{ij}=0$) and such that the metric function $g_{\theta\theta}$ is monotonic, then we will have $\pm\Omega_{\pm}>0$. This means that no blow-ups can develop for outgoing modes: Only ingoing modes can produce a gauge pathology. Of course, whether they will or not depends on the precise form of the initial data.

In Ref. [1] it was shown that pathologies appeared for a particular choice of initial data in a black hole spacetime. Here we give another example of how pathologies can form from apparently simple initial data. We choose the standard initial data for the metric and extrinsic curvature that has been successfully used for most black hole simulations to date [8–10]. It corresponds to an isotropically sliced Schwarzschild black hole with time symmetry. The key difference is that, instead of choosing an initial lapse that satisfies the maximal slicing condition, we choose the following ''Gaussian'' profile:

$$\alpha = 1 - A\ \exp\{[(r-r_0)/\sigma]^p\}. \tag{12}$$

We have used three independent numerical codes to evolve the system, one based on the standard Arnowitt-Deser-Misner (ADM) formulation [11] that uses a simple leapfrog scheme, and two based on the Bona-Masso hyperbolic formulation [2] that use shock capturing methods [6]. All three codes produce similar results with one important difference: the ADM code crashes soon after the pathologies start to develop, while the other two codes are capable of continuing past this point.

We report the results of a particular simulation obtained using one of the hyperbolic codes. The first plot in Fig. 2 shows the initial (dotted line) and final (solid line) values of the lapse at time $t=15M$ for the case $\{A=1,\ r_0=M/2,\ \sigma=6M,\ p=2\}$, with M the mass of the black hole. The next plot shows the same for the conformal metric function g_{rr}. Notice how both the lapse and the radial metric develop large spikes, the sizes of which increase with resolution. The bottom plots show the global convergence rate of the lapse

and the Hamiltonian constraint obtained from runs at 4000, 8000 and 16000 grid points, with the outer boundary located at $40M$. Clearly, we have a gauge pathology, whose blow-up time seems to be $t=(14\pm 1)M$.

Although the simulation presented here corresponds to harmonic slicing, gauge pathologies also develop with other hyperbolic gauge choices. An important difference is that in the harmonic case, the lapse becomes infinite, which indicates that the time slicing becomes null. In the other cases, the lapse becomes discontinuous, but remains finite, and the time slicing develops a kink instead. We have performed many simulations studying the parameter space $\{A,\sigma,p\}$ and found similar results. Details of all these studies will be given elsewhere [12].

Another crucial aspect of the problem is that of the numerical resolution of the gauge pathologies. We have found that if we evolve the previous system with only 200 grid points, the pathologies do not seem to form. The lapse grows until a certain value is reached and then it propagates out in a smeared manner due to the large numerical viscosity. The solution ''looks good,'' although it is non-physical, as a proper convergence test reveals.

The impact of these results on three dimensional (3D) numerical relativity should not be underestimated. Note that in the previous examples we have used thousands of points to be able to show very sharp fall-offs of the convergence rate and have a good estimate of the blow-up time. Even if this sharpness will not be possible at the resolutions currently available for 3D computations, at medium and low resolutions one can already see that the convergence fails at late times.

We should stress again the fact that the development of these pathologies depends crucially on the form of the initial data. For different choices of the initial lapse function, one can find that harmonic slicing is perfectly well behaved, as the simulations presented in Ref. [9] show. In fact, one can even find explicitly a harmonic slicing of a black hole in which the metric is static [13].

''*Cures*''—The appearance of gauge pathologies might seem to put into question the practical value of non-linear hyperbolic gauges in numerical studies of gauge field theories. After all, in a general situation it might be very difficult to know *a priori* if our initial data will develop such a pathology.

For ED, the solution is clear: use a gauge that decouples the characteristic speeds from the dynamics, *i.e.* use the Lorentz gauge. Unfortunately, this will not work in relativity where the characteristic speeds cannot be decoupled from the dynamics. We can think of at least two different ways to solve the problem. Both involve changing the character of the equations for the *gauge only*: all other equations (the ''physics'') remain hyperbolic.

The first approach implies using an elliptic gauge condition of which maximal slicing is the best known example. One can then either use an elliptic gauge always or, in cases where it might be of interest, use a hyperbolic gauge for some time and then switch to an elliptic gauge when a pathology is about to form. We have tested this idea and found that it works very well in practice [12].

The second approach consists of adding dissipation to our gauge condition. We will then have a parabolic equation and intuition tells us that this should prevent the pathologies. We then propose the gauge condition

$$\partial_t\alpha=-\alpha^2[f(\alpha)\mathrm{tr}\,K-\xi(\alpha)\nabla^2\alpha], \tag{13}$$

with $f,\xi>0$ but otherwise arbitrary. Notice that the numerical treatment of the diffusion term requires either a very stringent Courant condition or the use of implicit techniques. Note also that this term should be kept with the same coefficient for all resolutions of a given simulation, as it does not correspond to a simple ''artificial viscosity'' [10], but rather to an explicit change of the character of the gauge condition. We have tried this condition in spherically symmetric GR and found that it also prevents the development of gauge pathologies [12].

In conclusion, we have presented for the first time a characterization of hyperbolic gauge pathologies in ED and GR. We have shown how the coupling of characteristic gauge speeds to the dynamics produces a nonlinear blow-up mechanism, and how a careful convergence analysis can indicate the appearance of such a blow-up in numerical simulations. The origin of these pathologies is in the finite speed of propagation of the gauge modes, and therefore a way to avoid them is the use of elliptic or parabolic gauges with infinite gauge speeds.

Though we have concentrated in the cases of ED and GR, we expect similar pathologies to arise in any other gauge field theory. Because of this we feel that further mathematical study of these phenomena will be of fundamental importance for future numerical simulations of non-linear field theories.

We would like to thank Gabrielle Allen, Carles Bona, Bernd Brügmann, Carsten Gundlach, Ed Seidel, Joan Stela, Wai-Mo Suen and Paul Walker for helpful discussions. Special thanks to Bernard Schutz and Richard Matzner for a very careful and critical revision of the original manuscript.

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

[2] C. Bona, J. Massó, E. Seidel, and J. Stela, Phys. Rev. Lett. **75**, 600 (1995); C. Bona, J. Massó, E. Seidel, and J. Stela, Phys. Rev. D **56**, 3405 (1997).

[3] S. Kichenassamy, *Nonlinear Wave Equations* (Marcel Dekker, New York, 1996).

[4] M. W. Choptuik, Phys. Rev. Lett. **70**, 9 (1993).

[5] M. W. Choptuik, Phys. Rev. D **44**, 3124 (1991).

[6] R. J. LeVeque, *Numerical Methods for Conservation Laws* (Birhäuser Verlag, Basel, Switzerland, 1992).

[7] S. Frittelli and O. Reula, Commun. Math. Phys. **166**, 221 (1994); A. Abrahams, A. Anderson, Y. Choquet-Bruhat, and J. W. York, Phys. Rev. Lett. **75**, 3377 (1995); H. Friedrich, Class. Quantum Grav. **13**, 1451 (1996); M. van Putten and D. Eardley, Phys. Rev. D **53**, 3056 (1996).

[8] D. Bernstein, D. Hobill and L. Smarr, in *Frontiers in Numeri-*

cal Relativity, edited by C. Evans, L. Finn and D. Hobill (Cambridge University Press, Cambridge, England, 1989).

[9] C. Bona, J. Massó, and J. Stela, Phys. Rev. D **51**, 1639 (1995).

[10] P. Anninos *et al.*, Phys. Rev. D **52**, 2059 (1995).

[11] R. Arnowitt, S. Deser and C. Misner, in *Gravitation: an Introduction to Current Research*, edited by L. Witten (Wiley, New York, 1962).

[12] M. Alcubierre and J. Massó (in preparation).

[13] C. Bona and J. Massó, Phys. Rev. D **38**, 2419 (1988).