Class. Quantum Grav. **25** (2008) 114010 (14pp) doi:10.1088/0264-9381/25/11/114010

Strong cosmic censorship in the case of T^3-Gowdy vacuum spacetimes

Hans Ringström

Department of Mathematics, KTH, 100 44 Stockholm, Sweden

Received 4 September 2007, in final form 9 November 2007
Published 15 May 2008
Online at stacks.iop.org/CQG/25/114010

Abstract
In 1952, Yvonne Choquet-Bruhat demonstrated that it makes sense to consider Einstein's vacuum equations from an initial value point of view; given initial data, there is a globally hyperbolic development. Since there are many developments, one does, however, not obtain uniqueness. This was remedied in 1969 when Choquet-Bruhat and Robert Geroch demonstrated that there is a unique maximal globally hyperbolic development (MGHD). Unfortunately, there are examples of initial data for which the MGHD is extendible, and, what is worse, extendible in inequivalent ways. Thus it is not possible to predict what spacetime one is in simply by looking at initial data and, in this sense, Einstein's equations are not deterministic. Since the examples exhibiting this behaviour are rather special, it is natural to conjecture that for generic initial data, the MGHD is inextendible. This conjecture is referred to as the *strong cosmic censorship conjecture* and is of central importance in mathematical relativity. In this paper, we shall describe this conjecture in detail, as well as its resolution in the special case of T^3-Gowdy spacetimes.

PACS numbers: 04.20.Dw, 04.20.Ex, 04.20.Ha

1. The initial value problem

Einstein introduced his equations in 1915, cf [12, 13], but it was not until much later that they were studied from a partial differential equation (PDE) point of view. There were natural reasons for this, one of them being that, due to the diffeomorphism invariance, writing down the equations in arbitrary coordinates does not yield a PDE for which there is, e.g., an initial or a boundary value formulation. However, there is something to be learnt by considering the equations in coordinates. Here, we shall restrict our attention to the vacuum equations, and they can be written

$$R_{\mu\nu} = 0, \tag{1}$$

where $R_{\mu\nu}$ are the components of the Ricci tensor. In coordinates, we have

$$R_{\mu\nu} = -\tfrac{1}{2} g^{\alpha\beta}\partial_\alpha\partial_\beta g_{\mu\nu} + \nabla_{(\mu}\Gamma_{\nu)} + g^{\alpha\beta}g^{\gamma\delta}[\Gamma_{\alpha\gamma\mu}\Gamma_{\beta\delta\nu} + \Gamma_{\alpha\gamma\mu}\Gamma_{\beta\nu\delta} + \Gamma_{\alpha\gamma\nu}\Gamma_{\beta\mu\delta}], \tag{2}$$

where

$$\Gamma_{\alpha\gamma\beta} = \frac{1}{2}(\partial_\alpha g_{\beta\gamma} + \partial_\beta g_{\alpha\gamma} - \partial_\gamma g_{\alpha\beta}), \qquad \Gamma_\alpha = g_{\alpha\beta} g^{\mu\nu}\Gamma^\beta_{\mu\nu}, \qquad \nabla_\mu \Gamma_\nu = \partial_\mu \Gamma_\nu - \Gamma^\alpha_{\mu\nu}\Gamma_\alpha,$$

and a parenthesis denotes symmetrization, i.e.

$$\nabla_{(\mu}\Gamma_{\nu)} = \frac{1}{2}(\nabla_\mu \Gamma_\nu + \nabla_\nu \Gamma_\mu).$$

The problem is the second term on the right-hand side of (2), which involves second derivatives of the metric. However, if we have a solution, we can choose local coordinates so that the second term on the right-hand side of (2) disappears. What then results is a system of second order hyperbolic PDEs; in other words, something similar to a system of wave equations. From a naive point of view, we are thus led to expect that

- the natural PDE problem is the initial value problem.
- a natural choice for initial data is the metric and the first time derivative of the metric.

Due to the diffeomorphism invariance, there is, however, a problem in general. Diffeomorphism invariance means that if (M, g) is a vacuum solution to Einstein's equations and ϕ is a diffeomorphism of M, then $(M, \phi^* g)$ is a vacuum solution to Einstein's equations. As a consequence, if one writes down Einstein's equations with respect to coordinates without imposing any additional restrictions, there will be no uniqueness; pulling back the metric by a diffeomorphism which is the identity on a neighbourhood of the initial hypersurface but which is not the identity everywhere will typically lead to a different metric with respect to fixed coordinates even though the initial data remain the same. It thus seems reasonable to hope for an initial value formulation, but it is also clear that we need a geometric way to phrase it.

Let us consider the problem of formulating an initial value problem from a different perspective. Given a vacuum spacetime (M, g), can it be uniquely determined by initial data, and if so, what should the initial data be? Note that this involves restrictions on the spacetime. The initial data should be specified on a hypersurface, say Σ, so the first question is what the restrictions on Σ should be. It seems natural to require that information not be allowed to travel from one part of Σ to another; otherwise there are consistency problems if one wants to, say, perturb the initial data. Furthermore, it seems natural to require that all the information that arrives at a spacetime point can be traced back to the initial hypersurface; if not, it is not so clear that it is possible to predict what happens at the point solely on the basis of initial data. Since information travels along causal curves in general relativity, it thus seems natural to at least require that

- *consistency*: no timelike curve intersects Σ twice.
- *predictability*: every inextendible timelike curve intersects Σ at least once.

One can of course sum up the above by demanding that every inextendible timelike curve intersect Σ exactly once. A hypersurface Σ with this property is called a *Cauchy hypersurface* and a Lorentz manifold admitting a Cauchy hypersurface is called *globally hyperbolic*. Note that the property of being globally hyperbolic is a property of the spacetime, as opposed to a property of the hypersurface. From now on, Σ will be assumed to be smooth and spacelike, i.e. the metric induced on Σ is positive definite. Simple examples of Cauchy hypersurfaces are the constant t hypersurfaces in Minkowski space and in the standard Robertson–Walker spacetimes.

What should the information on Σ be? Natural candidates are the induced metric, say h, and the induced second fundamental form, say k. First of all, these data are geometric and secondly, considering these objects in local coordinates, the second fundamental form is roughly speaking the first time derivative of the metric. There is, however, one problem: h

and k are not independent of one another—they have to satisfy certain equations called the vacuum constraint equations:

$$r - k_{ij}k^{ij} + (\mathrm{tr}\, k)^2 = 0, \tag{3}$$

$$D^j k_{ji} - D_i(\mathrm{tr}\, k) = 0, \tag{4}$$

where D is the Levi-Civita connection of h, r is the associated scalar curvature and indices are raised and lowered by h. These equations follow from the fact that Σ is a spacelike hypersurface in a Lorentz manifold satisfying Einstein's vacuum equations. What we have said so far was carried out starting with a solution to Einstein's vacuum equations, but at this stage we can consider Σ to be an abstract manifold and consider h to be a Riemannian metric and k to be a symmetric covariant two-tensor on Σ satisfying the constraint equations. And we can do this without any reference to an ambient space. This leads to the following formulation of the initial value problem.

Definition 1. *Initial data to Einstein's vacuum equations consist of a three-dimensional manifold Σ, a Riemannian metric h and a covariant symmetric two-tensor k on Σ, both assumed to be smooth and to satisfy (3) and (4). Given initial data, the initial value problem is that of finding a four-dimensional manifold M with a Lorentz metric g such that (1) is satisfied, and an embedding $i : \Sigma \to M$ such that $i^*g = h$ and that if κ is the second fundamental form of $i(\Sigma)$, then $i^*\kappa = k$. Such a Lorentz manifold (M, g) is called a development of the data. If, furthermore, $i(\Sigma)$ is a Cauchy hypersurface in (M, g), then (M, g) is referred to as a globally hyperbolic development of the initial data. In both cases, the existence of an embedding i is tacit.*

The seminal result in this field is due to Yvonne Choquet-Bruhat, and the statement is as follows, cf [11].

Theorem 1. *Given initial data (Σ, h, k) to Einstein's vacuum equations, there is a globally hyperbolic development.*

This is a fundamental result, but in some sense it is disappointing; one obtains a globally hyperbolic development, but there are in fact infinitely many. The problem is that if person A has one development and person B has another development with completely different properties, there is no contradiction because there is no uniqueness. To obtain uniqueness, the development needs to be maximal in some sense.

Definition 2. *Given initial data to (1), a maximal globally hyperbolic development (MGHD) of the data is a globally hyperbolic development (M, g), with embedding $i : \Sigma \to M$, such that if (M', g') is any other globally hyperbolic development of the same data, with embedding $i' : \Sigma \to M'$, then there is a map $\psi : M' \to M$ which is a diffeomorphism onto its image such that $\psi^*g = g'$ and $\psi \circ i' = i$.*

Uniqueness is an immediate consequence of this definition. However, existence is far from obvious. For this reason, the following result of Yvonne Choquet-Bruhat and Robert Geroch from 1969, cf [3], is of fundamental importance.

Theorem 2. *Given initial data to (1), there is a maximal globally hyperbolic development of the data which is unique up to isometry.*

Remark. When we say that (M, g) is unique up to isometry, we mean that if (M', g') is another maximal globally hyperbolic development, then there is a diffeomorphism $\psi : M \to M'$ such that $\psi^* g' = g$ and $\psi \circ i = i'$, where i and i' are the embeddings of Σ into M and M' respectively. The theorem is a consequence of theorem 1, local uniqueness, Zorn's lemma and some additional geometric arguments.

This is a rather abstract result, but considering the initial value problem in general relativity without it would be meaningless. On the other hand, it does not say anything about causal geodesic completeness or curvature blow up.

2. Strong cosmic censorship

The maximal globally hyperbolic development is of course maximal in the class of all globally hyperbolic developments, but there is no reason to believe it to be maximal in the class of all developments. Unfortunately, there are initial data for which the MGHD is extendible and, what is worse, extendible in inequivalent ways. The canonical example is Taub–NUT, cf [4]. As a consequence, the initial data do not uniquely determine the spacetime. A natural reaction to this fact would be to say that the initial value problem in general relativity is nonsense and one should not consider it at all. However, the examples for which the MGHD is extendible are very special and one is led to the following conjecture.

Conjecture 1. *For generic initial data to Einstein's vacuum equations, the maximal globally hyperbolic development is inextendible.*

Remark. Note that the statement is a bit vague; to get a precise statement, one has to specify what one means by generic and what differentiability class one has in mind when speaking of extendibility.

This is what will be referred to in these notes as the *strong cosmic censorship conjecture* and the particular formulation is due to Eardly and Moncrief, cf [10]. If this conjecture were true, then it would still make sense to consider Einstein's equations from an initial value point of view. The statement is rather abstract, but it is connected to a question of more apparent physical interest, namely the question of curvature blow up at singularities.

Due to the work of Hawking and Penrose, we have been led to identify the existence of singularities with causal geodesic incompleteness. Of course, for this to make sense, the spacetime one considers has to be maximal in some natural sense; one can for instance consider the MGHD corresponding to initial data. The singularity theorems give quite general conditions ensuring that the MGHD is causally geodesically incomplete, but it is also of interest to know if the gravitational field (curvature) blows up as one approaches a singularity (in practice, we shall here use the Kretschmann scalar,

$$R_{\alpha\beta\gamma\delta}R^{\alpha\beta\gamma\delta}$$

as an indicator for curvature blow up). Unfortunately, there are examples in which there are singularities in the sense of causal geodesic incompleteness but the curvature remains perfectly bounded. Again, the Taub–NUT spacetimes are the canonical examples. One is thus led to the following conjecture.

Conjecture 2. *For generic initial data to Einstein's vacuum equations, curvature blows up in the incomplete directions of causal geodesics in the MGHD.*

Remark. This statement implies strong cosmic censorship.

To hope to prove this conjecture or strong cosmic censorship in all generality is not realistic. A more realistic problem would be to consider the same conjectures but to restrict one's attention to a specific symmetry class of solutions, the hope in the long run being that it will be possible to reduce the symmetry conditions gradually.

In cosmology, the simplest starting point would be to consider spatially homogeneous models. In the case of vacuum, such models are well understood as far as this question is concerned, with the exception of Bianchi $\text{VI}_{-1/9}$. The natural next step after spatial homogeneity is to look at the Gowdy spacetimes. There is a special subcase of this called polarized Gowdy, where the equations reduce to a linear PDE. For this case, the above questions have been sorted out, cf [6]. What will be discussed here concerns general T^3 Gowdy. In that case, the relevant equations are a system of nonlinear hyperbolic PDEs.

It should perhaps be pointed out that the motivation for studying these models is that they are the simplest ones in which one can study Einstein's vacuum equations in a cosmological, inhomogenous, nonlinear setting, and the desire to understand the equations in such a situation supercedes the desire to have a physically realistic model.

Finally, let us mention that there are other results concerning strong cosmic censorship in a cosmological setting with surface and T^2 symmetry, cf [7–9]. However, the methods used in these papers do not yield any conclusions concerning curvature blow up.

3. The T^3-Gowdy spacetimes

The Gowdy spacetimes were first introduced in [14] (see also [5]), and in [18] the fundamental questions concerning global existence were answered. These spacetimes can be characterized by geometric conditions, but since we wish to avoid the technical details, we shall take the Gowdy vacuum metrics on $\mathbb{R} \times T^3$ to be the metrics of the form

$$g = \mathrm{e}^{(\tau-\lambda)/2}(-\mathrm{e}^{-2\tau}\,\mathrm{d}\tau^2 + \mathrm{d}\theta^2) + \mathrm{e}^{-\tau}[\mathrm{e}^P d\sigma^2 + 2\,\mathrm{e}^P Q\,\mathrm{d}\sigma\,\mathrm{d}\delta + (\mathrm{e}^P Q^2 + \mathrm{e}^{-P})\,\mathrm{d}\delta^2]. \tag{5}$$

Here, $\tau \in \mathbb{R}$ and (θ, σ, δ) are coordinates on T^3. The functions P, Q and λ only depend on τ and θ. Consequently, translations in σ and δ constitute isometries, so that we have a T^2 group of isometries acting on the spacetime. The Einstein vacuum equations become

$$P_{\tau\tau} - \mathrm{e}^{-2\tau} P_{\theta\theta} - \mathrm{e}^{2P}\left(Q_\tau^2 - \mathrm{e}^{-2\tau} Q_\theta^2\right) = 0, \tag{6}$$

$$Q_{\tau\tau} - \mathrm{e}^{-2\tau} Q_{\theta\theta} + 2(P_\tau Q_\tau - \mathrm{e}^{-2\tau} P_\theta Q_\theta) = 0, \tag{7}$$

and

$$\lambda_\tau = P_\tau^2 + \mathrm{e}^{-2\tau} P_\theta^2 + \mathrm{e}^{2P}\left(Q_\tau^2 + \mathrm{e}^{-2\tau} Q_\theta^2\right), \tag{8}$$

$$\lambda_\theta = 2(P_\theta P_\tau + \mathrm{e}^{2P} Q_\theta Q_\tau). \tag{9}$$

Obviously, (6) and (7) do not depend on λ, so the idea is to solve these equations and then find λ by integration. There is, however, one obstruction to this; the integral of the right-hand side of (9) has to be zero. This is a restriction to be imposed on the initial data for P and Q, which is then preserved by the equations. In the end, the equations of interest are, however, the two nonlinear coupled wave equations (6) and (7). In the above parametrization, the singularity corresponds to $\tau \to \infty$. Note that $P = \tau$, $Q = 0$ and $\lambda = \tau$ is a solution to (6)–(9). The Riemann curvature tensor of the corresponding metric is identically zero. In fact, the corresponding spacetime is a part of Minkowski space after suitable identifications have been carried out. The existence of this special solution is, in part, the reason why the Gowdy spacetimes are interesting; since there is a solution with a singularity in the sense of causal geodesic incompleteness for which the curvature remains perfectly bounded, it is necessary to prove that this solution is unstable under perturbations.

4. Main result

In order for us to be able to make a formal statement of strong cosmic censorship, let us introduce the following terminology.

Definition 3. *Let $\mathcal{S}_{i,p}$ denote the set of smooth initial data to (6) and (7) on $\mathbb{R} \times S^1$, and let $\mathcal{S}_{i,p,c}$ denote the subset of $\mathcal{S}_{i,p}$ such that the corresponding solutions obey*

$$\int_{S^1} (P_\tau P_\theta + e^{2P} Q_\tau Q_\theta)\, d\theta = 0. \tag{10}$$

Remark. The left-hand side of (10) is independent of τ due to the equations.

Definition 4. *Let (M, g) be a connected Lorentz manifold which is at least C^2. Assume there is a connected C^2 Lorentz manifold $(\hat{M}, \hat{g})$ of the same dimension as M and an isometric embedding $i : M \to \hat{M}$ such that $i(M) \neq \hat{M}$. Then M is said to be C^2-extendible. If (M, g) is not C^2 extendible, it is said to be C^2 inextendible.*

Finally, we are able to give a precise statement of strong cosmic censorship in the class of T^3 Gowdy spacetimes, cf [26].

Theorem 3. *There is a subset $\mathcal{G}_{i,c}$ of $\mathcal{S}_{i,p,c}$ with the following properties*

- *$\mathcal{G}_{i,c}$ is open with respect to the $C^2 \times C^1$ topology on $\mathcal{S}_{i,p,c}$,*
- *$\mathcal{G}_{i,c}$ is dense with respect to the C^∞ topology on $\mathcal{S}_{i,p,c}$,*
- *every MGHD corresponding to initial data in $\mathcal{G}_{i,c}$ has the property that in one time direction, it is causally geodesically complete, and in the opposite time direction, the Kretschmann scalar $R_{\alpha\beta\gamma\delta}R^{\alpha\beta\gamma\delta}$ is unbounded along every inextendible causal curve,*
- *every MGHD corresponding to initial data in $\mathcal{G}_{i,c}$ is C^2 inextendible.*

Remark. All T^3-Gowdy spacetimes have the property that every causal geodesic is complete to the future and incomplete to the past, cf [23].

Note that the first two points of the statement clarify what is meant by generic in this context (openness and denseness) and $\mathcal{G}_{i,c}$ is the set of generic initial data. The openness means that the property of being generic is stable; starting with generic initial data, there is a ball with positive radius centered at this initial data contained in the set of generic initial data. The denseness means that arbitrary initial data can be arbitrarily well approximated by generic initial data.

5. Expanding direction

Let us describe some of the aspects of the asymptotic behaviour of solutions. It is natural to divide the problem of analyzing the asymptotics into a consideration of the expanding direction and a consideration of the direction towards the singularity. In the present section, we shall use the time coordinate $t = e^{-\tau}$, since $t \to \infty$ then corresponds to the expanding direction. With respect to this time coordinate, the equations take the form

$$P_{tt} + \frac{1}{t} P_t - P_{\theta\theta} - e^{2P}\left(Q_t^2 - Q_\theta^2\right) = 0 \tag{11}$$

$$Q_{tt} + \frac{1}{t} Q_t - Q_{\theta\theta} + 2\left(P_t Q_t - P_\theta Q_\theta\right) = 0. \tag{12}$$

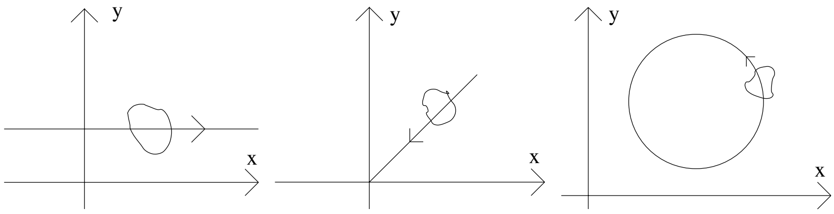

Figure 1. The three types of non-degenerate circles together with the solution at a fixed point in time, as seen in the upper half plane.

These equations can be viewed as a wave map equation with hyperbolic space as a target, cf [24]. The representation of hyperbolic space naturally associated with the equations is

$$g_R = \mathrm{d}P^2 + \mathrm{e}^{2P}\,\mathrm{d}Q^2 \tag{13}$$

on $\mathbb{R}^2$ and the map taking (Q, P) to (Q, e^{-P}) defines an isometry from $(\mathbb{R}^2, g_R)$ to the upper-half plane model. One reason for making this observation is that, when studying the behaviour of solutions, it is important to keep in mind that the natural geometry with which to measure the length of tangent vectors etc. in the target space is the hyperbolic one. Another reason is that due to the wave map structure, isometries of hyperbolic space map solutions to solutions; it is often convenient to use a suitable isometry to make the solution as simple as possible.

In order to visualize a solution, it is natural to think of $[Q(t,\cdot), P(t,\cdot)]$ for each time t as a loop in hyperbolic space. With this perspective, a solution is a loop evolving with time and a natural object to consider is the length of the loop, which is given by

$$\ell(t) = \int_{S^1} \left[\left(P_\theta^2 + \mathrm{e}^{2P} Q_\theta^2\right)(t,\theta)\right]^{1/2} \mathrm{d}\theta.$$

It turns out that, for a given solution, there is a constant $C > 0$ such that

$$\ell(t) \leqslant C t^{-1/2}$$

for all $t \geqslant 1$ (this result, as well as most other results quoted in this section are to be found in [23], but see also [2] for numerical results that influenced the mathematical studies). As a consequence, the spatial variation of the solutions dies out. Thus, it seems natural to expect the solutions to behave as spatially homogeneous solutions to the equations asymptotically. The orbits of the spatially homogeneous solutions are the geodesics of hyperbolic space. In other words, they are circles intersecting the boundary at a right angle (where we use the terminology, as we shall below as well, that straight lines in the upper-half plane are called circles). However, when analyzing the asymptotics, one can see that solutions generally asymptote to circles that need not be the orbits of geodesics. There are four types of circles in the upper half plane: a point (a degenerate circle), a non-degenerate circle that does not intersect the boundary, a circle that touches the boundary (a horocycle) and a circle that intersects the boundary transversally. After applying a suitable isometry of hyperbolic space, the non-degenerate situations can be reduced to one of the cases illustrated in figure 1. The figures depicted should be interpreted in the following way: the wiggly line represents the solution at one point in time (i.e. a loop), the length of the loop shrinks to zero and the solution converges to the circles depicted, going to infinity in the first case, to the origin in the second case, and oscillating around the circle forever in the third case. The first case is a borderline case between the last two; if one perturbs the initial data of a solution as depicted on the left, one will typically end up with a solution behaving as depicted in the center or on the right.

However, the other two situations are stable under perturbations. In all the cases, the solutions are arbitrarily well approximated by solutions to an ODE. After applying a suitable isometry of hyperbolic space, a non-degenerate spatially homogeneous solution has the y-axis as its orbit. For a solution, say $\mathbf{x}$, there is a circle, say Γ, such that the distance from $\mathbf{x}$ to Γ tends to zero as $t^{-1/2}$. The most interesting situation is when Γ is a non-degenerate circle which does not intersect the boundary. What happens in that case is that the solution oscillates around the circle for ever and is asymptotically periodic in a logarithmic time coordinate. In fact, $(Q - \alpha/(2\beta), P)$ can in the limit be arbitrarily well approximated by a solution (u, v) to

$$2\dot{u} = \frac{\beta}{t}\left(\mathrm{e}^{-2v} + \frac{\alpha^2 + 4\beta\gamma}{4\beta^2} - u^2\right), \tag{14}$$

$$\dot{v} = \frac{\beta}{t}u \tag{15}$$

where

$$\alpha = \frac{1}{2\pi}\int_{S^1}\{2Q(tQ_t)\,\mathrm{e}^{2P} - 2(tP_t)\}\,\mathrm{d}\theta \tag{16}$$

$$\beta = \frac{1}{2\pi}\int_{S^1}\mathrm{e}^{2P}(tQ_t)\,\mathrm{d}\theta \tag{17}$$

$$\gamma = \frac{1}{2\pi}\int_{S^1}\{(tQ_t)(1 - \mathrm{e}^{2P}Q^2) + 2Q(tP_t)\}\,\mathrm{d}\theta \tag{18}$$

are quantities that are preserved by the evolution, i.e. they are constants (in the case that the solution converges to a non-degenerate circle which does not intersect the boundary, we have $\alpha^2 + 4\beta\gamma < 0$, which is impossible for a spatially homogeneous solution). In fact, for every $\epsilon > 0$ there is a t_0 such that for $t \geqslant t_0$,

$$(Q - \alpha/(2\beta) - u)^2 + (P - v)^2 \leqslant \epsilon^2$$

where (u, v) is a solution of (14) and (15) with suitably chosen initial data.

There are several remarks worth making in the present context. First of all, the solution is, in the expanding direction, arbitrarily well approximated by a solution to an ODE, but the relevant ODE is *not* the ODE one obtains by dropping the spatial derivatives in the original equation. In fact, the behaviour of solutions to the relevant ODE is qualitatively completely different from the behaviour of spatially homogeneous solutions. We are led to make the following observations.

- Say that we have a solution to a PDE and that it, asymptotically, is arbitrarily well approximated by a solution to an ODE, does the ODE solution have to be a spatially homogeneous solution to the PDE? No!
- Say that we wish to model some phenomenon with a nonlinear evolutionary PDE, but we do not know which PDE is the relevant one. Say, furthermore, that we know that, asymptotically, the phenomenon is arbitrarily well approximated by a solution to an ODE. Does the PDE we choose have to have the property that it allows the given ODE solution as a spatially homogeneous solution? No!
- Say that we wish to understand the asymptotic behaviour of solutions to a PDE, the spatial variations of which die out asymptotically. Is it enough to consider perturbations of spatially homogeneous solutions to the equations to get an impression of what types of qualitative behaviour is possible? No! There are examples where interesting and important phenomena are missed if one takes the perturbation point of view.

All this having been said, it should be pointed out that the spacetime *metrics* corresponding to these 'funny' solutions are rather similar to spatially homogeneous vacuum metrics. Whether the above observations are of any relevance to cosmology is debatable, but they are worth keeping in mind. In particular, just because the spatial variation of a solution to a PDE dies out asymptotically does not mean that it is well approximated by a spatially homogeneous solution to the PDE.

Finally, let us make some observations concerning the asymptotic behaviour that might be of relevance to those interested in making a numerical analysis of this class of spacetimes. Let us consider the polarized subcase, i.e. the equations one obtains when $Q = 0$. Considering (11) and (12), it is clear that the equations then reduce to the single, linear equation

$$P_{tt} + \frac{1}{t}P_t - P_{\theta\theta} = 0. \tag{19}$$

Due to results of [16], there is, given a solution P to (19), two constants α and β and a solution ν to the flat space wave equation, i.e.

$$\nu_{tt} - \nu_{\theta\theta} = 0,$$

with the property that its mean value is always zero, i.e.

$$\int_{S^1} \nu(t,\theta)\,\mathrm{d}\theta = 0$$

for all $t > 0$, such that

$$P = \alpha \ln t + \beta + t^{-1/2}\nu + \psi, \tag{20}$$

where $\psi = O(t^{-3/2})$, and similarly for the first derivatives of ψ, and the mean value of ψ is zero. Due to results of [25], one can also go in the other direction, i.e. given constants α and β and a solution to the flat space wave equation ν with zero mean value, there is a unique ψ with the above mentioned properties such that P given by (20) satisfies (19). In other words, one can consider α, β and ν as data at the moment of infinite expansion.

The point is that, considering (20), it is clear that the most important object, determining the overall behaviour, is α. This constant can be obtained as

$$\alpha = \frac{1}{2\pi}\int_{S^1} t P_t(t,\theta)\,\mathrm{d}\theta. \tag{21}$$

The right-hand side is, needless to say, conserved by the evolution. However, considering the integrand, one sees that

$$t P_t = \alpha + t^{1/2}\nu_t - \tfrac{1}{2}t^{-1/2}\nu + t\psi_t. \tag{22}$$

Note that the last two terms converge to zero as $t \to \infty$. However, ν_t is bounded but no better, so that the second term tends to infinity as $t^{1/2}$ as $t \to \infty$ (recall that we are free to specify ν under the conditions mentioned above, so that we can take it to be, e.g., $\sin t \sin\theta$). Let us sum up the above observations

- The constant α given by (21) determines the overall behaviour of P.
- The relevant integrand appearing in (21) is given by (22).
- In the limit as $t \to \infty$, only the first two terms on the right-hand side of (22) are of relevance.
- The second term in (22) vanishes upon integration, but from a numerical point of view it seems natural to expect the first term to be noise for t large enough. Consequently, it seems natural to expect that the object appearing on the right-hand side of (21) not to be preserved by the numerical evolution and, since α determines the overall behaviour, to expect that the numerical solution might deviate significantly from the real solution as $t \to \infty$.

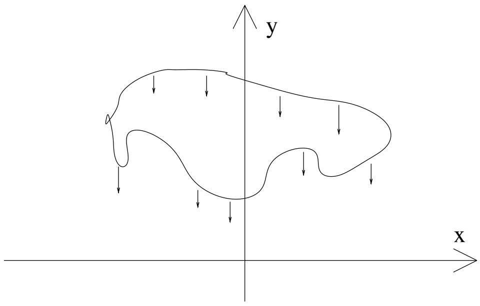

Figure 2. The asymptotic behaviour of a solution satisfying asymptotics of the form (23) and (24) as seen in the upper-half plane model.

The crucial point is that there are objects appearing in the expression for P_t which are on two different scales: there is the 'spatially homogeneous scale' which is on the level of t^{-1} and determines the overall behaviour and the 'spatially inhomogenous scale' which is on the level $t^{-1/2}$. It seems reasonable to expect, for reasons mentioned above, that this difference of scales might cause problems from a numerical point of view. However, and this is the important point, it is exactly this difference of scales that makes the general (non-polarized) inhomogenous problem tractable from a mathematical point of view. In other words, the difference in scales which might be expected to be a nuisance from a numerical point of view, is of a great help in the mathematical analysis.

6. The direction towards the singularity

The natural starting point for discussing the asymptotic behaviour in the direction of the singularity is the asymptotic expansions first proposed by Grubišić and Moncrief, cf [15]. In our setting, the natural expansions are

$$P(\tau,\theta) = v_a(\theta)\tau + \phi(\theta) + u(\tau,\theta) \tag{23}$$

$$Q(\tau,\theta) = q(\theta) + e^{-2v_a(\theta)\tau}[\psi(\theta) + w(\tau,\theta)] \tag{24}$$

where $w, u \to 0$ as $\tau \to \infty$ and $0 < v_a(\theta) < 1$. Note that if we have a solution with such expansions, then $Q(\tau,\theta)$ converges and $P(\tau,\theta)$ tends to infinity as $\tau \to \infty$. Viewing this in the upper-half plane, where $x = Q$ and $y = e^{-P}$, we see that for a fixed θ the solution roughly speaking goes to the boundary along a geodesic in the upper-half plane model, cf figure 2. In [17, 19], the authors developed methods for proving that given v_a, ϕ, q, ψ with a suitable degree of regularity and $0 < v_a < 1$, there are unique solutions to (6) and (7) with asymptotics of the form (23) and (24). It is of interest to note that if q is constant, the condition on v_a can be relaxed to $v_a > 0$. In [21, 22], we proved results going in the other direction, i.e. we provided conditions on initial data which lead to asymptotic expansions of the form (23) and (24).

6.1. Asymptotic velocity

According to our experience, the most important part of the expansions (23) and (24) is the function v_a. This object may seem to be arbitrary and devoid of geometric content. That this is not the case can be seen in the following way. Define the potential and kinetic energy densities by

$$\mathcal{P}(\tau,\theta) = \mathrm{e}^{-2\tau}\big(P_\theta^2 + e^{2P}Q_\theta^2\big)(\tau,\theta) \tag{25}$$

$$\mathcal{K}(\tau,\theta) = \big(P_\tau^2 + \mathrm{e}^{2P}Q_\tau^2\big)(\tau,\theta). \tag{26}$$

Note that these objects are geometric in nature, cf (13). Differentiating the expansions, assuming $u_\tau, w_\tau \to 0$, and computing $\mathcal{K}$, one sees that this expression converges to v_a^2. In this sense, v_a^2 has a geometric significance. One can prove that the pointwise limit of the kinetic energy density always exists. This naturally leads to the following definition.

Definition 5. *Let (Q, P) be a solution to (6) and (7) and let $\theta_0 \in S^1$. Then we define the asymptotic velocity at θ_0 to be*

$$v_\infty(\theta_0) = \Big[\lim_{\tau\to\infty}\mathcal{K}(\tau,\theta_0)\Big]^{1/2}.$$

The importance of the asymptotic velocity is partly due to the fact that it can be used as an indicator for curvature blow up. The reason is that the S^1 coordinate of an inextendible causal curve has to converge to something, say θ_0, in the direction of the singularity (see below) and if $v_\infty(\theta_0) \neq 1$, then the curvature blows up along the causal curve. Note that the solution $P = \tau$, $Q = 0$ has the property that $v_\infty = 1$. Furthermore, the corresponding metric, with $\lambda = \tau$, has a curvature tensor which is identically zero. In other words, if $v_\infty(\theta_0) = 1$, the curvature need not necessarily blow up along a causal curve ending at θ_0. Another reason why the asymptotic velocity is of importance is that it can be used as an indicator for the existence of expansions of the form (23) and (24). An example of such a statement would be if $0 < v_\infty(\theta_0) < 1$ and $P_\tau(\tau,\theta_0) \to v_\infty(\theta_0)$, then there are expansions of the form (23) and (24) in a neighbourhood of θ_0, cf [24].

6.2. *Generic solutions*

To give a formal definition of the set of generic initial data would require an extensive technical digression. Since we wish to avoid that here, we refer the reader to [26] for the details and simply describe some aspects of the asymptotic behaviour of the corresponding solutions. Let (Q, P) be a smooth solution corresponding to generic initial data and let the associated asymptotic velocity be denoted by v_∞. Except for a finite number of exceptional points,

$$\lim_{\tau\to\infty} P_\tau(\tau,\theta) = v_\infty(\theta), \qquad v_\infty(\theta) \in (0,1).$$

For a θ satisfying these two conditions, there is an open neighbourhood of θ such that we have expansions of the form (23) and (24) in this neighbourhood, as was noted above. The exceptional points are of two types, so let us denote the corresponding coordinates on S^1 by $\theta_1,\dots,\theta_l$ and $\theta_1',\dots,\theta_m'$. For $i = 1,\dots,l$, we have

$$\lim_{\tau\to\infty} P_\tau(\tau,\theta_i) = v_\infty(\theta_i) \qquad v_\infty(\theta_i) \in (1,2).$$

These points are referred to as true spikes, since v_∞, which is a geometric quantity, is discontinuous at θ_i, $i = 1,\dots,l$. For $i = 1,\dots,m$, we have

$$\lim_{\tau\to\infty} P_\tau(\tau,\theta_i') = -v_\infty(\theta_i') \qquad v_\infty(\theta_i') \in (0,1).$$

These points are referred to as false spikes, since v_∞ is perfectly smooth across a false spike, but the limit of P_τ is discontinuous. The reader interested in more details concerning the asymptotics is referred to [24, 26].

The reader interested in numerical studies of the singularity is referred to, e.g. [1], and a good reference for a discussion of spikes is [20], see figure 3.

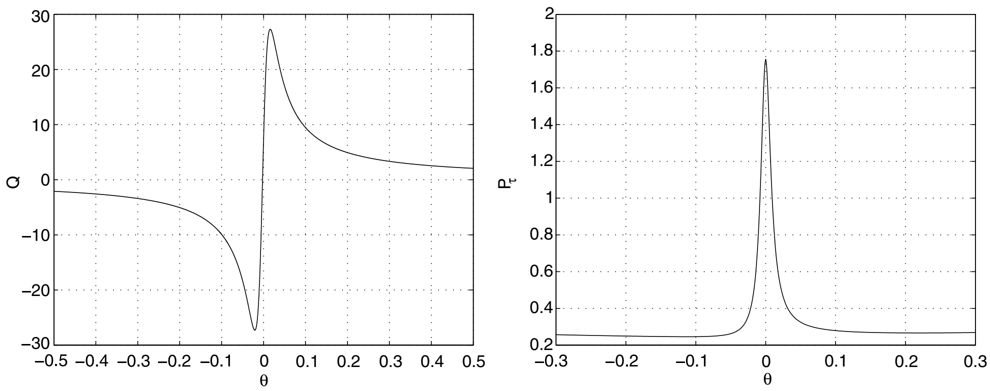

Figure 3. The first picture depicts Q for a false spike about to form and the second one P_τ for a true spike about to form.

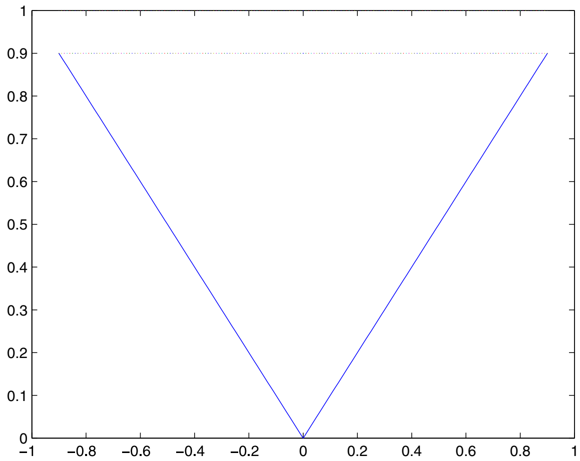

Figure 4. A causal triangle.

(This figure is in colour only in the electronic version)

6.3. Causal structure

Consider the Gowdy metric (5) in the $\tau\theta$ directions. Changing the time coordinate to $t = e^{-\tau}$, the singularity corresponds to $t = 0$ and the Gowdy metric in the $t\theta$ directions is conformal to the Minkowski metric. One consequence of this is that the S^1 coordinate of an inextendible causal curve converges in the direction of the singularity. Say, furthermore, that one wants to be able to predict what happens at $(t, \theta) = (0, \theta_0)$. Then it is enough to know what the initial data are at $t = T$ for $\theta \in [\theta_0 - T, \theta_0 + T]$, cf figure 4.

6.4. Conclusions

As has already been pointed out above, the causal picture in the $t\theta$ directions is conformal to that of Minkowski space after suitable identifications have been made in the spatial direction. If one is interested in determining the behaviour of the metric along causal curves, it is thus enough to focus on causal triangles (by a causal triangle we here mean a triangle with one vertex at the singularity such that the two sides intersecting at that point form 45° angles to the left and to the right of the vertical axis and such that the base of the triangle is parallel to the horizontal axis, see figure 4). It is important to note that the behaviour is quite simple in a causal triangle; the potential energy density $\mathcal{P}$ converges to zero, P_τ converges, $e^P Q_\tau$ converges to zero and the spatial variation of P_τ converges to zero. Furthermore, the spikes are only visible if one considers larger regions and in the proof of the existence of the limit defining the asymptotic velocity it is of crucial importance to consider causal triangles and nothing more. The drawback is of course that one only obtains information concerning one spatial point on the singularity. However, there are solutions for which the asymptotic velocity has infinitely many points of discontinuity. Widening the causal triangle to a region which includes an open subset of the singularity, the situation can thus become extremely complicated; there could be infinitely many points of discontinuity of the asymptotic velocity, the potential energy density need not converge to zero in the widened region etc. In other words, there are many reasons for focusing on a causal triangle; this is exactly the region one needs to control in order to predict the behaviour along causal curves and considering anything larger makes the situation much harder to analyze. This observation is of potential importance when considering classes of spacetimes with less symmetry. If, in the class of Gowdy spacetimes, the situation becomes almost unmanageable when taking the step from one point at the singularity to an open interval, then it seems unreasonable to expect the situation to be easier to deal with in a more general class. As a consequence, it seems natural to try to limit one's attention to as small a region as possible which is still large enough that one can predict what happens up to the singularity along some causal curve. One way of doing so would be to consider a past intextendible causal curve, take the causal future of it and intersect it with some Cauchy hypersurface. Knowing the initial data on the resulting subset of the Cauchy hypersurface, say $\mathcal{D}_\Sigma$, would be sufficient for predicting the behaviour of the gravitational field in the direction toward the singularity along the causal curve with which we started. However, for this to be useful, one needs to know that $\mathcal{D}_\Sigma$ is not 'too large', but *a priori*, there is no reason to assume it to be smaller than Σ itself. Note that in the BKL picture and similar proposals, it is assumed that $\mathcal{D}_\Sigma$ tends to a point as the distance from Σ to the singularity tends to zero (this distance is of course not canonically defined, but for each point on Σ, we can compute the maximum length of past directed causal curves starting at that point, then we can take the infimum over Σ of all the lengths and finally define the result to be the distance). From this point of view, it is of central importance to understand the causal structure close to the singularity. It is of interest to note that is not thoroughly understood even in the case of spatially homogeneous spacetimes; the causal structure of Bianchi VIII and IX in the direction of the singularity remains a mystery.

Finally, let us note that there are infinite dimensional families of initial data, the asymptotic behaviour in the direction of the singularity of which is unstable under perturbations. In fact, as we noted in connection with (23) and (24), we are allowed to specify v_a, ϕ and ψ freely under the condition that $v_a > 0$. However, if $v_a \geqslant 1$ at some point, one is not free to specify q; it has to be constant. In other words, we can demand that $v_a = 4711$ on the entire circle and we are still free to specify ϕ and ψ as we wish. However, the generic solutions have asymptotic velocity strictly less than 2, so that a slight perturbation of the initial data of the solutions with $v_a = 4711$ suffices in order to yield solutions with drastically different asymptotic behaviour. There are also solutions with an infinite number of true spikes, cf remark 8.6 of [24]. However, this behaviour is unstable as well, since no generic solution has this property.

Acknowledgments

The research was supported by the Swedish Research Council and the Göran Gustafsson Foundation.

References

[1] Berger B K and Garfinkle D 1998 Phenomenology of the Gowdy universe on $T^3 \times \mathbb{R}$ *Phys. Rev.* D **57** 1767–77

[2] Berger B K 2002 Asymptotic behaviour of a class of expanding Gowdy spacetimes *Preprint* gr-qc/0207035

[3] Choquet-Bruhat Y and Geroch R 1969 Global aspects of the Cauchy problem in general relativity *Commun. Math. Phys.* **14** 329–35

[4] Chruściel P T and Isenberg J 1993 Nonisometric vacuum extensions of vacuum maximal globally hyperbolic spacetimes *Phys. Rev.* D **48** 1616–28

[5] Chruściel P T 1990 On spacetimes with $U(1) \times U(1)$ symmetric compact Cauchy surfaces *Ann. Phys., NY* **202** 100–50

[6] Chruściel P T, Isenberg J and Moncrief V 1990 Strong cosmic censorship in polarised Gowdy spacetimes *Class. Quantum Grav.* **7** 1671–80

[7] Dafermos M and Rendall A D 2005 Inextendibility of expanding cosmological models with symmetry *Class. Quantum Grav.* **22** L143–7

[8] Dafermos M and Rendall A D 2006 Strong cosmic censorship for T^2-symmetric cosmological spacetimes with collisionless matter *Preprint* gr-qc/0610075

[9] Dafermos M and Rendall A D 2007 Strong cosmic censorship for surface-symmetric cosmological spacetimes with collisionless matter *Preprint* gr-qc/0701034

[10] Eardly D and Moncrief V 1981 The global existence problem and cosmic censorship in general relativity *Gen. Rel. Grav.* **13** 887–92

[11] Fourès-Bruhat Y 1952 Théorème d'existence pour certains systèmes d'équations aux derivées partielles non linéaires *Acta Math.* **88** 141–225

[12] Einstein A 1915 Zur Allgemeinen Relativitätstheorie *Preuss. Akad. Wiss. Berlin, Sitzber.* 778–86

[13] Einstein A 1915 Zur Allgemeinen Relativitätstheorie *Preuss. Akad. Wiss. Berlin, Sitzber.* 844–7

[14] Gowdy R H 1974 Vacuum spacetimes with two-parameter spacelike isometry groups and compact invariant hypersurfaces: topologies and boundary conditions *Ann. Phys., NY* **83** 203–41

[15] Grubišić B and Moncrief V 1993 Asymptotic behaviour of the $T^3 \times \mathbb{R}$ Gowdy space-times *Phys. Rev.* D **47** 2371–82

[16] Jurke T 2003 On the future asymptotics of polarized Gowdy $\mathbb{T}^3$-models *Class. Quantum Grav.* **20** 173–91

[17] Kichenassamy S and Rendall A D 1998 Analytic description of singularities in Gowdy spacetimes *Class. Quantum Grav.* **15** 1339–55

[18] Moncrief V 1981 Global properties of Gowdy spacetimes with $T^3 \times \mathbb{R}$ topology *Ann. Phys., NY* **132** 87–107

[19] Rendall A D 2000 Fuchsian analysis of singularities in Gowdy spacetimes beyond analyticity *Class. Quantum Grav.* **17** 3305–16

[20] Rendall A D and Weaver M 2001 Manufacture of Gowdy spacetimes with spikes *Class. Quantum Grav.* **18** 2959–76

[21] Ringström H 2004 On Gowdy vacuum spacetimes *Math. Proc. Camb. Phil. Soc.* **136** 485–512

[22] Ringström H 2004 Asymptotic expansions close to the singularity in Gowdy spacetimes *A Spacetime Safari: Essays in honour of Vincent Moncrief* ed B Berger and J Isenberg *Class. Quantum Grav.* **21** S305–22

[23] Ringström H 2004 On a wave map equation arising in General Relativity *Commun. Pure Appl. Math.* **57** 657–703

[24] Ringström H 2006 Existence of an asymptotic velocity and implications for the asymptotic behaviour in the direction of the singularity in T^3-Gowdy *Commun. Pure Appl. Math.* **59** 977–1041

[25] Ringström H 2005 Data at the moment of infinite expansion for polarized Gowdy *Class. Quantum Grav.* **22** 1647–53

[26] Ringström H Strong cosmic censorship in T^3-Gowdy, accepted *Ann. Math.*, http://www.math.kth.se/∼ hansr