The Cauchy Problem in General Relativity*

Hans Ringström

Department of Mathematics,
KTH, Royal Institute of Technology,
Stockholm, Sweden

After a brief introduction to classical relativity, we describe how to solve the Cauchy problem in general relativity. In particular, we introduce the notion of gauge source functions and explain how they can be used in order to reduce the problem to that of solving a system of hyperbolic partial differential equations. We then go on to explain how the initial value problem is formulated for the so-called Einstein-Vlasov system, and describe a recent future global non-linear stability result in this setting. In particular, this result applies to models of the universe which are consistent with observations.

PACS numbers: 98.80.Jk

1. Introduction

In the period following Einstein's introduction of his equations [4, 5], explicit solutions, such as the Schwarzschild solution, were found. Since some of the spacetimes found appeared to have singularities, it was natural to ask: are the singularities physical or a consequence of a high degree of symmetry of the solutions? Einstein himself thought that singularities would not appear in less symmetric solutions. An additional complication that occurred in the attempts to understand the Schwarzschild solution was the appearance of the $r = 2M$ hypersurface with respect to the standard coordinates. Did this hypersurface represent a singularity?

In the field of cosmology, Einstein suggested the solution which has since been named the Einstein static universe as a model. However, by doing so he overlooked the fact that expanding solutions are more natural in the general theory of relativity (for stability reasons), so that he failed to make a theoretical prediction which would otherwise have served as a

* Presented at 53 Cracow School of Theoretical Physics

confirmation of his theory; he had to introduce a cosmological constant in order to obtain a static solution (something which, at the time, was not yet necessary); and the solution which he advocated was unstable, and therefore not a reasonable model of the universe.

In the 50's and 60's, the work of Hawking and Penrose shed important light on the issue of singularities; cf., e.g., [10, 11, 15, 16]. What Hawking and Penrose proved was, roughly speaking, that spacetimes generally exhibit singularities. The results have therefore been dubbed the 'singularity theorems'. However, it is important to remember that the notion of a singularity used in the results is that of causal geodesic incompleteness. In particular, the results do not state that there are singularities in the sense of curvature blow up, something it would be desirable to prove. Moreover, there are examples of solutions which have singularities in the sense of causal geodesic incompleteness, but which do not have curvature singularites. The perhaps simplest example of this behaviour is the so-called flat Kasner solution

$$g_F = -dt^2 + t^2dx^2 + dy^2 + dz^2 \tag{1}$$

on $(0,\infty)\times R^3$. This solution is past causally geodesically incomplete, and Hawking's theorem applies to it; cf., e.g., [14, Theorem 55A, p. 431]. However, the Riemann curvature tensor of (1) is identically zero. From this point of view, the $t = 0$ hypersurface can therefore not be considered to be a singularity. In fact, the above solution can be embedded into Minkowski space, and due to this embedding it is possible to extend the solution beyond the $t = 0$ hypersurface without any singularities appearing. In short, the singularity theorems are important, but they do leave questions of interest unanswered. In connection with the appearance of the singularity theorems, the subject of Lorentz geometry developed into an important field in its own right. The geometric perspectives it provided made it possible to understand some of the fundamental solutions (such as the Schwarzschild solution) in a satisfactory way. In particular, the $r = 2M$ hypersurface in Schwarzschild obtained its proper interpretation.

In parallel with the Lorentz geometric perspective, the idea of formulating Einstein's equations as an initial value problem was developed. The seminal work, illustrating that Einstein's equations can be formulated as an initial value problem, is due to Yvonne Choquet-Bruhat; cf. [6]. The main result of the paper is a proof of the fact that, given initial data, there is a corresponding solution to the equations. However, there are infinitely many (inequivalent) developments corresponding to a given initial data set. It would therefore be preferable to prove that there is a development which is uniquely associated with the intial data. In order to obtain such a development, it is necessary to demand some form of maximality. It turns out that demanding maximality in the class of all developments does not lead to a unique object; cf., e.g., [3]. However, demanding maximality in the class of globally hyperbolic developments does. That this is the case was proven in the work of Choquet-Bruhat and Robert Geroch; cf. [2]. It is important to note that there are developments which are maximal in the class of globally hyperbolic developments but which can nevertheless be extended. In fact, there are sometimes inequivalent maximal extensions of the maximal globally hyperbolic development; cf. [3]. As a consequence, Einstein's general theory of relativity is not a deterministic theory; there is a lack of predictability. On the other hand, the existing examples are very special, an observation which leads to the **strong cosmic censorship conjecture**. One formulation of this conjecture is that for generic initial data (in the asymptotically flat or spatially compact setting), the associated maximal globally hyperbolic development is inextendible.

In the mathematics community, the focus of attention has gradually shifted towards the initial value point of view in the study of Einstein's equations. Numerical studies of the equations are also largeley based on this perspective. Some of the fundamental mathematical problems associated with the Cauchy problem are the following:

- **Stability:** most of the solutions used by physicists to model the universe/isolated systems have a high degree of symmetry. It is therefore of interest to ask if small perturbations of the corresponding initial data yield solutions which are similar. The Einstein static universe is an example of a solution which is unstable. In the isolated systems setting, the fundamental problem is that of proving stability of Kerr. In cosmology, proving future stability of standard models is of central importance. Moreover, understanding what features of singularities are robust is an important long term goal.

- **Predictability:** As described above, it is of central importance to prove the strong cosmic censorship conjecture. If this conjecture is not true, the value of considering Einstein's equations as an initial value problem is very limited (it would then be impossible to predict the behaviour of the system on the basis of initial data).

- **Singularities:** As already noted, singularities occur in many of the highly symmetric standard solutions used to model isolated systems and the universe. Moreover, singularities in the sense of causal geodesic incompleteness are generic. However, it is not clear that, for generic initial data, the curvature blows up in the incomplete directions of causal geodesics in the corresponding maximal Cauchy development. Note that if one were able to prove the corresponding conjecture, then the strong cosmic censorship conjecture would follow.

- **Homogenization/isotropization:** The standard starting point in cosmology is the cosmological principle; i.e., the assumption that the universe is spatially homogeneous and isotropic. However, it would be preferable to deduce this as a consequence of the evolution associated with Einstein's equations, rather than putting it in by hand as an assumption. In the presence of a positive cosmological constant, there is a conjecture, which goes under the name of the cosmic no-hair conjecture, which states that from the point of view of late time observers, the universe should approach de Sitter space asymptotically (in regimes where there are no black holes etc.). There are also conjectures in the absence of a cosmological constant, but they are harder to formulate.

It is of some interest to compare the mathematical perspective with the numerical. The advantage of doing numerical computations is that it is possible to compute specific numbers that can be compared with observations. This is typically not possible when taking the mathematical perspective. One disadvantage of numerical studies is that it is only possible to consider a finite number of solutions. Since the space of initial data is infinite dimensional, there is thus the risk that something may be overlooked, and there are examples of this. The advantage of the mathematical perspective is that there is no problem with considering an infinite number of initial data sets at the same time. Analyzing asymptotics is also something which, at least in some special situations, is actually easier to do mathematically than numerically. Moreover, the idea is that when it is possible to prove results, the proofs yield a deeper understanding of what is going on. On the other hand, there are many sitations in which a mathematical analysis cannot be expected to be of use; computing the gravitational wave signal resulting from inspiralling black holes mathematically, e.g., is not realistic. Needless to say, there are plenty of other examples as well. In the end, combining the two perspectives is with all probability the most fruitful approach.

2. Solving the initial value problem

In this section, we describe how to solve Einstein's equations, given initial data. We would also like to give a rough, heuristic justification of the choice of initial data. But in order to develop some intuition, it is natural to start by considering a simple example.

Example: Consider a spatially homogeneous and isotropic, and spatially flat spacetime, say (M, g). Then $M = I \times R^3$, where I is an open interval, and the metric takes the form

$$g = -dt^2 + a^2(t)\bar{g}, \tag{2}$$

where $\bar{g}$ is the standard flat Euclidean metric on R^3 and a is a positive scalar function on I. Say now that we want to find solutions to Einstein's equations with this type of symmetry and matter of dust type. In that case, the stress energy tensor takes the form

$$T = \rho dt^2,$$

where ρ is the energy density (recall that dust corresponds to the pressure being zero). Einstein's equations then take the following form:

$$3\left(\frac{\dot{a}}{a}\right)^2 = \rho, \tag{3}$$

$$0 = 0, \tag{4}$$

$$2\frac{\ddot{a}}{a} + \left(\frac{\dot{a}}{a}\right)^2 = 0, \tag{5}$$

$$\dot{\rho} + 3\frac{\dot{a}}{a} = 0. \tag{6}$$

These equations are, respectively, $G_{00} = T_{00}$, $G_{0i} = T_{0i}$, $G_{ij} = T_{ij}$ and $\text{div}T = 0$, where we use the convention that Latin indices range from 1 to 3; later on we shall use Greek indices, and they will range from 0 to 3. The question is: how do you construct solutions to these equations? In some respects, there seem to be too many equations. In the end, there are many ways to proceed, but one way is to

- choose initial data (in other words, a, $\dot{a}$ and ρ at some time, say t_0) so that (3) holds,
- choose a and ρ to be the solution to (5) and (6),
- hope that the resulting solution is also a solution to (3) for $t \neq t_0$.

A priori, it is not obvious why the solution constructed in the above way should also solve (3) for $t \neq t_0$. Let us, nevertheless, consider

$$f = 3\left(\frac{\dot{a}}{a}\right)^2 - \rho.$$

Using the equations (5) and (6), it can then be computed that

$$\dot{f} = -3\frac{\dot{a}}{a}f.$$

In particular, if $f(t_0) = 0$, then the solution equals zero for all times such that the solution is defined. Note also that f, depending on the sign of $\dot{a}/a$, either increases or decreases exponentially. If f is non-zero initially (which can certainly happen in a numerical simulation), then it can certainly grow exponentially.

Summarizing the above simple example, it is useful to note the following:

- We have broken the diffeomorphism invariance by demanding that the geometry and the coordinates be related according to (2).
- As a result we have obtained a system of evolution equations and constraints; (3) and (4) are the constraints and (5) and (6) are the evolution equations.

The idea is to start with initial data solving the constraints and then to solve the evolution equations. Finally, you then prove that the constraints are propagated. In what follows, we wish to illustrate that these are general aspects of solving the initial value problem for Einstein's equations.

2.1. Gauge source functions

Let us now consider the equations without any symmetry assumptions. For the sake of convenience, we shall restrict our attention to the vacuum case. The vacuum equations are given by

$$R_{\alpha\beta} = 0. \tag{7}$$

Considering the Ricci tensor with respect to local coordinates, we can consider it to be a differential operator acting on the components of the metric. In fact, one way of writing the Ricci tensor is

$$R_{\mu\nu} = -\frac{1}{2} g^{\alpha\beta}\partial_\alpha\partial_\beta g_{\mu\nu} + \nabla_{(\mu}\Gamma_{\nu)} + g^{\alpha\beta}g^{\gamma\delta}[\Gamma_{\alpha\gamma\mu}\Gamma_{\beta\delta\nu} + \Gamma_{\alpha\gamma\mu}\Gamma_{\beta\nu\delta} + \Gamma_{\alpha\gamma\nu}\Gamma_{\beta\mu\delta}]. \tag{8}$$

In this equation

$$\begin{aligned} \Gamma_{\alpha\gamma\beta} &= \frac{1}{2}(\partial_\alpha g_{\beta\gamma} + \partial_\beta g_{\alpha\gamma} - \partial_\gamma g_{\alpha\beta}), \\ \Gamma_\alpha &= g^{\mu\nu}\Gamma_{\mu\alpha\nu}, \\ \nabla_\mu\Gamma_\nu &= \partial_\mu\Gamma_\nu - g^{\alpha\beta}\Gamma_{\mu\alpha\nu}\Gamma_\beta, \\ \nabla_{(\mu}\Gamma_{\nu)} &= \frac{1}{2}(\nabla_\mu\Gamma_\nu + \nabla_\nu\Gamma_\mu). \end{aligned}$$

From a PDE point of view, the last term on the right hand side of (8) is harmless. If it were not for the second term on the right hand side, Einstein's vacuum equations would be a system of hyperbolic PDE's to which the standard local existence theory would apply. On the other hand, we would then obtain uniqueness of a form we know cannot hold; due to the diffeomorphism invariance of the equations, coordinate representations are not unique. In order to obtain an equation with unique solutions, we need to break the diffeomorphism invariance, as in the model example. One way of doing this is by introducing so called *gauge source functions*. The idea is to replace Γ_ν in the second term on the right hand side of (8) with another function, say F_ν. The function F_ν is allowed to depend on the coordinates and on the metric, but not on the derivatives of the metric. In other words, we introduce the modified Ricci tensor

$$\hat{R}_{\mu\nu} = R_{\mu\nu} + \nabla_{(\mu}\mathcal{D}_{\nu)},$$

where $\mathcal{D}_\mu = F_\mu - \Gamma_\mu$. Then

$$\hat{R}_{\mu\nu} = 0 \tag{9}$$

is a system of hyperbolic PDE's to which the standard theory applies. On the other hand, this is not the equation we wish to solve. It is therefore of interest to relate the two equations. Let us assume that we have a solution to (9) (since this is the equation we know how to solve). Due to the definition of $\hat{R}_{\mu\nu}$, we then conclude that

$$R_{\mu\nu} = -\nabla_{(\mu}\mathcal{D}_{\nu)}.$$

Taking the trace of this equation, we can compute the scalar curvature in terms of $\mathcal{D}$. As a consequence, we obtain

$$G_{\mu\nu} = -\nabla_{(\mu}\mathcal{D}_{\nu)} + \frac{1}{2}\nabla^\gamma\mathcal{D}_\gamma g_{\mu\nu}, \tag{10}$$

where $G_{\mu\nu} = R_{\mu\nu} - Rg_{\mu\nu}/2$ is the Einstein tensor. Taking the divergence of (10), the left hand side vanishes (due to the Bianchi identities) and we obtain

$$\nabla^\mu\nabla_\mu\mathcal{D}_\nu + R_\nu{}^\mu\mathcal{D}_\mu = 0. \tag{11}$$

Note that this is a homogeneous wave equation for $\mathcal{D}_\mu$. If we can arrange for the initial data to be such that $\mathcal{D}_\mu$ and $\nabla_\nu\mathcal{D}_\mu$ are zero initially, then $\mathcal{D}_\mu$ is identically zero, wherever the solution is defined. As a consequence, we obtain a solution to the original equation, just as in the example described above.

The crucial question is then: how should we choose the intial data? It is also of interest to ask: where should the initial data be specified? Since the equation (9) is a system of wave equations for the metric components, it is natural to recall the initial value problem for the ordinary wave equation in order to develop some intuition. One natural hypersurface on which to specify initial data for the wave equation is a $t =$ const. hypersurface. Recalling that the standard wave equation is the wave equation with respect to the Minkowski metric, it is of interest to note that the $t = \text{const.}$ hypersurfaces are spacelike hypersurfaces with respect to the Minkowski metric (in fact, they are spacelike Cauchy hypersurfaces). In what follows, we shall also restrict ourselves to spacelike hypersurfaces. That is not to say that it is impossible to consider other situations, but the case with spacelike hypersurfaces constitutes the simplest case. Schematically, the equation we are interested in solving is of the form

$$g^{\mu\nu}\partial_\mu\partial_\nu g_{\alpha\beta} = F_{\alpha\beta}(g, \partial g), \tag{12}$$

where g is a shorthand for all the components of g, and ∂g is a shorthand for all first derivatives of all components of g. In the case of the wave equation, we specify the function and its first time derivative as initial data. In the present situation it thus seems reasonable to specify the metric as well as its normal derivative on a hypersurface, say Σ. Since we want Σ to be a spacelike hypersuface with respect to the metric, the induced metric on Σ should be Riemannian. Moreover, the normal should be timelike. Note, however, that in order to obtain a unique solution to (12), we need more than just the induced metric; we need all the components of the metric initially. The same goes for the normal derivative; we need all the components. On the other hand, if we want a geometric initial value problem, this is not a good choice of initial data; they are coordinate dependent. However, there are geometric candidates:

Geometric initial data: the induced metric and the second fundamental form.

These data constitute information of the above type, and they are geometric in nature. However, they are insufficient in order to guarantee a unique solution to (12). The question is then how the remaining data should be specified. In part there is a freedom; the initial lapse and shift can be chosen freely; i.e., roughly speaking, g_{00} and g_{0i}. This choice simply corresponds to how you parametrize the time coordinate initially (lapse) and how you shift the coordinates around with time (shift). However, once these components have been chosen, the remaining components are uniquely fixed by the requirement that $\mathcal{D}_\mu = 0$ on Σ. Clearly, we thus have a problem: in order to obtain a solution to Einstein's vacuum equations, we need to know that $\nabla_\nu \mathcal{D}_\mu = 0$ on Σ, but there is no freedom left in choosing the initial data. This problem naturally leads us to the topic of the constraint equations.

2.2. Constraint equations

Assuming (M,g) to be a solution to Einstein's vacuum equations and Σ to be a spacelike hypersurface in (M,g), we know that

$$G(N,N) = 0,$$
$$G(N,X) = 0,$$

where N is normal and X is tangential to Σ. Using the so-called Gauss and Codazzi equations, the right hand sides of these equations can be expressed soleley in terms of the induced metric and second fundamental form on Σ; they are given by

$$\bar{R} - \bar{k}^{ij}\bar{k}_{ij} + (\mathrm{tr}\bar{k})^2 = 0, \tag{13}$$
$$\bar{\nabla}^j\bar{k}_{ji} - \bar{\nabla}_i\mathrm{tr}\bar{k} = 0. \tag{14}$$

In these equations, $\bar{R}$ and $\bar{\nabla}$ are the scalar curvature and the Levi-Civita connection of the induced metric $\bar{g}$ respectively, and $\bar{k}$ is the induced second fundamental form. Moreover, indices are raised and lowered with $\bar{g}$. In order for it to be possible to obtain a solution, the constraint equations consequently have to be satisfied. Finally, we are thus able to summarize how to solve Einstein's equations, given initial data.

2.3. Solving the vacuum equations

Step 1: Let Σ (a manifold) be given; it should be thought of as the initial hypersurface.

Step 2: Let $\bar{g}$ and $\bar{k}$ be a Riemannian metric and a symmetric covariant 2-tensor field on Σ; $\bar{g}$ and $\bar{k}$ should be thought of as the induced metric and second fundamental form on the initial hypersurface.

Step 3: Assume $\bar{g}$ and $\bar{k}$ to solve the vacuum constraint equations (13) and (14).

Step 4: Choose local coordinates $(\bar{x}, U)$ on Σ. These coordinates yield local coordinates $(x, R \times U)$ on $R \times \Sigma$.

Step 5: Choose gauge source functions F_μ on $R \times U$; the gauge source functions are also allowed to depend on the metric components (though not on the derivatives).

Step 6: Choose initial data $g_{\mu\nu}|_{t=0}$ and $\partial_t g_{\mu\nu}|_{t=0}$ which induce $\bar{g}$ and $\bar{k}$ and are such that $\mathcal{D}_\mu|_{t=0} = 0$.

Step 7: Solve (9) in a neighbourhood of $\{0\} \times U$, given these initial data; this is a matter of standard PDE theory (which is, however, a subject in its own right).

Step 8: Recall that (10) holds; this equation is a consequence of (9). Contracting this identity with the normal to the hypersurface twice, the left hand side is zero, since the constraint equation (13) is fulfilled. Contracting the equation with one normal vector and one vector which is tangential to the initial hypersurface yields zero, since (14) is satisfied. Combining these observations with the fact that $\mathcal{D}_\mu|_{t=0} = 0$, it is possible to conclude that $(\nabla_\mu \mathcal{D}_\nu)|_{t=0} = 0$; this is essentially a matter of algebraic manipulations. Since $\mathcal{D}_\alpha$ satisfies the homogeneous wave equation (11), we conclude that $\mathcal{D}_\mu = 0$ wherever the solution is defined. As a consequence, the solution to (9) is actually a solution to (7); we have a solution to Einstein's vacuum equations with the specified initial data in a neighbourhood of $\{0\} \times U$.

Step 9: Patching together the local solutions leads to a development of the initial data set $(\Sigma, \bar{g}, \bar{k})$.

To conclude the above discussion, there is a globally hyperbolic development, given initial data.

2.4. Maximal Cauchy development

Even though it is of interest to know that there is a globally hyperbolic development, given initial data, the result is unsatisfactory in one respect: there are infinitely many inequivalent globally hyperbolic developments of the same initial data. In order to speak of a solution, it is necessary to construct a development which is uniquely determined by the initial data. In order to obtain something which is unique, it is necessary to demand some form of maximality. As has already been mentioned, requiring maximality in the class of globally hyperbolic developments is an appropriate choice. In fact, as mentioned, there is a unique maximal globally hyperbolic development associated with initial data, due to the work of Yvonne Choquet-Bruhat and Robert Geroch; cf. [2]. This object is, in a sense, *the* development of the initial data.

The above description is a bit brief. A more extensive discussion, which is nevertheless of an overview character, is to be found in [20, Chapter 2]; this chapter also contains a brief discussion of non-linear wave equations.

In [19] there is a complete proof of the existence of a maximal Cauchy development in the Einstein-non-linear scalar field setting, and in [20] the Einstein-Vlasov-non-linear scalar field setting is treated.

3. On the topology and future stability of the universe

The remainder of this article is devoted to a discussion of the topology and future stability of models of the universe. The first topic is that of future stability of standard models. As a consequence, it is natural to begin by describing the standard perspective. The starting point is the cosmological principle; the assumption that the universe is spatially homogeneous and isotropic. As a consequence of this assumption, the only freedom left in Einstein's theory is the choice of one of three spatial geometries and a scalar function of time; the scale factor. The currently preferred spatial geometry is flat.

Turning to the matter content, it consists of ordinary matter (such as dust and radiation), dark matter (often modeled by dust, though there are many options), and dark energy (again, there are many options for modeling this type of matter, but we shall here simply equate it with a positive cosmological constant). Mathematically, the matter content is usually described by perfect fluids, but we shall here use kinetic theory; we shall return to this topic in greater detail below.

In short, the currently preferred picture is illustrated in Figure 1: there is a big bang, inflation, decoupling, a period of slow expansion, and finally an onset of accelerated expansion caused by dark energy.

3.1. Questions

The above description yields a very nice picture, but it does lead to a few questions.

Stability: The universe may, or may not, be almost spatially homogeneous and isotropic, but it is clear that it is not exactly spatially homogeneous and isotropic. As a consequence, it is natural to ask if the standard models are future stable. In other words, if we perturb initial data corresponding to a standard solution, do we obtain a spacetime which is globally similar to the future?

Topology: Another question concerns the global topology of the universe. The standard models only admit a very limited class of topologies (even if one is prepared to admit locally spatially homogeneous and isotropic solutions). However, it is not so clear what happens if we only assume every observer to consider the universe to be almost spatially homogeneous and isotropic.

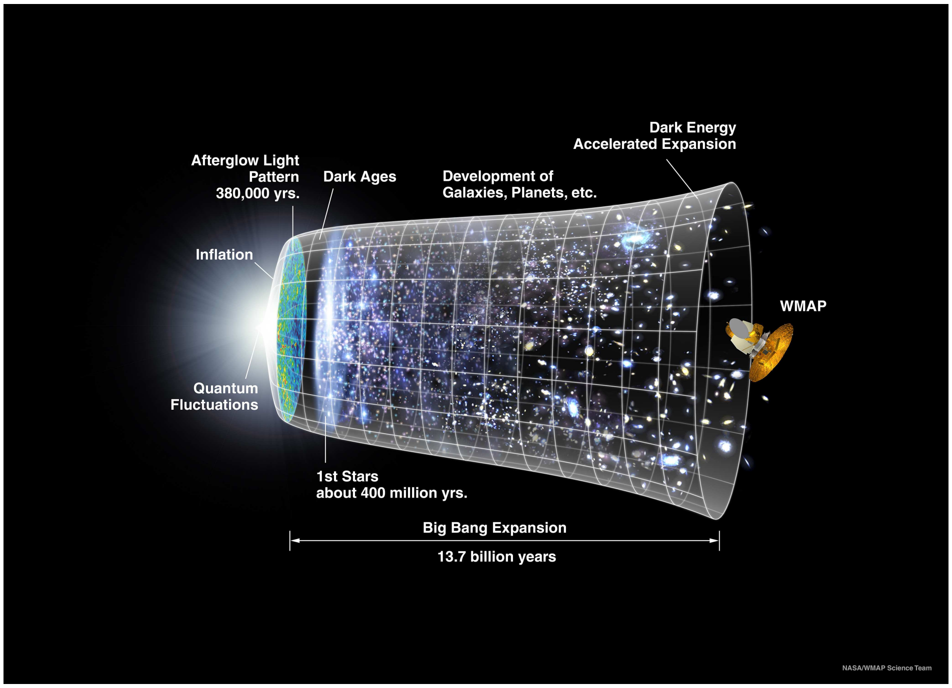

Fig. 1. Currently preferred model of the universe, courtesy of NASA/WMAP Science Team.

3.2. Matter models

Let us now turn to the subject of the matter models in greater detail. In the standard models, the matter is usually described by perfect fluids. In the spatially homogeneous and isotropic setting, perfect fluids are characterized by an energy density, say ρ and a presure, say p. In order to obtain equations for the matter, an equation of state, expressing p in terms of ρ, is often imposed. Two common choices are dust, in which case the equation of state is $p = 0$, and radiation, in which case the equation of state is $p = \rho/3$.

Vlasov matter. Here we shall prefer kinetic theory, or, more specifically, Vlasov matter in order to model the matter content of the universe. The intuitive interpretation of Vlasov matter is the following: consider a collection of particles, all having unit mass. At a spacetime point, such a particle is represented by a future directed unit timelike vector. Assuming collisions to be sufficiently rare that they can be neglected, the particles travel along timelike geodesics. On the other hand, the particles collectively generate a gravitational field which influences the geometry and, thereby, the geodesics. In order to describe the collection of particles, we use a distribution function. It should be thought of as representing the average properties of an ensemble of collections of particles. Turning to a mathematical description, the relevant mathematical structures are the following. To begin with, the *mass shell* is defined to be the set of future directed unit timelike vectors in the spacetime (M, g) and is denoted by P. An element of P represents the time, space and momentum coordinates of a particle; i.e., P is the space of states of particles. In order to describe the matter, we use a *distribution function*, say f, which is simply a non-negative function on the mass shell; though in practice, we need to impose regularity and fall off conditions on the distribution function in order for the theory to make sense. In order to connect the matter model with Einstein's general theory of relativity, it is necessary to describe how to construct a stress energy tensor out of the distribution function. The relevant prescription is

$$T_{\alpha\beta}|_{\xi} = \int_{P_{\xi}} f p_{\alpha} p_{\beta} \mu_{P_{\xi}}. \tag{15}$$

In this expression, $\xi \in M$; α and β refer to components with respect to some coordinates, say x; P_{ξ} denotes the mass shell above ξ (i.e., the set of future directed unit timelike vectors based at ξ); for $p \in P_{\xi}$, p^{α} denotes the components of p with respect to the coordinates x; $p_{\alpha} = g_{\alpha\beta} p^{\beta}$; and $\mu_{P_{\xi}}$ is defined as follows: the Lorentz metric g induces a Lorentz metric on $T_{\xi}M$, the tangent space at ξ; this Lorentz metric, in its turn, induces a Riemannian metric on P_{ξ} (note that, with respect to a suitable orthonormal frame, P_{ξ} is simply the hyperboloid sitting inside Minkowski space); and this Riemannian metric induces a volume form on P_{ξ}, denoted $\mu_{P_{\xi}}$. Note that in order for this definition to make sense, it is necessary to require some degree of fall off of the distribution function f. Turning to the equation the Vlasov matter has to satisfy, it is given by

$$\mathcal{L} f = 0. \tag{16}$$

Here $\mathcal{L}$ is the vector field induced by the geodesic flow on the mass shell. Another way of formulating the equation (16) is to say that f is constant along future directed unit timelike geodesics, the intuitive motivation being that collisions are neglected. It is of interest to note that (16) is equivalent to (15) being divergence free.

The Einstein-Vlasov system. To summarize: the Einstein-Vlasov system is given by the equations

$$\begin{aligned} G + \Lambda g &= T, \\ \mathcal{L} f &= 0, \end{aligned}$$

where G is the Einstein tensor, Λ is the (positive) cosmological constant, g is the metric, and T is given by (15). This is a system of geometric, non-linear integro-partial differential equations.

3.3. Standard models, approximating fluids

Before proceeding, it is useful to say a few words concerning the model we are going to prove the stability of. It is also of interest to compare kinetic theory with the matter models usually used in the standard models.

The assumption of spatial homogeneity, isotropy and spatial flatness imply that the metric takes the form

$$g = -dt^2 + a^2(t)\bar{g}.$$

In this expression, $\bar{g}$ is the standard flat metric on R^3 and a is a positive function of time; the metric is defined on $I \times R^3$, where I is an open interval. Under these circumstances, the stress energy tensor of the Vlasov matter takes perfect fluid form (as a consequence of the symmetry requirements). However, there is, in general, no equation of state relating the pressure and energy density. On the other hand, it is possible to express the energy density and pressure in terms of the initial datum for the distribution function and the scale factor a:

$$\begin{aligned} \rho_{\mathrm{Vl}}(t) &= \int_{R^3} \bar{\mathsf{f}}\,(a(t)\bar{q})\,(1+|\bar{q}|^2)^{1/2} d\bar{q}, \\ p_{\mathrm{Vl}}(t) &= \frac{1}{3}\int_{R^3} \bar{\mathsf{f}}\,(a(t)\bar{q})\,\frac{|\bar{q}|^2}{(1+|\bar{q}|^2)^{1/2}} d\bar{q}, \end{aligned}$$

where $\bar{\mathsf{f}}$, a function on R^3, is a symmetry reduced version of the initial datum for the distribution function.

The matter models that are commonly used in the standard models are dust and radiation. It is therefore of interest to ask: what is the relation between such models and Vlasov matter? Replacing $\bar{\mathsf{f}}$ by a Dirac δ-function at the origin yields zero pressure, but non-zero energy density; in fact, we recover dust. In the case of massless particles, the expressions $(1+|\bar{q}|^2)^{1/2}$ appearing in the above formulae are replaced by $|\bar{q}|$. As a consequence, $p_{\mathrm{Vl}} = \rho_{\mathrm{Vl}}/3$, and we recover radiation. As an alternative, it is possible to demand that $\bar{\mathsf{f}}$ only be non-zero for very large $|\bar{q}|$; the behaviour of the matter is then in practice as that of a radiation fluid. However, it is of interest to note that $\bar{\mathsf{f}}$ naturally behaves as a radition fluid close to the singularity and as dust in the expanding direction, something which, in a sense, has to be put in by hand in the standard models. If one wants to approximate both dust and a radiation fluid with matter of Vlasov type, the initial datum for the distribution function can be taken to be the sum of two parts: one which approximates a Dirac δ-function and one which has support very far from the origin.

3.4. Initial data

In order to formulate a stability result, it is necessary to explain what is meant by initial data, as well as what is meant by two initial data sets being 'close'. In order to develop some intuition, let us assume that we have a solution (M, g, f) to the Einstein-Vlasov system. Just as in the vacuum setting, we shall also focus on the case when the initial data are specified on a spacelike hypersurface. We shall therefore assume Σ to be a spacelike hypersurface in (M, g). Just as in the vacuum case, the initial data for the geometry are given by the induced (Riemannian) metric and the induced second fundamental form; say $(\bar{g}, \bar{k})$. Concerning the distribution function, we only need its initial datum, since the Vlasov equation is a first order equation. However, there is a slight technical problem; the distribution function is defined on the mass shell, and the mass shell is not intrinsic to the initial hypersurface. However, there is a diffeomorphism from the mass shell above Σ to the tangent space of Σ obtained by projecting orthogonally to the normal. We shall therefore think of the initial datum for the distribution function, say $\bar{f}$, to be defined on $T\Sigma$. Due to the fact that (M, g, f) is a solution to the Einstein-Vlasov system, the initial data $(\bar{g}, \bar{k}, \bar{f})$ have to satisfy the corresponding constraint equations; they are similar to (13) and (14); the only difference is that there is right hand side which can be computed in terms of $\bar{f}$ and $\bar{g}$.

So far, we have assumed that we have a solution and then defined initial data induced on a spacelike hypersurface. However, we would like to go in the other direction. Given $(\Sigma, \bar{g}, \bar{k}, \bar{f})$, where $\bar{g}$ is a Riemannian metric on Σ, $\bar{k}$ is a symmetric covariant 2-tensor field on Σ and $\bar{f}$ is a non-negative function on $T\Sigma$ (all having a suitable degree of regularity and, in the case of $\bar{f}$, appropriate fall off), satisfying the Einstein-Vlasov constraint equations, the question is: is there a corresponding globally hyperbolic development; i.e., a solution to the Einstein-Vlasov system inducing the given initial data on a hypersurface diffeomorphic to Σ? The answer to this question is yes. In fact, just as in the vacuum case, there is a unique maximal Cauchy development associated with the initial data.

What remains to be explained is how to measure the distance between initial data sets. In the case of $\bar{g}$ and $\bar{k}$, we shall use so called Sobolev spaces. In the case of $\bar{f}$, we shall use slight generalizations thereof (which include weights in the tangential directions). We omit the details of how these spaces are defined; the interested reader is referred to [19, 20] for the details.

3.5. Previous results

Before proceeding to a statement of stability, it is natural to say a few words concerning previous results that have been obtained in the case of accelerated expansion. To the best of our knowledge, the first results were obtained by Helmut Friedrich. In [7], he proved among other things, stability of $3+1$-dimensional de Sitter space. Later, he generalized his results to the presence of Yang-Mills and Maxwell fields; cf. [9]. Michael Anderson continued in the same vein by proving stability of the even dimensional de Sitter spaces; cf. [1]. All of these results are based on using a conformal reformulation of the equations so that the question of future global non-linear stability becomes a question of local stability in the conformally rescaled picture. However, in order for this to be possible, it is necessary to reformulate the equations appropriately. Helmut Friedrich succeeded in finding such a reformulation in [8].

Since the methods used by Friedrich and Anderson seem to be dependent on conformal invariance properties of the equations, we developed a different approach to proving stability in [17]. The methods developed in this paper have since been used in several proofs of future global non-linear stability; cf., e.g., [23, 18, 21, 22, 13, 12]. Moreover, the stability result in the Einstein-Vlasov setting that is the subject of the present section, cf. [20], is also based on the methods developed in [17]. Finally, let us note that the results of [17] immediately imply that de Sitter space is stable in all spacetime dimensions ≥ 4; the issue of odd and even spatial dimensions does not arise.

3.6. Statement of the stability result

Let us begin by describing the solutions we are going to prove stability of. There are several ways of doing so, but we shall here prefer to characterize the solutions in terms of their initial data. Since we are interested in proving stability of spatially homogeneous solutions, there are two natural cases:

- either you have left invariant initial data on a 3-dimensional Lie group, or
- you have initial data on $S^2 \times R$ which are invariant under the isometry group of the standard metric on $S^2 \times R$.

The latter case is somewhat special, and we are going to ignore it for the rest of the talk (which is not to say that it is not possible to prove stability of some solutions in this category). Instead, we shall focus on left invariant initial data on Lie groups, and we shall refer to such data as Bianchi initial data. Formally, Bianchi initial data consist of

- a 3-dimensional Lie group, say G,

- a left invariant Riemannian metric, say $\bar{g}_{\mathrm{bg}}$ and symmetric covariant 2-tensor field on G, say $\bar{k}_{\mathrm{bg}}$,
- a left invariant initial datum $\bar{f}_{\mathrm{bg}}$ on TG with appropriate fall off conditions in the momentum directions.

Then $(G, \bar{g}_{\mathrm{bg}}, \bar{k}_{\mathrm{bg}}, \bar{f}_{\mathrm{bg}})$ are referred to as *Bianchi initial data* to the Einstein-Vlasov system, assuming they satisfy the constraints. It should also be pointed out that there are solutions to the Einstein-Vlasov system which are

- consistent with the observations, and
- induce initial data of this type.

In fact, it is sufficent to choose $G = R^3$, $\bar{g}_{\mathrm{bg}}$ and $\bar{k}_{\mathrm{bg}}$ to be suitable multiples of the standard Euclidean metric, and to make an appropriate choice of $\bar{f}_{\mathrm{bg}}$. In short, models consistent with observations are a very small subclass of the situation considered above.

Stability. In order to formulate a stability result, let $(G, \bar{g}_{\mathrm{bg}}, \bar{k}_{\mathrm{bg}}, \bar{f}_{\mathrm{bg}})$ be Bianchi initial data to the Einstein-Vlasov system. In order to prove future stability, we need to impose two conditions. First of all:

- the universal covering group of G should not be isomorphic to SU(2),
- the mean curvature of the initial data should be strictly positive initially.

Why do we impose these conditions? To begin with: what is so special with SU(2)? The point is that SU(2) is the only simply connected 3-dimensional Lie group which admits a left invariant metric of positive scalar curvature. Moreover, positive scalar curvature plays an important role in the analysis. In fact, there are initial data on SU(2) such that the solution has a big bang and a big crunch. But there are also solutions that have a big bang and then expand forever. In the cases that you obtain sufficiently fast expansion, there should be no problem in proving stability, and in fact, there are stability results in the SU(2)-case; cf. [20]. However, in order to obtain a clean statment, we have here decided to exclude SU(2) completely. The condition on the mean curvature is simply there to ensure that the expanding direction is to the future.

In addition to these requirements, we also demand that there be a co-compact subgroup of the isometry group of the initial data, say Γ, and we shall think of the initial data as being defined on the quotient, say Σ. The question is then: why restrict to a compact hypersurface? We expect that it should be possible to prove results in the non-compact setting, but the statement would certainly be much more technical; it would be necessary to introduce much more complicated norms. The statement is then the following: Given Bianchi initial data satisfying the above conditions, there is an $\epsilon > 0$ such that if $(\Sigma, \bar{g}, \bar{k}, \bar{f})$ are initial data satisfying

$$\|\bar{g} - \bar{g}_{\mathrm{bg}}\|_{H^5} + \|\bar{k} - \bar{k}_{\mathrm{bg}}\|_{H^4} + \|\bar{f} - \bar{f}_{\mathrm{bg}}\|_{H^4_{\mathrm{V}l,\mu}} \leq \epsilon,$$

then the maximal Cauchy development of $(\Sigma, \bar{g}, \bar{k}, \bar{f})$ is future causally geodesically complete (the reader interested in the definition of the relevant norms is referred to [20]). Moreover, the solution is asymptotically de Sitter like (in other words, the cosmic no-hair conjecture holds). In addition, it is also possible to derive detailed asymptotics; the interested reader is referred to [20] for a more complete description.

3.7. On the topology of the universe

Let us now turn to the question of the global topology phrased at the beginning: what are the limitations on the global topology imposed by the constraint that every observer considers the universe to be close to one of the standard models? To be more precise, assume that

- the observational data indicate that, to our past, the universe is well approximated by one of the standard models,
- interpreting the data in this model, we only have information concerning the universe for $t \geq t_0$ (here it is natural to think of this time as representing decoupling; the models we are discussing here are only relevant after decoupling),
- there is a big bang (at least in the sense of past causal geodesic incompleteness),
- analogous statements apply to all observers in the universe (with the same standard model and t_0).

The question is then: what conclusions are we allowed to draw concerning the global spatial topology of the universe? We here want to argue that we are not allowed to draw any conclusions. It is of interest to note that if you assume every observer to see something which is exactly like a standard model, then there are very strong restrictions on the topology. However, if you only assume every observer to see something which is very close to a standard model, then there are no restrictions. In this sense, the conclusions depend discontinuously on the assumptions. The statement is still a bit vague, but let us take one more step in the direction towards a mathematical statement.

Assume that we are given

- a standard model, characterized by an existence interval I, a scale factor a etc.; this is how the universe should appear to an observer,
- a $t_0 \in I$, which represents the time to the future of which we wish the approximation to be valid; it can be thought of as representing decoupling,
- a positive integer l, specifying the norm with respect to which we measure proximity to the standard model; once we are given an l, the distance will be measured using the C^l-norm,
- an $\epsilon > 0$, characterising the size of the distance,
- a closed 3-manifold Σ; this will be the spatial topology of the constructed spacetime.

Given this information, there is then a solution (M, g, f) with the following properties:

- (M, g, f) is a maximal Cauchy development of initial data,
- (M, g) is future causally geodesically complete; in this sense there is an expanding direction,
- there is a Cauchy hypersurface, say $\bar{S}$, in (M, g), diffeomorphic to Σ; $\bar{S}$ should be thought of corresponding to the $t = t_0$ hypersurface in the background model, or decoupling. As a consequence, M is diffeomorphic to $R \times \Sigma$, so that the spatial topology is Σ.

The crucial question is now: how do you compare the constructed solution with the standard solution? The point is that you only want to compare regions seen by observers. Let us, for that reason, start with an observer in the spacetime, in other words with a causal curve, say γ, in (M, g). Note that the largest region that an observer can see is its causal past, denoted $J^-(\gamma)$. On the other hand, γ can, by assumption only observe things that occurred to the future of $\bar{S}$ (decoupling), and the future of the set $\bar{S}$ is denoted by $J^+(\bar{S})$. To summarize, the largest region that can be seen by γ is $J^-(\gamma) \cap J^+(\bar{S})$. At this stage we want to compare what is seen by γ with a standard solution. In order to be able to do so, we need a local diffeomorphism of regions containing the portion of the spacetime that observers can see. In fact, we have the following statement:

- given an observer γ in (M, g), there is a neighbourhood, say U, of

$$J^-(\gamma) \cap J^+(\bar{S})$$

and a diffeomorphism from U to an open neighbourhood, say V, of a solid cylinder of the form $[t_0, \infty) \times \bar{B}_R(0)$ in the specified standard model,

- pulling back the solution (M, g, f) to V using this diffeomorphism and computing the difference between the resulting solution and the pre-specified background solutions yields an error which is smaller than ϵ with respect to the pre-specified norm,
- all timelike geodesics in (M, g) are past incomplete; in this sense there is a big bang,
- the solution is stable with these properties; in other words, the construction is robust, it is not simply some special example.

To conclude the above discussion, there is no limitation on the global topology; any closed 3-manifold will do. It would perhaps be desirable to prove the same result in the case of any 3-manifold topology. It seems reasonable to expect that it should be possible to to so, but no complete proof has yet been written down.

3.8. References

The description given in the present section is a short summary of the results obtained in [20]. Interested readers is referred to this book for more details.

Acknowledgements

The author acknowledges the support of the Göran Gustafsson Foundation, the Göran Gustafsson Foundation for Research in Natural Sciences and Medicine, and the Swedish Research Council. The author would also like to thank the organizers of the '53 Cracow School of Theoretical Physics' for arranging a very stimulating summer school.

REFERENCES

[1] Anderson, M. T.: Existence and Stability of even-dimensional asymptotically de Sitter spaces. Ann. Henri Poincaré **6**, 801–820 (2005)

[2] Choquet-Bruhat, Y. Geroch R.: Global aspects of the Cauchy problem in General Relativity. Commun. Math. Phys., **14**, 329–335 (1969)

[3] Chruściel, P. T., Isenberg, J.: Non-isometric vacuum extensions of vacuum maximal globally hyperbolic spacetimes. Phys. Rev. D, **48**, 1616–1628 (1993)

[4] Einstein, A.: Zur Allgemeinen Relativitätstheorie. Preuss. Akad. Wiss. Berlin, Sitzber., 778–786 (1915)

[5] Einstein, A.: Der Feldgleichungen der Gravitation. Preuss. Akad. Wiss. Berlin, Sitzber., 844–847 (1915)

[6] Fourès-Bruhat, Y.: Théorème d'existence pour certains systèmes d'équations aux derivées partielles non linéaires. Acta Mathematica, **88**, 141-225 (1952)

[7] Friedrich, H.: On the existence of n-geodesically complete or future complete solutions of Einstein's field equations with smooth asymptotic structure. Commun. Math. Phys. **107**, 587–609 (1986)

[8] Friedrich, H.: Conformal Einstein evolution. In: The Conformal Structure of Space-Time, J. Frauendiener and H. Friedrich, Eds., Lecture Notes in Physics **604**, Springer Verlag, Berlin, 1–50 (2002)

[9] Friedrich, H.: On the global existence and the asymptotic behavior of solutions to the Einstein–Maxwell–Yang–Mills equations. J. Differential Geom. **34**, no. 2, 275–345 (1991)

[10] Hawking, S. W.: The Occurrence of Singularities in Cosmology. III. Causality and Singularities. Proc. Roy. Soc. Lond. **A300**, 182–201 (1967)

[11] Hawking, S. W., Penrose, R.: The Singularities of Gravitational Collapse and Cosmology. Proc. Roy. Soc. Lond. **A314**, 529–548 (1970)

[12] Hadzic, M.; Speck, J.: The Global Future Stability of the FLRW Solutions to the Dust-Einstein System with a Positive Cosmological Constant. arXiv:1309.3502

[13] Luo, X.; Isenberg, J.: Power Law Inflation with Electromagnetism. arXiv:1210.7566

[14] O'Neill, B.: Semi Riemannian Geometry. Academic Press, Orlando (1983)

[15] Penrose, R.: Gravitational Collapse and Space-Time Singularities. Phys. Rev. Lett. **14**, 57–59 (1965)

[16] Penrose, R.: Gravitational collapse – the role of general relativity, Riv. del Nuovo Cim. **1 (numero speziale)**, 252–276 (1969)

[17] Ringström, H.: Future stability of the Einstein non-linear scalar field system. Invent. math. **173**, 123–208 (2008)

[18] Ringström, H.: Power law inflation. Commun. Math. Phys. **290**, 155–218 (2009)

[19] Ringström, H.: The Cauchy problem in General Relativity, European Mathematical Society, Zürich (2009)

[20] Ringström, H.: On the Topology and Future Stability of the Universe, Oxford University Press, Oxford (2013)

[21] Rodnianski, I., Speck, J.: The Stability of the Irrotational Euler–Einstein System with a Positive Cosmological Constant. arXiv:0911.5501v1

[22] Speck, J.: The nonlinear future stability of the FLRW family of solutions to the Euler–Einstein system with a positive cosmological constant. Selecta Math. (N.S.) **18**, no. 3, 633–715 (2012)

[23] Svedberg, C.: Future Stability of the Einstein–Maxwell–Scalar Field System. Ann. Henri Poincaré **12**, No. 5, 849–917 (2011)