General Relativity

Alan D. Rendall

July 22, 2008

1 Introduction

General relativity (GR) is a physical theory which links gravity, space and time. The aim of these notes is to explain the mathematical structures which are used in GR and to present interesting mathematical results in this area. The point of view in these notes is that of the mathematician but the choice of topics is influenced by the physical applications. The techniques which are used come from differential geometry and the theory of partial differential equations (PDE).

Gravity, which leads to a mutual attraction of all material bodies, is one of the fundamental physical interactions. The first general theory which decribes gravity is that of Newton. It is still sufficient for many purposes. There is, however, since 1916 also another theory of gravity which under certain circumstances gives more accurate predictions. This is Einstein's theory of general relativity. This theory led to a new picture of space and time, which are unified to spacetime. Spacetime is a concept which already occurs in special relativity (SR), a theory which does not incorporate gravity. In GR gravity is related to the curvature of spacetime. Special relativity, which does not describe gravity, corresponds to a flat spacetime called Minkowski space. When treating a problem in GR it is often useful to compare with analogous problems in SR or in Newtonian gravity since these simpler models often provide useful intuition. Another useful comparison is with Maxwell's theory of electromagnetism.

Under which circumstances is it necessary to use GR to describe physical phenomena? When matter moves with a velocity which is comparable with the speed of light then special relativity is required. If gravity is also involved then GR is needed. Whenever strong gravitational fields occur then it is imperative to use GR to describe the system. These conditions, very high velocities of self-gravitating matter and very strong gravitational fields, occur primarily in astrophysics. Characteristic phenomena which occur are black holes and the initial singularity of the universe (Big Bang). GR also plays a role under milder conditions when measurements of exceptionally high precision are required. There is one practical application: the navigation system GPS (Global Positioning System) would be impossible without GR.

The fundamental equations of GR are the Einstein equations. These constitute a system of nonlinear PDE which describe the interaction between the geometry of spacetime and its matter content. Since the theory describes above all the interactions of masses it is necessary, in order to have a complete theory, to have a suitable description of matter. Curved spacetime can only be described using the tools of differential geometry. In fact the Einstein equations are relations between geometrical objects in the sense of differential geometry. In order to work with these objects in a concrete way it is often necessary to introduce coordinates and work with the components of the geometrical objects in these coordinates. Only in this way do we get a system of PDE in the usual sense out of the Einstein equations. A beautiful mathematical property of the Einstein equations is that they are independent of the choice of coordinates. It is however the case that when we have to choose coordinates in order to solve a specific problem this beautiful abstract property often leads to headaches. Since we can in principle choose any coordinate system we are in the uncomfortable situation of being forced to make a choice. We will encounter many examples of this in the following.

A good source of information on general relativity which pays attention to mathematical issues is [12]. Relevant background on differential geometry can be found for instance in [13].

In the first part of these notes various concepts which are needed in GR are introduced. This also allows notation and terminology to be fixed. Of key importance is the geometry of Lorentzian metrics, since these play the role of the unknown in the Einstein equations. The Einstein equations themselves will of course also be introduced. After that some simple solutions of the Einstein equations will be described. These are vacuum solutions, i.e. ones without matter. In order to go further we need to give an introduction to the description of matter in GR. Then simple solutions of the Einstein equations with matter can be introduced. These solutions play a central role in cosmology, where models of the Universe on the largest scales are formulated.

2 Basics

2.1 Lorentzian algebra

Let V be a real vector space of dimension $n+1$ with $n \geq 1$ and g a symmetric bilinear form on V. The quadratic form defined by g determines g uniquely. A basis e_i can be chosen so that $g(e_i, e_j) = \epsilon_{ij}$, where $\epsilon_{ij} = 0$ for $i \neq j$ and ϵ_{ii} is equal to 1, 0 or -1 for each value of i. The specification of the ϵ_{ii} is called the signature of g. When $\epsilon_{ij} \neq 0$ for $i = j$ the quadratic form is called non-degenerate. In this case $|\epsilon_{ii}| = 1$ for all i. The best known case in mathematics is that where all ϵ_{ii} are equal to $+1$, so that g is positive definite. For many applications, and in particular for those occurring in the following, it does not make a significant difference if g is replaced by $-g$. In this context there is no essential difference between positive and negative definite. What is important is the number of ϵ_{ii} with the less common sign. The case which is relevant for GR is that where this quantity is equal to one. The corresponding signature is called Lorentz signature. In this case the sequence of signs of the ϵ_{ii} is, after a permutation, either $(-,+,\ldots,+)$ or $(-,\ldots,-,+)$. Both possibilities are widespread in the literature on GR. Here we always make the choice of signature $(-,+,\ldots,+)$.

If a quadratic form g is given we can consider the linear transformations A of V which leave g invariant, i.e. $g(Ax,Ax)=g(x,x)$ for all vectors $x\in V$. These transformations form a group G. When g is positiv (or negative) definite G is the orthogonal group $O(n+1)$. When g has Lorentz signature G is the Lorentz group $O(n,1)$. The elements of this group are called Lorentz transformations.

When a quadratic form is positive definite $g(v,v)>0$ for any non-zero vector v. When g has Lorentz signature $g(v,v)$ can be positive, negative or zero. A vector with $g(v,v)>0$ is called spacelike, a vector with $g(v,v)<0$ timelike and a vector with $g(v,v)=0$ null. The vectors with $g(v,v)=0$ (i.e. the null vectors together with the zero vector) form a double cone, the light cone. The spacelike vectors are outside the light cone and the timelike vectors inside it. The set of timelike vectors has two connected components. Often one of these components is singled out and called the future component. Then we can distinguish between future-pointing and past-pointing timelike vectors. The set of null vectors also decomposes into future- and past-pointing vectors. For $n=1$ the set of spacelike vectors has two components while it is connected in higher dimensions so that there is an essential geometrical difference between timelike and spacelike vectors. Vectors which are timelike or null are called causal. The character of a vector as spacelike, null or timelike is called its causal character. A basis of V which casts g into canonical form is called an orthonormal basis. It consists of one timelike and n spacelike vectors. The sets of timelike, spacelike and null vectors are invariant under Lorentz transformations.

Let v and w be vectors in V with components (v^0,v^i) and (w^0,w^i) in an orthonormal basis. Then $g(v,w)=-v^0w^0+\delta_{ij}v^iw^j$. Here we use the Einstein summation convention that repeated indices are to be summed over. From this point on this convention will always be used unless an explicit exception is made. Suppose that v and w are causal and future-pointing. Then $v^0\geq(\delta_{ij}v^iv^j)^{1/2}\geq 0$ and $w^0\geq(\delta_{ij}w^iw^j)^{1/2}\geq 0$. It follows that

$$g(v,w)=-v^0w^0+\delta_{ij}v^iw^j\leq -v^0w^0+(\delta_{ij}v^iv^j)^{1/2}(\delta_{kl}w^kw^l)^{1/2}\leq 0 \tag{1}$$

The inner product of two future-pointing causal vectors is always non-positive. It is only zero when v and w are null and parallel.

Let W be a subspace of V and g a non-degenerate quadratic form. The orthogonal complement of W with respect to g consists of those vectors with $g(v,w)=0$ for all $w\in W$. Because g is non-degenerate the dimension of the orthogonal complement of W is equal to $\dim V-\dim W$. On the subspace W there is an induced quadratic form defined by $g_W(v_1,v_2)=g(v_1,v_2)$ for v_1 and v_2 in W. When the induced metric on W is nondegenerate the induced quadratic form on the orthogonal complement is also nondegenerate. The space V is the direct sum of W and its orthogonal complement. The signature of the induced quadratic form on the complement of W is determined by the signature of g_W.

An important special case is where W is spanned by a vector w. In that case the signature of the orthogonal complement of W is determined by the causal character of w.
Lemma Let g be a quadratic form on V with Lorentz signature, $w \in V$ a non-zero vector and W the orthogonal complement of w. If w is timelike then g_W is positive definite and W is called spacelike. If w is spacelike then g_W is of Lorentz signature and W is called timelike. If w is null then g_W is degenerate and W is called null. The signature of g_W is $(0, +, \ldots, +)$. In this case $w \in W$.
Proof The statements for w timelike or spacelike are very easy to prove. In that case the space is the direct sum of the space spanned by w and its orthogonal complement W. It then suffices to choose a basis of W which puts g_W into canonical form. There remains the case where w is null. Since $w \in W$ it follows that g_W is degenerate. There are no timelike vectors in W. We can assume w.l.o.g. that w is future-pointing. Let w' be a future-pointing timelike vector. Let Z be the set of all vectors which are orthogonal to v and v'. Then Z is a subspace of dimension $n-1$ and $Z \subset W$. All vectors in Z are orthogonal to w' and are therefore spacelike. The statement about the signature follows.

Because of the physical background the case $n = 3$ is particularly interesting. For this reason we will often, although not always, restrict consideration to this case. The character of one- and three-dimensional subspaces of a four-dimensional vector space with a quadratic form of Lorentzian signature has already been discussed. In the case $n = 3$ only one case remains, namely the case of two-dimensional subspaces. These are called spacelike, null or timelike when the induced metric is positive definite, degenerate or of Lorentz signature, respectively.

2.2 Lorentzian geometry

A knowledge of basic concepts of differential geometry such as manifold, coordinate system, vector field and diffeomorphism is assumed here. All vectors at all points of a smooth manifold M of dimension n form a manifold TM of dimension $2n$, the tangent bundle of M. In the following smooth means C^∞. The mapping $\pi : TM \to M$ which takes a tangent vector at a point p to p is called the projection and a vector field is a mapping V from M to TM such that $\pi \circ V$ is the identity. All vectors at a point p form a vector space T_pM, the tangent space at p and $T_pM = \pi^{-1}(\{p\})$. The dual space (in the sense of linear algebra) of T_pM is the cotangent space T_p^*M. An element of T_p^*M is called a one-form at p.

A pseudo-Riemannian metric g on M is a choice of a non-degenerate quadratic form g_p on T_pM for all $p \in M$ which depends smoothly on p in a suitable sense. One possibility to define smoothness is to require that for all smooth vector fields V and W the scalar functions $g_p(V_p, W_p)$ are smooth. When a manifold M is connected the signature of a pseudo-Riemannian metric is independent of the point $p \in M$. When this signature is $(+, +, \ldots, +)$ we speak of a Riemannian metric. When the signature is $(-, +, \ldots, +)$ we have a Lorentzian metric.

The pair (M, g) is called a Lorentzian manifold. This is the fundamental object in GR.

Let x^α be a coordinate system on the Lorentzian manifold (M, g). On the domain of definition of the coordinate system vector fields $\partial/\partial x^\alpha$ are defined. They form a basis of T_pM at each point p. The components V^α of a vector field V in the given coordinate system are defined by the relation $V = V^\alpha \partial/\partial x^\alpha$. The components of the Lorentz metric g are defined by $g_{\alpha\beta} = g(\partial/\partial x^\alpha, \partial/\partial x^\beta)$. The smoothness of the metric is equivalent to the condition that the components $g_{\alpha\beta}$ of g in each coordinate system are smooth. The coordinate system also defines one-forms dx^α. These are dual to the vector fields defined above in the sense that $dx^\alpha(\partial/\partial x^\beta) = \delta^\alpha_\beta$, where δ^α_β is the Kronecker δ.

It is important that certain indices (e.g. V^α) are upper indices and others (e.g. $g_{\alpha\beta}$) lower indices. The Einstein summation convention is usually only applied to repeated indices which occur once as upper and once as lower indices. For instance in the expression $g_{\alpha\beta}V^\alpha W^\beta$ summation over α and β is understood and the result is nothing other than $g(V, W)$. The matrix $g_{\alpha\beta}$ is invertible since g is non-degenerate. The inverse of this matrix is denoted $g^{\alpha\beta}$. It defines the so-called inverse metric. This is a smooth family of quadratic forms on the cotangent spaces T^*_pM. When these quadratic forms are evaluated on the basis vectors dx^α the result is the components $g^{\alpha\beta}$. The relation $g^{\alpha\beta}g_{\beta\gamma} = \delta^\alpha_\gamma$ holds. In GR the same notation (e.g. $g_{\alpha\beta}$) is used for a mathematical object and for the components of this object in a given coordinate system. This procedure may seem somewhat confusing at first glance but is easy to get used to. When the notation is meant in the first sense the term abstract indices is often used. It is important to note that with the interpretation of abstract indices $g_{\alpha\beta}$ and $g^{\alpha\beta}$ are two different mathematical objects, although the kernel g is the same.

Now we come back to the question of the position of indices. Let V^α be a vector field. Define $V_\alpha = g_{\alpha\beta}V^\beta$. V_α is the contravariant form of V^α. It follows that $V^\alpha = g^{\alpha\beta}V_\beta$. These constructions can be defined for any pseudo-Riemannian metric. All that is important is that g is non-degenerate. Similar conventions hold for tensors. A tensor T is a smooth family of objects T_p where T_p is a multilinear mapping $(T^*_pM)^k \times (T_pM)^l \to \mathbf{R}$. The components of the tensor T are denoted by $T^{\alpha\ldots\beta}{}_{\gamma\ldots\delta}$ and are obtained by evaluating T on k covectors $dx^\alpha, \ldots, dx^\beta$ and l vectors $\partial/\partial x^\gamma, \ldots, \partial/\partial x^\delta$. This notation can be interpreted in the sense of abstract indices. The metric can be used to raise and lower indices of arbitrary tensors. Consider for example a $(1, 2)$-tensor $T^\alpha{}_{\beta\gamma}$. The associated $(0, 3)$-tensor is defined by $T_{\sigma\beta\gamma} = g_{\sigma\alpha}T^\alpha{}_{\beta\gamma}$.

Since the definition of associated tensors given here apparently depends essentially on coordinates it might seem doubtful if these tensors are well-defined independently of coordinates. This is the case, although we do not prove it here. Another question which arises is why we want to use abstract indices at all. The reason is that in GR complicated and lengthy computations using tensors are needed and that the use of abstract indices minimises the effort required in such calculations. It is in a sense very convenient to work in an index-free manner. It is, however, a luxury which we cannot permit ourselves in complicated problems.

A curve γ in the manifold M is a mapping from an interval $I \subset \mathbf{R}$ to M. When the mapping is continuously differentiable there is a tangent vector $\dot{\gamma}(\lambda)$ defined at the point $\gamma(\lambda)$ for each $\lambda \in I$. When this vector is always timelike the curve is called timelike. Spacelike, null and causal curves can be defined similarly. A hypersurface S in M is called spacelike, null or timelike when the tangent space to S at each point has this character. In particular, a hypersurface is spacelike iff the induced metric is positive definite.

The basic equations of GR are the Einstein equations. They are equations where the unknown is a Lorentzian metric called the spacetime metric. Before we can write down these equations we need the important concept of curvature. This subject will be treated in the next section. At this point we can already define the simplest solution of the Einstein equations. This is the Minkowski metric $\eta_{\alpha\beta}$. This metric is defined on the manifold $\mathbf{R}^4$ and the matrix of its components in Cartesian coordinates is diagonal with entries $(-1, 1, 1, 1)$. In another notation the Minkowski metric is $-dt^2 + dx^2 + dy^2 + dz^2$. Of course a Minkowski metric can be defined in any dimension in an analogous way.

2.3 Connection and curvature

Let f be a real-valued function on M. The exterior derivative df of f is a one-form and has components $\partial_\alpha f$. On the other hand the partial derivatives $\partial_\alpha V^\beta$ of the components V^β of a vector field do not define a tensor. In the presence of a pseudo-Riemannian metric we can introduce a derivative which maps tensors into tensors. This is the covariant derivative. When a metric $g_{\alpha\beta}$ is given the Christoffel symbols are defined by

$$\Gamma^\alpha_{\beta\gamma} = \frac{1}{2} g^{\alpha\delta}(\partial_\beta g_{\gamma\delta} + \partial_\gamma g_{\beta\delta} - \partial_\delta g_{\beta\gamma}) \tag{2}$$

The covariant derivative of the vector field V^α is

$$\nabla_\alpha V^\beta = \partial_\alpha V^\beta + \Gamma^\beta_{\alpha\gamma} V^\gamma \tag{3}$$

and is a tensor. This formula can be extended to general tensors.

$$\nabla_\alpha T^{\beta\ldots\gamma}{}_{\delta\ldots\epsilon} = \partial_\alpha T^{\beta\ldots\gamma}{}_{\delta\ldots\epsilon} + \Gamma^\beta_{\alpha\sigma} T^{\sigma\ldots\gamma}{}_{\delta\ldots\epsilon} + \ldots - \Gamma^\sigma_{\alpha\delta} T^{\beta\ldots\gamma}{}_{\sigma\ldots\epsilon} - \ldots \tag{4}$$

The right hand side is a sum over all indices. In the case of a scalar function f we define $\nabla_\alpha f = \partial_\alpha f$. The covariant derivative of the metric is zero and this is one of the most important properties of covariant differentiation.

Covariant derivatives commute when applied to scalar functions, i.e. $\nabla_\alpha \nabla_\beta f = \nabla_\beta \nabla_\alpha f$. This is not the case for vectors. Instead

$$\nabla_\gamma \nabla_\delta V^\alpha - \nabla_\delta \nabla_\gamma V^\alpha = R^\alpha{}_{\beta\gamma\delta} V^\beta \tag{5}$$

Here $R^\alpha{}_{\beta\gamma\delta}$ is a tensor, the Riemann curvature tensor. The components of the Riemann tensor can be expressed in terms of the Christoffel symbols in the following way:

$$R^\alpha{}_{\beta\gamma\delta} = \partial_\gamma \Gamma^\alpha_{\beta\delta} - \partial_\delta \Gamma^\alpha_{\beta\gamma} + \Gamma^\alpha_{\gamma\sigma}\Gamma^\sigma_{\beta\delta} - \Gamma^\alpha_{\delta\sigma}\Gamma^\sigma_{\beta\gamma} \tag{6}$$

Taking a trace gives the Ricci tensor $R_{\alpha\beta} = R^{\gamma}{}_{\alpha\gamma\beta}$. Using the metric we can take another trace to get the scalar curvature $R = g^{\alpha\beta}R_{\alpha\beta}$. In the literature of GR and Riemannian geometry there are various conventions used when defining these curvature quantities. The order of the indices plays an essential role. Changing the order of the indices leads in some cases to changes in the signs occurring in the definitions. This follows from the symmetry properties of the curvature tensor. The Riemann tensor with all indices down is symmetric under interchange of the two index pairs

$$R_{\alpha\beta\gamma\delta} = R_{\gamma\delta\alpha\beta} \tag{7}$$

and antisymmetric under interchange of the indices within one of these index pairs

$$R_{\alpha\beta\gamma\delta} = -R_{\beta\alpha\gamma\delta} \tag{8}$$

In addition there is the algebraic, or first, Bianchi identity

$$R_{\alpha\beta\gamma\delta} + R_{\alpha\delta\beta\gamma} + R_{\alpha\gamma\delta\beta} = 0 \tag{9}$$

A simple consequence of these symmetry properties is that the Ricci tensor is symmetric, $R_{\alpha\beta} = R_{\beta\alpha}$. There is also an important symmetry property of the covariant derivative of the Riemann tensor known as the differential, or second, Bianchi identity

$$\nabla_{\gamma}R_{\alpha\beta\delta\epsilon} + \nabla_{\epsilon}R_{\alpha\beta\gamma\delta} + \nabla_{\delta}R_{\alpha\beta\epsilon\gamma} = 0 \tag{10}$$

Taking a trace gives the identity

$$\nabla^{\gamma}(R_{\beta\gamma} - \frac{1}{2}Rg_{\beta\gamma}) = 0 \tag{11}$$

If we define $G_{\alpha\beta} = R_{\alpha\beta} - \frac{1}{2}Rg_{\alpha\beta}$ then as a consequence of this identity $G_{\alpha\beta}$ is divergence-free. $G_{\alpha\beta}$ is the Einstein tensor and plays a key role in the formulation of the Einstein equations. Since $G_{\alpha\beta}$ is symmetric it has ten algebraically independent components.

2.4 Geodesics

Let $\gamma(\lambda)$ be a smooth curve in a spacetime. If t^{α} denotes the tangent vector to γ then we can consider the expression $t^{\alpha}\nabla_{\alpha}t^{\beta}$. In order for this expression to be well-defined it is in principle necessary to have a vector field t^{α} on a neighbourhood of γ. This can be got around as follows. Choose a smooth extension of the vector field on γ to a neighbourhood and notice that the restriction of $t^{\alpha}\nabla_{\alpha}t^{\beta}$ to γ is independent of the extension chosen. A different extension would be of the form $t^{\alpha} + fx^{\alpha}$, where f vanishes on γ. The desired result follows by a simple calculation.

The curve γ is called a geodesic when the tangent vector t^{α} satisfies $t^{\alpha}\nabla_{\alpha}t^{\beta} = ft^{\beta}$ for a function f. In this case we can introduce a new parameter on γ so that the new tangent vector satisfies $t^{\alpha}\nabla_{\alpha}t^{\beta} = 0$. This is called an affine parameter and is unique up to an affine transformation. Integrating the expression

$(g_{\alpha\beta}t^\alpha t^\beta)^{1/2}$ with respect to the parameter along a spacelike curve gives the arc length. In the case of a timelike curve it is possible to do something similar. There the quantity $(-g_{\alpha\beta}t^\alpha t^\beta)^{1/2}$ is integrated. The result is sometimes called arc length and sometimes proper time. The origin of the latter terminology is in physics and is as follows. An observer in spacetime has a world line, a timelike curve, which is given by his position at different times. If he carries an accurate clock the time it measures is the proper time along his world line. The proper time is a preferred affine parameter along timelike geodesics. In the case of a null curve the arc length is zero and so cannot be used as a parameter. The world line of a particle with zero rest mass (for example a photon) is a null curve. The world line of a particle of positive rest mass is a timelike curve. If we imagine that causal influences in spacetime can only take place via the action of particles then these influences can only propagate along timelike or null curves. This is the reason why these curves are called causal. The world line of a free particle, i.e. a particle which is not affected by any forces other than gravity, is a geodesic.

A geodesic is called complete when an affine parameter is defined for all real values. Since it is problematic physically when a particle stops existing after a finite time, complete geodesics are particularly important. A spacetime where all geodesics are complete is called geodesically complete. A spacetime which is not geodesically complete is called singular. In this case it is also said that a singularity is present, but without identifying a mathematical object which could be said to be a singularity.

2.5 The Einstein equations

In GR we consider a Lorentzian metric which satisfies the Einstein equations. The metric describes the geometry of space and time and also the gravitational field. The Einstein equations are

$$G_{\alpha\beta} = 8\pi(G/c^4)T_{\alpha\beta} \tag{12}$$

Here $G_{\alpha\beta}$ is the Einstein tensor of a Lorentzian metric $g_{\alpha\beta}$ and $T_{\alpha\beta}$ is the energy-momentum tensor. The tensor $T_{\alpha\beta}$ describes the energy, momentum and stress of the matter which is present in spacetime. It depends on certain matter fields and also on the metric. Details of the description of matter will be treated late. In equation (12) the symbol G without indices denotes the Newtonian gravitational constant and c is the speed of light. The numerical values of G and c depend on the choice of units which are used to describe the physical system. In the following we use units such that the numerical values of G and c are equal to unity. This is a frequent choice in the literature on GR. In this case (12) simplifies to

$$G_{\alpha\beta} = 8\pi T_{\alpha\beta} \tag{13}$$

The number 8π is natural in four dimensions. In other dimensions this is not so clear.

The equation (13) only has some content after the form of $T_{\alpha\beta}$ has been specified. Some examples will show what is meant by this. First consider the case of a region of spacetime where there is no matter present. In that case the energy-momentum tensor vanishes and the Einstein equations become $G_{\alpha\beta} = 0$. This condition is equivalent (in any dimension $n > 1$) to the condition $R_{\alpha\beta} = 0$. A second example is the scalar field. In fact, in order to distinguish this matter model from others we should say a linear, massless, minimally coupled scalar field but we will not take this trouble for the moment. This matter field has no simple concrete physical meaning. However it is often used for theoretical purposes by both mathematicians and physicists. It is one of the simplest of all matter models. The scalar field is a real-valued function ϕ and the energy-momentum tensor is in this case

$$T_{\alpha\beta} = \nabla_\alpha \phi \nabla_\beta \phi - \frac{1}{2}(\nabla_\gamma \phi \nabla^\gamma \phi) g_{\alpha\beta} \tag{14}$$

When we consider a matter field which interacts with the gravitational field we must consider not only the Einstein equations but also the equations of motion of the matter. The equation of motion of the scalar field is $\nabla_\alpha \nabla^\alpha \phi = 0$ where the covariant derivative is defined by the metric $g_{\alpha\beta}$. The equation of motion of the scalar field follows from the Einstein equations as a consequence of the Bianchi identity. It should be emphasized that this is not typical. In general the divergence-free property of the Einstein tensor implies that of the energy-momentum tensor. The equations of motion of the matter also imply the divergence-free property of the energy-momentum tensor, but not conversely.

Consider the Einstein vacuum equations. In local coordinates these evidently consist of a system of ten second order partial differential equations. What kind of equations are they? Unfortunately they do not belong to one of the types of equations to be found in elementary textbooks on PDE. They are closely related to hyperbolic PDE but unfortunately the relation is complicated. To find out more about the nature of these equations it is natural to look at the principal part. A short calculation gives

$$g^{\alpha\beta}(\partial_\alpha \partial_\beta g_{\gamma\delta} + \partial_\gamma \partial_\delta g_{\alpha\beta} - \partial_\alpha \partial_\gamma g_{\beta\delta} - \partial_\beta \partial_\delta g_{\alpha\gamma}) + \ldots = 0 \tag{15}$$

The lower order terms which are omitted in this formula are all bilinear in the components $\partial_\alpha g_{\beta\gamma}$ and in the components $g^{\alpha\beta}$. The first term in the equation (15) looks like the principal part of the wave equation in the given geometry $g_{\alpha\beta}$. Here we see the hyperbolic influence. It is, however, unclear what we should do with the other terms.

2.6 Group actions

A Lie group G is a group G which is at the same time a manifold, where multiplication is a smooth mapping $G \times G \to G$ and the operation of taking the inverse defines a smooth mapping $G \to G$. An action of the group G on the manifold M is a smooth mapping $\phi : G \times M \to M$ with the property that $\phi(g, \phi(h, x)) = \phi(gh, x)$ for all g and h in G and all $x \in M$. A simple example is where the group is the group $SO(3)$ of rotations in $\mathbf{R}^3$, $M = \mathbf{R}^3$ and $\phi(A, x) = Ax$.

2.7 Covering spaces

Let M and N be connected manifolds of the same dimension. A smooth mapping ϕ from M onto N is called a covering map when each point $x \in N$ has an open neighbourhood U with the property that $\phi^{-1}(U)$ is the disjoint union of subsets U_α so that the restriction of ϕ to U_α is a diffeomorphism of U_α onto U. Any manifold has a universal cover, which itself has no non-trivial covering. For example, the mapping $\theta \mapsto e^{i\theta}$ is a covering mapping from $\mathbf{R}$ to the unit circle in the complex plane. In fact it defines the universal covering.

3 Explicit vacuum solutions

3.1 Minkowski space

Minkowski space has the property that its Riemann curvature tensor vanishes and so it is flat. It follows, in particular, that this metric (this spacetime) satisfies the vacuum Einstein equations. If $\gamma(s)$ is a future-directed timelike curve parametrized by proper time then $dt/ds \geq 1$. For an affinely parametrized null curve a constant lower bound for the derivative of t with respect to the parameter along the curve can also be obtained. It follows that any inextendible causal curve meets the spacelike hypersurface $t = 0$ exactly once. A spacelike hypersurface in a spacetime which has this property is called a Cauchy surface. A spacetime which contains a Cauchy surface is called globally hyperbolic. The property of global hyperbolicity has to do with the fact that a spacetime of this kind can be determined by initial data on a spacelike hypersurface. This relates to the initial value problem for the Einstein equations, a somewhat complicated subject. That it is natural to consider such an initial value problem is suggested by the observation that the Einstein equations have similarities to hyperbolic equations. In any case, Minkowski space is globally hyperbolic.

Consider now the hyperboloid H with equation $t^2 = 1 + \delta_{ij} x^i x^j$, $t > 0$. It is easy to see that H is not a Cauchy surface for Minkowski space. The region of Minkowski space with $t > \sqrt{\delta_{ij} x^i x^j}$ is a spacetime in its own right. This new spacetime is globally hyperbolic and has H as a Cauchy surface. The hypersurfaces with equation $t^2 = c^2 + \delta_{ij} x^i x^j$ are spacelike and form a foliation of the spacetime. The spacetime with this foliation is a kind of cosmological model and is sometimes called the Milne model. The foliation can be interpreted in such a way that the tangent spaces to the leaves are the rest spaces of galaxies. We get another spacetime if we cut the origin out of Minkowski space. The spacetime defined in this way is not globally hyperbolic. By trivial procedures like this many pathological spacetimes can be produced. They are mentioned here in order to emphasize that to obtain physically reasonable spacetimes certain pitfalls must be excluded. The initial value problem is a good way of doing this.

Each element of the Lorentz group defines a smooth transformation of Minkowski space. It is an isometry. A diffeomorphism ϕ of a Lorentzian manifold can be used to transport tensors. If V is a vector at the point p then ϕ defines a vector $\phi_* V$ at the point $q = \phi(p)$. In local coordinates we have

$$(\phi_* V)^\alpha = \frac{\partial \phi^\alpha}{\partial x^\beta} V^\beta \tag{16}$$

The metric g_q can be transported to p.

$$(\phi^* g)_p(V, W) = g_q(\phi_* V, \phi_* W) \tag{17}$$

If $\phi^* g = g$ then ϕ is called an isometry of g. The translations of Minkowski space are also isometries and generate together with the Lorentz transformations the Poincaré group.

Now consider elements of the Lorentz group which map future-pointing vectors to future-pointing vectors. They form a subgroup of the Lorentz group, the isochronous Lorentz group. The isochronous Lorentz group maps the subset of Minkowski space which defines the Milne model to itself. Hence all elements of the isochronous Lorentz group can be considered as isometries of the Milne model. The preferred foliation is also left invariant.

The geodesics of Minkowski space are the straight lines and the affine parameters are affine functions of the Cartesian coordinates. Hence Minkowski space is geodesically complete. The Milne model is not complete. It is, however, in an obvious sense complete in the future.

Each of the preferred spacelike hypersurfaces in the Milne model has an induced Riemannian metric. This metric is well known in Riemannian geometry and is the hyperbolic space. The curvature tensor of hyperbolic space is very simple and satisfies

$$R_{abcd} = k(g_{ac} g_{bd} - g_{ad} g_{bc}) \tag{18}$$

for a constant k. The isochronous Lorentz group defines a group of isometries of the Milne model. It is known that there are discrete subgroups with the property that the corresponding quotient of the hyperbolic space is a compact manifold. In this way it is possible to make a spatially compact cosmological model, i.e. a spacetime with a compact Cauchy surface, out of the Milne model

3.2 De Sitter space

There is a generalization of the Einstein equations (13), whose popularity has varied during the history of GR. The equations are

$$G_{\alpha\beta} + \Lambda g_{\alpha\beta} = 8\pi T_{\alpha\beta} \tag{19}$$

where Λ is a constant, the cosmological constant. Present astronomical observations strongly indicate that in our universe $\Lambda > 0$. In this case the role of Minkowski space is taken over by de Sitter space.

De Sitter space can be represented as a hypersurface in five-dimensional Minkowski space. This higher-dimensional flat space is in this context a mathematical tool and has no physical meaning. Consider $\mathbf{R}^5$ with Cartesian coordinates (v, w, x, y, z) and the Lorentzian metric

$$-dv^2 + dw^2 + dx^2 + dy^2 + dz^2 \tag{20}$$

The hyperboloid with the equation $-v^2 + w^2 + x^2 + y^2 + z^2 = \alpha^2$ and $\alpha > 0$ is a timelike hypersurface with topology $S^3 \times \mathbf{R}$. Its induced metric has Lorentzian signature and defines de Sitter space. It is possible to define coordinates on de Sitter space by the following relations:

$$\alpha \sinh(\alpha^{-1}t) = v, \quad \alpha \cosh(\alpha^{-1}t) \cos\chi = w, \tag{21}$$
$$\alpha \cosh(\alpha^{-1}t) \sin\chi \cos\theta = x, \quad \alpha \cosh(\alpha^{-1}t) \sin\chi \sin\theta \cos\phi = y, \tag{22}$$
$$\alpha \cosh(\alpha^{-1}t) \sin\chi \sin\theta \sin\phi = z. \tag{23}$$

In these coordinates the metric has the form

$$-dt^2 + \alpha^2 \cosh^2(\alpha^{-1}t)[d\chi^2 + \sin^2\chi(d\theta^2 + \sin^2\theta d\phi^2)] \tag{24}$$

Here $t \in \mathbf{R}$, $\chi \in [-\pi, \pi]$, $\theta \in [-\pi, \pi]$ and $\phi \in [0, 2\pi]$. These are not globally regular coordinates but their singularities are harmless like the origin of polar coordinates in the plane. The metric appears singular. In the example of the plane the metric $dr^2 + r^2 d\theta^2$ looks singular because the determinant of the matrix of metric components vanishes at the origin. In this situation we speak of a coordinate singularity of the metric. We know that the singularity at $\chi = 0$ in the metric (24) is a coordinate singularity because we started from a regular geometry. When a metric is presented in coordinates it is not always easy to decide whether an apparent singularity is in fact a coordinate singularity and whether the metric could not perhaps be smoothly extended beyond it. This metric satisfies the Einstein equations with $\alpha = \sqrt{3/\Lambda}$.

Another set of coordinates can be introduced in the following way.

$$\hat{t} = \log\frac{w+v}{\alpha}, \hat{x} = \frac{\alpha x}{w+v}, \hat{y} = \frac{\alpha y}{w+v}, \hat{z} = \frac{\alpha z}{w+v} \tag{25}$$

In these coordinates the metric has the form

$$-d\hat{t}^2 + \exp(2\alpha^{-1}\hat{t})(d\hat{x}^2 + d\hat{y}^2 + d\hat{z}^2) \tag{26}$$

These coordinates only cover the half of the hyperboloid with $v + w < 0$. The problem of coordinate singularities can also arise in the context of infinity in a coordinate system. For example the de Sitter solution can be extended beyond $\hat{t} = -\infty$.

The only non-vanishing Christoffel symbols in this case are

$$\Gamma^0_{ab} = \alpha^{-1}\exp(2\alpha^{-1}\hat{t})\delta_{ab} \tag{27}$$
$$\Gamma^a_{0b} = \Gamma^a_{b0} = \alpha^{-1}\delta^a_b \tag{28}$$

The non-vanishing components of the curvature tensor are

$$R_{abcd} = \frac{\Lambda}{3}(g_{ac}g_{bd} - g_{ad}g_{bc}) \tag{29}$$

$$R_{0abc} = 0 \tag{30}$$

$$R_{0a0b} = -\frac{\Lambda}{3}g_{ab} \tag{31}$$

These equations can be combined to give $R_{\alpha\beta\gamma\delta} = \frac{\Lambda}{3}(g_{\alpha\gamma}g_{\beta\delta} - g_{\alpha\delta}g_{\beta\gamma})$. It follows that $R_{\alpha\beta} = \Lambda g_{\alpha\beta}$. The curvature invariant $R_{\alpha\beta\gamma\delta}R^{\alpha\beta\gamma\delta}$ is called the Kretschmann scalar. One way of showing that a spacetime cannot be extended is to show that the Kretschmann scalar blows up. In the case of the de Sitter spacetime, which is extendible through $\hat{t} = -\infty$, the Kretschmann scalar is $\frac{24\Lambda^2}{9}$ and does not blow up.

Consider a future-directed causal geodesic with affine parameter τ. The geodesic equations are

$$\frac{d^2\hat{t}}{d\tau^2} + \alpha^{-1}e^{2\alpha^{-1}\hat{t}}\delta_{ab}\frac{d\hat{x}^a}{d\tau}\frac{d\hat{x}^b}{d\tau} = 0 \tag{32}$$

$$\frac{d^2\hat{x}^a}{d\tau^2} + 2\alpha^{-1}\frac{d\hat{x}^a}{d\tau}\frac{d\hat{t}}{d\tau} = 0 \tag{33}$$

Integrating the equation for $\hat{x}^a$ shows that $\frac{d\hat{x}^a}{d\tau} = A^a e^{-2\alpha^{-1}\hat{t}}$ for constants A^a. Substituting this into the equation for $\hat{t}$ gives

$$\frac{d^2\hat{t}}{d\tau^2} = -\alpha^{-1}e^{-2\alpha^{-1}\hat{t}}\delta_{ab}A^aA^b \tag{34}$$

In particular this shows that $d\hat{t}/d\tau$ is non-increasing and that $d\tau/d\hat{t}$ is non-decreasing. This is enough to show future geodesic completeness. Let $\epsilon = 0$ for a null geodesic and $\epsilon = 1$ for a timelike geodesic. Then

$$-\epsilon = -\left(\frac{d\hat{t}}{d\tau}\right)^2 + g_{ab}\frac{d\hat{x}^a}{d\tau}\frac{d\hat{x}^b}{d\tau} \tag{35}$$

The last term in this expression is a constant multiple of $\exp(-2\alpha^{-1}\hat{t})$. In the timelike case $d\tau/d\hat{t} = (1 + Ce^{-2\alpha^{-1}\hat{t}})^{-1/2}$ and it follows that asymptotically for $\hat{t}$ large τ is approximately equal to $\hat{t}$. In the null case $d\tau/d\hat{t} = Ce^{\alpha^{-1}\hat{t}}$ for a positive constant C and so the affine parameter has an exponential dependence on $\hat{t}$.

3.3 Anti-de Sitter space

There is an analogue of de Sitter space for the Einstein vacuum equations with $\Lambda < 0$. This anti-de Sitter space will not play much of a role in the following but for the sake of completeness we will describe it briefly. To describe this space we consider a hypersurface in $\mathbf{R}^5$ with a pseudo-Riemannian metric of signature

$(-,-,+,+,+)$. The metric can be written in the form $-du^2-dv^2+dx^2+dy^2+dz^2$ and the hypersurface is described by the equation $-u^2-v^2+x^2+y^2+z^2=1$. The induced metric has Lorentz signature. The hypersurface has the topology $S^1\times\mathbf{R}^3$. This spacetime has closed timelike curves, which means that causality is violated. For this reason it is common to consider the universal cover which does not suffer from this problem. The covering space is the anti-de Sitter space. In suitable coordinates the metric has the form

$$-\cosh^2 r dt^2+dr^2+\sinh^2 r(d\theta^2+\sin^2\theta d\phi^2) \tag{36}$$

This spacetime is not globally hyperbolic.

3.4 The Kasner solution

Now we return to the vacuum Einstein equations with $\Lambda=0$. The perhaps simplest solution except for Minkowski space is the Kasner solution. The metric is

$$-dt^2+t^{2p_1}dx^2+t^{2p_2}dy^2+t^{2p_3}dz^2 \tag{37}$$

Here $t\in(0,\infty)$ and x, y and z are Cartesian coordinates on $\mathbf{R}^3$. The constants p_1, p_2 and p_3 satisfy the relations

$$p_1+p_2+p_3=1,\quad p_1^2+p_2^2+p_3^2=1 \tag{38}$$

These are called the first and second Kasner relations. An alternative interpretation is obtained when the coordinates x, y and z are identified periodically. Then the spatial slices t=const. have topology T^3. All these solutions have a three-dimensional group of isometries which is defined by translations in x, y and z. When the spatial topology is $\mathbf{R}^3$ the isometry group is also $\mathbf{R}^3$. When the spatial topology is T^3 the isometry group is also T^3 or, in another notation, $U(1)\times U(1)\times U(1)$.

The solution set of the Kasner relations is a circle in $\mathbf{R}^3$. The solutions where p_1 attains its extremal value are $(-1/3,2/3,2/3)$ and $(1,0,0)$. When the spatial manifold is taken to be $\mathbf{R}^3$ these solutions have in addition to the symmetries defined by translations a further symmetry, which is defined by rotations in the (y,z) plane. In the case of the topology T^3 we can call this a local symmetry. In that case there only exists a global symmetry on the universal covering space. The solutions with the additional symmetry are called LRS (locally rotationally symmetric).

Consider the Kasner solution with topology T^3 as a cosmological model. The volume of the spatial slices is proportional to t and this volume grows monotonically with time from zero to infinity. This is typical for an always expanding cosmological model. The only non-vanishing Christoffel symbols are

$$\Gamma^0_{11}=p_1t^{2p_1-1},\quad \Gamma^0_{22}=p_2t^{2p_2-1},\quad \Gamma^0_{33}=p_3t^{2p_3-1} \tag{39}$$

$$\Gamma^1_{01}=p_1t^{-1},\quad \Gamma^2_{02}=p_2t^{-1},\quad \Gamma^3_{03}=p_3t^{-1} \tag{40}$$

and the components obtained from these by using the symmetry of the Christoffel symbols in the lower indices. The non-vanishing components of the curvature tensor are

$$R_{1212} = p_1 p_2 t^{2(p_1+p_2-2)}, \quad R_{1313} = p_1 p_3 t^{2(p_1+p_3-2)}, \tag{41}$$
$$R_{2323} = p_2 p_3 t^{2(p_2+p_3-2)} \tag{42}$$
$$R_{0abc} = 0 \tag{43}$$
$$R_{0101} = p_1(1-p_1)t^{2p_1-2}, \quad R_{0202} = p_2(1-p_2)t^{2p_2-2}, \tag{44}$$
$$R_{0303} = p_3(1-p_3)t^{2p_3-2} \tag{45}$$

and the components obtained from these by using the antisymmetry in each index pair. The Kretschmann scalar is

$$R_{abcd}R^{abcd} + 4R_{0a0b}R^{0a0b} = (p_1^2 p_2^2 + p_2^2 p_3^2 + p_3^2 p_1^2)t^{-4} \tag{46}$$

It follows that unless two of the p_i vanish the Kretschmann scalar tends to infinity uniformly as $t \to 0$. We say that there is a curvature singularity at $t = 0$. The spacetime cannot be extended through $t = 0$. What happens in the exceptional case that two of the p_i vanish? We can assume w.l.o.g. that $p_1 = 1$ and $p_2 = p_3 = 0$. The metric is

$$-dt^2 + t^2 dx^2 + dy^2 + dz^2 \tag{47}$$

The curvature tensor vanishes identically and so this spacetime is flat. For this reason it is called the flat Kasner solution. In fact it can be identified with a piece of flat space. With a periodic identification in x it is also known as Misner space. To see the relation to flat space let $\hat{t} = t\cosh x$ and $\hat{x} = t\sinh x$. Then $-dt^2 + t^2 dx^2$ is equal to $-d\hat{t}^2 + d\hat{x}^2$.

Consider now a future-directed causal geodesic in a general Kasner solution. The geodesic equations are

$$\frac{d^2 t}{d\tau^2} + \sum_i p_i t^{2p_i-1}\left(\frac{dx^i}{d\tau}\right)^2 = 0 \tag{48}$$
$$\frac{d^2 x^i}{d\tau^2} + 2p_i t^{-1}\frac{dx^i}{d\tau}\frac{dt}{d\tau} = 0 \tag{49}$$

Integrating the equation for x^i shows that (no sum) $\frac{dx^i}{d\tau} = A^i t^{-2p_i t}$ for constants A^i. Defining ϵ as in Section 3.2 gives

$$-\epsilon = -\left(\frac{dt}{d\tau}\right)^2 + \sum_i (A^i)^2 t^{-2p_i} \tag{50}$$

In order to ensure geodesic completeness it is enough to ensure that t^{p_i} is not integrable for any i. Since $p_i \geq -1/3$ this condition is satisfied.

To close this section we define the generalized Kasner exponents (GKE). Let $(M, g_{\alpha\beta})$ be a spacetime with a foliation S_t by spacelike hypersurfaces. For each spacelike hypersurface S_t we can define the second fundamental form k_{ab} as follows. It is a symmetric tensor on S_t and satisfies

$$k_{ab}x^a y^b = X^\alpha \nabla_\beta Y^\beta n^\beta \qquad (51)$$

Here x^a and y^b are any vectors tangent to S_t, X^α and Y^β are smooth vector fields defined on a neighbourhood of S_t which agree with x^a and y^b respectively on S_t, and n^α is the unit normal vector to S_t. The result does not depend on the particular extensions chosen. The trace $\mathrm{tr}k = g^{ab}k_{ab}$ is called the mean curvature of S_t. Now consider the special case of Gauss coordinates. This means that the hypersurfaces S_t are the level surfaces of t, $g_{00} = -1$ and the spatial coodinates are chosen so that $g_{0i} = 0$. In this case it can be computed that $\partial_t g_{ab} = -2k_{ab}$. This allows us to easily obtain k_{ab} in the case of the Kasner solution. Next consider the eigenvalues of the second fundamental form with respect to the induced metric, i.e. the solutions of $\det(k_{ab} - \lambda g_{ab}) = 0$. There are three solutions λ_1, λ_2, λ_3. Let $p_i = \lambda_i / \sum_j \lambda_j$, assuming that the mean curvature is non-zero. The p_i are by definition the generalized Kasner exponents and in the case of the Kasner solutions they agree with the Kasner exponents, hence the name. The GKE always satisfy the first Kasner relation but do not in general satisfy the second Kasner relation. They are useful invariants for describing the dynamics of solutions of the Einstein equations.

4 Matter models

4.1 Energy conditions

Before we come to specific matter models some general properties common to many physically reasonable types of matter will be listed. These are inequalities which should be satisfied by the energy-momentum tensor. The weak energy condition says that $T_{\alpha\beta}V^\alpha V^\beta \geq 0$ for all causal vectors V^α. Physically this condition says that each observer measures a positive energy density of matter in his rest frame. The dominant energy condition says that for each future-pointing causal vector V^α the vector $T_{\alpha\beta}V^\beta$, if non-zero, is causal and past-pointing. An equivalent condition is that for all future-pointing vector fields V^α and W^α the inequality $T_{\alpha\beta}V^\alpha W^\beta \geq 0$ holds. The strong energy condition says that for all future-pointing causal vectors V^α the inequality

$$[T_{\alpha\beta} - \frac{1}{2}(g^{\gamma\delta}T_{\gamma\delta})g_{\alpha\beta}]V^\alpha V^\beta \geq 0 \qquad (52)$$

holds. When the Einstein equations with $\Lambda = 0$ hold this condition is equivalent to the geometrical condition $R_{\alpha\beta}V^\alpha V^\beta \geq 0$ and means in a certain sense that the gravitational force is attractive. The dominant energy condition implies the weak energy condition. Unfortunately the strong energy condition does not imply the weak energy condition. Another condition satisfied by many matter models is the non-negative sum pressures condition. This says that if V^α is a unit timelike vector then $T^{\alpha\beta}(g_{\alpha\beta} + V_\alpha V_\beta) \geq 0$.

4.2 The scalar field

In section 2.5 we already introduced one example of a matter field in addition to the vacuum Einstein equations. This was the scalar field. In a broader context this should be referred to as a linear massless scalar field. A more general possibility is a nonlinear scalar field ϕ with a potential $V(\phi)$. Here V is a non-negative function. In this case the energy-momentum tensor is of the form

$$T_{\alpha\beta} = \nabla_\alpha\phi\nabla_\beta\phi - \left[\frac{1}{2}(\nabla_\gamma\phi\nabla^\gamma\phi) + V(\phi)\right] g_{\alpha\beta} \tag{53}$$

The equation of motion of the scalar field is

$$\nabla^\alpha\nabla_\alpha\phi - V'(\phi) = 0 \tag{54}$$

In the special case where $V(\phi) = (1/2)m^2\phi^2$ for a positive constant m we get a massive scalar field of mass m.

To verify that the dominant energy condition is satisfied it is enough to examine $T_{\alpha\beta}X^\alpha Y^\beta$ for two arbitrary future-pointing unit timelike vectors X^α and Y^β. By a suitable choice of orthonormal basis we can assume that X^α has components $(1, 0, 0, 0)$. In addition $Y^0 = \sqrt{1 + \delta_{ij}Y^iY^j}$. Now

$$T_{\alpha\beta}X^\alpha Y^\beta = \frac{1}{2}(\partial_0\phi)^2 Y^0 + \partial_0\phi\partial_i\phi Y^i + \frac{1}{2}\delta^{ij}\partial_i\phi\partial_j\phi Y^0 + V(\phi)Y^0 \tag{55}$$

The last term is evidently non-negative. By the Cauchy-Schwarz inequality $|\partial_i\phi Y^i| \leq (\delta^{ij}\partial_i\phi\partial_j\phi)^{1/2}(\delta^{kl}Y_kY_l)^{1/2}$ and the last factor on the right hand side of this inequality is bounded by Y^0. Hence the sum of the first three terms in the expression for $T_{\alpha\beta}X^\alpha Y^\beta$ is non-negative and the dominant energy condition holds. Consider next the strong energy condition.

$$T_{\alpha\beta} - \frac{1}{2}Tg_{\alpha\beta} = \nabla_\alpha\phi\nabla_\beta\phi + V(\phi)g_{\alpha\beta} \tag{56}$$

If $V = 0$ then clearly the strong energy condition is satisfied. For a non-zero potential, on the other hand, it can be violated. It suffices to consider the case where the gradient of ϕ is spacelike and to contract with a timelike vector orthogonal to $\nabla_\alpha\phi$. This is related to inflationary expansion, as will be discussed in more detail later. A positive cosmological constant also leads to a violation of the condition $R_{\alpha\beta}V^\alpha V^\beta \geq 0$ for timelike vectors V^α and also has an inflationary effect. The non-negative sum pressures condition can fail even in the case of a vanishing potential. To see this, suppose that the gradient of ϕ is spacelike and take a timelike vector V^α orthogonal to $\nabla_\alpha\phi$. Then, with $V(\phi) = 0$,

$$T_{\alpha\beta}(g^{\alpha\beta} + V^\alpha V^\beta) = -\frac{1}{2}(\nabla_\gamma\phi\nabla^\gamma\phi) \tag{57}$$

4.3 The Maxwell field

In this case the fundamental object is the electromagnetic field tensor $F_{\alpha\beta}$ which is antisymmetric. The energy-momentum tensor is

$$T_{\alpha\beta} = F_{\alpha\gamma}F_\beta{}^\gamma - (1/4)(F^{\gamma\delta}F_{\gamma\delta})g_{\alpha\beta} \tag{58}$$

This tensor is trace-free, $g_{\alpha\beta}T^{\alpha\beta}=0$. The equations of motion are the Maxwell equations and read

$$\partial_\alpha F_{\beta\gamma}+\partial_\gamma F_{\alpha\beta}+\partial_\beta F_{\gamma\alpha} = 0 \tag{59}$$
$$\nabla_\alpha F^{\alpha\beta} = 4\pi j^\beta \tag{60}$$

In the case that the four-current j^α vanishes these equations are the source-free Maxwell equations. If we want to describe charged matter we must introduce additional matter fields in order to describe this. The four-current will then depend on these other fields. They also contribute to the energy-momentum tensor.

The relation of $F_{\alpha\beta}$ to the classical description of electromagnetism by the electric field E^a and the magnetic field B^a is (in Minkowski space) as follows:

$$F_{01}=-E^1,\quad F_{02}=-E^2,\quad F_{03}=-E^3 \tag{61}$$
$$F_{12}=-B^3,\quad F_{13}=B^2,\quad F_{23}=-B^1 \tag{62}$$

An algebraic analysis of the tensor $F^\alpha{}_\beta$, which will not be reproduced here, shows that $F_{\alpha\beta}$ can be brought into one of the two forms:

$$F_{\alpha\beta} = Al_{[\alpha}n_{\beta]}+Bx_{[\alpha}y_{\beta]} \tag{63}$$
$$F_{\alpha\beta} = Cl_{[\alpha}x_{\beta]} \tag{64}$$

Here the square brackets denote the antisymmetric form of a tensor, e.g. $T_{[\alpha\beta]}=(1/2)(T_{\alpha\beta}-T_{\beta\alpha})$. Note for future reference that round brackets denote the symmetric form of a tensor, e.g. $T_{(\alpha\beta)}=(1/2)(T_{\alpha\beta}+T_{\beta\alpha})$. The vectors l^α and n^α are null and not parallel while x^α and y^α are spacelike. The null vectors are orthogonal to the spacelike vectors. The quantities A, B and $C$are constants. By choosing these constants appropriately we can assume that $l_\alpha n^\alpha=-1$ and that x^α and y^α are unit. Then we get in the two cases

$$T_{\alpha\beta} = \tfrac{1}{4}[A^2(l_\alpha n_\beta+n_\alpha l_\beta)+B^2(x_\alpha x_\beta+y_\alpha y_\beta)]+\tfrac{1}{8}(A^2-B^2)g_{\alpha\beta} \tag{65}$$
$$T_{\alpha\beta} = \tfrac{1}{4}C^2 l_\alpha l_\beta \tag{66}$$

In the second case the dominant and strong energy conditions are evidently satisfied. The non-negative sum pressures condition also holds. In the first case the dominant energy condition is a little more complicated to verify. The metric can be written in the form

$$g_{\alpha\beta}=-l_\alpha n_\beta-n_\alpha l_\beta+x_\alpha x_\beta+y_\alpha y_\beta \tag{67}$$

As a consequence

$$T_{\alpha\beta}=\frac{1}{2}(A^2+B^2)(l_\alpha n_\beta+n_\alpha l_\beta+x_\alpha x_\beta+y_\alpha y_\beta) \tag{68}$$

Thus it suffices to verify the dominant energy condition for the tensor $l_\alpha n_\beta+n_\alpha l_\beta+x_\alpha x_\beta+y_\alpha y_\beta$. Suppose that $V^\alpha=Ml^\alpha+Nn^\alpha+Px^\alpha+Qy^\alpha$ for some constants M, N, P and Q. Then

$$W_\alpha=(l_\alpha n_\beta+n_\alpha l_\beta+x_\alpha x_\beta+y_\alpha y_\beta)V^\beta=-Ml_\alpha-Nn_\alpha+Px_\alpha+Qy_\alpha \tag{69}$$

and

$$V^\alpha V_\alpha = -2LM + P^2 + Q^2 = W^\alpha W_\alpha \tag{70}$$

Hence $T_{\alpha\beta}$ maps causal vectors to causal vectors. That they have the appropriate time orientation follows from the relation $T_{\alpha\beta}V^\alpha l^\beta = -V_\alpha l^\alpha$. It follows that the dominant ebergy condition is satisfied in this case. Due to the fact that the energy-momentum tensor is traceless the strong energy condition and non-negative sum pressures conditions also hold.

4.4 The perfect fluid

The perfect fluid is described by the energy density ρ, the pressure p and the four-velocity U^α. The relation $g_{\alpha\beta}U^\alpha U^\beta = -1$ holds. The energy-momentum tensor is

$$T^{\alpha\beta} = (\rho + p)U^\alpha U^\beta + pg^{\alpha\beta} \tag{71}$$

It is assumed that ρ and p are non-negative. The nature of the fluid is given by an equation of state $p = f(\rho)$. A minimal condition that this equation of state should satisfy is that f is continuously differentiable for $\rho > 0$ and that there $f'(\rho) \geq 0$. A physical explanation for this is that $f'(\rho) \geq 0$ is equal to the square of the velocity of sound. Mathematically the condition (in the stronger form $f'(\rho) > 0$) ensures that the equations have a well-posed initial value problem. A physically desirable condition is that $f'(\rho) \leq 1$ since this has the interpretation that the speed of sound is no greater than the speed of light. The equations of motion, the Euler equations, are $\nabla_\alpha T^{\alpha\beta} = 0$. This relation must hold for any energy-momentum tensor. What is special in this case is that it is equivalent to the equations of motion. A special case of the perfect fluid is that where the pressure is identically zero. This case is often called dust. Solutions of the Einstein equations with dust are frequently considered, although they sometimes have unpleasant properties. A linear equation of state $p = (\gamma - 1)\rho$ with $1 < \gamma \leq 2$ is also a common choice. The case $\gamma = 4/3$ is known as a radiation fluid. When space is filled with radiation then it is possible to introduce an effective description by a fluid with this equation of state.

To check the energy conditions for a perfect fluid it is useful to use a frame which is adapted to the four-velocity. Then

$$T_{\alpha\beta}V^\alpha W^\beta = \rho V^0 W^0 + p\delta_{ij}V^i W^j \tag{72}$$

Thus the reasonable condition that $p \leq \rho$ is enough to ensure that the dominant energy condition is satisfied.

$$(T_{\alpha\beta} - \frac{1}{2}Tg_{\alpha\beta})V^\alpha V^\beta = \frac{1}{2}(\rho + 3p)(V^0)^2 + \frac{1}{2}(\rho - p)\delta_{ij}V^i V^j \tag{73}$$

Provided $p \leq \rho$ and $\rho + 3p \geq 0$ then the strong energy condition is satisfied.

$$T_{\alpha\beta}(g^{\alpha\beta} + V^\alpha V^\beta) = \rho[(V^0)^2 - 1] + p(3 + \delta_{ij}V^i V^j) \tag{74}$$

Thus the non-negative sum pressures condition holds.

4.5 Collisionless matter

Collisionless matter consists of particles which move without direct mutual interaction and are described statistically. The fundamental quantity is a function $f(x^\alpha, p^\alpha)$. It describes the number density of particles at a given point of spacetime with given four-momentum and is non-negative. The energy-momentum tensor is obtained by integrating over momenta:

$$T^{\alpha\beta}(x^\alpha) = \int p^\alpha p^\beta f(x^\alpha, p^\alpha)(\det g)^{1/2} dp^\alpha \tag{75}$$

The equation of motion is the Vlasov equation

$$p^\alpha \partial f/\partial x^\alpha + \Gamma^\alpha_{\beta\gamma} p^\beta p^\gamma \partial f/\partial p^\alpha = 0 \tag{76}$$

Often consideration is restricted to particles of unit mass. Then $g_{\alpha\beta}p^\alpha p^\beta = -1$ and p^0 can be expressed as a function of p^a. Then the energy-momentum tensor is

$$T^{\alpha\beta}(t, x^a) = -\int p^\alpha p^\beta f(t, x^a, p^a)/p_0 (\det g)^{1/2} dp^a \tag{77}$$

The Vlasov equation in that case takes the form

$$\partial f/\partial t + (p^a/p^0)\partial f/\partial x^a + (\Gamma^a_{\beta\gamma} p^\beta p^\gamma/p^0)\partial f/\partial p^a = 0 \tag{78}$$

In the case of collisionless matter all the energy conditions introduced above and the non-negative sum pressures condition are satisfied.

5 Cosmological models

A cosmological model is a model of the universe on the largest scales we can observe. Structures on smaller scales such as galaxies are more or less neglected. The Kasner solution which we have already seen is a kind of cosmological model. This solution does not contain any matter and so is not a good description of our universe. It is better to include matter. The simplest and most popular model is the Einstein-de Sitter model (which has nothing to do with the de Sitter model apart from having been discovered by the same person). The metric is

$$-dt^2 + t^{4/3}(dx^2 + dy^2 + dz^2) \tag{79}$$

The metric satisfies the Einstein equations with dust. The four-velocity has components $(1, 0, 0, 0)$ and the energy density depends only on time. This metric has the full Euclidean group as its symmetry group. It is invariant under translations in x, y and z. We say that it is homogeneous (or, more preciely, spatially homogeneous). In each time slice, i.e. in each level surface of the function t, each point is equivalent to each other point. This is what 'homogeneous' means. The metric is also invariant under all rotations about an arbitrary point of the (x, y, z) plane. We say that the spacetime is isotropic. In each point of the time slice all directions in the tangent space are equivalent. They are related by symmetries of spacetime. Physically the property of isotropy means that the sky looks the same in all directions.

In contrast to this solution, the Kasner solution is homogeneous but not isotropic. The induced metric on each time slice is flat and we say that the model is spatially flat. Models which are homogeneous and isotropic are called FLRW models (Friedmann-Lemaître-Robertson-Walker) and the Einstein-de Sitter model is a special FLRW model. Sometimes it is called the flat FLRW model but this is misleading. The curvature of this spacetime does not vanish.

If we identify in x, y und z we get a spatially compact model. The spatial topology is that of a three-dimensional torus. The volume of the spatial slices is equal to t^2. The volume of space grows steadily with t and we call this an expanding universe. The integral curves of the four-velocity can be identified with the world-lines of galaxies and the distance between these world-lines also grows with t. Information about more general cosmological FLRW models can be found in my notes on cosmological models.

6 Gowdy spacetimes

This section is concerned with the Gowdy spacetimes, a class of inhomogeneous solutions of the vacuum Einstein equations. They have a two-dimensional isometry group which acts on spacelike surfaces. After factoring out this symmetry we are left with an effective problem in one space and one time dimension. This class of spacetimes serves as a laboratory for studying the dynamical properties of solutions of the Einstein equations under relatively simple conditions. This special class of solutions of the vacuum Einstein equations displays a remarkable variety of behaviour.

In suitable coordinates the metric can be written in the form:

$$t^{-1/2}e^{\lambda/2}(-dt^2+dx^2)+t(e^P(dy+Qdz)^2+e^{-P}dz^2) \tag{80}$$

Here $t>0$ and P, Q and λ are functions of t and x which are periodic in x. The Einstein equations are equivalent to the following five equations in this case:

$$P_{tt}+t^{-1}P_t-P_{xx} = e^{2P}(Q_t^2-Q_x^2) \tag{81}$$
$$Q_{tt}+t^{-1}Q_t-Q_{xx} = -2(P_tQ_t-P_xQ_x) \tag{82}$$
$$\lambda_t = t[P_t^2+P_x^2+e^{2P}(Q_t^2+Q_x^2)] \tag{83}$$
$$\lambda_x = 2t(P_tP_x+e^{2P}Q_tQ_x) \tag{84}$$
$$\lambda_{tt}+t^{-1}\lambda_t-\lambda_{xx} = 2(P_x^2+e^{2P}Q_x^2) \tag{85}$$

If the first four equations are satisfied then a straightforward calculation shows that the fifth one is also satisfied. Note that the first two equations are independent of λ and so a natural strategy when studying these equations is to first solve the wave equations for P and Q and then substitute the result into the equations for λ. The appropriate initial data for the wave equations consists of the restrictions to a hypersurface $t = t_0 > 0$ of the functions P, P_t, Q and Q_t. Note that we cannot specify initial data at $t = 0$ since the equations are singular there. In order to be consistent with all Einstein equations the data on $t = t_0$ cannot quite be specified freely. Integrating the equation for λ_θ in θ and using periodicity gives the restriction

$$\int_0^{2\pi} P_t(t_0, x)P_x(t_0, x) + e^{2P(t_0,x)}Q_t(t_0, x)Q_x(t_0, x)dx = 0 \tag{86}$$

If P and Q are functions which satisfy the wave equations above then the following identity is satisfied

$$\partial_x(t[P_t^2 + P_x^2 + e^{2P}(Q_t^2 + Q_x^2)]) = \partial_t[2t(P_tP_x + e^{2P}Q_tQ_x)] \tag{87}$$

Integrating this relation in x shows that if the compabitibility condition (86) is satisfied for some t_0 it is satisfied for all t. Moreover, it shows that if the equation for λ_x is satisfied for $t = t_0$ and the equation for λ_t is satisfied everywhere then the equation for λ_x is satisfied everywhere. Thus, in order to solve the Einstein equations in the case of Gowdy spacetimes we can proceed as follows. First solve the wave equations for P and Q with initial data at $t = t_0$ satisfying the restriction (86). Determine λ at $t = t_0$ by integrating the equation for λ_x there. Then determine λ away from the initial hypersurface by integrating the equation for λ_t in time starting at $t = t_0$. After this has been done all of the Einstein equations are satisfied. This means that in Gowdy spacetimes the essential part of the Einstein equations are the wave equations for P and Q and we will refer to these as the Gowdy equations.

A special case of the Gowdy equations is obtained by setting $Q = 0$. The resulting equation for P is called the polarized Gowdy equation and is linear. It is easy to write down explicit solutions of the polarized Gowdy equation. Suppose that $P(t, x) = p(t)\sin kx$ for some function $p(t)$ and a positive integer k. Then

$$p_{tt} + t^{-1}p_t + k^2p = 0 \tag{88}$$

The solutions of this ordinary differential equation are Bessel functions, whose qualitative beaviour for $t \to 0$ and $t \to \infty$ is well understood. Since the equation is linear, finite linear combinations of these explicit solutions are also solutions. A possible strategy to study the general solution would be to try and expand it as a series in these explicit solutions but this does not turn out to be very useful in practice. In any case, it is an approach which cannot be extended to the general Gowdy equations, which are nonlinear.

To study general solutions of the Gowdy equations we consider the initial value problem. The first question is that of local existence and uniqueness. If initial data is prescribed on $t = t_0$ does there exist a corresponding solution of the equations at least on a short time interval? For concreteness suppose that the data are smooth (C^∞) and that we are looking for a smooth solution. The equations are semilinear hyperbolic and standard results from the theory of hyperbolic equations give a positive answer to the question of local existence and uniqueness. They also give global uniqueness.

The next question is that of global existence. This depends on the detailed structure of the equations. Consider for comparison the equation

$$u_{tt} - u_{xx} = u_t^2 \tag{89}$$

A simple solution of this equation is $u(t,x) = -\log(2-t)$. Taking $t_0 = 1$ this solution has initial data $u(x) = 0$, $u_t(x) = 1$ at $t = t_0$. The unique corresponding solution becomes singular at $t = 2$ and so global existence fails. For the Gowdy equations things look better. It was shown by V. Moncrief [6] in 1981 that any data for the Gowdy equations at $t = t_0 > 0$ gives rise to a solution on the whole time interval $(0,\infty)$.

Once a solution is known to exist globally on the time interval of interest we may ask about its asymptotics as $t \to 0$ and $t \to \infty$. In the following we will prove the form of the asymptotics for general polarized Gowdy solutions near $t = 0$. The limiting behaviour as $t \to \infty$ is harder to control. The method to be used to determine the behaviour near $t = 0$ is that of energy estimates. It is one of the most important techniques in the theory of hyperbolic partial differential equations.

For the proof we need the following inequality for periodic functions of one variable

$$|f(x)| \le C\left[\int (f(y))^2 + (f_x(y))^2 dy\right]^{1/2} \tag{90}$$

Now this will be proved.

$$|f(x_2) - f(x_1)| \le \int_{x_1}^{x_2} |f_x(y)|dy \le (2\pi)^{1/2}\left[\int_0^{2\pi} |f_x(y)|^2 dy\right]^{1/2} \tag{91}$$

where the Cauchy-Schwarz inequality has been used. At some point x_0 the value of f is equal to its average $\bar{f} = \frac{1}{2\pi}\int_0^{2\pi} f(y)dy$. Taking $x_1 = x_0$ and x_2 to be an arbirary point x we see from the above that

$$|f(x)| \le |\bar{f}| + (2\pi)^{1/2}\left[\int_0^{2\pi} |f_x(y)|^2 dy\right]^{1/2} \tag{92}$$

On the other hand, by the Cauchy-Schwarz inequality $|\bar{f}| \le (2\pi)^{-1/2}[\int f^2(y)dy]^{1/2}$. Combining these results gives the inequality (90). It is a simple example of a Sobolev inequality. Sobolev inequalities which give uniform estimates on a function in terms of integral norms play a very important role in the theory of partial differential equations and, in particular, in the theory of hyperbolic equations.

If P is a solution of the polarized Gowdy equation let

$$E(t) = t^2 \int_0^{2\pi} (P_t^2 + P_x^2)(t,x)dx \tag{93}$$

Then $dE/dt = 2t\int_0^{2\pi}(P_x(t,x))^2 dx$. In particular $E_t \ge 0$ and so $E(t) \le E(t_0)$ for all $t \le t_0$. Now P_x satisfies exactly the same equation as P and so if we define $E_1 = t^2 \int_0^{2\pi} (P_{tx}^2 + P_{xx}^2)(t,x)dx$ then $E_1(t)$ can be bounded by its value at t_0 for any $t \le t_0$. We can also define analogous quantities E_n by replacing P by $\partial_x^n P$ in the definition of E. The fact that all quantities E_n are bounded for $t \le t_0$ implies, via the Sobolev inequality, that tP and its spatial derivatives of all orders are bounded.

The polarized Gowdy equation can be written in the form $\partial_t(tP_t) = tP_{xx}$. The right hand side is bounded on the interval $(0, t_0]$. Integrating this relation once gives

$$t_0 P_t(t_0) - tP_t(t) = \int_t^{t_0} sP_{xx}(s,x)ds \tag{94}$$

Since the quantity in the integral is bounded this can be rewritten as

$$tP_t(t) = t_0 P_t(t_0) - \int_0^{t_0} sP_{xx}(s,x)ds + \int_0^t sP_{xx}(s,x)ds \tag{95}$$

Let $k(x) = t_0 P_t(t_0) - \int_0^{t_0} sP_{xx}(s,x)ds$. Then $P_t(t,x) = k(x)t^{-1} + O(1)$ as $t \to 0$. This can be integrated again to give

$$P(t_0) - P(t) = k(x)[\log t_0 - \log t] + \int_0^{t_0} r(s,x)ds - \int_0^t r(s,x)ds \tag{96}$$

where $r(t,x)$ is remainder in the expression for P_t which is $O(1)$. Hence if $\phi = P(t_0) - k \log t_0 - \int_0^{t_0} r(s,x)ds$ then $P(t,x) = k(x)\log t + \phi(x) + O(t)$. The relations obtained by differentiating these expansions term by term to any order with respect to t and x are also valid, as can be shown by similar techniques. Every solution of the polarized Gowdy equation can be written in the form just given. Conversely it can be shown that given any pair of smooth functions π and ω there is a unique solution of the polarized Gowdy equation whose asymptotic expansion gives exactly those functions as coefficients. This means that in effect all solutions of these equations can be parametrized by the two functions π and ω which play the role of data on the singularity.

The asymptotics of solutions of the polarized Gowdy equations for $t \to \infty$ will now be described without proof. We have

$$P(t,x) = A \log t + B + t^{-1/2}\nu(t,x) + o(t^{-1/2}) \tag{97}$$

where A and B are constants and ν satisfies the flat space wave equation $\nu_{tt} = \nu_{xx}$. Since the equation for ν has wavelike solutions whose amplitude remains constant the solutions have persistent oscillations. The amplitude of the oscillations in P decays like $t^{-1/2}$ but the solution does not allow an expansion of the form $P(t,x) = \sum_k A_k(x)t^{-p_k}$. The validity of this expansion was proved in [4]. This suffices to show future geodesic completeness of polarized Gowdy spacetimes.

To conclude this section we make some remarks about the non-polarized case. The proofs are too complicated to be included here. Given any four functions k, ϕ, Q_0 and ψ with $0 < k(x) < 1$ for all x there exists a unique solution of the Gowdy equations with the following asymptotic expansion:

$$P(t,x) = k(x)\log t + \phi(x) + o(1) \tag{98}$$
$$Q(t,x) = Q_0(x) + t^{2k(x)}\psi(x) + o(t^{2k(x)}) \tag{99}$$

This was proved in the case of analytic data in [5] and generalized to the smooth case in [9]. The condition $k < 1$, which is known as the low velocity condition, has a very important role to play and it cannot expected that it can be dropped. The question of which initial data at $t = t_0$ give rise to a solution with the asymptotics given above is more difficult. It has been proved that all Gowdy spacetimes are future geodesically complete [10] and that for generic initial data the Kretschmann scalar blows up uniformly as $t \to 0$ [11].

7 Spherical symmetry

7.1 The asymptotically flat case

In this section we consider metrics of the form

$$-e^{2\mu(t,r)}dt^2 + e^{2\lambda(t,r)}dr^2 + r^2(d\theta^2 + \sin^2\theta d\phi^2) \tag{100}$$

The simplest example is that where the functions μ and λ vanish identically, in which case we get Minkowski space in spherical polar coordinates. The metric (100) is invariant under all rotations in the angular variables. The spacetime has the full symmetry of the standard sphere and is therefore called spherically symmetric. It is written in a special coordinate system, sometimes known as Schwarzschild coordinates. This is characterized by two conditions. Firstly, the factor in front of the metric of the standard sphere is given by the square of the coordinate function r. In this case r is called area radius since it is determined by the fact that the area of the spheres of symmetry is proportional to r^2. Secondly the curves of constant (t,θ,ϕ) are orthogonal to those of constant (r,θ,ϕ). This is equivalent to the fact that $g_{01} = 0$. Of course it is also the case that g_{02}, g_{03}, g_{12} and g_{13} vanish but that is such a natural condition in spherically symmetric spacetimes that it is almost always assumed.

We wish to consider spacetimes which describe isolated physical systems. A good example of an isolated system is the solar system. The motion of the planets under the influence of their mutual gravitational interaction is scarcely affected by what happens outside the solar system. Thus we can call it an isolated system and use an idealized model where the solar system is placed in an otherwise empty universe. Far away from the region where the solar system is located there is no matter and the gravitational field becomes weaker and weaker. A single star like the sun can also be described as an isolated system. There it is reasonable to use a spherically symmetric model. This would fall into the class of models treated in this section. Similarly a globular cluster, which is a spherical collection of a large number of stars (perhaps a hundred thousand) can be modelled in this way.

In general relativity the notion of an isolated system is formalized using asymptotically flat spacetimes. These are spacetimes where at large spatial distances the energy-momentum tensor is zero, or at least very small, and the metric approaches the Minkowski metric. For the spherically symmetric metric given above this is translated into the condition that $\mu(t,r)$ and $\lambda(t,r)$ should tend to zero as r tends to infinity for each fixed value of t. There is another important boundary condition which is that of a regular centre. This means that $\lambda(t,0)=0$ for each t. This ensures that for small circles at constant distance from the origin in the hypersurfaces of constant t the ratio of the circumference to the radius approaches 2π. This is necessary if the metric written in coordinates is to describe a geometry which is smooth at $r=0$. It ensures that there is only a coordinate singularity there and not a geometric singularity. If λ approached a different constant value as $r\to 0$ then there would be a conical point at $r=0$ instead of a smooth metric.

As a consequence of spherical symmetry the only non-vanishing components of the energy-momentum tensor are T_{00}, $T_{01}=T_{10}$, T_{11}, $T_{22}=T_{33}$. To see this consider the orthonormal basis $e_0=e^{-\mu}\partial/\partial t$, $e_1=e^{-\lambda}\partial/\partial r$, $e_2=r^{-1}\partial/\partial\theta$, $e_3=r^{-1}(\sin\theta)^{-1}\partial/\partial\phi$. Since there is a rotation which fixes e_0 and e_1 and replaces e_2 by $-e_2$ it follows that T_{02}, T_{03}, T_{12} and T_{13} vanish. A rotation which takes e_2 to e_3 and e_3 to $-e_2$ shows that T_{23} vanishes and that $T_{22}=T_{33}$. We introduce the following notation to parametrize the non-vanishing energy-momentum components. $\rho=T(e_0,e_0)$, $j=-T(e_0,e_1)$, $p=T(e_1,e_1)$, and $q=T(e_2,e_2)+T(e_3,e_3)$. The Einstein equations then take the form:

$$e^{-2\lambda}(2r\lambda'-1)+1 = 8\pi r^2\rho \qquad (101)$$

$$e^{-2\lambda}(2r\mu'+1)-1 = 8\pi r^2 p \qquad (102)$$

$$\dot\lambda = -4\pi r e^{\lambda+\mu} j \qquad (103)$$

$$e^{-2\lambda}(\mu''+(\mu'-\lambda')(\mu'+1/r))-e^{-2\mu}(\ddot\lambda+\dot\lambda(\dot\lambda-\dot\mu)) = 4\pi q \qquad (104)$$

Here the dot and prime denote derivatives with respect to t and r respectively.

We now consider the vacuum case, ignoring for a moment the boundary conditions. It follows from one of the field equations that $(re^{-2\lambda})'=1$. Hence $re^{-2\lambda}=r-2m$ for some constant m and $e^{-2\lambda}=1-2m/r$. The constant m might a priori have depended on t but this is ruled out by the field equation for $\dot\lambda$. Combining two of the field equations gives $r(\lambda'+\mu')=0$. Thus $\lambda+\mu=C(t)$ for some constant $C(t)$ depending on t. It follows that $e^{\mu(t,r)}=(1-2m/r)^{1/2}e^{C(t)}$. By a reparametrization of t we can set $C(t)$ to unity. There results the metric

$$-(1-2m/r)dt^2+(1-2m/r)^{-1}dr^2+r^2(d\theta^2+\sin^2 d\phi^2) \qquad (105)$$

This metric defines the Schwarzschild spacetime, which is the standard model for a black hole in general relativity.

Note that there are two places where the Schwarzschild metric appears to become singular, namely when $r=0$ and when $r=2m$. To find out more about the nature of these singularities we look at the Kretschmann scalar. It is given by $48m^2/r^6$. Hence the singularity at $r=0$ is a geometric singularity and it is not possible to extend the metric through it. On the other hand the Kretschmann scalar remains bounded as r approaches $2m$ and this might suggest that this is a coordinate singularity. This turns out to be the case. To see this we introduce a new coordinate system, Eddington-Finkelstein coordinates. Define $r^* = r + 2m \log(r - 2m)$ and $v = t + r^*$. Then in the coordinates (v, r, θ, ϕ) the Schwarzschild metric takes the form

$$-(1 - 2m/r)dv^2 + 2dvdr + r^2(d\theta^2 + \sin^2 d\phi^2) \tag{106}$$

This shows that the metric has a regular extension through $r = 2m$. The coordinate singularity in the Schwarzschild coordinates arises as follows. The level surfaces of the coordinate r become null as r approaches $2m$. The level curves of t in the (t, r) surface are orthogonal to the level curves of r there. Thus both become parallel when the gradient of r becomes null. At that point the functions t and r no longer form a coordinate system.

When $m = 0$ the Schwarzschild solution reduces to the Minkowski metric. It is possible to consider geodesics of the Schwarzschild solution for large values of r and see that they are approximated by the trajectories of Newtonian particles moving in the gravitational field of a point mass m. Thus the Schwarzschild solution is the general relativistic equivalent of a point mass in Newtionian gravity and m should be non-negative in order to be physically reasonable. In fact it can be shown that the Schwarzschild solution with negative mass has undesirable properties. It is not globally hyperbolic and cannot be made so by any extension. This is connected with the absence of a hypersurface $r = 2m$ for $m < 0$.

In the case $m > 0$ the hypersurface $r = 2m$ is called the event horizon. It has the property that no causal curve which starts within the region $r \leq 2m$ can ever reach the exterior region $r > 2m$. The region $r \leq 2m$ is called the black hole and the mathematical fact just mentioned has the physical interpretation that no particle, or observer, can ever escape from a black hole; the event horizon is a boundary which can only be crossed in one direction, namely from outside to inside. It protects the exterior region from feeling any influence of the singularity at $r = 0$. This is sometimes called cosmic censorship. The opposite of this, a singularity which can be seen from far away like that of the negative mass Schwarzschild solution, is called a naked singularity.

The Schwarzschild solution is asymptotically flat. On the other hand it does not have a regular centre. Physically it represents an eternal black hole which has always been there. More physically interesting is the situation where a black hole arises dynamically. In spherically symmetric vacuum spacetimes there is no dynamics. We have seen that after a suitable choice of coordinates the metric components are time independent. A more invariant way of expressing this is to say that there is a family of isometries with timelike trajectories (translations in t). A spacetime with this property is called static. The statement that any spherically symmetric vacuum spacetime is static is known as Birkhoff's theorem. To describe the formation of a black hole from an initially regular configuration we must use non-static spacetimes. If we wish to stay in the context of spherically symmetry it is necessary to introduce some kind of matter. The standard spacetime which as used as a simple model for the formation of a black hole is the Oppenheimer-Snyder solution which will be discussed later. In that case the matter model is dust.

The Oppenheimer-Snyder model is the source of many ideas about gravitational collapse and the formation of black holes. However as soon as we go away from the very special assumptions of that model dust turns out to have pathological properties. Therefore for the general study of gravitational collapse it is better to choose a better-behaved matter model. The two which have been studied most are the massless scalar field and collisionless matter. In the following we will discuss what is known about the case of collisionless matter.

A general picture of what happens in the spherical gravitational collapse of collisionless matter can be obtained from numerical investigations. For small initial data the matter disperses. For large initial data a black hole is often formed. It is difficult to confirm these features by rigorous proofs. At least the case of small initial data has been treated analytically [8].

The proof of the small data global existence theorem for the spherically symmetric Einstein-Vlasov system relies on the fact that after a long time the geodesics are almost straight lines in the given coordinates. Since straight lines are more easily described in Cartesian than in polar coordinates it is convenient to write the Vlasov equation in Cartesian coordinates. The Cartesian coordinates are defined in terms of the given polar coordinates just as in Euclidean space, i.e. $x = r\sin\theta\cos\phi$, $y = r\sin\theta\sin\phi$, $z = r\cos\theta$. In these new coordinates we have $g_{ab} = \delta_{ab} + (e^{2\lambda} - 1)x_a x_b/r^2$. The Vlasov equation is

$$\partial_t f + \frac{p^a}{p^0}\partial_{x^a} f - \frac{1}{p^0}\Big[e^{2(\mu-\lambda)}\mu'(p^0)^2 + 2\dot{\lambda}\frac{x\cdot p}{r}p^0 \tag{107}$$

$$+ \lambda'\left(\frac{x\cdot p}{r}\right)^2 + \frac{1-e^{-2\lambda}}{r}\left(|p|^2 - \left(\frac{x\cdot p}{r}\right)^2\right)\Big]\frac{x^a}{r}\partial_{p^a} f = 0 \tag{108}$$

It turns out to be more convenient to express the Vlasov equation in components v^α with respect to an orthonormal frame instead of the components p^α with respect to a coordinate frame. The transformation of the spatial coordinates is given by

$$v^a = p^a + (e^{-\lambda} - 1)\frac{x\cdot p}{r}\frac{x^a}{r} \tag{109}$$

When we consider spherically symmetric solutions of the Einstein-Vlasov system the distribution function is required to be spherically symmetric. This means that $f(t, Ax, Av) = f(t, x, v)$ for any orthogonal matrix A. It implies that f satisfies the following condition:

$$(r^2 v - (x\cdot v)x)\partial_x f = (|v|^2 x - (x\cdot v)v)\partial_v f \tag{110}$$

Replacing p^α by v^α in the Vlasov equation and using (110) shows that the Vlasov equation is equivalent to the simplified equation

$$\partial_t f + e^{\mu-\lambda}\frac{v}{\sqrt{1+|v|^2}}\cdot\partial_x f - \left(\dot{\lambda}\frac{x\cdot v}{r} + e^{\mu-\lambda}\mu'\sqrt{1+|v|^2}\right)\frac{x}{r}\cdot\partial_v f = 0 \tag{111}$$

We continue to use polar coordinates for the Einstein equations. The use of orthonormal frame components makes the expressions for the components of the energy-momentum tensor particularly simple. They are

$$\rho(t,x) = \int \sqrt{1+|v|^2} f(t,x,v)dv \tag{112}$$

$$p(t,x) = \int \left(\frac{x\cdot v}{r}\right)^2 f(t,x,v)\frac{dv}{\sqrt{1+|v|^2}} \tag{113}$$

$$j(t,x) = \int \frac{x\cdot v}{r} f(t,x,v)dv \tag{114}$$

$$q(t,x) = \int \left|\frac{x\times v}{r}\right|^2 f(t,x,v)\frac{dv}{\sqrt{1+|v|^2}} \tag{115}$$

Now the whole Einstein-Vlasov system has been obtained in a suitable form for proving the global existence theorem. The proof is too long and complicated to be explained here. We simply state the result.

Theorem Let $f_0(x,v)$ be a non-negative C^1 function of compact support which is spherically symmetric. Then there exists a unique local solution of the Einstein-Vlasov system $f(t,x,v)$ with $f(0,x,v) = f_0(x,v)$. If f_0 is sufficiently small then the solution exists globally in t, the spacetime is geodesically complete and the components of the Riemann tensor converge uniformly to zero as $t \to \infty$.

It remains to explain what is meant by 'sufficiently small'. For instance it is enough to assume that for a fixed support of f_0 the maximum value of f_0 is sufficiently small.

More recently it has been possible to prove some theorems on the long-time behaviour of solutions corresponding to certain classes of initial data which are not small. The first case is where there is a large amount of collisionless matter and the particles are all initially moving outwards with sufficiently high velocity [1]. With suitable technical assumptions it is shown that the solution exists globally in the future with asymptotics which are similar to those in the case of small data. The second case concerns initial data representing matter in an annular region (shell) where the particles are moving inwards sufficiently fast [2]. With suitable technical assumptions it is shown that a black hole forms before the shell reaches the centre. For comparison, it may noted that even more recently a theorem on the formation of a black hole in vacuum was proved by Christodoulou [3]. Of course, due to Birkhoff's theorem, these solutions are not spherically symmetric.

7.2 The cosmological case

This section is concerned with spherically symmetric spacetimes satisfying different boundary conditions from those considered so far. They are such that they admit a compact Cauchy hypersurface, i.e. they are spatially compact. The related class of plane symmetric spacetimes will also be considered. These solutions of the Einstein equations can be interpreted as cosmological models. In appropriate coordinates the spherically symmetric metric takes the form

$$-e^{2\mu(t,r)}dt^2+e^{2\lambda(t,r)}dr^2+t^2(d\theta^2+\sin^2\theta d\phi^2) \qquad (116)$$

There is a great similarity to (100) with the difference that now instead of an area radius we have an areal time coordinate. The metric is invariant under rotations in the angular coordinates as before and so is spherically symmetric. The coordinate r is supposed to be periodic. The plane symmetric metric is

$$-e^{2\mu(t,r)}dt^2+e^{2\lambda(t,r)}dr^2+t^2(d\theta^2+d\phi^2) \qquad (117)$$

with r, θ and ϕ periodic. The spatial manifold is that of a three dimensional torus. Translations in θ and ϕ are isometries. On the universal cover rotations in the (θ,ϕ) plane are also isometries. Thus the spacetime has local symmetries which coincide with those of the Euclidean plane. This leads to the name plane symmetry. For convenience in writing certain formulae we introduce a parameter k which takes the value one for spherical symmetry and the value zero for plane symmetry. The plane and spherically symmetric spacetimes where the functions μ and λ do not depend on r are spatially homogeneous. A further specialization is obtained by setting $\lambda=\log t$ in the plane symmetric case. This leads to a FLRW model. It is written in a different time coordinate from what we have seen so far. Thus the plane symmetric cosmological models are anisotropic and inhomogeneous generalizations of the spatially flat FLRW models. The models of this class which are homogeneous are called Bianchi type I models. The homogeneous spherically symmetric models are called Kantowski-Sachs models.

The non-vanishing components of the energy-momentum tensor are as in the last section. We introduce an orthonormal frame adapted to the given coordinates and quantities ρ, j, p and q as before. The Einstein equations take the form

$$\begin{aligned} e^{-2\mu}(2t\dot\lambda+1)+k &= 8\pi t^2\rho && (118)\\ e^{-2\mu}(2t\dot\mu-1)-k &= 8\pi t^2 p && (119)\\ \mu' &= -4\pi t e^{\lambda+\mu}j && (120)\\ e^{-2\lambda}(\mu''+(\mu'-\lambda')\mu')-e^{-2\mu}(\ddot\lambda+(\dot\lambda+1/t)(\dot\lambda-\dot\mu)) &= 4\pi q && (121)\end{aligned}$$

Consider the vacuum case. Then one of the field equations shows that μ only depends on t. It follows that λ is the sum of a function of r and a function of t. By a transformation of the coordinate r the r dependence in λ can be eliminated. This means that the vacuum solutions with plane and spherical symmetry are very simple. This has the physical interpretation that these symmetries are incompatible with the presence of gravitational waves. Another field equation shows that $d/dt(te^{-2\mu(t)})=-k$. Hence $e^{2\mu(t)}=(k+C/t)^{-1}$ for a constant C. In the case $k=1$ we can write $C=-2m$ and then $e^{2\mu(t)}=(1-2m/t)^{-1}$. The remaining equation can be integrated to give $e^{2\lambda}=1-2m/t$. Thus we see that this is nothing other than a part of the Schwarzschild solution with

t and r interchanged in the notation. The solution is defined for $t \in (0, 2m)$; a non-trivial solution is only obtained if $m > 0$. This solution has an initial singularity for $t = 0$ but also exists only for a finite time in the future. What happens as $t \to 2m$? This is a coordinate singularity. It marks the end of the largest globally hyperbolic spacetime having the hypersurfaces t=const. as Cauchy hypersurfaces. It is however possible to extend the spacetime further. The hypersurface $t = 2m$ where global hyperbolicity breaks down is called a Cauchy horizon. Consider now the plane symmetric case. There $e^{2\mu(t)} = Ct$. By rescaling the time coordinate we can assume $C = 1$. Hence $\mu = \frac{1}{2}\log t$. The remaining field equation then gives $\lambda = -\frac{1}{2}\log t$. The metric is

$$-tdt^2 + t^{-1}dr^2 + t^2(d\theta^2 + d\phi^2) \tag{122}$$

Putting $\tau = t^{3/2}$ gives the metric

$$-\frac{4}{9}d\tau^2 + \tau^{-2/3}dr^2 + \tau^{4/3}(d\theta^2 + d\phi^2) \tag{123}$$

A constant rescaling of the coordinates shows that this is the Kasner solution with Kasner exponents $(2/3, 2/3, -1/3)$.

Next cosmological spacetimes with spherical and plane symmetry and matter will be considered. In particular the case of collisionless matter will be treated. The Vlasov equation can be written in the explicit form

$$\partial_t f + \frac{e^{\mu-\lambda}v^1}{(1+|v|^2)^2}\partial_r f - (\dot{\lambda} + e^{\mu-\lambda}\mu'(1+|v|^2)^{1/2}\partial_{v^1} f = 0 \tag{124}$$

in an orthonormal frame. The global properties of spacetimes of this type have been studied in [7].

What kind of global properties are of interest? Firstly there is the question of the interval on which solutions of the equations exist. The equations become singular as t tends to zero and so the maximal possible interval of existence of the solutions is $(0, \infty)$. In the spherically symmetric case solutions may recollapse and the areal time is not a good global coordinate. From this point of view the plane symmetric case is more favourable and in that case there is a global existence theorem for the Einstein-Vlasov system in an areal time coordinate (global on the interval $(0, \infty)$). An interesting open question is whether these spacetimes are future geodesically complete. This seems to be hard to decide. Is there a physical phenomenon behind this mathematical difficulty? A possible candidate for such a phenomenon is the Jeans instability. This arises if we consider linearized perturbations of a spatially flat FLRW model. It is found that the density contrast $\delta\rho/\rho$ can grow. This is a fundamental element of the usual explanations of the formation of galaxies. Unfortunately the issue is mathematically poorly understood. If a cosmological constant is added to the plane symmetric model then geodesic completeness can be proved. The inflationary behaviour of the model simplifies the mathematics.

Also of interest is the question of the nature of the initial singularity at $t = 0$ in these models (supposing that no singularity occurs before $t = 0$ is reached).

This is not completely understood but Rein [7] has obtained a number of results for a large class (large open set) of solutions. For these solutions (with spherical or plane symmetry) it is found that the solution exists up to $t = 0$. The generalized Kasner exponents converge to the values $(2/3, 2/3, -1/3)$. Note that these satisfy the second Kasner relation and this corresponds to a vacuum solution. In a certain sense the given solution is approximated by a vacuum solution near the singularity. This does not mean that the energy density is small near the singularity. In fact it diverges. It is just that other terms in the Einstein equations grow faster than the matter terms and dominate the dynamics near the singularity. There is a conjectural picture due to Belinskii, Khalatnikov and Lifschitz (BKL) that this kind of vacuum dominance is a typical feature of spacetime singularities. These spacetimes are also such that near their singularities the inhomogeneous spacetimes look like homogeneous spacetimes depending on the spatial coordinates as parameters. It is as if the evolution at different spatial points becomes independent. This decoupling is another aspect of the BKL picture. The validity of the BKL picture in general is still a matter of intense investigation. Coming back to the solutions analysed by Rein, he was able to show that the Kretschmann scalar blows up as $t = 0$ is approached so that this is a genuine geometric singularity and not just a coordinate effect.

References

[1] Andréasson, H., Kunze, M. and Rein, G. 2006 Global existence for the spherically symmetric Einstein–Vlasov system with outgoing matter. Preprint gr-qc/0611115.

[2] Andréasson, H., Kunze, M. and Rein, G. 2007 The formation of black holes in spherically symmetric gravitational collapse. Preprint arXiv:0706.3787.

[3] Christodoulou, D. 2008 The formation of black holes in general relativity. Preprint arXiv:0805.3880.

[4] Jurke, T. 2003 On future asymptotics of polarized Gowdy T^3-models. Class. Quantum Grav. 20, 173-192.

[5] Kichenassamy, S. and Rendall, A. D. 1998 Analytic description of singularities in Gowdy spacetimes. Class. Quantum Grav. 15, 1339-1355.

[6] Moncrief, V. 1981 Global properties of Gowdy spacetimes with $T^3 \times R$ symmetry. Ann. Phys. (NY) 132, 87-107.

[7] Rein, G. 1996 Cosmological solutions of the Vlasov–Einstein system with spherical, plane, or hyperbolic symmetry. Math. Proc. Camb. Phil. Soc. 119, 739–762.

[8] Rein, G. and Rendall, A. D. 1992 Global existence of solutions of the spherically symetric Vlasov-Einstein system with small initial data. Commun. Math. Phys. 150, 561-583.

[9] Rendall, A. D. 2000 Fuchsian analysis of singularities in Gowdy spacetimes beyond analyticity. Class. Quantum Grav. 17, 3305-3316.

[10] Ringström, H. 2004 On a wave map equation arising in general relativity. Commun. Pure Appl. Math. 57, 657-703.

[11] Ringström, H. 2003 Strong cosmic censorship in T^3-Gowdy spacetimes. Ann. Math. (to appear)

[12] Wald, R. M. 1984 General relativity. University of Chicago Press, Chicago.

[13] Wasserman, R. H. 1992 Tensors and manifolds. Oxford University Press, Oxford.