Introduction to Numerical Relativity

Miguel Alcubierre
Instituto de Ciencias Nucleares
Universidad Nacional Autónoma de México
(malcubi@nucleares.unam.mx)

May 2, 2006

Contents

Abstract

General relativity is a highly successful theory. It not only has radically modified our understanding of gravity, space and time, but it also possesses an enormous predictive power. To this date, it has passed with extraordinary precision all the experimental and observational tests that it has been subjected too. Among its more important results are the predictions of exotic objects such as neutron stars and black holes, and the cosmological model of the Big Bang. Also, general relativity has predicted the existence of gravitational waves, which might very well be detected directly for the first time before this decade is out.

General relativity, for all its conceptual simplicity and elegance, turns out to be in practice a highly complex theory. The Einstein field equations are a system of ten coupled and non-linear partial differential equations in four dimensions. Written in fully expanded form in a general coordinate system they have thousands of terms. Because of this complexity, exact solutions of the Einstein equations are only known in cases with high symmetry, either in space or in time: solutions with spherical or axial symmetry, static or stationary solutions, homogeneous and/or isotropic solutions, etc. If one is interested in studying systems with astrophysical relevance, which involve strong and dynamical gravitational fields with little or no symmetry, it is simply impossible to solve the field equations exactly. The need to study these type of systems has given birth to the field of numerical relativity, which tries to solve the Einstein field equations using numerical techniques and complex computational codes.

Numerical relativity appeared as an independent field of research in the mid 1960's with the pioneering efforts of Hahn and Lindquist [32], but it wasn't until the late 1970's when the first truly successful simulations where carried out by Smarr and Eppley [54, 24, 25] in the context of the head-on collision of two black holes. At that time, however, the power of the available computers was very modest, and the simulations that could be performed where limited to either spherical symmetry or very low resolution axial symmetry. This situation has changed, and during the decades of the eighties and nineties a true revolution has taken place in numerical relativity. Researchers have studied ever more complex problems in many different aspects of general relativity, from the simulation of rotating stars and black holes, to the study of topological defects, gravitational collapse, and the collisions of compact objects. Maybe the most influential result coming from numerical relativity has been the discovery by M. Choptuik of critical phenomena in gravitational collapse [20, 30]. A summary of the history of numerical relativity and its more recent developments can be found in [37].

Numerical relativity has now reached a state of maturity. The appearance of powerful super-computers, together with an increased understanding of the underlying theoretical

issues, and the development of robust numerical techniques, has finally allowed the simulation of fully three-dimensional systems with strong and highly dynamic gravitational fields. And all this activity is happening at precisely the right time, as a new generation of advanced gravitational wave detectors is finally coming on-line [39, 62, 29, 59]. The expected gravitational wave signals, however, are so weak that even with the amazing sensitivity of the new detectors it will be necessary to extract them from the background noise. As it is much easier to extract a signal if one knows what to look for, numerical relativity has become badly needed in order to provide the detectors with precise predictions of the type of gravitational waves produced by the most common astrophysical sources. We are living a truly exciting time in the development of this field.

These notes present an introduction to the main ideas behind the field of numerical relativity. The notes are organized as follows: In section 1 I present a brief introduction to general relativity. Section 2.2 presents the Cauchy problem of general relativity and introduces the 3+1 formalism, which is the most commonly used in numerical relativity (though not the only one). In section 3 I present alternative formulations of the 3+1 evolution equations, and I discuss the important concept of hyperbolicity. Section 4 presents an introduction to the numerical techniques of finite differencing. Finally, section 5 presents applications of the techniques of numerical relativity to some simple systems. In particular, I consider 1+1 relativity and the case of spherical symmetry.

1 General relativity

1.1 Introduction

The theory of general relativity, postulated by Albert Einstein at the end of 1915 [22, 23], is the modern theory of gravitation. According to this theory, gravity is not a force as it used to be considered in Newtonian physics, but rather a manifestation of the "curvature" of space-time. A massive object produces a distortion in the geometry of space-time around it, and in turn this distortion controls (or alters) the movement of physical objects. In the words of John A. Wheeler: Matter tells space-time how to curve, and space-time tells matter how to move [64].

When Einstein introduced special relativity in 1905, he was already fully aware that Newton's theory of gravity would have to be modified. The main reason for this was that Newton's theory implies that the gravitational interaction was transmitted between different bodies at infinite speed, in clear contradiction with one of the fundamental results of special relativity: no physical interaction can travel faster than the speed of light. It is interesting to note that Newton himself was never happy with the existence of this "action at a distance", but he considered that it was a necessary hypothesis to be used until a more adequate explanation of the nature of gravity was found. In the years from 1905 to 1915, Einstein focused his efforts in finding such an explanation.

The basic ideas that guided Einstein in his quest towards general relativity where the "principle of general covariance", which says that the laws of physics must take the same form for all observers, the "principle of equivalence" that says that all objects fall with the same acceleration in a gravitational field regardless of their mass, and "Mach's principle", formulated by Ernst Mach at the end of the XIX century, which states that the local inertial properties of physical objects must be determined by the total distribution of matter in the Universe. The principle of general covariance lead Einstein to ask for the equations of physics to be written in tensor form, the principle of equivalence lead him to the conclusion that the natural way to describe gravity was identifying it with the geometry of space-time, and Mach's principle lead him to the idea that such geometry should be fixed by the distribution of mass and energy.

In the following sections we will see how one can express these ideas in mathematical language. Before going further I will mention some basic conventions followed in these notes. I will use throughout the so-called "geometric units", in which the speed of light c and Newton's gravitational constant G are taken to be equal to one. When working in such units, distance, time, mass and energy all have the same dimensions (of length). Conventional (SI) units can always be recovered by multiplying a given quantity with as many powers of c and G as are needed in each case in order to get back to their correct

SI dimensions (for example, making the substitutions $t \to ct$ and $M \to GM/c^2$).

There are many excellent introductory books to general relativity. The presentation given here is based mainly on references [42, 50, 63].

1.2 Metric and geodesics

As we have already mentioned, the equivalence principle lead Einstein to the idea that gravity could be identified with the curvature of space-time. Mathematically this means that the gravitational theory should be a "metric theory", in which gravity manifests itself exclusively through the distortion in the geometry of space-time.

Let us consider space-time to be a four-dimensional (three space dimensions and one time dimension) continuous space, or in mathematical terms, a four-dimensional differentiable manifold. Let now x^α be the coordinates of a given "event" in this space-time, where the index α takes the values $\{0, 1, 2, 3\}$: four numbers that indicate the moment in time and the position in space of this event (in these notes we will take the 0 component to refer to the time coordinate, and the components $\{1, 2, 3\}$ as those that refer to space, also we will use the convention that greek indices take values from 0 to 3, while latin indices take only values from 1 to 3).

Between two infinitesimally close events with coordinates x^α and $x^\alpha + dx^\alpha$, it is possible to define an invariant "distance" in the following way:

$$ds^2 = \sum_{\alpha,\beta=1}^{4} g_{\alpha\beta}\, dx^\alpha dx^\beta \equiv g_{\alpha\beta}\, dx^\alpha dx^\beta \,, \tag{1.1}$$

where $g_{\alpha\beta}$ is known as the "metric tensor", or simply the "metric", and where the last equality defines the "Einstein summation convention": indices that appear twice in an expression are understood to be summed over all their possible values. The fact that this "distance" is invariant means that it is absolute in a physical sense, all observers will find the same value for this quantity, regardless of their respective motions and the coordinate systems they are using to measure it. At each point of space-time, the components of the metric tensor form a symmetric non-singular 4×4 matrix, whose eigenvalues have signs $(-, +, +, +)$, that is, one negative eigenvalue associated with the time dimension, and three positive eigenvalues associated with the spatial dimensions. In special relativity, the metric tensor reduces to the so-called Minkowski metric:

$$ds^2 = -dt^2 + dx^2 + dy^2 + dz^2 \equiv \eta_{\alpha\beta}\, dx^\alpha dx^\beta \,, \tag{1.2}$$

which corresponds to a flat space-time. In general relativity, on the other hand, the metric tensor changes from point to point.

The Lorentz transformations imply that the interval ds^2 has the same value as measured by any observer. This is a direct consequence of the postulate (or better, the empirical fact) of the invariance of the speed of light. Note that due to the presence of a negative eigenvalue, the invariant distance is not positive definite. In fact, from the metric one can distinguish events related to each other in three different ways:

$$ds^2 > 0 \qquad \text{spacelike separation}\,, \tag{1.3}$$

$$ds^2 < 0 \qquad \text{timelike separation}\,, \tag{1.4}$$

$$ds^2 = 0 \qquad \text{null separation}\,. \tag{1.5}$$

Spacelike separations correspond to events separated in such a way that one would have to move faster than light to reach one from the other (they are separated mainly in "space"), timelike separations correspond to events that can be reached one from the other moving slower than light (they are separated mainly in "time"), and null separations correspond to events that can be reached one from the other moving precisely at the speed of light. All material objects move following timelike trajectories, and light moves along null trajectories. Null trajectories also define the "light-cone" (see Figure 1), which indicates which events can have a physical influence with each other and therefore defines causality.

The metric tensor can also be used to define an invariant scalar product between two four-dimensional vectors ("4-vectors"). Let $\vec{u}$ and $\vec{v}$ be two 4-vectors with components u^α and v^α respectively, their scalar product is then defined as:

$$\vec{u} \cdot \vec{v} := g_{\alpha\beta} u^\alpha v^\beta \,, \tag{1.6}$$

where we are again using the summation convention. In particular, the square-magnitude of a 4-vector can be defined as

$$u^2 := \vec{u} \cdot \vec{u} = g_{\alpha\beta} u^\alpha u^\beta \,. \tag{1.7}$$

Notice that again this is not positive definite, so we usually don't take its square root. We can then classify vectors just as intervals into spacelike, timelike and null. Is is now clear that the interval ds^2 is nothing more than the square-magnitude of the (infinitesimal) displacement vector itself.

In special relativity, in the absence of external forces objects move in straight lines through space-time (corresponding to uniform-rectilinear motion). A straight line is the trajectory of extremal length according to the Minkowski metric (not necessarily minimal length because of the minus sign in the coefficient of dt^2). Einstein's idea was precisely to think that in the presence of a gravitational field objects would still follow straight

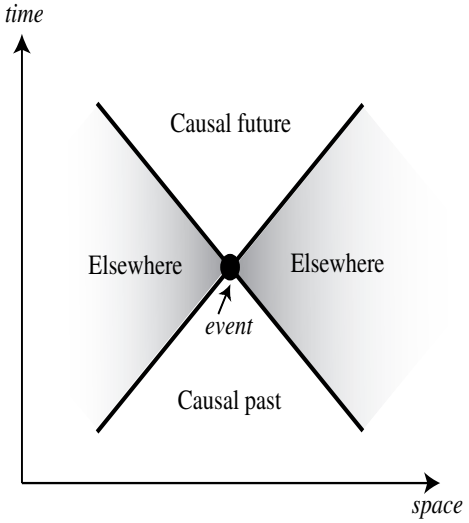

Figure 1: The light cone of a given event defines its causal relationship with other events, and divides space into three regions: the causal past (those events that can influence the event under consideration), the causal future (those events that can be influenced by the event under consideration), and "elsewhere" (those events with which there is no causal contact).

lines *locally*, that is extremal trajectories, but now in a different metric. This extremal trajectories are called "geodesics". In this way gravity is not seen as external force, but only as a distortion of the geometry, in this distorted geometry objects simply follow the straightest paths possible, that is, a geodesic.

It is usual to parametrize a geodesic using the so-called "proper time" $d\tau^2 := -ds^2$, which corresponds to the time experienced by the object itself (the time measured by an ideal clock attached to the object). Note that proper time does not have to be equal to dt, since t is only a coordinate, that is, an arbitrary label used to identify events. The geodesic equation has the general form:

$$\frac{d^2x^\alpha}{d\tau^2} + \Gamma^\alpha_{\beta\gamma}\frac{dx^\beta}{d\tau}\frac{dx^\gamma}{d\tau} = 0\,, \tag{1.8}$$

where the quantities $\Gamma^\alpha_{\beta\gamma}$ are known as the "Christoffel symbols" and are defined in terms of the metric as

$$\Gamma^\alpha_{\beta\gamma} := \frac{g^{\alpha\mu}}{2}\left[\frac{\partial g_{\beta\mu}}{\partial x^\gamma} + \frac{\partial g_{\gamma\mu}}{\partial x^\beta} - \frac{\partial g_{\beta\gamma}}{\partial x^\mu}\right]. \tag{1.9}$$

In the previous equation we have introduced the coefficients $g^{\alpha\beta}$ which are defined as the components of the inverse matrix to $g_{\alpha\beta}$: $g^{\alpha\mu}g_{\beta\mu} = \delta^{\alpha}_{\beta}$.

The geodesic equation can also be written as

$$\frac{du^{\alpha}}{d\tau} + \Gamma^{\alpha}_{\beta\gamma} u^{\beta} u^{\gamma} = 0 \,, \tag{1.10}$$

where we have defined the 4-vector $u^{\alpha} := dx^{\alpha}/d\tau$. This vector is know as the "4-velocity", and is the relativistic generalization of the ordinary concept of velocity. It is important to notice that:

$$u^2 = \vec{u} \cdot \vec{u} = g_{\alpha\beta} \frac{dx^{\alpha}}{d\tau} \frac{dx^{\beta}}{d\tau} = \frac{ds^2}{d\tau^2} = -1 \,, \tag{1.11}$$

that is, the 4-velocity is a unit timelike vector.

Given a gravitational field, that is, given the metric of space-time, the geodesic equation (1.8) given us the trajectories of objects: space-time tells objects how to move.

1.3 Curvature

As we have seen, the metric of space-time allows to know the trajectory of objects in the gravitational field. Still, the metric tensor is not in itself the most convenient way of describing the presence of a gravitational field. In order to see this, it is enough to consider that even on a flat space, one can always change the form of the metric tensor with a simple coordinate transformation. For example, the metric of a flat three-dimensional space in Cartesian coordinates $\{x, y, z\}$ is given simply by the Pythagoras rule

$$dl^2 = dx^2 + dy^2 + dz^2 \,, \tag{1.12}$$

while the metric of the same flat space in spherical coordinates $\{r, \theta, \phi\}$ turns out to be

$$dl^2 = dr^2 + r^2 d\theta^2 + r^2 \sin^2\theta d\phi^2 \,, \tag{1.13}$$

which can be easily derived from the coordinate transformation

$$x = r\, \sin\theta\, \cos\phi \,, \qquad y = r\, \sin\theta\, \sin\phi \,, \qquad z = r\, \cos\theta \,. \tag{1.14}$$

that implies

$$dx = dr\, \sin\theta\, \cos\phi - r\, \sin\theta\, \sin\phi\, d\phi + r\, \cos\theta\, \cos\phi\, d\theta \,, \tag{1.15}$$

$$dy = dr\, \sin\theta\, \sin\phi + r\, \sin\theta\, \cos\phi\, d\phi + r\, \cos\theta\, \sin\phi\, d\theta \,, \tag{1.16}$$

$$dz = dr\, \cos\theta - r\, \sin\theta\, d\theta \,. \tag{1.17}$$

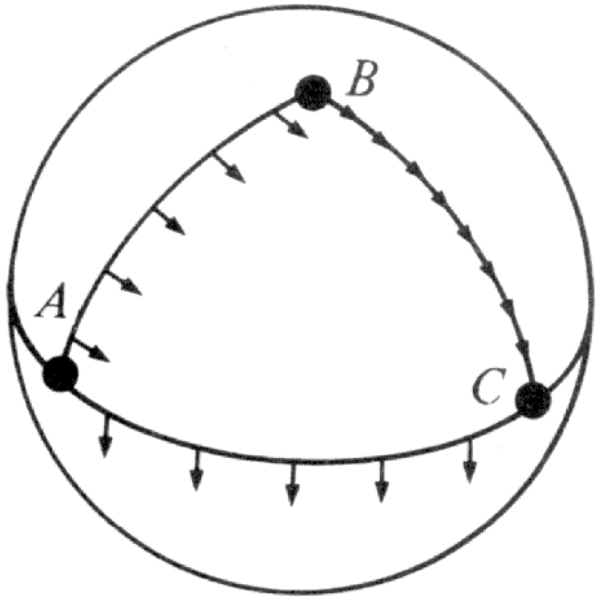

Figure 2: Parallel transport on the Earth surface.

From the metric (1.13) it is possible to calculate the Christoffel symbols for spherical coordinates (they are not all equal to zero!), and verify that the equation for a straight line (geodesic) in these coordinates is no longer trivial. By making even more complex coordinate transformations it is possible to end up with very complicated metrics. We must then find a way to distinguish with certainty between spaces that are flat, and those that are not.

The way to do this is through the "Riemann curvature tensor". This tensor measures the change of a vector as it is transported around a closed circuit while keeping it always parallel to itself ("parallel transport"). On a flat space, the vector does not change when this is done, while on a curved space it does change. This can be clearly seen if one thinks about moving a vector on the surface of the Earth. Assume that we start somewhere on the equator (point A) with a vector pointing east. We then move north up to the north pole (point B) following a meridian, then we come back south along another meridian that is ninety degrees to the east of the first one until we get back to the equator (point C). Finally, we move back to our staring point following the equator itself. If we do this we will find that our vector now points south! (see figure 2).

In these notes we will not derive the Riemann tensor from first principles, and we will

limit ourselves to giving its expression:

$$R^{\sigma}{}_{\mu\nu\rho} := \partial_\nu \Gamma^\sigma_{\mu\rho} - \partial_\mu \Gamma^\sigma_{\nu\rho} + \left(\Gamma^\alpha_{\mu\rho} \Gamma^\sigma_{\alpha\nu} - \Gamma^\alpha_{\nu\rho} \Gamma^\sigma_{\alpha\mu} \right) , \tag{1.18}$$

where ∂_μ is a shorthand for $\partial/\partial x^\mu$, and where $\Gamma^\alpha_{\mu\nu}$ are the Christoffel symbols defined above. Note that the Riemann tensor has 4 indices, that is $4 \times 4 \times 4 \times 4 = 256$ components. However, its definition implies that it has many symmetries, and in the end it only has 20 independent components (in 4 dimensions). It is possible to show that the Riemann tensor vanishes if and only if the corresponding space is flat.

From the Riemann tensor one can define the "Ricci tensor" by summing (contracting) over two of its free indices:

$$R_{\mu\nu} := \sum_\lambda R^\lambda{}_{\mu\lambda\nu} = R^\lambda{}_{\mu\lambda\nu} \, . \tag{1.19}$$

The Ricci tensor is symmetric in its two indices, and hence has only 10 independent components. Notice that having the the Ricci tensor vanish *does not* mean that spacetime is flat, as the remaining 10 components of Riemann can still be different from zero.

It is important to notice that in some cases we have written indices as super-indices (up), and other times as sub-indices (down). This is not an error, tensors with up or down indices are not the same, but they are related. The rule is as follows: the indices of tensors are raised or lowered by contracting with the metric tensor $g_{\mu\nu}$ or its inverse $g^{\mu\nu}$. For example:

$$\begin{aligned} v_\mu &= g_{\mu\nu} \, v^\nu \, , \\ v^\mu &= g^{\mu\nu} \, v_\nu \, , \\ T^{\mu\nu} &= g^{\mu\alpha} g^{\nu\beta} \, T_{\alpha\beta} \, , \\ R_{\sigma\mu\nu\rho} &= g_{\sigma\lambda} R^\lambda{}_{\mu\nu\rho} \, . \end{aligned}$$

1.4 Coordinate basis and covariant derivatives

When one considers the change in a vector field (objects with one index) or a tensor field (objects with more than one index) as we move in space-time, one must take into account not only that the components of those objects might change, but also that the basis in which those components are calculated might change from one point to another. One must remember that when we consider the components of a geometric object (vector or tensor), those components are always given with respect to a very specific basis.

In flat three-dimensional space in Cartesian coordinates, the basis is usually the "canonical basis": $\vec{i} = (1,0,0), \vec{j} = (0,1,0), \vec{k} = (0,0,1)$. In a curved space, or a flat

space with curvilinear coordinates, it is necessary first to choose a basis before one can talk about the components of a geometric object. A common basis choice (though certainly not the only possibility) is the so-called "coordinate basis", in which one takes as the basis those vectors that have a component equal to 1 along the direction of a given coordinate, and have all other components equal to zero.

For example, for a flat space in spherical coordinates $\{r, \theta, \phi\}$ the coordinate basis is simply:

$$\vec{e}_r = (1, 0, 0)\,, \qquad \vec{e}_\theta = (0, 1, 0)\,, \qquad \vec{e}_\phi = (0, 0, 1)\,. \tag{1.20}$$

When written in Cartesian coordinates, these vectors turn out to be:

$$\vec{e}_r = (\sin\theta\cos\phi, \sin\theta\sin\phi, \cos\theta)\,, \tag{1.21}$$

$$\vec{e}_\theta = (r\cos\theta\cos\phi, r\cos\theta\sin\phi, -r\sin\phi)\,, \tag{1.22}$$

$$\vec{e}_\phi = (-r\sin\theta\sin\phi, r\sin\theta\cos\phi, 0)\,. \tag{1.23}$$

Note that not all these basis vectors are unitary, their square-magnitudes are:

$$|\vec{e}_r|^2 = (\sin\theta\cos\phi)^2 + (\sin\theta\sin\phi)^2 + (\cos\theta)^2 = 1 \tag{1.24}$$

$$|\vec{e}_\theta|^2 = (r\cos\theta\cos\phi)^2 + (r\cos\theta\sin\phi)^2 + (r\sin\phi)^2 = r^2\,, \tag{1.25}$$

$$|\vec{e}_\phi|^2 = (r\sin\theta\sin\phi)^+(r\sin\theta\cos\phi)^2 = r^2\sin^2\theta\,, \tag{1.26}$$

where in order to calculate those magnitudes we have simply summed the squares of their Cartesian components. As we have seen, the magnitude of a vector can also be calculated directly from the components in spherical coordinates using the metric tensor:

$$|\vec{v}|^2 = g_{\alpha\beta}\, v^\alpha v^\beta\,. \tag{1.27}$$

It is not difficult to see that if we use the above equation and the expression for the metric in spherical coordinates (equation (1.13)), we find the same square-magnitudes for the vectors of the spherical coordinate basis given above.

As the basis vectors have by definition components equal to 1 for the corresponding coordinate and zero for all others, one can find that the scalar product of the basis vectors is given by:

$$\vec{e}_\alpha \cdot \vec{e}_\beta = g_{\alpha\beta}\,. \tag{1.28}$$

That is, the components of the metric tensor are precisely the scalar product of the basis vectors.

It is important to notice that in the previous expressions v^α refers to the components of the vector $\vec{v}$ with respect to the coordinates x^α, while $\vec{e}_\alpha$ refers to the basis vector that points in the direction x^α. This implies that, in general:

$$\vec{v} = v^\alpha\, \vec{e}_\alpha\,, \tag{1.29}$$

equation that expresses the vector $\vec{v}$ as a linear combination of the basis vectors $\{\vec{e}_\alpha\}$ with coefficients v^α.

The coordinate basis, though not unique, is the most commonly used, and the expressions for the Riemann tensor given in the previous section are precisely those obtained in this basis.

Consider now the derivative of a vector with respect to a given coordinate:

$$\frac{\partial \vec{v}}{\partial x^\alpha} = \frac{\partial}{\partial x^\alpha}\left(v^\beta \vec{e}_\beta\right) = \frac{\partial v^\beta}{\partial x^\alpha}\,\vec{e}_\beta + v^\beta\,\frac{\partial \vec{e}_\beta}{\partial x^\alpha}\,. \tag{1.30}$$

This equation shows that the derivative of a vector is more that just the derivative of its components. One must also take into account the change in the basis vectors.

The derivative $\partial\vec{e}_\beta/\partial x^\alpha$ is in itself also a vector, which means that it can be expressed as a linear combination of the basis vectors. We introduce the symbols $\Gamma^\mu_{\alpha\beta}$ to denote the coefficients of such linear combination:

$$\frac{\partial \vec{e}_\beta}{\partial x^\alpha} = \Gamma^\mu_{\alpha\beta}\,\vec{e}_\mu\,. \tag{1.31}$$

The coefficients $\Gamma^\mu_{\alpha\beta}$ turn out to be precisely the Christoffel symbols defined above. Using these coefficients we find

$$\frac{\partial \vec{v}}{\partial x^\alpha} = \frac{\partial v^\beta}{\partial x^\alpha}\,\vec{e}_\beta + v^\beta\,\Gamma^\mu_{\alpha\beta}\,\vec{e}_\mu\,. \tag{1.32}$$

Changing now the name of the summed (dummy) indices, we can rewrite this as

$$\frac{\partial \vec{v}}{\partial x^\alpha} = \left(\frac{\partial v^\beta}{\partial x^\alpha} + v^\mu\,\Gamma^\beta_{\alpha\mu}\right)\vec{e}_\beta\,. \tag{1.33}$$

This equation gives us the components of the vector $\partial\vec{v}/\partial x^\alpha$. We now define the "covariant derivative" of the vector $\vec{v}$ as:

$$v^\alpha{}_{;\beta} \equiv \nabla_\beta\, v^\alpha := \frac{\partial v^\alpha}{\partial x^\beta} + v^\mu\,\Gamma^\alpha_{\beta\mu}\,. \tag{1.34}$$

Note that we have introduced two alternative notations for the covariant derivative: In one case we use a semicolon, and in the other the nabla symbol. Both notations are common and are used indistinctively.

The covariant derivative tells us how the components of a vector change as we move in a general space, and includes the change in the basis vectors themselves. Notice that the covariant derivative reduces to the partial derivative when the Christoffel symbols vanish,

which happens in flat space with Cartesian coordinates but not with spherical coordinates. However, it is always possible to find coordinates such that the Christoffel symbols vanish at a given point (but not at other points unless the space is flat). This is because every curved space is "locally flat": in an infinitesimal region around an arbitrary point the geometry can be approximated by the flat one (that is why flat maps of a city are usually very accurate, while flat maps of the whole planet introduce serious distortions).

The covariant derivative of a vector with the indices down v_α (a "covariant" vector or simply "co-vector") can be shown to be given by:

$$v_{\alpha;\beta} = \frac{\partial v_\alpha}{\partial x^\beta} - \Gamma^\mu_{\alpha\beta} v_\mu \,. \tag{1.35}$$

The same concept of covariant derivative can also be applied to tensors with many indices, the rule is to add a term with Christoffel symbols for each free index, with the adequate sign depending on whether the index is down or up. For example:

$$\begin{aligned} T^{\mu\nu}{}_{;\alpha} &= \partial_\alpha T^{\mu\nu} + \Gamma^\mu_{\alpha\beta} T^{\beta\nu} + \Gamma^\nu_{\alpha\beta} T^{\mu\beta} \,, \\ T_{\mu\nu;\alpha} &= \partial_\alpha T_{\mu\nu} - \Gamma^\beta_{\alpha\mu} T_{\beta\nu} - \Gamma^\beta_{\alpha\nu} T_{\mu\beta} \,, \\ T^\mu{}_{\nu;\alpha} &= \partial_\alpha T^\mu{}_\nu + \Gamma^\mu_{\alpha\beta} T^\beta{}_\nu - \Gamma^\beta_{\alpha\nu} T^\mu{}_\beta \,. \end{aligned}$$

Using this rule one can show that the covariant derivative of the metric tensor is in fact always zero:

$$g_{\mu\nu;\alpha} = 0 \,, \qquad g^{\mu\nu}{}_{;\alpha} = 0 \,, \tag{1.36}$$

which implies that the operation of raising and lowering indices commutes with the covariant derivative:

$$v^\mu{}_{;\alpha} = \left(g^{\mu\nu}\, v_\nu\right)_{;\alpha} = g^{\mu\nu}\, \left(v_{\nu;\alpha}\right) \,. \tag{1.37}$$

1.5 The Einstein field equations

The element we are still missing from Einstein's theory of gravity is the one that relates the geometry of space-time with the distribution of mass and energy. This final element is contained in the Einstein field equations, or simply Einstein's equations. These equations can be derived in several different ways, either looking for a consistent relativistic generalization of Newton's theory (the road followed by Einstein), or deriving them in a formal way from a variational principle staring with an adequate Lagrangian (the road followed by Hilbert). Here I will limit myself to writing the equations down. In their most compact form, the Einstein equations are:

$$G_{\mu\nu} = 8\pi T_{\mu\nu} \,, \tag{1.38}$$

where $G_{\mu\nu}$ is the so-called "Einstein tensor" that is related with the Ricci curvature tensor (see below), and $T_{\mu\nu}$ is the "stress-energy tensor" of matter. That is, the left hand side represents the geometry of space-time, and the right hand side the distribution of mass-energy. The factor 8π is simply an adequate normalization necessary to obtain the correct Newtonian limit. Note that these are in fact 10 equations, as the indices μ and ν take values from 0 to 3, that is $4 \times 4 = 16$, and the Einstein and stress-energy tensors are symmetric, reducing the number of independent equations to only 10.

The Einstein equations that we have just introduced could not appear to be more simple. This simplicity, however, is only apparent since each term is in fact a short-hand for considerably more complex objects. Written in their most general way, in an arbitrary coordinate system, and with all the terms expanded out, the Einstein equations become a system of 10 coupled non-linear partial differential equations with thousands of terms.

Let us consider each term in Einstein's equations separately. The Einstein tensor is defined in terms of the Ricci tensor as:

$$G_{\mu\nu} := R_{\mu\nu} - \frac{1}{2}\, g_{\mu\nu} R\,, \tag{1.39}$$

with $R := g^{\mu\nu}\, R_{\mu\nu}$ the trace of the Ricci tensor, also know as the "scalar curvature".

The other side of the Einstein equations, the stress-energy tensor, describes the energy density, momentum density and momentum flux of matter ($i, j = 1, 2, 3$):

$$T^{00} = \text{energy density}\,, \tag{1.40}$$
$$T^{0i} = \text{momentum density}\,, \tag{1.41}$$
$$T^{ij} = \text{flux of } i \text{ momentum in direction } j. \tag{1.42}$$

For example, for a perfect fluid with no pressure (known as "dust" in relativity) and in flat space we have:

$$T^{00} = \rho/(1 - v^2)\,, \tag{1.43}$$
$$T^{0i} = \rho\, v^i/(1 - v^2)\,, \tag{1.44}$$
$$T^{ij} = \rho\, v^i v^j/(1 - v^2)\,, \tag{1.45}$$

where ρ is the energy density measured in the rest frame of the fluid, v^i is the velocity field of the fluid, and $v^2 = v_x^2 + v_y^2 + v_z^2$.

It is important to mention that in relativity, all fields expect gravity itself are considered as a type of matter. For example, an electromagnetic field has an associated stress-energy tensor that acts as the source of gravity, i.e. electric and magnetic fields produce gravity.

Before finishing our discussion of Einstein's equations a final comment is in order. In the case of vacuum the stress-energy tensor vanishes, and Einstein's equations reduce to:

$$G_{\mu\nu} = 0\,, \tag{1.46}$$

or equivalently

$$R_{\mu\nu} = 0\,. \tag{1.47}$$

Note that, as mentioned before, the fact that the Ricci tensor vanishes does not imply that space-time is flat. This is as it should be, since we know that the gravitational field of an object extends beyond the object itself, which means that the curvature of spacetime in the empty space around a massive object can not be zero. The Einstein equations in vacuum have another important consequence, they describe the way in which the gravitational field propagates in empty space and, in an analogous way to electromagnetism, they predict the existence of gravitational waves: perturbations in the gravitational field that propagate at the speed of light. The prediction of the existence of gravitational waves tells us that in Einstein's theory gravitational interactions do not propagate at an infinite speed, but rather at the speed of light.

1.6 Bianchi identities and conservation laws

Starting from the definition of the Riemann tensor it is possible to show that this tensor has the following property:

$$R_{\alpha\beta\mu\nu;\lambda} + R_{\alpha\beta\lambda\mu;\nu} + R_{\alpha\beta\nu\lambda;\mu} = 0\,. \tag{1.48}$$

These relations are known as the "Bianchi identities", and play a very important role in general relativity. One of their most important consequences is the fact that the covariant divergence of the Einstein tensor vanishes:

$$G^{\mu\nu}{}_{;\nu} = 0\,, \tag{1.49}$$

a result that comes from contracting the Bianchi identities twice.

The Einstein tensor can be shown to be the unique combination that can be obtained from the Ricci tensor with vanishing divergence, and it is precisely because of this fact that the field equations involve the the Einstein tensor and not the Ricci tensor directly. If we now use the field equations we see that the previous property immediately implies:

$$T^{\mu\nu}{}_{;\nu} = 0\,. \tag{1.50}$$

This last equation (4 equations in fact) is of fundamental importance, since it represents the local laws of conservation of energy and momentum, and guarantees that the loss of energy an momentum in a given region of space is compensated by the flux of energy and momentum out of that region. In the case of a fluid, for example, the $\mu = 0$ component of this equation transforms into the well known "continuity equation". We then see that the Einstein field equations automatically imply the laws of conservation of energy and momentum.

2 The formalism of numerical relativity

2.1 Introduction

There exist several different formalisms used in numerical relativity. In each case, what needs to be done is to separate the Einstein field equations in a way that allows us to give certain initial data, and from there obtain the subsequent evolution of the gravitational field. The different formalisms differ in the specific way in which these separation is carried out. In these notes I will concentrate on the "3+1 formalism", where one splits spacetime into three-dimensional space on the one hand, and time on the other. The 3+1 formalism is the most commonly used in numerical relativity, but it is certainly not the only one. Alternative formalisms are known as the "characteristic formalism" [65], where spacetime is separated into light-cones emanating from a central timelike world-tube, and the "conformal formalism" [27], where a conformal transformation is used that brings the boundary of spacetime, that is "asymptotic null infinity", to a finite distance in coordinate space. The different formalisms have advantages and disadvantages depending on the specific physical system under study.

In the following sections I will introduce the 3+1 formalism of general relativity. The discussion found here can be seen in more detail in references [42, 67].

2.2 The 3+1 formalism

In order to study the evolution in time of any physical system, the first thing that needs to be done is to formulate such an evolution as an "initial value" or "Cauchy" problem: Given adequate initial (and boundary) conditions, the fundamental equations must predict the future (or past) evolution of the system.

When trying to write Einstein's equations as a Cauchy problem, we immediately encounter a stumbling block: the field equations are written in such a way that space and time are treated in an equal footing. This "covariance" is very important (and quite elegant) from a theoretical point of view, but is does not allow one to think clearly about the evolution of the gravitational field in time. Therefore, the first thing we need to do in order to rewrite Einstein's equations as a Cauchy problem is to split the roles of space and time in a clear way. The formulation of general relativity that results from this split is known as the "3+1 formalism".

Let us start by considering a spacetime with metric $g_{\alpha\beta}$. If such a spacetime can be completely foliated (that is, sliced into three-dimensional cuts) in such a way that each three-dimensional (3D) slice is spacelike, then we say that the spacetime is "globally hyperbolic". A globally hyperbolic spacetime has no closed timelike curves, which means

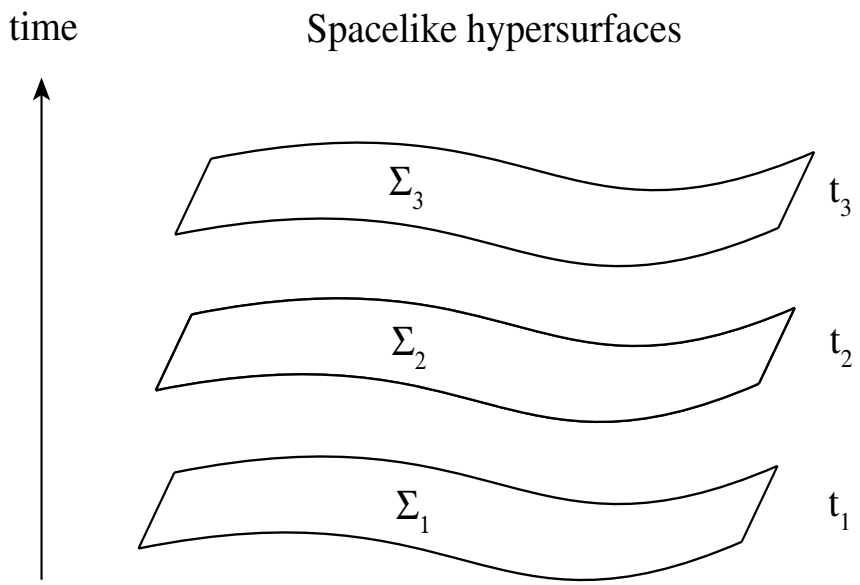

Figure 3: Foliation of spacetime into three-dimensional spacelike hypersurfaces.

that it does not allow time travel to the past. Not all possible spacetimes have this property, but the 3+1 formalism assumes that all physically reasonable spacetimes are of this type.

Once we have a globally hyperbolic spacetime, by definition we can slice it into a series of 3D spacelike hypersurfaces (see Figure 3). This foliation is, of course, not unique. We can parametrize the foliation with a parameter t that identifies each of the slices, t can then be considered as a "universal time function" (but one must be careful, t does not necessarily coincides with the proper time of any observer). Consider now a specific foliation, and take two adjacent hypersurfaces Σ_t and Σ_{t+dt}. The geometry of the region of spacetime contained between these two hypersurfaces can be determined from the following three basic ingredients (see Figure 4):

- The three-dimensional metric γ_{ij} $(i,j = 1,2,3)$ that measures proper distances within the hypersurface itself:

$$dl^2 = \gamma_{ij}\, dx^i dx^j \,. \tag{2.1}$$

- The lapse of proper time between both hypersurfaces measured by those observers moving along the direction normal to the hypersurfaces (the Eulerian observers):

$$d\tau = \alpha(t, x^i)\, dt \,. \tag{2.2}$$

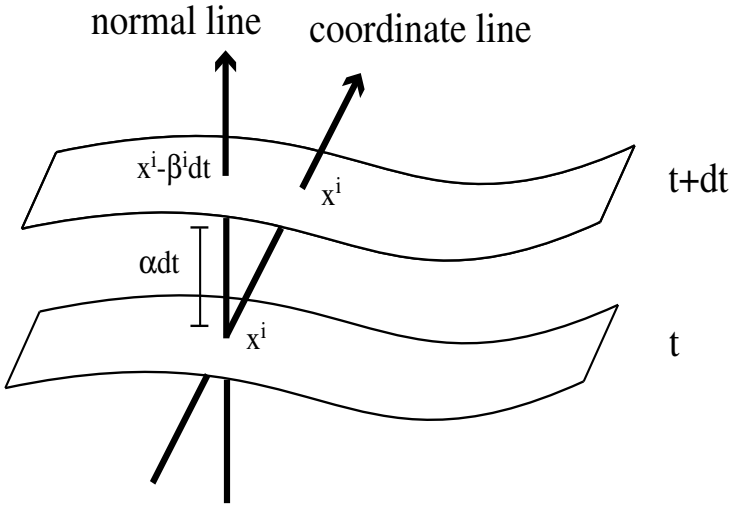

Figure 4: Two adjacent spacelike hypersurfaces. The figure shows the definitions of the lapse function α and the shift vector β^i.

Here α is known as the "lapse function".

- The relative velocity β^i between the Eulerian observers and the lines that correspond to constant spatial coordinates:

$$x^i_{t+dt} = x^i_t - \beta^i(t, x^j)\, dt\,, \qquad \text{for Eulerian observers} \tag{2.3}$$

The 3-vector β^i is known as the "shift vector" (notice the minus sign in the above definition).

The way in which spacetime is foliated is clearly not unique, and also the way in which the spatial coordinates propagate from one hypersurface to the next is not unique. This means that both the lapse function α and the shift vector β^i are functions that can be freely specified. These functions determine our choice of coordinate system and are known as "gauge functions".

In terms of the functions $\{\alpha, \beta^i, \gamma_{ij}\}$, the metric of spacetime takes the following form:

$$ds^2 = \left(-\alpha^2 + \beta_i\beta^i\right) dt^2 + 2\,\beta_i\, dt dx^i + \gamma_{ij}\, dx^i dx^j\,, \tag{2.4}$$

where we have defined $\beta_i := \gamma_{ij}\,\beta^j$ (from here on we will assume that indices of purely spatial tensors are raised and lowered with the spatial metric γ_{ij}).

More explicitly we have:

$$g_{\mu\nu} = \begin{pmatrix} -\alpha^2 + \beta_k\beta^k & \beta_i \\ \beta_j & \gamma_{ij} \end{pmatrix}, \tag{2.5}$$

$$g^{\mu\nu} = \begin{pmatrix} -1/\alpha^2 & \beta^i/\alpha^2 \\ \beta^j/\alpha^2 & \gamma^{ij} - \beta^i\beta^j/\alpha^2 \end{pmatrix}. \tag{2.6}$$

In the same way, it is not difficult to show that the components of the unit normal vector n^μ to the spatial hypersurfaces are given by:

$$n^\mu = \left(1/\alpha, -\beta^i/\alpha\right), \qquad n_\mu = (-\alpha, 0), \qquad n^\mu n_\mu = -1\,. \tag{2.7}$$

Note that this unit normal vector corresponds by definition to the 4-velocity of the Eulerian observers.

2.3 Extrinsic curvature

When talking about the spatial hypersurfaces that form the foliation of spacetime, one needs to distinguish between the "intrinsic" curvature of those hypersurfaces coming from their internal geometry, and the "extrinsic" curvature coming from the way in which those hypersurfaces are immersed in four-dimensional spacetime.

The intrinsic curvature is given by the three-dimensional Riemann tensor defined in terms of the 3-metric γ_{ij}. The extrinsic curvature, on the other hand, is defined in terms of what happens to the normal vector n^α as it is parallel-transported from one point in the hypersurface to another. In general, one will find that as we parallel-transport this vector to a nearby point, the new vector will not be normal the hypersurface anymore. The "extrinsic curvature tensor" $K_{\alpha\beta}$ is a measure of the change of the normal vector under parallel transport.

In order to define the extrinsic curvature, we first need to introduce the "projection" operator P^α_β onto the spatial hypersurfaces:

$$P^\alpha_\beta := \delta^\alpha_\beta + n^\alpha n_\beta\,, \tag{2.8}$$

where n^α is the unit normal vector. It is easy to show that for any vector v^α we have:

$$\left(P^\alpha_\beta\, v^\beta\right) n_\alpha = 0\,, \tag{2.9}$$

(it is enough to remember that $n^\alpha n_\alpha = -1$). This means that every vector projected onto the hypersurface will be orthogonal to n^α. It is also possible to project tensors of arbitrary rank, one simply has to contract all free indices with the projection operator:

$$P\,T_{\alpha\beta} \equiv P^\mu_\alpha P^\nu_\beta\, T_{\mu\nu}\,. \tag{2.10}$$

Using the projection operator, the extrinsic curvature tensor is defined as:

$$K_{\alpha\beta} := -P\,\nabla_\alpha n_\beta = -\left(\nabla_\alpha n_\beta + n_\alpha n^\mu \nabla_\mu n_\beta\right)\,, \tag{2.11}$$

where ∇_α is the covariant derivative that we introduced before given in terms of the Christoffel symbols as:

$$\nabla_\alpha n_\beta = \partial_\alpha n_\beta - \Gamma^\lambda_{\alpha\beta}\, n_\lambda\,. \tag{2.12}$$

As defined above, $K_{\alpha\beta}$ can in fact be shown to be symmetric: $K_{\alpha\beta} = K_{\beta\alpha}$. It is also clearly a purely spatial tensor, that is, $n^\alpha K_{\alpha\beta} = n^\alpha K_{\beta\alpha} = 0$ (notice that this means that $K^{00} = K^{0i} = 0$, but K_{00} and K_{0i} are in general *not* zero). Because of this, from now on we will only consider its spatial components K_{ij}.[1]

Substituting now the explicit form of the normal vector (2.7) in the definition of the extrinsic curvature, one can show that K_{ij} is given in terms of the spatial metric as

$$K_{ij} = \frac{1}{2\alpha}\left[-\partial_t \gamma_{ij} + D_i\beta_j + D_j\beta_i\right]\,, \tag{2.13}$$

where D_i represents the three-dimensional covariant derivative, that is, the one associated with the 3-metric γ_{ij}. The previous equation can be rewritten as:

$$\partial_t \gamma_{ij} = -2\alpha K_{ij} + D_i\beta_j + D_j\beta_i\,. \tag{2.14}$$

We then see that the extrinsic curvature K_{ij} gives us the change in time of the spatial metric. This association can be done in a more precise way by using the concept of the

[1] A couple of remarks are in order regarding the definition of $K_{\alpha\beta}$. First, notice that the projection of $\nabla_\alpha n_\beta$ is important in order to make $K_{\alpha\beta}$ purely spatial. One might argue that because n^α is unitary its gradient is always orthogonal to it. This is true in the sense that $n^\alpha \nabla_\beta\, n_\alpha = 0$, but $\nabla_\alpha n_\beta$ is in general not symmetric and $n^\alpha \nabla_\alpha\, n_\beta \neq 0$ unless the normal lines are geodesic (which is not always the case). Consider now the symmetry of $K_{\alpha\beta}$, as just mentioned $\nabla_\alpha n_\beta$ is not in general symmetric even though n^α is hypersurface orthogonal. The reason for this is that n^α is a unitary vector and thus in general is not equal to the gradient of the time function t, except when the lapse is equal to 1. However, once we project onto the hypersurface, it turns out that $P\nabla_\alpha n_\beta$ is indeed symmetric (the non-symmetry of $\nabla_\alpha n_\beta$ has to do with the lapse, which is not intrinsic to the hypersurface).

"Lie derivative", which for scalars, vectors and co-vectors is defined as

$$\begin{aligned} \pounds_{\vec{u}}\,\phi &= u^\mu \partial_\mu \phi \,, & (2.15)\\ \pounds_{\vec{u}}\,v^\alpha &= u^\mu \partial_\mu v^\alpha - v^\mu \partial_\mu u^\alpha \,, & (2.16)\\ \pounds_{\vec{u}}\,\omega_\alpha &= u^\mu \partial_\mu \omega_\alpha + \omega_\mu \partial_\alpha u^\mu \,. & (2.17) \end{aligned}$$

The Lie derivative is a concept of derivative different from covariant differentiation, independent of the metric and the Christoffel symbols, that is defined in terms of a congruence of lines with tangent vector $\vec{u}$. The congruence defines a mapping of basis vectors in the following way: 1) at a given point, consider lines that have as tangent vectors the basis vectors, 2) now drag those lines along the congruence by moving the same "distance", measured by the change in the value of the parameter defining the congruence, 3) the new basis vectors are the tangent vectors to the dragged curves. Generally, the dragged basis vectors will not coincide with the basis vectors at the new point. The Lie derivative of a tensor tells us the difference between the value of the tensor applied to the dragged basis vectors and the value of the tensor applied to the original basis vectors at that point.

The Lie derivative can also be understood in a different way. If we construct a coordinate system such that the coordinate lines associated with a give coordinate coincide with our congruence, then the Lie derivative is simply the partial derivative with respect to that coordinate. The Lie derivative is the coordinate-independent way of defining a partial derivative.

For a tensor of rank two with indices down, the Lie derivative turns out to be:

$$\pounds_{\vec{u}}\,T_{\alpha\beta} = u^\mu \partial_\mu T_{\alpha\beta} + T_{\alpha\mu}\partial_\beta u^\mu + T_{\mu\beta}\partial_\alpha u^\mu \,. \qquad (2.18)$$

In particular, for the metric tensor we have:

$$\pounds_{\vec{u}}\,g_{\alpha\beta} = u^\mu \partial_\mu g_{\alpha\beta} + g_{\alpha\mu}\partial_\beta u^\mu + g_{\mu\beta}\partial_\alpha u^\mu \,. \qquad (2.19)$$

On the other hand:

$$\begin{aligned} \nabla_\alpha u_\beta + \nabla_\beta u_\alpha &= \partial_\alpha u_\beta + \partial_\beta u_\alpha - 2\Gamma^\mu_{\alpha\beta} u_\mu \\ &= \partial_\alpha u_\beta + \partial_\beta u_\alpha + u^\mu \partial_\mu g_{\alpha\beta} - u^\mu \partial_\alpha g_{\beta\mu} - u^\mu \partial_\beta g_{\alpha\mu} \\ &= u^\mu \partial_\mu g_{\alpha\beta} + g_{\beta\mu}\partial_\alpha u^\mu + g_{\alpha\mu}\partial_\beta u^\mu \,, \qquad (2.20) \end{aligned}$$

from which we find [2]:

$$\pounds_{\vec{u}}\,g_{\alpha\beta} = \nabla_\alpha u_\beta + \nabla_\beta u_\alpha \,. \qquad (2.21)$$

[2]This result can be derived much faster by noticing that for the calculation of Lie derivatives, partial derivatives can always be substituted with covariant derivatives as the Christoffel symbols cancel out. Remembering then that the covariant derivative of the metric is zero we find the desired result.

By comparing this last result with (2.14) we see that this equation can be rewritten as:

$$\left(\partial_t - \pounds_{\vec{\beta}}\right) \gamma_{ij} = -2\alpha K_{ij} \ . \tag{2.22}$$

But, from our discussion of the Lie derivative we know that $\partial_t = \pounds_{\vec{t}}$ with $\vec{t}$ the tangent vector to the time lines. From figure 4 this vector can be seen to be given by $\vec{t} = \alpha\vec{n} + \vec{\beta}$. Our final equation is then:

$$\pounds_{\alpha\vec{n}} \gamma_{ij} = -2\alpha K_{ij} \ . \tag{2.23}$$

This equation shows that the extrinsic curvature is precisely the change in time of the spatial metric as seen by the Eulerian observers.

This brings us half-way to our goal of writing Einstein's equations as a Cauchy problem: We already have an evolution equation for the spatial metric γ_{ij}. In order to close the system we still need an evolution equation for K_{ij}. It is important to notice that until now we have only worked with purely geometric concepts, and we have not used the field equations at all. It is precisely from the field equations that we will obtain the evolution equations for K_{ij}. In other words, the evolution equation (2.14) for the 3-metric is purely kinematic, while the dynamics of the system will be contained in the evolution equations for K_{ij}.

2.4 The Einstein field equations in the 3+1 formalism

In order to write down the Einstein equations in 3+1 form, we will use the projection operator defined in the last section, together with the normal vector, to separate Einstein's equations in three groups:

- Normal projection (1 equation):

$$n^\alpha n^\beta \left(G_{\alpha\beta} - 8\pi T_{\alpha\beta}\right) = 0 \, . \tag{2.24}$$

- Mixed projections (3 equations):

$$P\ \left[n^\alpha \left(G_{\alpha\beta} - 8\pi T_{\alpha\beta}\right)\right] = 0 \, . \tag{2.25}$$

- Projection onto the hypersurface (6 equations):

$$P\ \left(G_{\alpha\beta} - 8\pi T_{\alpha\beta}\right) = 0 \, . \tag{2.26}$$

In order to rewrite these groups of equations in 3+1 language one needs a long algebra that I will not show in these notes. Here I will just write down the final results. From the normal projection we find the following equation:

$$R^{(3)} + (\mathrm{tr}\, K)^2 - K_{ij}K^{ij} = 16\pi\rho\,. \tag{2.27}$$

where $R^{(3)}$ is the Ricci scalar associated with the 3-metric, $\mathrm{tr}\, K \equiv \gamma^{ij}K_{ij}$ is the trace of the extrinsic curvature tensor, and ρ is the energy density of matter as measured by the Eulerian observers:

$$\rho := n_\alpha n_\beta\, T^{\alpha\beta}\,. \tag{2.28}$$

Equation (2.27) contains no time derivatives (though it does contain second spatial derivatives of γ_{ij} in the Ricci scalar). Because of this, the equation is not a dynamical equation but rather a "constraint" of the system. As it is related with the energy density, it is known as the "energy" or "hamiltonian" constraint.

From the mixed projection of the field equations we find:

$$D_j\left[K^{ij} - \gamma^{ij}\mathrm{tr}\, K\right] = 8\pi j^i\,, \tag{2.29}$$

where now j^i is the momentum flux of matter as measured by the Eulerian observers:

$$j^i := P^i_\beta\left(n_\alpha T^{\alpha\beta}\right)\,. \tag{2.30}$$

Equation (2.29) again has no time derivatives, so it is another constraint (or rather, three constraints). These equations are known as the "momentum" constraints.

The existence of the constraints implies that in general relativity it is not possible to specify arbitrarily all 12 dynamical quantities $\{\gamma_{ij}, K_{ij}\}$ as initial conditions. The initial data must already satisfy the constraints, otherwise we will not be solving Einstein's equations.

The remaining 6 equations are obtained from the projection onto the hypersurface and contain the true dynamics of the system. These equations take the form:

$$\begin{aligned}\partial_t K_{ij} \;=\;& \beta^a\partial_a K_{ij} + K_{ia}\partial_j\beta^a + K_{ja}\partial_i\beta^a \\ & -D_iD_j\alpha + \alpha\left[R^{(3)}_{ij} - 2K_{ia}K^a_j + K_{ij}\,\mathrm{tr}\, K\right] \\ & +4\pi\alpha\left[\gamma_{ij}\left(\mathrm{tr}\, S - \rho\right) - 2S_{ij}\right]\,,\end{aligned} \tag{2.31}$$

where S_{ij} is the stress tensor of matter, defined as:

$$S_{ij} := P\, T_{ij}\,. \tag{2.32}$$

Notice that, just as it happened with the evolution equations for the 3-metric, the last equations can be rewritten as:

$$\begin{aligned}\left(\partial_t - \pounds_{\vec{\beta}}\right) K_{ij} &= -D_i D_j \alpha + \alpha \left[R^{(3)}_{ij} - 2K_{ia}K^a_j + K_{ij} \operatorname{tr} K \right] \\ &\quad +4\pi\alpha \left[\gamma_{ij} \left(\operatorname{tr} S - \rho \right) - 2S_{ij} \right] , \end{aligned} \tag{2.33}$$

Equations (2.14) and (2.31) form a closed system of evolution equations, they are known as the Arnowitt-Deser-Misner (ADM) equations [13, 67]. These equations finally allow us to write down the field equations for general relativity as a Cauchy problem.

It is important to note that we do not have evolution equations for the gauge quantities $\{\alpha, \beta^i\}$. As we have mentioned before, these quantities represent our coordinate freedom and can therefore be chosen freely.

It is also possible to show, by using the Bianchi identities, that the evolution equations guarantee that if the constraints are satisfied initially, then they will continue to be satisfied during the evolution: The evolution equations propagate the constraints.

2.5 The initial data problem

The existence of the constraints in general relativity implies that it is not possible to choose arbitrarily the 12 dynamical quantities $\{\gamma_{ij}, K_{ij}\}$ as initial data. The initial data has to be chosen in such a way that the constraints are satisfied from the beginning. This means that before starting an evolution, it is necessary to first solve the initial data problem to obtain adequate values of $\{\gamma_{ij}, K_{ij}\}$ that represent the physical situation that we are interested in.

The constraints form a system of four coupled partial differential equations of elliptic type, and in general they are difficult to solve. Still, there exists several well know procedures to solve these equations in specific circumstances. Until a few years ago, the most common procedure was the "conformal decomposition" of York-Lichnerowicz [38, 67]. More recently, the so-called "thin sandwich formalism" [66] has become more and more popular when solving the constraints. In these notes I will concentrate on the first of these procedures since it is better known and understood. A recent review of the different ways to find initial data can be found in [21].

The York-Lichnerowicz conformal decomposition starts from considering a transformation of the 3-metric of the form:

$$\gamma_{ij} = \phi^4 \tilde{\gamma}_{ij} \, , \tag{2.34}$$

where $\tilde{\gamma}_{ij}$ is considered known. Such a transformation is called a "conformal transformation" because is can be easily shown to preserve angles. The metric $\tilde{\gamma}_{ij}$ is called the "conformal metric".

At the same time, the extrinsic curvature is first separated into its trace:

$$\mathrm{tr}K = \gamma^{ij}\, K_{ij}\ , \tag{2.35}$$

and its trace free part A_{ij}:

$$A_{ij} := K_{ij} - \frac{1}{3}\gamma_{ij}\,\mathrm{tr}K\,. \tag{2.36}$$

A conformal transformation is also performed on A_{ij} in the following way:

$$A_{ij} = \phi^{-10}\tilde{A}_{ij}\,, \tag{2.37}$$

where the tenth power is chosen for convenience since it simplifies the equations obtained later. We can now use a general algebraic result that states that any symmetric and trace-free tensor can be decomposed in the following way:

$$\tilde{A}_{ij} = \tilde{A}^{*}_{ij} + (\hat{L}W)_{ij}\,, \tag{2.38}$$

where $\tilde{A}^{*}_{ij}$ is a zero divergence tensor $\tilde{D}^i\tilde{A}^{*}_{ij} = 0$ (and also trace-less), W_i is a vector, and $\hat{L}$ is a differential operator defined as:

$$(\hat{L}W)_{ij} = \tilde{D}_iW_j + \tilde{D}_jW_i - \frac{2}{3}\,\tilde{\gamma}_{ij}\tilde{D}_kW^k\,, \tag{2.39}$$

with $\tilde{D}_i$ the covariant derivative associated with the conformal metric γ_{ij}.

In order to find our initial data, we now assume that $\tilde{\gamma}_{ij}$, $\mathrm{tr}K$ and $\tilde{A}^{*}_{ij}$ are given, and use the four constraints to determine the four quantities $\{\phi, W_i\}$.

It is not difficult to show that the hamiltonian constraint takes the form

$$8\tilde{D}^2\phi - \tilde{R}\,\phi + \phi^{-7}\left(\tilde{A}_{ij}\tilde{A}^{ij}\right) - \frac{2}{3}\,\phi^5(\mathrm{tr}K)^2 + 16\pi\tilde{\phi}^{-3}\rho = 0\,, \tag{2.40}$$

which provides us with an elliptic equation for ϕ (basically Poisson's equation). It is important to remember that the Laplacian operator $\tilde{D}^2$ that appears in the previous equation must be the one associated with the conformal metric.

On the other hand, the momentum constraints become:

$$\tilde{\Delta}^2_{\hat{L}}\,W^i - \frac{2}{3}\psi^6\tilde{D}^iK - 8\pi\psi^{10}j^i = 0\ , \tag{2.41}$$

where we have defined

$$\tilde{\Delta}^2_{\hat{L}}\,W^i := \tilde{D}_j(\hat{L}W)^{ij} = \tilde{D}^2W^i + \frac{1}{3}\,\tilde{D}^i\tilde{D}_jW^j + \tilde{R}^i_jW^i\ , \tag{2.42}$$

which are three coupled elliptic equations for W^i.

The equations just found are the most general way of writing the constraints in the York-Lichnerowicz approach. Note how all four equations are coupled with each other.

A way to simplify the problem considerably and decouple the equations is to simply choose $\mathrm{tr}K = 0$. If, moreover, we also take the conformal metric to be the one corresponding to flat space $\tilde{\gamma}_{ij} = \delta_{ij}$, then the constraints reduce (in vacuum) to:

$$8\, D^2_{\rm flat}\phi + \phi^{-7}\left(\tilde{A}_{ij}\tilde{A}^{ij}\right) = 0\,, \tag{2.43}$$

$$D^2_{\rm flat}W^i = 0\,, \tag{2.44}$$

where now $D^2_{\rm flat}$ is just the standard flat space Laplacian. The second equation is in fact linear, and can be solved analytically in many cases. Once we have a solution for W^i, we can use it to reconstruct $\tilde{A}_{ij}$, and from there one can solve Poisson's equation for ϕ.

2.6 Black hole initial data

As a first example of a solution to the initial data problem, consider the following case: we have a conformally flat metric $\tilde{\gamma}_{ij} = \delta_{ij}$, and a moment of time-symmetry, that is, $K_{ij} = 0$. In that case the momentum constraints are satisfied trivially, and the hamiltonian constraint takes the simple form:

$$D^2\phi = 0\,, \tag{2.45}$$

which is nothing more that the standard Laplace equation. The boundary conditions correspond to an asymptotically flat spacetime (far away the gravitational field goes to zero), so that at infinity we must have $\phi = 1$. The simplest solution to this equation that satisfies the boundary conditions is then:

$$\phi = 1\,, \tag{2.46}$$

which implies that the spatial metric is simply:

$$dl^2 = dx^2 + dy^2 + dz^2\,. \tag{2.47}$$

That is, we have recovered initial data for Minkowski spacetime (though through a rather elaborate route).

The next interesting solution is clearly:

$$\phi = 1 + k/r\,, \tag{2.48}$$

with k an arbitrary constant, so the spatial metric (in spherical coordinates) now takes the form:

$$dl^2 = (1 + k/r)^4 \left[dr^2 + r^2 d\Omega^2\right] . \tag{2.49}$$

It is not difficult to see that this is nothing more than the spatial metric for a Schwarzschild black hole in so-called "isotropic coordinates" for which the radial coordinate is related to the standard Schwarzschild radius R as $R = r(1 + M/2r)^2$, with the mass of the black hole given by $M = 2k$. We have therefore found the solution to the initial data problem corresponding to a single Schwarzschild black hole.

As Laplace's equation is linear, we can simply superpose solutions to obtain new solutions. For example, the solution:

$$\phi = 1 + \sum_{i=1}^{N} \frac{M_i}{2|r - r_i|} , \tag{2.50}$$

will represent N black holes at the points r_i that are momentarily at rest, and is known as Brill-Lindquist initial data. It is possible to generalize this solution to the case when the black holes are not initially at rest, constructing for example initial data for orbiting black holes. However, in this case the solution of the hamiltonian constraint is no longer so simple and one has to obtain it using numerical techniques.

It is important to notice that this type of initial data for black holes are in fact topological models: the black holes are represented by wormholes (Einstein-Rosen bridges) that join our Universe with other universes. This also happens in Schwarzschild's solution for an "eternal" black hole (that is, not formed by gravitational collapse). In the real world, it is assumed that black holes form by collapse, so the Einstein-Rosen bridge will not be present.

2.7 Gauge conditions

As we have already mentioned, in the 3+1 formalism the choice of the coordinate system is given in terms of the "gauge variables": the lapse function α and the shift vector β^i. These functions appear in the evolution equations (2.14) and (2.31) for the metric and extrinsic curvature. However, Einstein's equations say nothing about how the gauge variables should be chosen. This is what should be expected, since it is clear that the coordinates can be chosen freely.

The freedom in choosing the gauge variables is a mixed blessing. On the one hand, it allows us to choose things in a way that simplifies the equations, or makes the solution better behaved. On the other hand, we are immediately faced with the question: What is a "good" choice for the functions α and β^i? Remember that this decision has to be made

even before we start the evolution. There are, of course, an infinite number of possible choices. In these notes we will just mention some of the better known possibilities.

2.7.1 Slicing conditions

Let us first consider the choice of the lapse function α. The choice of this function determines the way in which proper time advances between adjacent hypersurfaces, that is, it determines the way in which spacetime is foliated into slices of constant t coordinate. Because of this, the choice of lapse is usually called a "slicing condition".

Before considering different slicing conditions, we must mention a couple of results that are very important when discussing this issue. First, consider the motion of the Eulerian observers, *i.e.* those that move following the normal direction to the hypersurfaces of the foliation. There is no reason to assume that these observers will be in free fall, so in general they should have a proper acceleration that measures the force that would be required to keep them in a trajectory different from free fall. This acceleration is given in terms of the directional derivative of the 4-velocity of the Eulerian observers (the normal vector n^μ) along itself:

$$a^\mu = n^\nu \nabla_\nu n^\mu \ . \tag{2.51}$$

Notice that this involves the four-dimensional covariant derivative ∇_μ. It is not difficult to show that this equation implies that the proper acceleration is given in terms of the lapse function as

$$a_i = \partial_i \ln \alpha \ . \tag{2.52}$$

Another important relationship comes from the evolution of the volume elements associated with the Eulerian observers. The change in time of these volume elements is simply given by the divergence of their 4-velocities $\nabla_\mu n^\mu$. Using now the definition of the extrinsic curvature we find:

$$\nabla_\mu n^\mu = -\mathrm{tr}K \ , \tag{2.53}$$

that is, the rate of change of the volume elements in time is just (minus) the trace of the extrinsic curvature.

Let us now go back to the problem of choosing the lapse function. The most obvious way to choose this function would be to ask for the coordinate time t to coincide everywhere with the proper time of the Eulerian observers. After all, what could be more natural? This choice corresponds to taking $\alpha = 1$ everywhere, and from (2.52) we see that it implies that the proper acceleration of the Eulerian observers is zero, that is, they are in free fall and move along geodesics. Because of this, this choice of lapse function is known as "geodesic slicing".

Geodesic slicing, though apparently very natural, turns out to be in practice a very poor choice. The reason for this is easy to see if we think for a moment about what will happen to observers in free fall in a non-uniform gravitational field. As the field is not uniform, different observers fall in different directions, and nothing can prevent them from eventually colliding with one another. When this happens, our coordinate system (tied to these observers) stops being one to one: One point has now more that one set of coordinates associated to it. This means that the coordinate system has become singular. Because of this, geodesic slicing is never used in practice.

In order to look for a better slicing condition, we must think about what precisely was the problem with geodesic slicing. The main problem is that observers in free fall are "focused" by the gravitational field. This means that they will get closer to each other, so the volume elements associated with them will become smaller until they become zero. We can then try to built a foliation where one demands that the volume elements remain constant. From equation (2.53) we can see that this is equivalent to asking for:

$$\mathrm{tr}K = \partial_t \mathrm{tr}K = 0 \,. \tag{2.54}$$

As can be seen from this expression, it is clear that we must require not only that $\mathrm{tr}K$ vanishes initially, but also that it remains zero during the evolution. On the other hand, from the evolution equations (2.14) and (2.31) we find:

$$\partial_t \mathrm{tr}K = \beta^i \partial_i \mathrm{tr}K - D^2\alpha + \alpha \left[K_{ij}K^{ij} + \frac{1}{2}\left(\rho + \mathrm{tr}S\right) \right] \,, \tag{2.55}$$

where we have used the hamiltonian constraint to eliminate the Ricci scalar. Asking now for condition (2.54) to be satisfied, we find that the lapse function must satisfy the following elliptic equation:

$$D^2\alpha = \alpha \left[K_{ij}K^{ij} + \frac{1}{2}\left(\rho + \mathrm{tr}S\right) \right] \,. \tag{2.56}$$

This condition is known as "maximal slicing" [56]. The name comes from the fact that one can prove that this condition guarantees that the volume of any region of the hypersurface is maximal with respect to small variations in the hypersurface itself.

Maximal slicing has been used over the years for many numerical simulations of a number of different systems, including black holes, and is still very popular today. It has the advantage of being given by a simple equation whose solution is smooth (because it comes from an elliptic equation), and also that it guarantees that the Eulerian observers will not focus. In particular, maximal slicing has the very important property of avoiding singularities, both coordinate singularities caused by the focusing of observers, and also

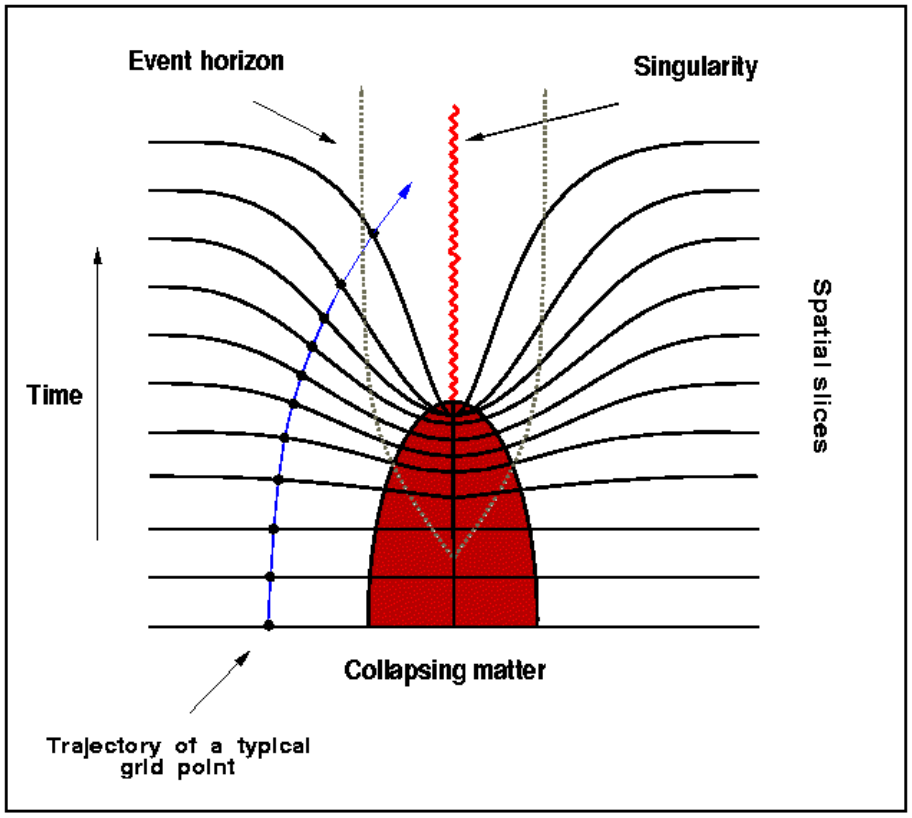

Figure 5: Collapse of the lapse. Time slows down in the region approaching the singularity, but continues advancing away from the singularity.

physical singularities as those found inside black holes. In the case of a black hole, the solution of the maximal slicing condition makes the lapse function fall exponentially to zero in the region inside the black hole, a phenomenon known as "collapse of the lapse" (see Figure 5). The collapse of the lapse turns out to be very useful as otherwise the hypersurfaces would march on directly to the physical singularity very quickly, where they would find infinite gravitational fields which would cause the computer code to crash.

One disadvantage of using maximal slicing is the fact that in 3D solving the elliptic equation (2.56) requires a lot of computer time. Because of this, some alternative slicing conditions have been proposed that share some of the properties of maximal slicing but are at the same time much easier to solve. One particular family of slicing conditions that

has these properties is the Bona-Masso slicing condition [16]. This family of conditions specifies an evolution equation for the lapse function of the form:

$$\left(\partial_t - \pounds_{\vec{\beta}}\right) \alpha = \partial_t \alpha - \beta^i \partial_i \alpha = -\alpha^2 f(\alpha)\, K \,, \tag{2.57}$$

where $f(\alpha)$ is a positive, but otherwise arbitrary, function of α (the reason why it has to be positive is related to the concept of hyperbolicity that will be discussed in Section 3). The term with the shift vector β^i is included to form a "total derivative" with respect to time, that is, a derivative along the normal direction.

The Bona-Masso condition includes some well known sub-families. For example, if we take $f = 1$ we obtain the so-called "harmonic slicing", for which the time coordinate turns out to obey a wave equation: $g^{\mu\nu}\nabla_\mu \nabla_\nu t = 0$. In this case the evolution equation for the lapse can be integrated to find:

$$\alpha = h(x^i)\, \gamma^{1/2} \,. \tag{2.58}$$

where $h(x^i)$ is a time independent function and γ is the determinant of the spatial metric, that is, $\gamma^{1/2}$ is the volume element.

Another important family is obtained by taking $f(\alpha) = N/\alpha$ with N a positive constant. This is known as the "1+log" family [1, 12], whose name comes from the fact that the lapse function now has the form:

$$\alpha = h(x^i) + \ln\left(\gamma^{N/2}\right) \,. \tag{2.59}$$

The 1+log family gives rise to foliations of spacetime that are very similar to those obtained with maximal slicing, and also has the property of singularity avoidance. In practice, it has been found that the particular case $N = 2$ is particularly well behaved [8].

2.7.2 Shift conditions

Considerably less is know about shift conditions than about slicing conditions. The main reason is that simply taking the shift equal to zero works well in many cases. Still, there are indeed proposals for how to chose the shift vector in some circumstances, and it is well known that in some cases taking a zero shift vector is a very poor choice.

The best know proposal for choosing a shift vector is know as "minimal distortion" [55]. The basic idea is to use the shift vector to eliminate as much as possible the distortion in the volume elements during the evolution. In order to achieve this we start from defining the "distortion tensor" Σ_{ij} in the following way:

$$\Sigma_{ij} = \frac{1}{2}\gamma^{1/2}\partial_t \tilde{\gamma}_{ij} \,, \tag{2.60}$$

where again γ is the determinant of the 3-metric, and $\tilde{\gamma}_{ij} = \gamma^{-1/3}\gamma_{ij}$. That is, $\tilde{\gamma}_{ij}$ is a conformal metric where the volume element has been factored out ($\tilde{\gamma} = 1$). The distortion tensor can therefore be understood as the change in the *shape* of the volume elements during the evolution (and not in their volume).

From the evolution equation for the metric (2.14), one can show that:

$$\Sigma_{ij} = -\alpha \left(K_{ij} - \frac{1}{3}\gamma_{ij} \mathrm{tr}K \right) + \frac{1}{2} \left(\hat{L}\beta \right)_{ij} , \tag{2.61}$$

where $\hat{L}$ is the same operator that we defined when discussing how to find initial data (equation (2.39)).

The minimal distortion condition is obtained by asking for the integral over all of space of the non-negative square of the distortion tensor, $\Sigma_{ij}\Sigma^{ij}$, to be as small as possible. In this way, this condition minimizes the distortion in a global sense. The condition we have just stated can be expressed mathematically as a variational principle, whose solution is:

$$D^j \Sigma_{ij} = 0 , \tag{2.62}$$

or equivalently:

$$D^j \left(\hat{L}\beta \right)_{ij} = 2 D^j \left[\alpha \left(K_{ij} - \frac{1}{3}\gamma_{ij} \mathrm{tr}K \right) \right] . \tag{2.63}$$

This is the final form of the minimal distortion condition.

There are several things that is important to mention about this shift condition. First, it is given by three differential equations (the index i is free), which in principle allow us to determine the three components of the shift β^i. Second, it as a system of coupled elliptic equations (generalizations of Poisson's equation). Because of this, the minimal distortion condition is hard to solve in 3D (though by no means impossible), which means that it has not been used extensively in practice.

Some shift conditions that are related with minimal distortion but are much easier to solve have been proposed in the last few years. In particular, proposals of the form:

$$\partial_t^2 \beta^i \propto D^j \Sigma^i_j , \tag{2.64}$$

that transform the system of elliptic equations into a system of (hyperbolic) evolution equations have been proposed in [9]. These conditions turn out to be easy to implement and very robust, but here we will not discuss them further.

To finish this section it is important to comment that, even if it is true that in many cases taking the shift vector to be equal to zero works well, there are cases when it is clear that this will not work. In particular, systems with angular momentum (rotating black

holes or neutron stars) produce an effect know as the "dragging of inertial frames", which in short means that observers close to the rotating object are dragged in its rotation. In these type of situations, the only way in which one can avoid ending up with a coordinate system that is completely tangled around the rotating object is by using a shift vector that counters the dragging effect. This type of situations also develop in the case of orbiting compact objects, and it is because of this that the study of shift conditions has become a research topic a great relevance.

3 Alternative formulations and hyperbolicity

3.1 Introduction

In the last section we introduced the ADM evolution equations (2.14) y (2.31). In fact, these equations are not written in the original form of Arnowitt, Deser and Misner [13], but they are rather a non-trivial rewriting due to York [67]. Even so, in the numerical relativity community, equations (2.14) and (2.31) are known simply as the ADM equations.

It is important to mention exactly what is the difference between the original ADM equations and the ADM equations *a la* York (which we will call from now on "standard" ADM). Both groups of equations differ in two main aspects. In the first place, the original ADM variables are the spatial metric γ_{ij} and its canonical conjugate momentum π_{ij}, obtained from the Hamiltonian formulation of general relativity, and which is related to the extrinsic curvature as [3]:

$$K_{ij} = -\gamma^{-1/2}\left(\pi_{ij} - \frac{1}{2}\,\gamma_{ij}\pi\right)\,, \tag{3.1}$$

with $\pi = \mathrm{tr}\,\pi_{ij}$.

This change of variables is a rather minor detail. Still, even if we rewrite them in terms of the extrinsic curvature, the original evolution equations of ADM for K_{ij} differ from (2.31) and have the form:

$$\begin{aligned}\partial_t K_{ij} \;=\;& \beta^a D_a K_{ij} + K_{ia} D_j \beta^a + K_{ja} D_i \beta^a \\ & - D_i D_j \alpha + \alpha\left[R^{(3)}_{ij} - 2K_{ia}K^a_j + K_{ij}\,\mathrm{tr}\,K\right] \\ & + 4\pi\alpha\left[\gamma_{ij}\left(\mathrm{tr}\,S - \rho\right) - 2S_{ij}\right] - \frac{\alpha}{4}\,\gamma_{ij} H\,,\end{aligned} \tag{3.2}$$

where H is the hamiltonian constraint (2.27):

$$H := R^{(3)} + \left(\mathrm{tr}\,K\right)^2 - K_{ij}K^{ij} - 16\pi\rho = 0\,. \tag{3.3}$$

It is clear that both sets of evolution equations for K_{ij} are physically equivalent since they differ only by the addition of a term proportional to the hamiltonian constraint which must vanish for any physical solution. However, the different evolution equations for K_{ij} are not *mathematically* equivalent. There are basically two reasons why this is so:

[3] One must remember that the original goal of ADM was to write a Hamiltonian formulation that could be used as a basis for quantum gravity, and not a system of evolution equations for dynamical simulations.

1. In the first place, the space of solutions to the evolution equations is different in both cases, and only coincides for physical solutions, that is, those that satisfy the constraints. In other words, both systems are only equivalent in a subset of the full space of solutions. This subset is called the "constraint hypersurface".

   Of course, one could always argue that since in the end we are only interested in physical solutions, the distinction just mentioned is irrelevant. This is strictly true only if one can solve the equations exactly. But in the case of numerical solutions there will always be some error that will take us out of the constraint hypersurface, and the question is then not only relevant but crucial: If we move slightly off the constraint hypersurface, does the subsequent evolution remain close to it, or does it diverge rapidly away from it?

2. The second reason why both systems of evolution equations differ mathematically is related to the last point and is of greater importance. Since the hamiltonian constraint has second derivatives of the metric (hidden inside the Ricci scalar R), then by adding a multiple of it to the evolution equations we are in fact altering the structure of the differential equations, that is, we can change the system from hyperbolic to elliptic, or even from well-posed to ill-posed.

These type of considerations take us to a fundamental observation that today has become one of the most active areas of research associated with numerical relativity: The 3+1 evolution equations are highly non-unique since one can always add to them arbitrary multiples of the constraints. The different systems of evolution equations will still coincide in the physical solutions, but might differ dramatically in their mathematical properties, and particularly in the way in which they react to small violations of the constraints (inevitable numerically). Which is the best way to write down the evolution equations, if any, is still an open problem.

3.2 The BSSN formulation

Even thought the ADM equations are the starting point of numerical relativity (at least in the 3+1 approach), in practice they have not turned out to be very well behaved with respect to small constraint violations. In the following section we will discuss a possible explanation for this fact, but for the moment we will take it as an empirical observation: the ADM equations are not very stable with respect to constraint violations.

From the early 1990's, a large number of alternative formulations of the 3+1 evolution equations have been proposed motivated by a variety of reasons. In these notes we can not possibly cover all these re-formulations, not even a small representative number, so

we will instead limit ourselves to discussing one particular reformulation that has become increasingly popular in the last few years precisely because of the fact that empirically it has turned out to be far better behaved than ADM. This formulation is known as "BSSN" (Baumgarte-Shapiro [14] and Shibata-Nakamura [52]).

The BSSN formulation differs from ADM in several aspects. In the first place, BSSN introduces several new variables similar to those used when solving the initial data problem discussed in Sec. 2.5. The spatial metric γ_{ij} is rewritten in terms of a conformal metric $\tilde{\gamma}_{ij}$ as:

$$\gamma_{ij} = \psi^4 \tilde{\gamma}_{ij}, \tag{3.4}$$

where the conformal factor ψ is chosen in such a way that the determinant of $\tilde{\gamma}_{ij}$ is equal to 1, that is:

$$\begin{aligned} \psi &= \gamma^{1/12} \,, & (3.5)\\ \tilde{\gamma}_{ij} &= \psi^{-4}\gamma_{ij} = \gamma^{-1/3}\gamma_{ij} \,, & (3.6)\\ \tilde{\gamma} &= 1 \,, & (3.7) \end{aligned}$$

with γ the determinant of γ_{ij} and $\tilde{\gamma}$ the determinant of $\tilde{\gamma}_{ij}$. On the other hand, the extrinsic curvature is separated into its trace $K := \mathrm{tr}K$ and its trace-free part:

$$A_{ij} = K_{ij} - \frac{1}{3}\gamma_{ij}K \,. \tag{3.8}$$

Instead of the ADM variables γ_{ij} and K_{ij}, BSSN uses the variables:

$$\begin{aligned} \phi &= \ln\psi = \frac{1}{12}\ln\gamma \,, & (3.9)\\ K &= \gamma_{ij}K^{ij} \,, & (3.10)\\ \tilde{\gamma}_{ij} &= e^{-4\phi}\gamma_{ij} \,, & (3.11)\\ \tilde{A}_{ij} &= e^{-4\phi}A_{ij} \,. & (3.12) \end{aligned}$$

Also, BSSN introduces three auxiliary variables known as the "conformal connection functions" and defined as:

$$\tilde{\Gamma}^i = \tilde{\gamma}^{jk}\tilde{\Gamma}^i{}_{jk} = -\partial_j\tilde{\gamma}^{ij} \,, \tag{3.13}$$

where $\tilde{\Gamma}^i{}_{jk}$ are the Christoffel symbols of the conformal metric, and where the second equality comes from the definition of the Christoffel symbols in the case when the determinant $\tilde{\gamma}$ is equal to 1 (which must be true by construction). We call the 17 variables ϕ, K, $\tilde{\gamma}_{ij}$, $\tilde{A}_{ij}$ and $\tilde{\Gamma}^i$ the "BSSN variables".

Up to this point all we have done is redefine variables and introduce three additional auxiliary variables. We still need to give the evolution equations for our new variables.

For simplicity, from here on we will assume that the shift vector vanishes and we are in vacuum (the general case can be recovered easily enough). From the evolution equation for the spatial metric (2.14) we find:

$$\partial_t \tilde{\gamma}_{ij} = -2\alpha \tilde{A}_{ij} \ , \tag{3.14}$$

$$\partial_t \phi = -\frac{1}{6}\alpha K \ , \tag{3.15}$$

while from the evolution equation for the extrinsic curvature (2.31) we obtain:

$$\partial_t \tilde{A}_{ij} = e^{-4\phi}\left[-D_i D_j \alpha + \alpha R_{ij}\right]^{\rm TF} + \alpha\left(K\tilde{A}_{ij} - 2\tilde{A}_{ik}\tilde{A}^k{}_j\right) \ , \tag{3.16}$$

$$\partial_t K = -D_i D^i \alpha + \alpha\left(\tilde{A}_{ij}\tilde{A}^{ij} + \frac{1}{3}K^2\right) \ , \tag{3.17}$$

where TF denotes the trace-free part of the expression inside the brackets. It is important to notice that in the evolution equation for K the hamiltonian constraint has been used in order to eliminate the Ricci scalar:

$$R = K_{ij}K^{ij} - K^2 = \tilde{A}_{ij}\tilde{A}^{ij} - \frac{2}{3}K^2 \ . \tag{3.18}$$

We then see how we have already started to play the game of adding multiples of constraints to evolution equations.

We are still missing an evolution equation for the $\tilde{\Gamma}^i$. This equation can be obtained directly from (3.13) and (2.14). We find:

$$\partial_t \tilde{\Gamma}^i = -2\left(\alpha\partial_j \tilde{A}^{ij} + \tilde{A}^{ij}\partial_j\alpha\right) \ , \tag{3.19}$$

Notice that in the evolution equations for $\tilde{A}_{ij}$ and K there appear covariant derivatives of the lapse function with respect to the physical metric γ_{ij}. These can be easily calculated by using the fact that the Christoffel symbols are related through:

$$\Gamma^k{}_{ij} = \tilde{\Gamma}^k{}_{ij} + 2\left(\delta^k_i\partial_j\phi + \delta^k_j\partial_i\phi - \tilde{\gamma}_{ij}\tilde{\gamma}^{kl}\partial_l\phi\right) \ , \tag{3.20}$$

where $\tilde{\Gamma}^k{}_{ij}$ are the Christoffel symbols of the conformal metric.

Also, in the equation for $\tilde{A}_{ij}$ we see the Ricci tensor associated with the physical metric, which can be separated into two contributions in the following way:

$$R_{ij} = \tilde{R}_{ij} + R^\phi_{ij} \ , \tag{3.21}$$

where $\tilde{R}_{ij}$ is the Ricci tensor associated with the conformal metric $\tilde{\gamma}_{ij}$:

$$\begin{aligned}\tilde{R}_{ij} &= -\frac{1}{2}\tilde{\gamma}^{lm}\partial_l\partial_m\tilde{\gamma}_{ij} + \tilde{\gamma}_{k(i}\partial_{j)}\tilde{\Gamma}^k + \tilde{\Gamma}^k\tilde{\Gamma}_{(ij)k} \\ &+ \tilde{\gamma}^{lm}\left(2\tilde{\Gamma}^k{}_{l(i}\tilde{\Gamma}_{j)km} + \tilde{\Gamma}^k{}_{im}\tilde{\Gamma}_{klj}\right) \ . \end{aligned} \tag{3.22}$$

and where R^{ϕ}_{ij} denotes additional terms that depend on ϕ:

$$R^{\phi}_{ij} = -2\tilde{D}_i\tilde{D}_j\phi - 2\tilde{\gamma}_{ij}\tilde{D}^k\tilde{D}_k\phi + 4\tilde{D}_i\phi\ \tilde{D}_j\phi - 4\tilde{\gamma}_{ij}\tilde{D}^k\phi\ \tilde{D}_k\phi \ , \tag{3.23}$$

with $\tilde{D}_i$ the covariant derivative associated with the conformal metric.

There are several motivations for the change of variables given above. The conformal transformation and the separating out of the trace of the extrinsic curvature are done in order to have better control over the slicing conditions that, as was already mention in Sec. 2.7.1, are generally related with the trace of K_{ij} and are chosen in a way that allow some degree of control over the evolution of the volume elements. On the other hand, the introduction of the variables $\tilde{\Gamma}^i$ is done because, when these are considered independent variables, the second derivatives of the conformal metric that appear on the right hand side of equation (3.16) (contained in the Ricci tensor (3.22)) reduce to the simple Laplace operator $\tilde{\gamma}^{lm}\partial_l\partial_m\tilde{\gamma}_{ij}$. All other terms with second derivatives of $\tilde{\gamma}_{ij}$ can be rewritten in terms of first derivatives of the $\tilde{\Gamma}^i$. This means that equations (3.14) and (3.16) now have the structure of a wave equation with a complicated source term.

We are still missing one key element of the BSSN formulation that we have not yet mentioned. In practice it turns out to be that, in spite of the motivations mentioned above, if one uses equations (3.14), (3.15), (3.16), (3.17), and (3.19) in a numerical simulation, the system turns out be violently unstable (much worse than ADM).

In order to fix this problem we first need to write down the momentum constraints, which in terms of the BSSN variables become:

$$\partial_j\tilde{A}^{ij} = -\tilde{\Gamma}^i{}_{jk}\tilde{A}^{jk} - 6\tilde{A}^{ij}\partial_j\phi + \frac{2}{3}\tilde{\gamma}^{ij}\partial_j K \ . \tag{3.24}$$

We can now use this equation to substitute the divergence of $\tilde{A}^{ij}$ that appears on the evolution equation (3.19) for the $\tilde{\Gamma}^i$. We find:

$$\partial_t\tilde{\Gamma}^i = -2\tilde{A}^{ij}\partial_j\alpha + 2\alpha\left(\tilde{\Gamma}^i{}_{jk}\tilde{A}^{jk} + 6\tilde{A}^{ij}\partial_j\phi - \frac{2}{3}\tilde{\gamma}^{ij}\partial_j K\right) \ . \tag{3.25}$$

The system of evolution equations is now: (3.14), (3.15), (3.16), (3.17), and (3.25). This new system not only does not present the violent instability mentioned before, but

at the same time turns out to be far better behaved than ADM in all cases studied until now. That this is so was first shown empirically (that is, through direct comparison of numerical simulations) by Baumgarte and Shapiro in [14], and was later explained by Alcubierre *et al.* in [4]. The explanation of why BSSN is much more robust that ADM is related with the concept of hyperbolicity that we will discuss in the next section.

3.3 The concept of hyperbolicity

In this section we will discuss a brief introduction to the concept of hyperbolicity of systems of evolution equations, a concept that turns out to be crucial in order to analyze the mathematical properties of the systems of evolution equations used in many different areas of physics, and particularly in numerical relativity. A more detailed description of the theory of hyperbolic systems can be found in [36].

Consider a system of evolution equations in one spatial dimension of the form:

$$\partial_t u_i \, + \, \partial_x F_i \, = \, q_i \qquad i \in \{1, ..., N_u\} \, , \tag{3.26}$$

where F_i and q_i are arbitrary functions, possibly non-linear, of the u's but not of their derivatives. This system can also be written as:

$$\partial_t \, u_i \, + \, \sum_j M_{ij} \partial_x u_j \, = \, q_i \qquad i \in \{1, ..., N_u\} \, , \tag{3.27}$$

with $M_{ij} = \partial F_i / \partial u_j$ the so-called "Jacobian matrix".

It is important to mention that most differential evolution equations in physics can be written in this form. In the case when there are higher order derivatives, one can always define auxiliary variables in order to obtain a first order system (we will see below how one can do this for the wave equation).

Let now λ_i be the eigenvalues of the Jacobian matrix M. The system of evolution equations is called "hyperbolic" if all the λ_i turn out to be real functions. Furthermore, the system is called "strongly hyperbolic" if there exist a complete set of eigenvectors. In the case when all eigenvalues are real, but a complete set of eigenvectors does not exist, the system is called "weakly hyperbolic". [4]

Let us now assume that we have a strongly hyperbolic system. In that case, we define the "eigenfields" w_i in the following way:

$$\mathbf{u} = R \, \mathbf{w} \quad \Rightarrow \quad \mathbf{w} = R^{-1} \, \mathbf{u} \, , \tag{3.28}$$

[4]The otherwise important distinction between "strongly" and "symmetric" hyperbolic systems does not arise in the case of one spatial dimension. In multiple dimensions, symmetric hyperbolic systems are those that can be diagonalized in a way that is independent of the direction.

where R is the matrix of column eigenvectors $\mathbf{e}_i$. It is possible to show that the matrix R is such that:

$$RMR^{-1} = \Lambda \ , \tag{3.29}$$

with $\Lambda = \mathrm{diag}(\lambda_i)$. That is, the matrix R diagonalizes the Jacobian matrix M (this is essentially a change of basis vectors).

The evolution equations for the eigenfields turn out to be:

$$\partial_t w_i + \lambda_i \, \partial_x w_i = q_i' \ , \tag{3.30}$$

with q_i' functions of the w's but not their derivatives. The system has not been transformed into a series of advection equations with characteristic velocities given by the eigenvalues λ_i. In other words, we now have a series of perturbations propagating with speeds λ_i.

Hyperbolicity is of fundamental importance in the study of the evolution equations associated with a Cauchy problem. In a physical sense, hyperbolicity implies that a system of equations is causal and local, that is, the solution at a given point in space-time depends only on information inside a compact region to the past of that point, the so-called "characteristic cone" (or light-cone in relativity). Mathematically, one can show that a strongly hyperbolic system is "well posed", that is, its solutions exist and are unique (at least locally), and further the solutions are stable in the sense that small changes in the initial data correspond to small changes in the solution.

The concept of hyperbolicity can be easily extended to three-dimensional systems by considering equations of the form:

$$\partial_t u_i \, + \, \partial_x F_i^x + \partial_y F_i^y + \partial_z F_i^z \, = \, q_i \ , \qquad i \in \{1, ..., N_u\} \ , \tag{3.31}$$

and analyzing the three Jacobian matrices $M_{ij}^k = \partial F_i^k / \partial u_j$ $(k = x, y, z)$.

In order to see a simple example of hyperbolicity, consider the one-dimensional wave equation

$$\frac{\partial^2 \phi}{\partial t^2} - c^2 \, \frac{\partial^2 \phi}{\partial x^2} = 0 \ , \tag{3.32}$$

where ϕ is the wave function and c the wave speed. This is a second order equation, but we can transform it into a first order system by introducing the auxiliary variables:

$$\Pi := \partial \phi / \partial t \ , \qquad \Psi := \partial \phi / \partial x \ . \tag{3.33}$$

In terms of these variables, the wave equation transforms into the following system:

$$\frac{\partial \phi}{\partial t} = \Pi \ , \qquad \frac{\partial \Pi}{\partial t} - c^2 \frac{\partial \Psi}{\partial x} = 0 \ , \qquad \frac{\partial \Psi}{\partial t} - \frac{\partial \Pi}{\partial x} = 0 \ , \tag{3.34}$$

where the first equation is just the definition of Π, the second is the wave equation itself, and the third is the requirement that the partial derivatives of ϕ should commute. The last system clearly already has the form (3.26). As the function ϕ does not appear in either of the two last equations, we can analyze the sub-system $\{\Pi, \Psi\}$. The Jacobian matrix in this case is a 2×2 matrix and has the following form:

$$\mathbf{M} = \begin{pmatrix} 0 & -c^2 \\ -1 & 0 \end{pmatrix} , \tag{3.35}$$

Its corresponding eigenvalues are real and are given by $\lambda_\pm = \pm c$, that is, the waves can travel both to the right and to the left with speed c. The matrix M has two independent eigenvectors, that is a complete set, so the system is strongly hyperbolic. The eigenvectors turn out to be $v_\pm = (\mp c, 1)$, and from this we can easily find the eigenfunctions $\omega_\pm = \Pi \mp c\Psi$ (note that the speed $+c$ corresponds to the eigenfunction $\Pi - c\Psi$ and vice-versa). The evolution equations for the eigenfunctions are clearly:

$$\frac{\partial \omega_\pm}{\partial t} \pm c \frac{\partial \omega_\pm}{\partial x} = 0 . \tag{3.36}$$

We see that the eigenfunctions propagate in opposite directions independently of each other.

3.4 Hyperbolicity of the 3+1 evolution equations

We will now study the hyperbolicity properties of the ADM evolution equations (2.14) and (2.31). As we are only interested in the basic idea, we will concentrate on a very simple case (the main conclusions are not fundamentally modified in the general case). In the first place, we will assume that we are in vacuum and also that the shift vector β^i vanishes. The equations we will use are then:

$$\partial_t \gamma_{ij} = -2\alpha K_{ij} , \tag{3.37}$$

$$\partial_t K_{ij} = -D_i D_j \alpha + \alpha \left[R^{(3)}_{ij} - 2K_{ia}K^a_j + K_{ij} \operatorname{tr} K \right] , \tag{3.38}$$

In order to simplify things even further, we will only consider linear perturbations of a flat space-time. In that case, the lapse function and spatial metric can be written as:

$$\alpha = 1 + a, \tag{3.39}$$

$$\gamma_{ij} = \delta_{ij} + h_{ij}, \tag{3.40}$$

with a and h_{ij} much smaller than 1. At first order in small quantities, the evolution equations take the simplified form (notice that under this circumstances the K_{ij} turn out to be small quantities themselves):

$$\partial_t h_{ij} = -2K_{ij}, \tag{3.41}$$

$$\partial_t K_{ij} = -\partial_i \partial_j a + R^{(1)}_{ij}, \tag{3.42}$$

where the linearized Ricci tensor is given by:

$$R^{(1)}_{ij} = -1/2 \left(\nabla^2_{\rm flat} h_{ij} - \partial_i \Gamma_j - \partial_j \Gamma_i \right) , \tag{3.43}$$

and where we have defined:

$$\Gamma_i := \sum_k \partial_k h_{ik} - 1/2 \, \partial_i h , \tag{3.44}$$

with $h = \sum_i h_{ii}$.

To continue the analysis we must now choose our slicing condition. The simplest choice is to take $a = 0$, but this is equivalent to geodesic slicing and we have already mentioned that it is not a good choice. We will take instead harmonic slicing, which in this approximation corresponds to:

$$\partial_t a = -K , \tag{3.45}$$

where now $K = \sum_i K_{ii}$.

The next step is to rewrite these equations as a first order system. In order to do that we introduce the auxiliary quantities:

$$A_i := \partial_i a , \qquad D_{ijk} := \frac{1}{2} \, \partial_i h_{jk} . \tag{3.46}$$

The system of equations now take the form:

$$\partial_t a = -K , \tag{3.47}$$

$$\partial_t h_{ij} = -2K_{ij} , \tag{3.48}$$

$$\partial_t A_i = -\partial_i K , \tag{3.49}$$

$$\partial_t D_{ijk} = -\partial_i K_{jk} , \tag{3.50}$$

$$\partial_t K_{ij} = -\partial_{(i} A_{j)} + \sum_k \left(2\partial_{(i} D_{kkj)} - \partial_{(i} D_{j)kk} - \partial_k D_{kij} \right) . \tag{3.51}$$

To carry out the hyperbolicity analysis we first notice that the variables a and h_{ij} evolve only through source terms (there are no derivatives in the right hand side of their

evolution equations). Moreover, they don't even appear in the other equations, so we can safely ignore them in the hyperbolicity analysis. We will then concentrate on the sub-system $\{A_i, D_{ijk}, K_{ij}\}$. As this is a three-dimensional system, we should in principle construct all three Jacobian matrices. Here we will only analyze the x direction, as the other two directions are entirely analogous.

If we consider only derivatives along the x direction, we notice that we only need to consider the components A_x and D_{xjk}, as the components A_q and D_{qjk}, for q different from x, can also be taken as fixed in this case (we can assume them to be zero without affecting the analysis). After some algebra we find that the system we need to analyze has the form:

$$\partial_t A_x + \partial_x \left(K_{xx} + K_{yy} + K_{zz}\right) = 0 \,, \tag{3.52}$$
$$\partial_t K_{xx} + \partial_x A_x + \partial_x \left(D_{xyy} + D_{xzz}\right) = 0 \,, \tag{3.53}$$
$$\partial_t K_{yy} + \partial_x D_{xyy} = 0 \,, \tag{3.54}$$
$$\partial_t K_{zz} + \partial_x D_{xzz} = 0 \,, \tag{3.55}$$
$$\partial_t K_{xy} = 0 \,, \tag{3.56}$$
$$\partial_t K_{xz} = 0 \,, \tag{3.57}$$
$$\partial_t K_{yz} + \partial_x D_{xyz} = 0 \,, \tag{3.58}$$
$$\partial_t D_{xjk} + \partial_x K_{jk} = 0 \,. \tag{3.59}$$

This is a system of thirteen variables:

$$u := \{A_x, K_{xx}, K_{yy}, K_{zz}, K_{xy}, K_{xz}, K_{yz}, D_{xxx}, D_{xyy}, D_{xzz}, D_{xxy}, D_{xxz}, D_{xyz}, \}$$

The Jacobian matrix turns out to be:

$$M = \begin{pmatrix}
0 & 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \\
0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0
\end{pmatrix} . \tag{3.60}$$

The previous matrix, with 13×13 elements, may seem large, but most of its entries are zero, so it is in fact not difficult to find its eigenvalues. These eigenvalues turn out to be: $+1$ with multiplicity 4, -1 with multiplicity 4, and 0 with multiplicity 5, for a total of 13. As all these eigenvalues are real, the system is clearly hyperbolic.

More interesting are the eigenvectors. Associated with the eigenvalue $\lambda = -1$ we have the following three eigenvectors:

$$\begin{aligned} v_1^- &= (1,-1,0,0,0,0,0,0,1,0,0,0,0,0), && (3.61)\\ v_2^- &= (0,0,1,-1,0,0,0,0,0,-1,1,0,0,0), && (3.62)\\ v_3^- &= (0,0,0,0,0,0,0,1,0,0,0,0,0,-1), && (3.63) \end{aligned}$$

associated with the eigenvalue $\lambda = +1$ we have the three eigenvectors:

$$\begin{aligned} v_1^+ &= (1,1,0,0,0,0,0,0,1,0,0,0,0,0), && (3.64)\\ v_2^+ &= (0,0,1,-1,0,0,0,0,0,1,-1,0,0,0), && (3.65)\\ v_3^+ &= (0,0,0,0,0,0,0,1,0,0,0,0,0,1), && (3.66) \end{aligned}$$

and associated with the eigenvalue $\lambda = 0$ the three eigenvectors:

$$\begin{aligned} v_1^0 &= (0,0,0,0,0,0,0,0,1,0,0,0,0,0), && (3.67)\\ v_2^0 &= (0,0,0,0,0,0,0,0,0,0,0,1,0,0), && (3.68)\\ v_3^0 &= (0,0,0,0,0,0,0,0,0,0,0,0,1,0), && (3.69) \end{aligned}$$

for a total of only nine eigenvectors. That is, we do not have a complete set of eigenvectors so the system is only weakly hyperbolic. A weakly hyperbolic system is in fact not mathematically well-posed, so we should expect problems when using the ADM evolution equations for numerical simulations.

In fact, there is no need to look at the full matrix to notice that there will be trouble with the eigenvectors. Take, for instance, the sub-system $\{K_{xy}, D_{xxy}\}$. These variables do not appear in other evolution equation, so one can analyze them separately. The Jacobian matrix for this sub-set has the form:

$$M = \begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix}, \qquad (3.70)$$

which clearly can not be diagonalized. A similar problem arises for the pair $\{K_{xz}, D_{xxz}\}$. There are other, more serious, problems with our matrix, but it was clear from the beginning that we should expect trouble.

Fortunately, it is possible to solve the problem of the weak hyperbolicity of ADM by adding multiples of the constraints in appropriate places. In order to see this, we first write the linear approximation to the constraints:

$$\sum_j \partial_j f_j = 0, \qquad \text{(hamiltonian)} \tag{3.71}$$

$$\sum_j \partial_j K_{ij} - \partial_i K = 0, \qquad \text{(momentum)} \tag{3.72}$$

where we have defined $f_i := \sum_k \partial_k h_{ik} - \partial_i h = 2\sum_k (D_{kki} - D_{ikk})$. It is not difficult to see that if we consider only derivatives in the direction x (and again ignore the components D_{qjk} with q different from x), the constraints take the form:

$$\partial_x (D_{xyy} + D_{xzz}) = 0, \qquad \text{(hamiltonian)} \tag{3.73}$$

$$\partial_x (K_{yy} + K_{zz}) = 0, \qquad \text{(momentum x)} \tag{3.74}$$

$$\partial_x K_{xy} = 0, \qquad \text{(momentum y)} \tag{3.75}$$

$$\partial_x K_{xz} = 0, \qquad \text{(momentum z)} \tag{3.76}$$

Substituting these into the evolution equations, we find that the Jacobian matrix now takes the form:

$$M = \begin{pmatrix}
0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \\
0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0
\end{pmatrix}. \tag{3.77}$$

The eigenvalues are the same as before, but now we do have a complete set of eigenvectors. In fact, from inspection one can even read the eigenfields from the structure of the above matrix: $\{A_x \pm K_{xx}, D_{xyy} \pm K_{yy}, D_{xzz} \pm K_{zz}, D_{xyz} \pm K_{yz}, K_{xy}, K_{xz}, D_{xxx}, D_{xxy}, D_{xxz}\}$. The lesson we have learned is the following: By substituting constraints into the evolution equations it is possible to transform the linearized ADM equations into a strongly hyperbolic system.

In the general case, when we have a strong field and a non-zero shift vector, one can perform a similar analysis. The problem in that in that case it is far from evident how the constraints should be added to the evolution equations in order to obtain a strongly hyperbolic system. In fact, there are many different ways of doing it, and several families of strongly hyperbolic formulations have been constructed, among which one should mention the formulations of Bona-Masso [16, 17], Frittelli-Reula [28], and Kidder-Scheel-Teukolsy [34]. Is is even possible to show that the BSSN formulation that was introduced in the previous section is strongly hyperbolic under certain assumptions [46].

Today, the problem of finding some criteria that can tell us when a given hyperbolic formulation is better than others has become a very important area of research, and several groups world-wide are actively working on this.

4 Finite differencing approximations

4.1 Introduction

Field theories play a fundamental role in modern physics. From Maxwell's classical electrodynamics, to quantum field theories, through the Schrödinger equation, hydrodynamics and general relativity, the notion of a field as a physical entity on its own right has had profound implications in our understanding of the Universe. Fields are continuous functions of space and time, and the mathematical description of their dynamics must be done in the context of partial differential equations.

The partial differential equations associated with physical theories are in general impossible to solve exactly except in very idealized cases. This difficulty can have different origins, from the presence of irregular boundaries, to the existence of non-linear terms in the equations themselves. In order to solve this type of equations in general dynamical situations it becomes inevitable to use numerical approximations.

There are many different ways in which one can solve partial differential equations numerically. The most popular methods are three: "finite differencing" [45], "finite elements" [43] and "spectral methods". In these notes I will limit myself to studying some simple finite differencing methods as this is the most commonly used approach in numerical relativity (though not the only one, in particular spectral methods have become increasingly popular in recent years [18, 33, 35]).

4.2 Fundamental ideas of finite differencing

When one studies a field in a continuous space-time, one is faced with considering an infinite and non-countable number of unknown variables: the value of the field at every point of space and for all times. In order to find the value of the field using numerical approximations, the first thing that needs to be done is to reduce the number of unknowns to a finite number. There are several different forms of doing this. Spectral methods, for example, expand the solution as a finite linear combination of some appropriate base functions. The variables to solve for are then the coefficients of such an expansion. A different approach is taken by finite differencing and finite elements. In both cases the number of variables is reduced by discretizing the domain of dependence of the functions, although using different strategies in each case.

The basic idea of finite differencing approximations is to substitute the continuous spacetime with a set of discrete points. This set of points is known as the computational "grid" or "mesh". The distances in space between points on the grid don't necessarily have to be uniform, but in these notes we will assume for simplicity that they are. The

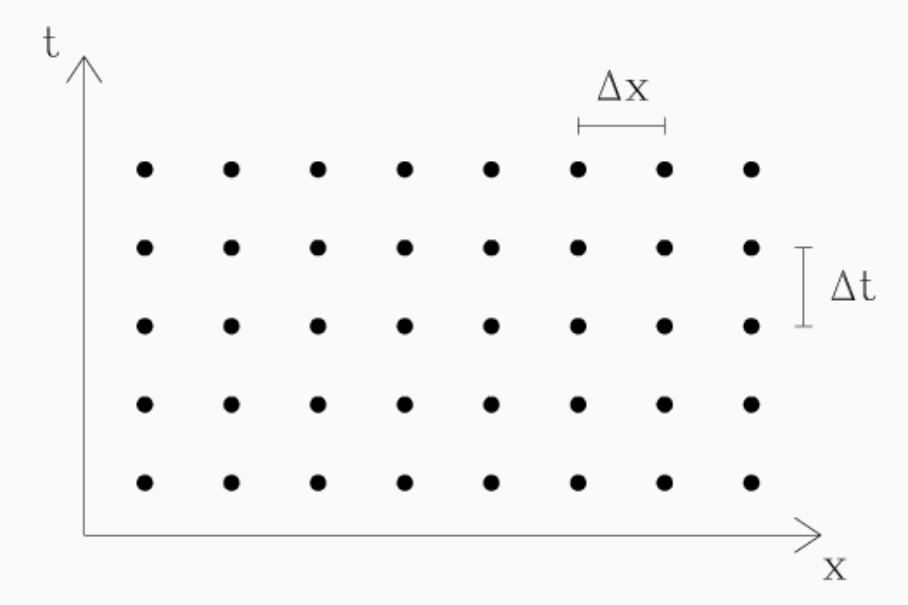

Figure 6: Discretization of space-time used in finite differencing.

time step between two consecutive levels is denoted by Δt, and the distance between two adjacent points by Δx. Figure 6 is a graphical representation of the computational grid.

Once we have established the computational grid, the next step is to substitute our differential equations with a system of algebraic equations. This is done by approximating the differential operators by finite differences between the values of our functions at nearby points on the grid. In this way one obtains an algebraic equation at each grid point for each differential equation. These algebraic equations involve the values of the functions at the point under consideration and its nearest neighbors. The system of algebraic equations can then be solved in a simple way, the price we have paid is that now we have lots of algebraic equations, so that a computer is needed in order to solve them all.

To see how this is done in practice, we will consider as a simple example the one-dimensional wave equation. This equation has a number of advantages. In the first place it can be solved exactly and this exact solution can be used to compare with the numerical solution. Also, most fundamental equations in modern field theory can be seen as generalizations of one type or another of the wave equation.

4.3 The one-dimensional wave equation

The one-dimensional wave equation (in flat space) has the following form:

$$\frac{\partial^2\phi}{\partial x^2} - \frac{1}{c^2}\frac{\partial^2\phi}{\partial t^2} = 0 \,, \tag{4.1}$$

where ϕ is the wave function and c the wave speed.

In order to find a finite differencing approximation to this equation we will start by introducing the following notation for the values of ϕ at the points of the computational grid:

$$\phi^n_m := \phi\left(n\Delta t,\, m\Delta x\right) \,, \tag{4.2}$$

We can now approximate the differential operators that appear in equation (4.1) by using Taylor expansions of ϕ around the point (n, m). Consider, for example, the value of ϕ at the points $(n, m+1)$ and $(n, m-1)$:

$$\phi^n_{m+1} = \phi^n_m + \left(\frac{\partial\phi}{\partial x}\right)\Delta x + \frac{1}{2}\left(\frac{\partial^2\phi}{\partial x^2}\right)(\Delta x)^2 + \frac{1}{6}\left(\frac{\partial^3\phi}{\partial x^3}\right)(\Delta x)^3 + \cdots \,, \tag{4.3}$$

$$\phi^n_{m-1} = \phi^n_m - \left(\frac{\partial\phi}{\partial x}\right)\Delta x + \frac{1}{2}\left(\frac{\partial^2\phi}{\partial x^2}\right)(\Delta x)^2 - \frac{1}{6}\left(\frac{\partial^3\phi}{\partial x^3}\right)(\Delta x)^3 + \cdots \,, \tag{4.4}$$

where all derivatives are to be evaluated at the point $(t = n\,\Delta t,\, x = m\,\Delta x)$. From these expressions it is easy to see that:

$$\left(\frac{\partial^2\phi}{\partial x^2}\right) = \frac{\phi^n_{m+1} - 2\,\phi^n_m + \phi^n_{m-1}}{(\Delta x)^2} + \frac{(\Delta x)^2}{12}\left(\frac{\partial^4\phi}{\partial x^4}\right) + \cdots \,. \tag{4.5}$$

We can then approximate the second derivative as:

$$\left(\frac{\partial^2\phi}{\partial x^2}\right) \simeq \frac{\phi^n_{m+1} - 2\,\phi^n_m + \phi^n_{m-1}}{(\Delta x)^2} \,. \tag{4.6}$$

How good is this approximation depends, of course, on the size of the grid spacing Δx. If Δx is small in the sense that the function ϕ changes very little in a region of that size, then the approximation can be very good indeed. The error involved in this approximation is known as "truncation error", and as we can see from (4.5) for small Δx its dominant part is proportional to Δx^2, so we say that this approximation is of second order.

The second derivative of ϕ with respect to t can be approximated in exactly the same way. In this way we find the following finite difference approximation for the wave equation:

$$\frac{\phi^n_{m+1} - 2\,\phi^n_m + \phi^n_{m-1}}{(\Delta x)^2} - \frac{1}{c^2}\,\frac{\phi^{n+1}_m - 2\,\phi^n_m + \phi^{n-1}_m}{(\Delta t)^2} = 0\ . \tag{4.7}$$

We can rewrite this equation in more compact form if we introduce the so-called "Courant parameter": $\rho := c\Delta t/\Delta x$. Our approximation then takes the final form:

$$\rho^2\,\left(\phi^n_{m+1} - 2\,\phi^n_m + \phi^n_{m-1}\right) - \left(\phi^{n+1}_m - 2\,\phi^n_m + \phi^{n-1}_m\right) = 0\ . \tag{4.8}$$

This equation has a very important property: it involves only one value of the wave function at the last time level, the value ϕ^{n+1}_m. We can then solve for this value in terms of values at previous time levels to obtain:

$$\phi^{n+1}_m = 2\,\phi^n_m - \phi^{n-1}_m\ +\ \rho^2\,\left(\phi^n_{m+1} - 2\,\phi^n_m + \phi^n_{m-1}\right)\ . \tag{4.9}$$

Because of this property, the last approximation is known as an "explicit" approximation. If we know the values of the function ϕ at the time levels n and $n-1$, we can use the last equation to calculate directly the values of ϕ at the new time level $n+1$. The process can then be iterated as many times as desired.

It is clear that all that is required in order to start the evolution is to know the values of the wave function at the first two time levels. But finding these two first levels is easy to do. As we are dealing with a second order equation, the initial data must include:

$$f\,(x) := \phi\,(0, x)\ , \qquad g\,(x) := \left.\frac{\partial\phi}{\partial t}\right|_{t=0}\ . \tag{4.10}$$

The knowledge of $f(x)$ evidently gives us the first time level:

$$\phi^0_m\ =\ f\,(m\,\Delta x)\ . \tag{4.11}$$

For the second time level it is enough to approximate the first time derivative using finite differences. One possible approximation is given by:

$$g\,(m\,\Delta x)\ =\ \frac{\phi^1_m - \phi^0_m}{\Delta t}\ , \tag{4.12}$$

from where we find:

$$\phi^1_m\ =\ g\,(m\,\Delta x)\ \Delta t\ +\ \phi^0_m\ . \tag{4.13}$$

The previous expression has one important drawback. From the Taylor expansion one can easily see that the truncation error for this expression is of order Δt, so our approximation is only first order. It is clear that is we start our evolution with a first order error, the second order accuracy of the whole scheme will be lost. However, this problem is quite easy to fix. A second order approximation to the first time derivative is:

$$g\left(m\,\Delta x\right) \;=\; \frac{\phi_m^1 - \phi_m^{-1}}{2\,\Delta t}\;. \tag{4.14}$$

The problem now is that this expression involves the value of the function ϕ_m^{-1} which is also unknown. But we already have one other equation that makes reference to ϕ_m^1 and ϕ_m^{-1}: the approximation to the wave equation (4.8) evaluated at $n = 0$. We can use these two equations to eliminate ϕ_m^{-1} and solve for ϕ_m^1. In this way we find the following second order approximation for the second time level:

$$\phi_m^1 \;=\; \phi_m^0 \;+\; \frac{\rho^2}{2}\left(\phi_{m+1}^0 - 2\,\phi_m^0 + \phi_{m-1}^0\right) \;+\; \Delta t\; g\left(m\,\Delta x\right)\;. \tag{4.15}$$

Equations (4.11) and (4.15) give us all the information we require in order to start our evolution.

There is another important point that must be mentioned here. In order to reduce the total number of variables to a finite number, it is also necessary to consider a finite region of space, known as the "computational domain", with a finite number of points N. It is therefore crucial to specify the boundary conditions that have to be applied at the edges of the computational domain. It is clear that the approximation to the wave equation (4.8) can not be used at the boundaries since it involves points outside the computational domain. There are many different ways to impose boundary conditions for the wave equation. However, for more complex systems like Einstein's equations, the choice of appropriate and consistent boundary conditions is still an open problem of great interest [57, 58]. For the purpose of these notes we will simple ignore the boundary issue and assume that we have a periodic space, so one can simply choose as boundary conditions:

$$\phi_0^n \;\equiv\; \phi_N^n\;. \tag{4.16}$$

This choice, apart from being extremely simple, is equivalent to using the interior approximation everywhere, so it allows us to concentrate on the properties of the interior scheme only, without having to worry about possible effects introduced by the boundaries.

4.4 Implicit approximations and computational molecules

In the last section we introduced the basic ideas of finite differencing approximations using the one dimensional wave equation as a working example. The approximation that

we found, however, is far from being unique. In principle, there are an infinite number of different ways to approximate the same differential equation using finite differences. Different approximations have different properties. In this section we will see which of those properties can make a particular approximation more useful than other.

In order to simplify things, I will now introduce a more compact notation for the finite differences. Let us define the "first centered difference" operators as:

$$\delta_x \, \phi_m^n \;\; := \;\; \frac{1}{2} \left(\phi_{m+1}^n \, - \, \phi_{m-1}^n \right) \, , \tag{4.17}$$

$$\delta_t \, \phi_m^n \;\; := \;\; \frac{1}{2} \left(\phi_m^{n+1} \, - \, \phi_m^{n-1} \right) \, , \tag{4.18}$$

and the "second centered difference" operators as:

$$\delta_x^2 \, \phi_m^n \;\; := \;\; \phi_{m+1}^n \, - \, 2 \, \phi_m^n \, + \, \phi_{m-1}^n \, , \tag{4.19}$$

$$\delta_t^2 \, \phi_m^n \;\; := \;\; \phi_m^{n+1} \, - \, 2 \, \phi_m^n \, + \, \phi_m^{n-1} \, . \tag{4.20}$$

(A word of caution: With this notation $(\delta_x)^2 \neq \delta_x^2$).

Having defined these operators, we can now go back to the approximations we used for the differential operators that appear on the wave equation. Staring again from the Taylor series, it is possible to show that the second spatial derivative can be approximated more generally as:

$$\left(\frac{\partial^2 \phi}{\partial x^2} \right) \simeq \frac{1}{(\Delta x)^2} \, \delta_x^2 \left[\frac{\theta}{2} \left(\phi_m^{n+1} \, + \, \phi_m^{n-1} \right) \, + \, (1 - \theta) \, \phi_m^n \right] \, , \tag{4.21}$$

with θ some arbitrary parameter. The expression we had before, equation (4.6), can be recovered by taking $\theta = 0$. This new approximation corresponds to taking an average, with a certain weight, of the finite difference operators acting on different time levels. In the particular case when $\theta = 1$, the contribution from the middle time level in fact completely disappears.

If we now use the new approximation for the second spatial derivative, but keep the same approximation as before for the time derivative, we find the following finite difference approximation for the wave equation:

$$\rho^2 \, \delta_x^2 \left[\frac{\theta}{2} \left(\phi_m^{n+1} \, + \, \phi_m^{n-1} \right) \, + \, (1 - \theta) \, \phi_m^n \right] \, - \, \delta_t^2 \, \phi_m^n \, = \, 0 \, . \tag{4.22}$$

This is one possible generalization of (4.8), but not the only one, it is clear that one can play this game in many ways to obtain even more general approximations, all equally

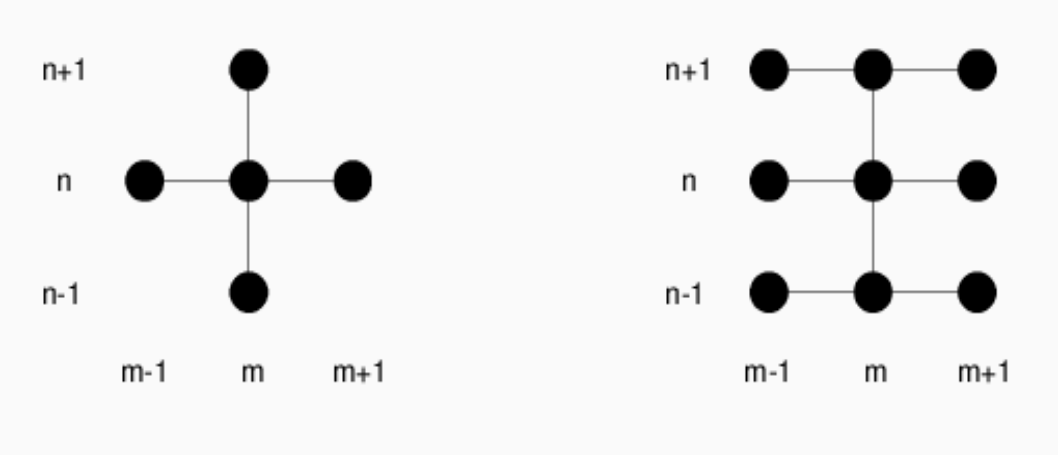

Figure 7: Computational molecules.

valid, and all second order (it is even possible to find ways to make the approximations fourth order or higher). The approximation given by (4.22) has a new very important property: it involves not one, but three different values of ϕ at the last time level. This means that it is not possible to solve for ϕ at the last time level explicitly in terms of its values in the two previous time levels. Because of this, the approximation (4.22) is known as an "implicit approximation".

When one considers the equations for all the points in the grid, including the boundaries, it is possible to solve the full system by inverting a non-trivial matrix, which is of course a more time consuming operation that the one needed for the explicit approximation. However, in many cases implicit approximations turn out to have better properties than explicit ones, in particular related to the stability of the numerical scheme, a concept that will be discussed in the following section.

The difference between an implicit and an explicit approximation can be seen graphically using the concept of a "computational molecule", that is nothing more than a diagram that shows the relationship between the different grid points used in the finite difference approximation. Figure 7 shows the computational molecules for the implicit and explicit approximations we have considered.

4.5 Consistency, convergence and stability

In the last section we encountered what is maybe one of the most important lessons in finite differencing: There are, in general, an infinite number of possible ways to approximate the same differential equation. This is even true if one is restricted to considering only

approximations with the same order of accuracy.

The multiplicity of possible approximations makes us immediately ask the following question: How can one know in which case to use a given approximation and not another? Unfortunately, there is no general answer to this question, which is why it is often said that finite differencing is more an art than a science. However, there are indeed some guidelines that allow us to choose among different approximations in some cases. These guidelines have to do with the concepts of consistency, convergence and stability.

Let us consider a given finite difference approximation to a differential equation. As the grid is refined (that is, as Δt and Δx become smaller), one would expect the approximation to become better in the sense that the truncation errors become smaller. We look then for an approximation that in the continuous limit approaches the original differential equation and not a different one. When this happens locally we say that our approximation is "consistent". In general, this property is quite easy to see from the structure of the finite differences, and can often by checked by "eye". The important exceptions are situations where the coordinate system becomes singular, since proving consistency at a singular point might be non-trivial. For example, it is common for "standard" finite difference approximations to fail at the point $r = 0$ when using spherical coordinates. Consistency is clearly fundamental for any finite difference approximation. When it fails, even if it is at just one point, we will not be able to recover the correct solution to the original differential equation.

Consistency is only a local property: A consistent finite difference approximation reduces *locally* to the differential equation in the continuum limit. In practice, one is really interested in a more global property. What one really looks for is an approximation that improves *after a finite time* T as the grid is refined. That is, the difference between the exact solution and the numerical solution at fixed time T must go to zero in the continuum limit. This condition is known as "convergence".

Convergence is clearly different from consistency: Perfectly consistent schemes can quite easily fail to be convergent. This is easy to understand if we think for a moment that in the limit when Δt becomes zero, a finite time T can only be reached after an infinite number of time steps. This implies that even if the error in each time step is infinitesimal, its total integral could well be finite. The numerical solution can even diverge and in fact become infinite in that limit! It is in general quite difficult to verify analytically if a given approximation scheme is convergent or not. Numerically, on the other hand, it is easy to check if the approximate solution is converging to something (that is, it is not diverging). The difficult part is to know if the numerical solution is converging to the exact solution and not to something else.

There is another very important property of finite differencing approximations. Independently of the behavior of the solution to the differential equation, we must ask that

the *exact solutions to the finite difference equations* should remain bounded after a given finite time T for any time step Δt. This requirement is known as "stability", and implies that no component of the initial data should be amplified arbitrarily. Stability is a property of the system of finite difference equations, and is essentially the discrete analog of the condition that a system of differential equations should be well posed. An unstable finite difference approximation is useless in practice.

A fundamental result of the theory of finite difference approximations is Lax theorem (for a proof see, for example, Ref. [45]):

THEOREM: *Given an initial value problem that is mathematically well posed, and a finite difference approximation to it that is consistent, then stability is a necessary and sufficient condition for convergence.*

This theorem is of great importance since it relates the final objective of any finite difference approximation, namely convergence to the exact solution, with a property that is much easier to prove: stability.

4.6 Von Newmann stability

A general method for studying the stability of systems of finite difference equations can be obtained directly from the definition of stability. We start by writing the finite difference equations as:

$$\mathbf{v}^{n+1} = \mathbf{B}\,\mathbf{v}^{n}\,, \tag{4.23}$$

where $\mathbf{v}^n$ is the solution vector at time level n, and $\mathbf{B}$ is a matrix (in general sparse). It is important to notice that all finite difference approximation can be written in this way, even those that involve more that two time levels (as the ones we have introduced for the wave equation in the previous sections), by simply introducing auxiliary variables.

Since the vector $\mathbf{v}^n$ can be written as a linear combination of the eigenvectors of $\mathbf{B}$, the requirement of stability can be reduced to asking for the matrix $\mathbf{B}$ not to amplify any of its eigenvectors, that is, we must ask for the magnitude of its largest eigenvalue to be less than or equal to 1. In other words, the "spectral radius" of $\mathbf{B}$ must be less than or equal to 1.

The stability analysis based on the idea just described is quite general, but it requires the knowledge of the entries of $\mathbf{B}$ over all space, including the boundary. There is, however, a very popular stability analysis method that, even though it can only be shown to give necessary conditions for stability, in many cases turns out to give also sufficient conditions. This method, originally introduced by von Newmann, is based on a Fourier decomposition.

To introduce von Newmann's method, we start by expanding the solution of (4.23) as a Fourier series:

$$\mathbf{v}^n(\mathbf{x}) = \sum_{\mathbf{k}} \tilde{\mathbf{v}}^n(\mathbf{k})\, e^{i\,\mathbf{k}\cdot\mathbf{x}} \,, \tag{4.24}$$

where the sum is over all wave vectors $\mathbf{k}$ that can be represented on the grid. [5] If we now substitute this into the original equation (4.23) we find:

$$\tilde{\mathbf{v}}^{n+1} = \mathbf{G}(\Delta x, \Delta t, \mathbf{k})\ \tilde{\mathbf{v}}^n \,. \tag{4.25}$$

The matrix $\mathbf{G}$ is known as the "amplification matrix". The stability condition now corresponds to asking that for no Fourier mode to be amplified, that is, for the spectral radius of $\mathbf{G}$ to be less than or equal to 1. This is von Newmann's stability condition.

It is important to stress that in order to use this stability criteria we have assumed two things: 1) The boundary conditions are periodic, since otherwise one can not make a Fourier expansion, and 2) the entries of the matrix $\mathbf{B}$ are constant (i.e. independent of position), since otherwise it is not possible to decouple the different Fourier modes.

As an example of von Newmann's stability analysis we can now study the stability of the implicit approximation to the wave equation we derived previously (equation (4.22)):

$$\rho^2\delta_x^2\left[(\theta/2)\left(\phi_m^{n+1}+\phi_m^{n-1}\right)+(1-\theta)\,\phi_m^n\right]-\delta_t^2\phi_m^n=0\,. \tag{4.26}$$

Let us consider a Fourier mode of the form:

$$\phi_m^n = \xi^n\, e^{imk\,\Delta x}\,. \tag{4.27}$$

If we substitute this back into the finite difference equation we find, after some algebra, the following quadratic equation for ξ:

$$A\xi^2 + B\xi + C = 0\,, \tag{4.28}$$

with coefficients given by:

$$A = \rho^2\theta\left[\cos(k\Delta x)-1\right]-1\,, \tag{4.29}$$

$$B = 2\rho^2(1-\theta)\left[\cos(k\Delta x)-1\right]+2\,, \tag{4.30}$$

$$C = \rho^2\theta\left[\cos(k\Delta x)-1\right]-1\,. \tag{4.31}$$

[5] The shortest wavelength that can be represented on the grid is clearly $2\,\Delta x$, also known as the "Niquist wavelength". This implies that the maximum value that any component of the wave vector can take is $\pi/\Delta x$.

The two roots of this quadratic equation are, clearly:

$$\xi_{\pm} = \frac{-B \pm \left(B^2 - 4AC\right)^{1/2}}{2A} \ , \tag{4.32}$$

and the general solution to the finite difference equation turns out to be:

$$\phi_m^n = \sum_k \left[Z_k^+ \left(\xi_+ \left(k\right)\right)^n + Z_k^- \left(\xi_- \left(k\right)\right)^n \right] e^{imk\,\Delta x} \ , \tag{4.33}$$

where Z_k^+ and Z_k^- are arbitrary constants.

On the other hand, from the fact that $A = C$ one can easily show that:

$$|\xi_+ \xi_-| = \left|\frac{C}{A}\right| = 1 \ . \tag{4.34}$$

This is a very important property, it implies that if the system is stable for all k, that is, if $|\xi_{\pm}(k)| \leq 1$, then the system will necessarily also be non-dissipative (the Fourier modes not only don't grow, they don't decay either). For the system to be stable we must then ask for:

$$\xi_+(k) = \xi_-(k) = 1 \ . \tag{4.35}$$

It is easy to see that this will happen as long as:

$$B^2 - 4AC \leq 0 \ . \tag{4.36}$$

Substituting now the values of the coefficients A, B and C into this expression we find the following stability condition:

$$\rho^2 \left(1 - 2\theta\right) \left[1 - \cos\left(k\Delta x\right)\right] - 2 \leq 0 \ . \tag{4.37}$$

As we want this to hold for all k, we must consider the case when the left hand side reaches its maximum value. If we take $\theta < 1/2$, this will happen for $k = \pi/\Delta x$, in which case the stability condition takes the simple form:

$$\rho^2 \leq 1/\left(1 - 2\theta\right) \ . \tag{4.38}$$

For the explicit scheme we have $\theta = 0$, and the stability condition reduces to the well known Courant-Friedrich-Lewy (CFL) condition:

$$c\Delta t \leq \Delta x \ . \tag{4.39}$$

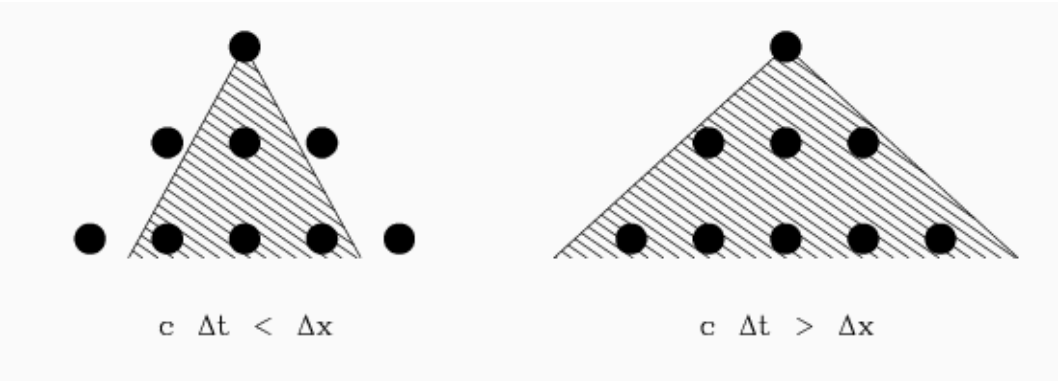

Figure 8: CFL stability condition. For $c\Delta t \leq \Delta x$, the numerical domain of dependence is larger than the physical domain of dependence (shaded region), and the system is stable. For $c\Delta t > \Delta x$ we have the opposite situation, and the system is unstable.

The CFL condition has a clear geometric interpretation: The numerical domain of dependence must be larger than the physical domain of dependence, and not the other way around (see Figure 8). If this wasn't the case, it would be impossible for the numerical solution to converge to the exact solution, since as the grid is refined there will always be relevant physical information that would remain outside the numerical domain of dependence. And, as we have seen, Lax theorem implies that if there is no convergence then the system is unstable.

The argument we have just given clearly only applies to explicit schemes. This is because for an implicit scheme, the numerical domain of dependence is in fact the whole grid. In that case there is no simple geometric argument that can tell us what the stability condition should be.

In order to obtain the stability condition (4.38), we assumed that $\theta < 1/2$. If, on the other hand, we take $\theta \geq 1/2$, then we must go back to the general condition (4.37). However, in this case it is easy to see that the condition is always satisfied. This means that an implicit scheme with $\theta \geq 1/2$ is stable for all value of ρ, that is, it is "unconditionally stable".

This takes us to one of the most important lessons of the theory of finite differencing: Simple schemes not always have the best stability properties. [6]

[6]This is even more so for systems of equations of "parabolic" type (like the heat equation), for which explicit schemes are practically useless since the CFL condition becomes $\Delta t < \Delta x^2$. The problem with this is that it means that if we reduce Δx by half, we must reduce Δt by a factor of four, so integrating to a desired finite time T quickly becomes prohibitive in terms of computer time.

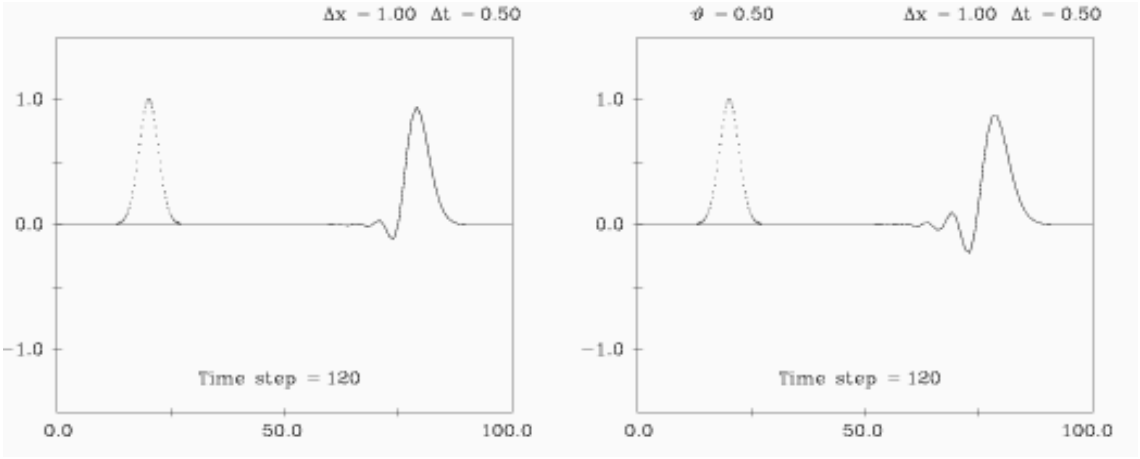

Figure 9: Numerical simulation for the case $\rho = 1/2$, using an explicit scheme (left), and an implicit scheme with $\theta = 1/2$ (right).

4.7 Examples

In order to illustrate the different concepts introduced so far, we will now consider some numerical experiments that use the wave equation and our approximation (4.22). In all the simulations shown here, we will take as computational domain the region $x \in [0, 100]$, and we will use periodic boundary conditions. We will also always take the wave speed c to be equal to 1.

Let us first consider the case when Δx and Δt are such that:

$$\Delta x = 1\,, \qquad \Delta t = 1/2\,. \tag{4.40}$$

The Courant parameter in that case is then $\rho = 1/2$. Figure 9 shows the results of two simulations using these grid parameters. The left panel corresponds to an explicit scheme, while the right panel uses an implicit scheme with $\theta = 1/2$. In both cases the initial data (dotted line) corresponds to a Gaussian wave packet moving to the right, and the results of the simulation are shown after 120 time steps

From the figure we see that both simulations are very similar. In both cases the initial packet has dispersed as it moves, in contrast with the exact solution for which the packet propagates keeping its original shape. Dispersion is a numerical effect, caused by the fact that, in the numerical approximation, different Fourier modes travel at different speeds.

The next simulation corresponds to the same situation, but now taking $\Delta x = \Delta t = 1$, for a Courant parameter of $\rho = 1$. Figure 10 shows the results of this simulation. Notice that since Δt is now twice as large, we only need 60 time steps to reach the same stage.

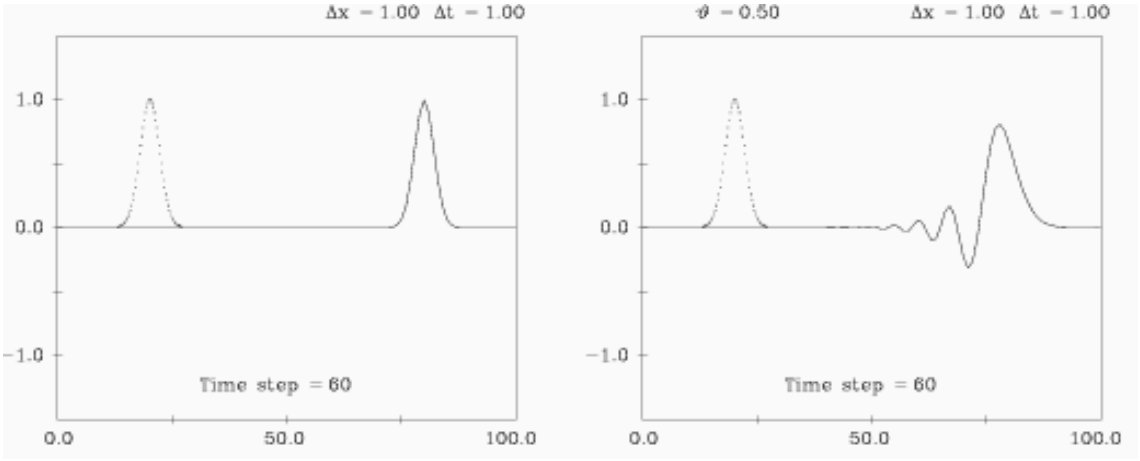

Figure 10: Numerical simulation for the case $\rho = 1$.

The first thing to note is that for the explicit scheme the dispersion has vanished. This is a surprising result, and it only happens for the one-dimensional wave equation: In this case one can in fact show that when $\rho = 1$, all the truncation errors cancel, and the numerical solution is in fact *exact*. The implicit scheme, on the other hand, is still dispersive (even more so than before).

Finally, Figure 11 shows the results of simulations with $\Delta x = 1$ and Δt =1.02, for a Courant parameter of $\rho = 1.02$. For the explicit scheme it is evident that a high frequency perturbation has appeared after only 55 time steps. If we continue the simulation, this perturbation grows exponentially and rapidly dominates the full solution. This is a classic numerical instability: a localized perturbation, generally high frequency (though not always), that grows exponentially without propagating. The implicit scheme, on the other hand, shows no sign of an instability. In fact, we can continue the integration in this case without ever encountering any problem. The dispersion, however, is still present and will eventually cause the numerical solution to be very different from the exact solution.

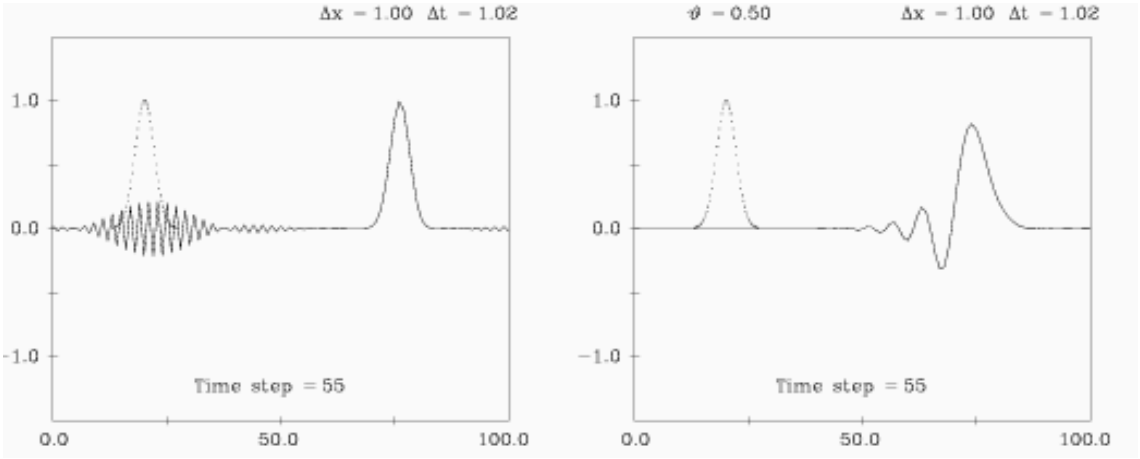

Figure 11: Numerical solution for the case $\rho = 1.02$.

5 Simple applications of numerical relativity

In this section we will consider some simple examples of simulations in numerical relativity. We will consider only two specific cases: relativity in 1+1 dimensions (that is, only one spatial dimension), and relativity in spherical symmetry. The case of three-dimensional systems with no symmetries is considerably more complex and would require a course on its own.

5.1 Toy relativity in 1+1 dimensions

The first example we will consider is that of relativity in 1+1 dimensions, that is, one spatial dimension and one time dimension. Now, differential geometry tells us that for a two-dimensional space as the one we are considering (that is, a surface), the Riemann curvature tensor has only one independent component given essentially by the Gaussian curvature of the surface. If, moreover, we assume that we are on vacuum, the Einstein field equations tell us that this unique component must be zero, which takes us to the well known result that there is no gravity in 1+1 dimensions.

One could then assume that studying this system is completely trivial, and that there is nothing to be learned from it. However, it turns out that this system is ideal for studying the properties of different gauge conditions. This is because even if in one spatial dimension we are limited to Minkowski space, nothing forces us to use the usual coordinates, so we can in fact have highly non-trivial evolutions of the coordinates that will reflect on the evolution of the spatial metric and the extrinsic curvature. For simplicity, in all the following discussion we will assume that the shift vector vanishes.

We start by considering the slicing condition. It is clear that in this case maximal slicing ($K = 0$) would force us to using only flat hypersurfaces, that is, it would restrict us to the standard Minkowski slicing. In order to obtain something less trivial it is preferable to use the Bona-Masso family of generalized harmonic slicing conditions, Eq. (2.57):

$$\partial_t \alpha = -\alpha^2 f(\alpha) K \,, \tag{5.1}$$

where in this case the trace of the extrinsic curvature is just given by $K = K^x_x$.

The ADM evolution equations in the 1+1 case, together with the slicing condition, can now be written in first order form as

$$\partial_t \alpha = -\alpha^2 f K \,, \tag{5.2}$$

$$\partial_t g = -2\alpha g K \,, \tag{5.3}$$

and

$$\begin{aligned}
\partial_t D_\alpha + \partial_x \left(\alpha f K\right) &= 0 \,, & (5.4)\\
\partial_t D_g + \partial_x \left(2\alpha K\right) &= 0 \,, & (5.5)\\
\partial_t K + \partial_x \left(\alpha D_\alpha / g\right) &= \alpha \left(K^2 - D_\alpha D_g / 2g\right) \,, & (5.6)
\end{aligned}$$

where we have defined $g := g_{xx}$, $D_\alpha := \partial_x \ln \alpha$ and $D_g := \partial_x \ln g$.

It turns out that the evolution equation for K can in fact be rewritten as a conservation law in the following way

$$\partial_t \left(g^{1/2} K\right) + \partial_x \left(\alpha D_\alpha / g^{1/2}\right) = 0 \,. \tag{5.7}$$

If we now define the vector $\vec{v} := (D_\alpha, D_g, \tilde{K})$, with $\tilde{K} := g^{1/2} K$, then the evolution equations for the first order variables can be written as a conservative system of the form

$$\partial_t \vec{v} + \partial_x \left(\mathbf{M}\, \vec{v}\right) = 0 \,, \tag{5.8}$$

with the characteristic matrix $\mathbf{M}$ given by:

$$\mathbf{M} = \begin{pmatrix} 0 & 0 & \alpha f / g^{1/2} \\ 0 & 0 & 2\alpha / g^{1/2} \\ \alpha / g^{1/2} & 0 & 0 \end{pmatrix} \,. \tag{5.9}$$

The characteristic matrix has the following eigenvalues

$$\lambda_0 = 0 \,, \qquad \lambda_\pm = \pm \alpha \left(f/g\right)^{1/2} \,, \tag{5.10}$$

with corresponding eigenvectors

$$\vec{e}_0 = (0, 1, 0) \,, \qquad \vec{e}_\pm = \left(f, 2, \pm f^{1/2}\right) \,. \tag{5.11}$$

Since the eigenvalues are real for $f > 0$ and the eigenvectors are linearly independent, the system (5.8) is strongly hyperbolic. The eigenfunctions are given by:

$$\vec{\omega} = \mathbf{R}^{-1} \vec{v} \,, \tag{5.12}$$

with $\mathbf{R}$ the matrix of column eigenvectors. We find (using an adequate choice of normalization):

$$\omega_0 = D_\alpha / f - D_g / 2 \,, \qquad \omega_\pm = \tilde{K} \pm D_\alpha / f^{1/2} \,, \tag{5.13}$$

which can be easily inverted to give:

$$\tilde{K} = \frac{(\omega_+ + \omega_-)}{2} , \tag{5.14}$$

$$D_\alpha = \frac{f^{1/2}\ (\omega_+ - \omega_-)}{2} , \tag{5.15}$$

$$D_g = \frac{(\omega_+ - \omega_-)}{f^{1/2}} - 2\omega_0 . \tag{5.16}$$

Clearly, the eigenfield ω_0 "propagates" with zero speed, while the other two propagate with the "gauge speed" $\lambda^f_\pm = \pm\alpha\sqrt{f/g}$. Given an initial perturbation, we expect that in general this will give rise to two pulses moving in opposite directions.

It is important to notice that with the eigenfunctions scaled as above, their evolution equations also turn out to be conservative and have the simple form:

$$\partial_t \vec{\omega} + \partial_x \left(\mathbf{\Lambda}\, \vec{\omega}\right) = 0 , \tag{5.17}$$

with $\mathbf{\Lambda} = \mathrm{diag}\,\{\lambda_i\}$. If, however, the eigenfunctions are rescaled in the way $\omega'_i = F_i(\alpha, g)\omega_i$, then the evolution equations for the ω'_i will in general no longer be conservative and non-trivial sources will be present. The crucial point is that there is in fact one normalization in which the equations are conservative, namely the one given in (5.13).

Let us now consider some numerical examples. In order to have a non-trivial evolution, we take initial data that corresponds to a curved surface in Minkowski spacetime given in terms of the usual coordinates (t_M, x_M) as:

$$t_M = h\left(x_M\right) . \tag{5.18}$$

If we take the spatial coordinate x in such a way that initially it coincides with the Minkowski coordinate x_M, it is then not difficult to show that the initial values for the metric and extrinsic curvature are:

$$g = 1 - h'^2 \quad \Rightarrow \quad D_g = -2h'h''/g , \tag{5.19}$$

$$K := K^x_x = -h''/g^{3/2} \quad \Rightarrow \quad \tilde{K} = -h''/g . \tag{5.20}$$

To complete the initial data, we take as initial lapse $\alpha = 1$.

In the simulations discussed below, the function $h(x)$ has a Gaussian profile:

$$h = A \exp\left[\left(x - x_c\right)^2 / \sigma^2\right] . \tag{5.21}$$

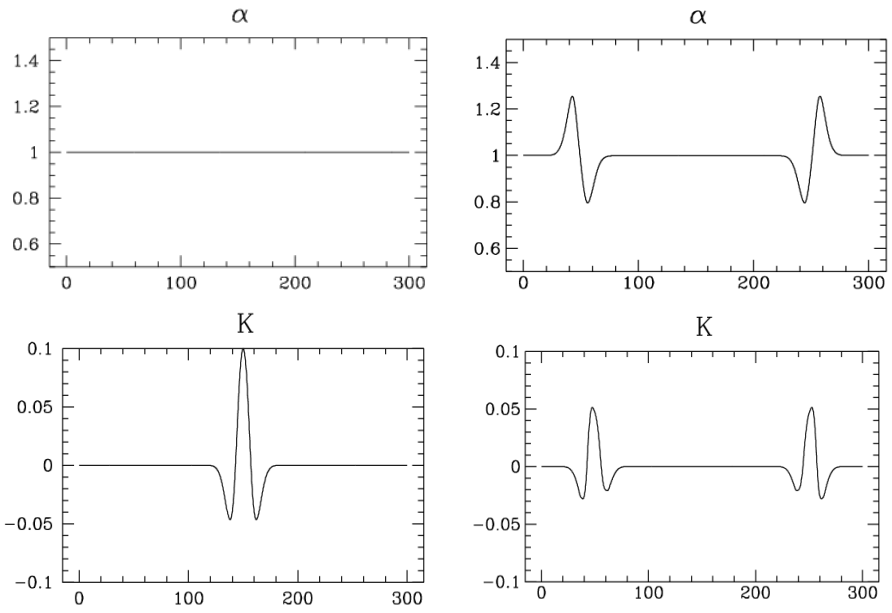

Figure 12: Evolution of the lapse α and the trace of the extrinsic curvature K for a simulation with $f = 1$. Left column: Initial data. Right column: Values at $t = 100$.

In particular, we will take $A = 5$, $\sigma = 10$ and will center the initial perturbation in the middle of the computational domain. Also, in all simulations shown here we will use a time step and a spatial interval given by $\{\Delta t = 0.125, \Delta x = 0.25\}$, and a computational domain in the interval $[0, 300]$.

Let us first consider the case $f = 1$, that is, harmonic slicing. Figure 12 shows the initial data for α and K, and their final values at $t = 100$. From the figure we can see that the initial pulse has separated into two pulses of smaller size traveling in opposite directions with constant speed and keeping their shape. This result is very similar to what one would have expected from the simple wave equation.

When f is a constant different from 1, the situation is no longer so simple. The initial perturbation separates into two pulses as before, but these pulses do not keep their shape as they propagate, and in fact after some time they stop being smooth (they develop infinite gradients). As an example of this, Figure 13 shows results from a simulation with $f = 1.69$ (corresponding to a gauge speed of $v \sim \sqrt{f} = 1.3$). At time $t = 75$ we can see how the pulses in the lapse function have developed very large gradients at the front and at the back, while the pulses in the extrinsic curvature K have developed very sharp spikes. The situation is similar for values of f smaller than 1 (but still larger than zero), except for the fact that in this case the large gradients develop in the middle of each pulse

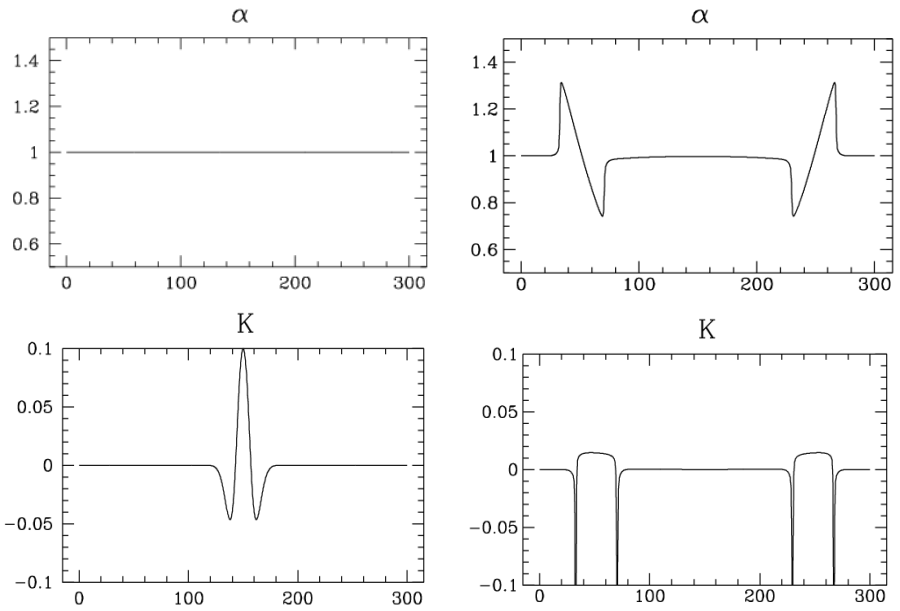

Figure 13: Evolution of the lapse α and the trace of the extrinsic curvature K for a simulation with $f = 1.69$. Left column: Initial data. Right column: Values at $t = 75$.

instead of at their edges.

What is happening for constant values of f different from 1 is a phenomenon know as "gauge shocks", in which the evolution of the coordinate system results in the development of a coordinate singularity at a finite time, associated with the crossing of the characteristic lines. This is a clear example of how perfectly reasonable initial data, together with a well posed set of evolution equations, can nevertheless still give rise to pathological situations. Gauge shocks are a property of the Bona-Masso family of slicing conditions and are not in any way a numerical artifact. A discussion of their origin is outside the scope of these notes, but interested readers can look at references [2, 7, 8, 3].

5.2 Relativity in spherical symmetry

As a second example we will consider the case of spacetimes with spherical symmetry. In this case gravitation is no longer trivial, after all a large number of interesting astrophysical systems (stars, black holes, etc.) have, to a very good approximation, spherical symmetry. However, something that is important to mention is the fact that due to the transverse character of gravitational waves, in spherical symmetry there is no emission of gravitational radiation. In fact, the same thing happens in electromagnetism, where

there are also no spherical (monopolar) electromagnetic waves.

A particularly interesting phenomenon that can be studied in spherical symmetry is the gravitational collapse of different types of matter, and the subsequent formation of black holes. In the early 1990's, this type of studies resulted in the discovery by Choptuik of critical phenomena associated with black hole formation [20].

To start our analysis we first write the spatial metric in spherical symmetry in the following form:

$$dl^2 = A(r,t)dr^2 + r^2 B(r,t)d\Omega^2 \ , \tag{5.22}$$

where A and B are positive functions and where $d\Omega^2$ is the standard solid angle element $d\Omega^2 = d\theta^2 + \sin^2\theta d\phi^2$. Notice that we have already factored out the r^2 from the angular metric, this will make it easier to regularize the equations near the origin.

As we are interested in writing down the evolution equations in first order form, we will introduce the following auxiliary quantities:

$$D_A := \partial_r \ln A \ , \qquad D_B := \partial_r \ln B \ . \tag{5.23}$$

We will also be working with the mixed components of the extrinsic curvature: $K_A := K^r_r$, $K_B := K^\theta_\theta = K^\phi_\phi$.

With this notation, the ADM evolution equations in spherical symmetry take the form (for vacuum and in the case of zero shift vector):

$$\partial_t A = -2\alpha A K_A \ , \tag{5.24}$$

$$\partial_t B = -2\alpha B K_B \ , \tag{5.25}$$

$$\partial_t D_A = -2\alpha[K_A D_\alpha + \partial_r K_A] \ , \tag{5.26}$$

$$\partial_t D_B = -2\alpha[K_B D_\alpha + \partial_r K_B] \ , \tag{5.27}$$

$$\begin{aligned} \partial_t K_A &= -\frac{\alpha}{A}\left[\partial_r(D_\alpha + D_B) + D_\alpha^2 - \frac{D_\alpha D_A}{2}\right. \\ &+ \frac{D_B^2}{2} - \frac{D_A D_B}{2} - A K_A(K_A + 2K_B) \\ &- \left.\frac{1}{r}(D_A - 2D_B)\right] \ , \end{aligned} \tag{5.28}$$

$$\begin{aligned} \partial_t K_B &= -\frac{\alpha}{2A}\left[\partial_r D_B + D_\alpha D_B + D_B^2 - \frac{D_A D_B}{2}\right. \\ &- \left.\frac{1}{r}(D_A - 2D_\alpha - 4D_B) - \frac{2\,(A-B)}{r^2 B}\right] \\ &+ \alpha K_B(K_A + 2K_B) \ , \end{aligned} \tag{5.29}$$

where, just as we did in the previous section, we have defined $D_\alpha := \partial_r \ln \alpha$.

For the hamiltonian and momentum constraints we find, respectively

$$\begin{aligned}\partial_r D_B &= \frac{1}{r^2 B}(A-B) + AK_B(2K_A+K_B) \\ &+ \frac{1}{r}(D_A - 3D_B) + \frac{D_A D_B}{2} - \frac{3D_B^2}{4} , \qquad (5.30)\\ \partial_r K_B &= (K_A - K_B)\left[\frac{1}{r} + \frac{D_B}{2}\right] , \qquad (5.31)\end{aligned}$$

At this point it is important to discuss a problem that is specific to the use of spherical coordinates (something similar happens in cylindrical coordinates). As we can see from the previous equations, at the origin $r = 0$ several terms become singular. In order to solve this problem one needs two ingredients. In the first place, one must fix the regularity conditions of the different dynamical variables, i.e. their behavior close to the origin. It is not difficult to see that if we want all variables to be well defined at the origin, we must necessarily have the following behavior for small r:

$$\begin{aligned}\alpha &\sim \alpha_0 + \mathcal{O}(r^2) , \qquad (5.32)\\ A &\sim A^0 + \mathcal{O}(r^2) , \qquad (5.33)\\ B &\sim B^0 + \mathcal{O}(r^2) , \qquad (5.34)\\ D_\alpha &\sim \mathcal{O}(r) , \qquad (5.35)\\ D_A &\sim \mathcal{O}(r) , \qquad (5.36)\\ D_B &\sim \mathcal{O}(r) , \qquad (5.37)\\ K_A &\sim K_A^0 + \mathcal{O}(r^2) , \qquad (5.38)\\ K_B &\sim K_B^0 + \mathcal{O}(r^2) , \qquad (5.39)\end{aligned}$$

with $\{\alpha_0, A^0, B^0, K_A^0, K_B^0\}$ functions of time, but not of r. These symmetry conditions are in fact very easy to implement numerically. For example, one can use a grid that staggers the origin, starting at $r = \Delta r/2$, and one can later put data on the fictitious point at $r = -\Delta r/2$ by demanding for $\{\alpha, A, B, K_A, K_B\}$ to be even functions at $r = 0$ and for $\{D_A, D_B\}$ to be odd. As $\{D_\alpha, D_A, D_B\}$ are proportional to r close to the origin, we see that terms of the form $D_{\{\alpha,A,B\}}/r$ in the ADM equations are regular and cause no problem (remember that we never evaluate exactly at $r = 0$).

However, there still remains a more serious problem. We can see that in both the hamiltonian constraint and the evolution equation for K_B there is a term of the form $(A-B)/r^2$, while in the momentum constraint there is a term of the form $(K_A - K_B)/r$. Given the behavior of the variables close to $r = 0$, these terms would appear to diverge.

The reason why this is in fact not the case is that, close to the origin, we must also ask for the extra regularity conditions

$$A - B \sim \mathcal{O}(r^2) \,, \qquad K_A - K_B \sim \mathcal{O}(r^2) \,, \tag{5.40}$$

that is, $A^0 = B^0$ and $K^0_A = K^0_B$. These extra conditions come directly from the fact that spacetime must remain locally flat near $r = 0$.

Imposing numerically all regularity conditions at $r = 0$ is not trivial. This is because we have in fact more regularity conditions than variables, so the system is over-determined. In the analytic case it is easy to see that if all conditions are satisfied initially, then they will remain satisfied during the time evolution. Numerically, however, this is no longer true because of truncation errors. This means that very rapidly one of the conditions fails to be satisfied and the evolution becomes ill behaved near the origin (the errors grow without control and the code eventually stops). There are in fact general methods for regularizing the equations in spherical symmetry, but describing them would take too long in these notes. Interested readers can look at [6].

It is important to mention, however, that there is a very simple way, and because of this often used in practice, of avoiding the regularity problem. This method consists in restricting the choice of slicing and using the so called "area gauge", in which one forces the angular metric variable to be $B = 1$. This implies that $g_{\theta\theta} = r^2$, so the area of spheres is always $4\pi r^2$. It is because of this that in this case r is called the "areal radius". If one makes this choice, it is easy to see that by taking the boundary condition $A(r = 0) = 1$, one can in fact solve for $A(r)$ by simply integrating the hamiltonian constraint outward taking $B = 1$ and $D_B = 0$ (and ignoring the evolution equations). In this case the term $(A - B)/r^2$ is no longer a problem. Asking for $B = 1$ during the whole evolution implies that one must have $K_B = \partial_t K_B = 0$, which gives rise to a differential equation for α in terms of r that must be solved at each time step, resulting in what is known as "polar slicing". The area gauge has some disadvantages, however. In the first place, it can only be used for situations in which the area of spheres is a monotonically increasing function of radius, that is, one can not consider "bag of gold" type geometries.[7] Also, this gauge has the disadvantage that it can not penetrate inside apparent horizons associated with black holes. This is because inside an apparent horizon it is impossible to maintain the area of spheres constant without using a non-trivial shift vector.

Another important point is the fact that the ADM evolution equations in spherical symmetry that we have just written down are not strongly hyperbolic when used in

[7] A "bag of gold" geometry corresponds to a space where, as we move away from a certain point, the area of spheres first increases, later decreases, and finally increases again, much as what happens to the circumference of circles as one moves through the inside of a bag that has been tied with a rope.

conjunction with common slicing conditions (maximal, Bona-Masso, etc.). This problem is not difficult to solve, and the interested reader can look at [6] where a simple example is given of how to construct a strongly hyperbolic formulation in spherical symmetry using the constraints (this solution is not unique, and it is not clear which choice is the best among the infinite number possible).

To finish, let us consider as an example a numerical simulation in the case of a simple Schwarzschild black hole. One could think that, since the Schwarzschild spacetime is static, there would be no evolution in a numerical simulation. This is not true for two reasons: In the first place, the Schwarzschild spacetime is only static outside the black hole horizon, inside the horizon the solution is dynamic and the singularity is reached in a finite proper time. Second, as we have already seen in the previous section, even in static spacetimes it is interesting to study the artificial evolution induced by a non-trivial gauge choice. Studying the case of Schwarzschild allows us to acquire valuable experience that can later be used in clearly dynamic situations, like gravitational collapse or the collision of compact objects, where one expects to find black holes during the evolution even if none were present initially.

As always, we must start by the choice of initial data. It is clear that simply taking the Schwarzschild metric at $t = 0$ in not a good choice since this metric is singular at the horizon. A much better choice is to use the metric in isotropic coordinates, which are obtained from the standard Schwarzschild coordinates through the change of radial coordinate $R = r(1 + M/2r)^2$, where R is the Schwarzschild radial coordinate and r is the so-called isotropic radius (see for example [42]). In fact, we have already found the solution for the spatial metric in this case when we discussed the conformal decomposition of York-Lichnerowicz for solving the constraints in sections 2.5 and 2.6. The solution was given by equation (2.49) which we rewrite here:

$$dl^2 = \psi^4 \left(dr^2 + r^2 d\Omega^2 \right) \,, \tag{5.41}$$

where the conformal factor ψ is simply:

$$\psi = 1 + M/2r \,, \tag{5.42}$$

with M the black hole mass. This is the most common choice for Schwarzschild initial data, but certainly not the only one. Another common choice are the so called "Kerr-Schild" coordinates which we will not discuss here (but see [41]).

It is important to mention that, even though the metric above is singular at $r = 0$, this singularity in fact does not correspond to the physical singularity of Schwarzschild spacetime. The physical singularity of Schwarzschild is in the future, and is not contained

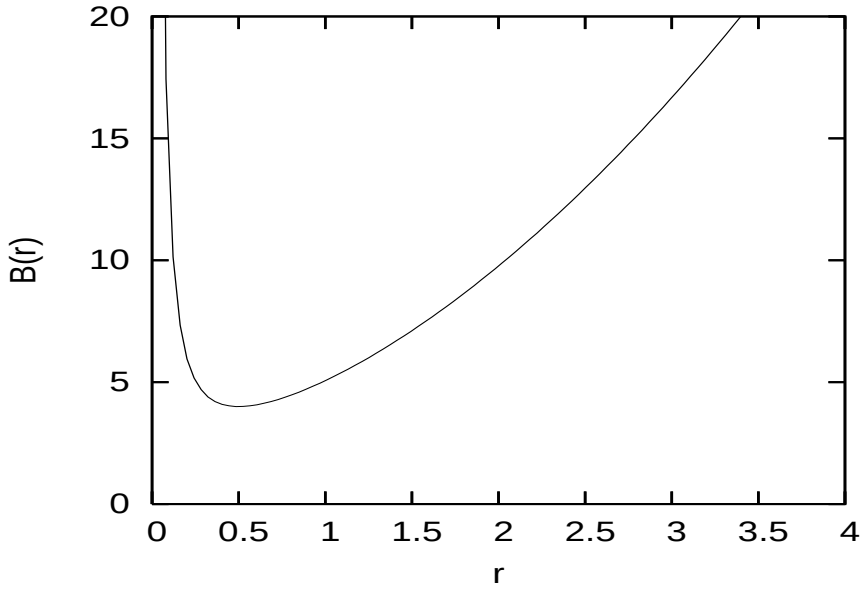

Figure 14: Angular metric function $B = r^2\psi^4$ for a Schwarzschild black hole with $M = 1$ in isotropic coordinates. The throat of the Einstein-Rosen bridge is at $r = 1/2$.

in our initial data. Remember that in isotropic coordinates a black hole is represented topologically by a wormhole (an Einstein-Rosen bridge) between two different universes. This is easy to see if one plots the angular metric $g_{\theta\theta} = r^2\psi^4$, which turns out to be increasing from $r = M/2$ toward infinity, and decreasing (from infinity) from $r = 0$ toward $r = M/2$ (see Figure 14). In these coordinates $r = M/2$ is in fact a coordinate singularity introduced by the fact that we have compactified the asymptotic infinity of the "other" universe to a single point.

In situations like this, when one has a singular conformal factor in the initial data, it is common to extract analytically this factor from the evolution in order to work with regular dynamical variables. We then define new variables as:

$$\tilde{A} := A/\psi^4 , \tag{5.43}$$

$$\tilde{B} := B/\psi^4 , \tag{5.44}$$

$$\tilde{D}_A := D_A - 4\,\partial_r \ln\psi , \tag{5.45}$$

$$\tilde{D}_B := D_B - 4\,\partial_r \ln\psi , \tag{5.46}$$

(the variables K_A and K_B are not rescaled). We now rewrite the ADM equations, or an equivalent strongly hyperbolic reformulation, in terms of the new variables. This tech-

nique based on the evolution of regular rescaled variables together with a static singular conformal factor is known as "puncture evolution", and the point $r = 0$ is commonly called the "puncture" (see for example [19, 9]).

Puncture evolution is not the only technique used to deal with the singularities that exist inside black holes. Another very important technique is based on the idea that the interior of a black hole, where the singularity is located, is causally disconnected from the exterior. This implies that one can simply cut this region from the computational domain without affecting the exterior evolution. This idea, know as "singularity excision" was first suggested in the 1980's (Thornburg attributes the original idea to Unruh in [60, 61]). For simulations with spherical symmetry, singularity excision has been very successful [51, 11, 48, 40, 31, 47, 49, 35], but in the case of fully 3D simulations with no symmetries it has turned out to be extremely complex to implement. However, it is still considered to be the most adequate technique to deal with the interior of a black hole and a lot of work is being done to find proper implementations [51, 11, 5, 10, 53].

In terms of our rescaled variables, the initial data for Schwarzschild are simply:

$$\tilde{A} = \tilde{B} = 1 \ , \qquad \tilde{D}_A = \tilde{D}_B = 0 \ . \tag{5.47}$$

As the Schwarzschild metric is static, and the shift vector is zero in isotropic coordinates, the extrinsic curvature is just:

$$K_A = K_B = 0 \ . \tag{5.48}$$

We still have to choose the gauge condition. For the shift vector we simply choose to keep it equal to zero, while for the lapse we will use maximal slicing (2.56), which in this case reduces to:

$$\frac{1}{\tilde{A}\psi^4}\left[\partial_r^2\alpha + \left(\frac{2}{r} + \tilde{D}_B - \frac{\tilde{D}_A}{2} + 2\partial_r \ln\psi\right)\partial_r\alpha\right] = K_A^2 + 2K_B^2 \ . \tag{5.49}$$

This equation must be integrated numerically every time step. As it is a second order differential equation in r, it requires two boundary conditions. The boundary condition at infinity (or in practice, at the exterior boundary of the computational domain) is just:

$$\partial_r\alpha\big|_{r\to\infty} = \frac{1-\alpha}{r} \ . \tag{5.50}$$

This is a "Robin" type boundary condition that demands that as $r \to \infty$, the lapse behaves as $\alpha = 1 + \mathcal{O}(r^{-1})$. That is, the lapse approaches 1 (its value in flat space) as $1/r$.

The other boundary condition has to be given in the interior, and there are three interesting possibilities. The first possibility is to ask for $\alpha(r=0) = -1$. In this case it is in fact possible to show that there is a exact solution to the maximal equation given by:

$$\alpha = \frac{1 - M/2r}{1 + M/2r} \, . \tag{5.51}$$

This lapse function is anti-symmetric with respect to the throat of the wormhole at $r = M/2$, and is precisely the lapse that gives us a static solution, that is, it is the standard Schwarzschild lapse written in terms of the isotropic radius. This choice is not very interesting for our purposes, since besides taking us to a completely static situation, it does not penetrate inside the black hole horizon.

Another possibility is to ask for:

$$\partial_r \alpha|_{r=0} = 0 \, , \tag{5.52}$$

This no longer results in the Schwarzschild lapse and gives us a dynamical situation (even if the dynamics are just a result of the gauge choice). Also, this case does penetrate the black hole horizon. This is in fact the choice we use in the numerical simulations shown below.

A third possibility is to ask for the lapse to be symmetric at the throat of the wormhole at $r = M/2$. This choice also results in dynamical evolution and penetrates the horizon. In the case of a puncture evolution the throat is not a natural boundary of the computational domain, and this makes the symmetric lapse hard to use. However, it is always possible to change things so that our computational domain starts precisely at $r = M/2$ (by using, for example, a logarithmic radial coordinate $\eta = \ln(2r/M)$), in this way the symmetric lapse becomes a natural choice. This choice of boundary condition is in fact very common in practice and corresponds to maintaining the isometry between both universes at all times.

It is interesting to note that the maximal slicing of Schwarzschild can in fact be studied analytically in the general case, and the differences between the three boundary conditions mentioned can be understood in detail (see [26, 15, 44]).

With the ingredients we have introduced so far we are now ready to do a numerical simulation. The simulation shown here has been performed using a grid with 1000 points in the radial direction, and an interval of $\Delta r = 0.01$. The time step used was $\Delta t = \Delta r/2 = 0.005$, and a total of 4000 time steps have been calculated in order to reach a final time of $t = 10$. The initial data correspond to a black hole of mass $M = 1$, so in these units the simulation reaches a time equal to $t = 10M$. The results of the simulation

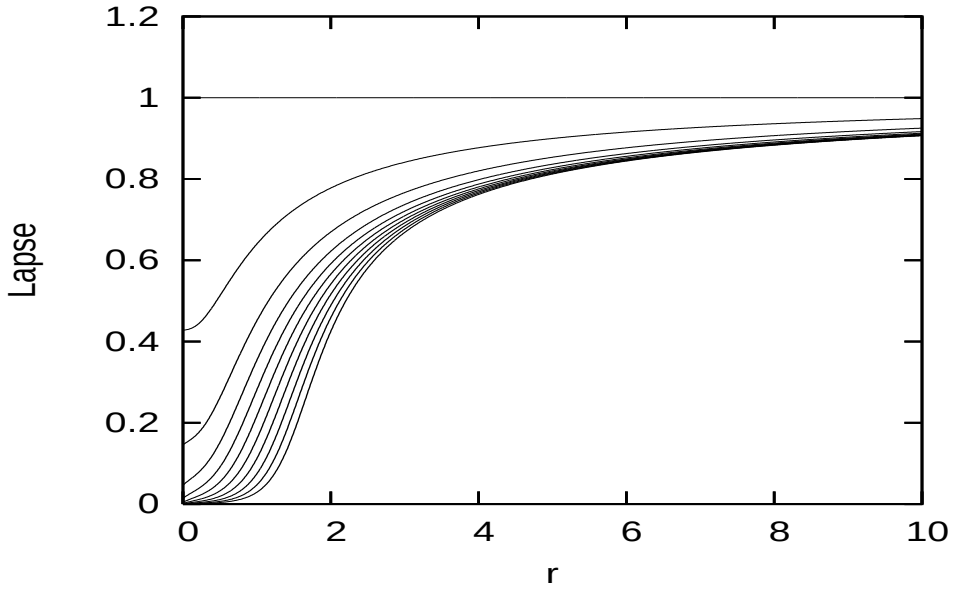

Figure 15: Evolution of the lapse function α. The value of α is shown every $t = 1$. The collapse of the lapse is clearly visible.

are shown in figures 15 and 16, where one can see the evolution of the lapse α and the radial metric A every $t = 1$.

Let us first consider the evolution of the lapse shown in figure 15. From the figure we can see how the lapse evolves rapidly toward zero in the region close to the puncture ($r = 0$). This phenomenon is the well known "collapse of the lapse", and happens because the maximal slicing condition does not allow the volume elements to change, and close to the singularity the only way to achieve this is to freeze the time evolution. In the central regions the lapse in fact approaches zero exponentially.

Figure 16 shows the evolution of the radial metric A. We see how the radial metric is growing in the region close to the black hole horizon. This phenomenon is also well known, and is called "slice stretching", or "grid stretching".[8] What happens here is a combination of two effects. In the first place, as the lapse has collapsed in the central regions time stops moving forward there while it keeps moving forward outside, so the result is that the hypersurfaces stretch. Also, even if the lapse remained equal to 1 during the whole

[8]Historically the name "grid stretching" has been most commonly used. However, the name is misleading, as the phenomenon has nothing to do with the existence of a numerical grid. Because of this more recently the name "slice stretching" has been preferred.

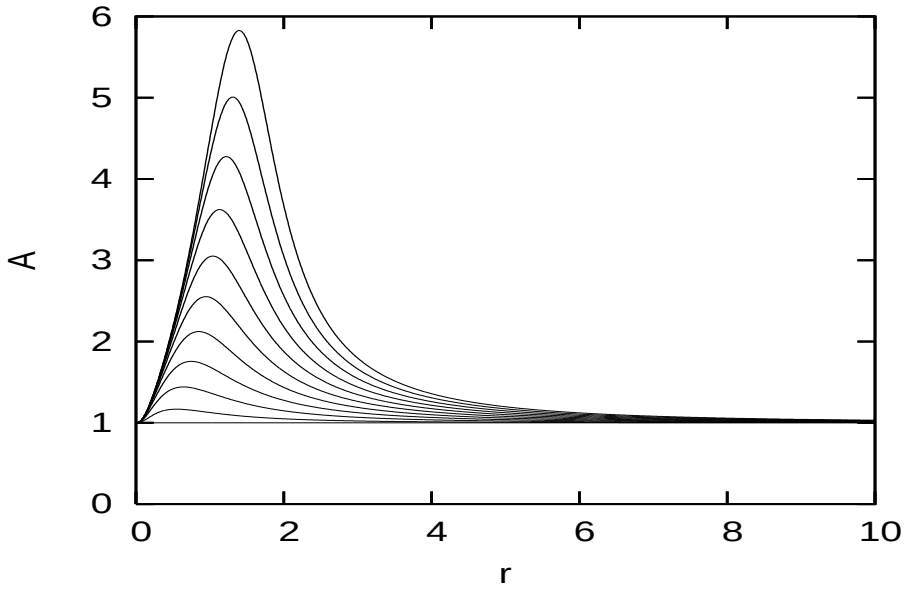

Figure 16: Evolution of the radial metric A. The value of A is shown every $t = 1$. The slice stretching is evident.

evolution, we should expect a growth in the radial metric due to the fact that the normal observers at different distances from the black hole fall with different accelerations, so the distance between them increases (remember that since the shift is zero, our coordinates are tied to these observers).

From the figures, it is clear that with the gauge conditions that we have chosen, Schwarzschild spacetime does not appear static. This can be seen even more dramatically if we study the position of the horizon during the evolution. Figure 17 shows the radius of the black hole horizon r_h as a function of time. At $t = 0$ the horizon is located at $r = 0.5$, as corresponds to the isotropic coordinates (for $M = 1$), but during the evolution the horizon moves outward, and by $t = 10$ it is already at $r \sim 1.8$. This apparent growth is not real, it is simply a coordinate effect caused by the fact that our radial coordinate r is tied to the normal observers, and these observers are falling into the black hole. To convince ourselves that the growth is only apparent, figure 18 shows the horizon area a_h as a function of time during the simulation. We see that the area remains constant at $a_h \simeq 50.2$ for the duration of the run, a value that corresponds to the area of the horizon of a Schwarzschild black hole of unit mass $a_h = 16\pi \sim 50.2$.

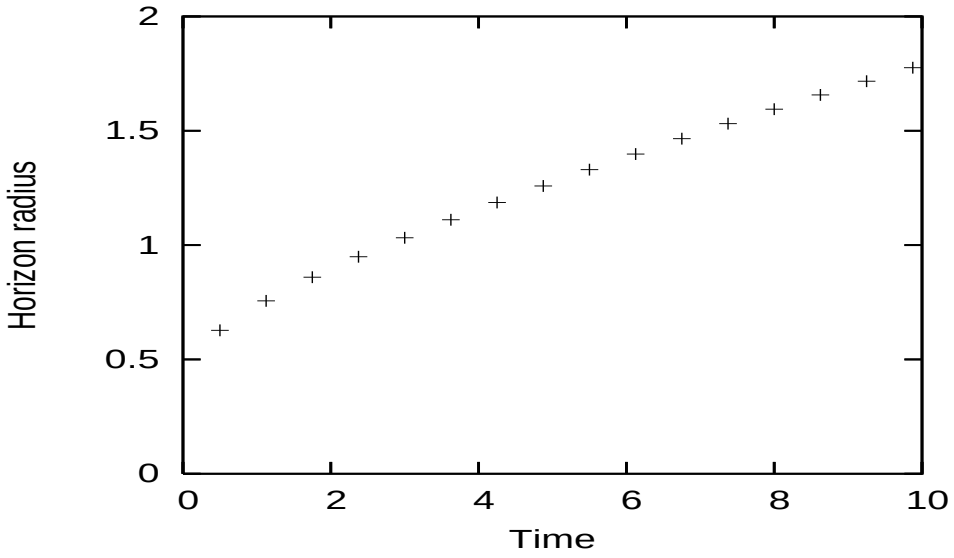

Figure 17: Horizon radius as a function of time. At $t = 0$ the horizon is located at $r = 0.5$, and by $t = 10$ it has moved out to $r = 1.8$.

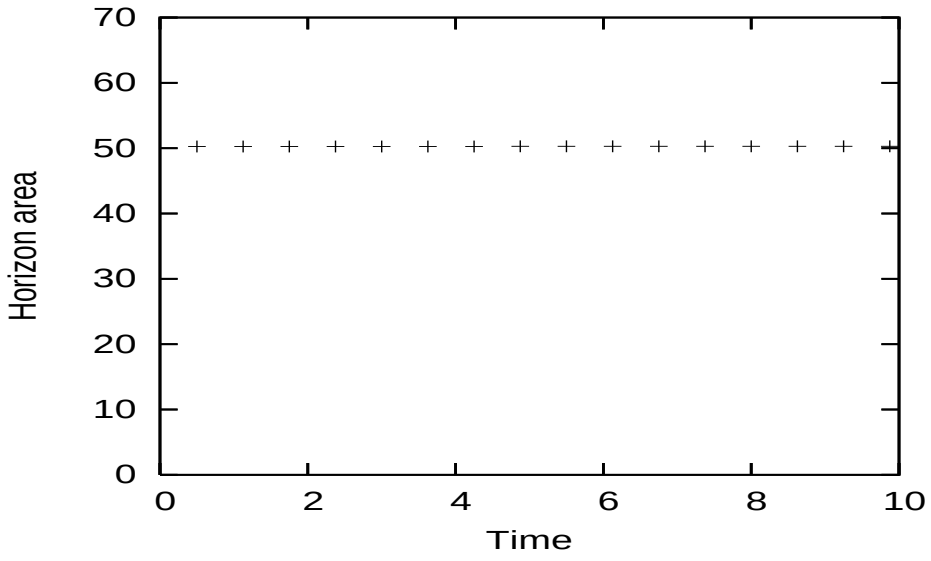

Figure 18: The area of the horizon remains constant during the evolution, with a value close to the theoretical prediction of $a_h = 16\pi \sim 50.2$.

References

[1] A. Abrahams, D. Bernstein, D. Hobill, E. Seidel, and L. Smarr. Numerically generated black hole spacetimes: Interaction with gravitational waves. *Phys. Rev. D*, 45:3544–3558, 1992.

[2] M. Alcubierre. The appearance of coordinate shocks in hyperbolic formulations of general relativity. *Phys. Rev. D*, 55:5981–5991, 1997.

[3] M. Alcubierre. Are gauge shocks really shocks? 2005. gr-qc/0503030.

[4] M. Alcubierre, G. Allen, B. Brügmann, E. Seidel, and W.-M. Suen. Towards an understanding of the stability properties of the 3+1 evolution equations in general relativity. *Phys. Rev. D*, 62:124011, 2000.

[5] M. Alcubierre and B. Brügmann. Simple excision of a black hole in 3+1 numerical relativity. *Phys. Rev. D*, 63:104006, 2001.

[6] M. Alcubierre and J.A. González. Regularization of spherically symmetric evolution codes in numerical relativity. *Comp. Phys. Comm.*, 167:76–84, 2005. gr-qc/0401113.

[7] M. Alcubierre and J. Massó. Pathologies of hyperbolic gauges in general relativity and other field theories. *Phys. Rev. D*, 57(8):4511–4515, 1998.

[8] Miguel Alcubierre. Hyperbolic slicings of spacetime: singularity avoidance and gauge shocks. *Class. Quantum Grav.*, 20(4):607–624, 2003.

[9] Miguel Alcubierre, Bernd Brügmann, Peter Diener, Michael Koppitz, Denis Pollney, Edward Seidel, and Ryoji Takahashi. Gauge conditions for long-term numerical black hole evolutions without excision. *Phys. Rev. D*, 67:084023, 2003.

[10] Miguel Alcubierre, Bernd Brügmann, Denis Pollney, Edward Seidel, and Ryoji Takahashi. Black hole excision for dynamic black holes. *Phys. Rev. D*, 64:061501(R), 2001.

[11] P. Anninos, G. Daues, J. Massó, E. Seidel, and W.-M. Suen. Horizon boundary conditions for black hole spacetimes. *Phys. Rev. D*, 51(10):5562–5578, 1995.

[12] P. Anninos, D. Hobill, E. Seidel, L. Smarr, and W.-M. Suen. Collision of two black holes. *Phys. Rev. Lett.*, 71(18):2851–2854, 1993.

[13] R. Arnowitt, S. Deser, and C. W. Misner. The dynamics of general relativity. In L. Witten, editor, *Gravitation: An Introduction to Current Research*, pages 227–265. John Wiley, New York, 1962.

[14] T. W. Baumgarte and S. L. Shapiro. Numerical integration of Einstein's field equations. *Phys. Rev. D*, 59:024007, 1999.

[15] R. Beig and N. Ó Murchadha. Late time behavior of the maximal slicing of the schwarzschild black hole. *Phys. Rev. D*, 57(8):4728–4737, 1998. gr-qc/9706046.

[16] C. Bona, J. Massó, E. Seidel, and J. Stela. New Formalism for Numerical Relativity. *Phys. Rev. Lett.*, 75:600–603, 1995.

[17] C. Bona, J. Massó, E. Seidel, and J. Stela. First order hyperbolic formalism for numerical relativity. *Phys. Rev. D*, 56:3405–3415, 1997.

[18] S. Bonazzola and J.-A. Marck. Pseudo-spectral methods applied to gravitational collapse. In C. Evans, L. Finn, and D. Hobill, editors, *Frontiers in Numerical Relativity*, pages 239–253. Cambridge University Press, Cambridge, England, 1989.

[19] B. Brügmann. Binary black hole mergers in 3D numerical relativity. *Int. J. Mod. Phys. D*, 8:85, 1999.

[20] M. W. Choptuik. Universality and scaling in gravitational collapse of massless scalar field. *Phys. Rev. Lett.*, 70:9, 1993.

[21] Gregory B. Cook. Initial data for numerical relativity. *Living Rev. Rel.*, 3:5, 2000. http://relativity.livingreviews.org/Articles/lrr-2000-5/index.html.

[22] A. Einstein. Die feldgleichungen der gravitation. *Preuss. Akad. Wiss. Berlin, Sitzber*, pages 844–847, 1915.

[23] A. Einstein. Zur algemeinen relativitätstheorie. *Preuss. Akad. Wiss. Berlin, Sitzber*, pages 778–786, 1915.

[24] K. Eppley. *The Numerical Evolution of the Collision of Two Black Holes.* PhD thesis, Princeton University, Princeton, New Jersey, 1975.

[25] K. Eppley. Evolution of time-symmetric gravitational waves: Initial data and apparent horizons. *Phys. Rev. D*, 16(6):1609–1614, 1977.

[26] Frank Estabrook, Hugo Wahlquist, Steve Christensen, Bryce DeWitt, Larry Smarr, and Elaine Tsiang. Maximally slicing a black hole. *Phys. Rev. D*, 7(10):2814–2817, 1973.

[27] H. Friedrich. The asymptotic characteristic initial value problem for einstein's vacuum field equations as an initial valur problem for a first order quasi-linear symmetric hyperbolic system. *Proc. Roy. Soc. London*, A 378:401–421, 1981.

[28] S. Frittelli and O. Reula. First-order symmetric-hyperbolic Einstein equations with arbitrary fixed gauge. *Phys. Rev. Lett.*, 76:4667–4670, 1996.

[29] GEO600 - http://www.geo600.uni-hannover.de/.

[30] C. Gundlach. Critical phenomena in gravitational collapse. *Living Rev. Rel.*, 2:4, 1999.

[31] C. Gundlach and P. Walker. Causal differencing of flux-conservative equations applied to black hole spacetimes. *Class. Quantum Grav.*, 16:991–1010, 1999. gr-qc/9809021.

[32] S. G. Hahn and R. W. Lindquist. The two body problem in geometrodynamics. *Ann. Phys.*, 29:304–331, 1964.

[33] L. E. Kidder and L. S. Finn. Spectral methods for numerical relativity. the initial data problem. *Phys. Rev. D*, 62:084026, 2000.

[34] L. E. Kidder, M. A. Scheel, and S. A. Teukolsky. Extending the lifetime of 3D black hole computations with a new hyperbolic system of evolution equations. *Phys. Rev. D*, 64:064017, 2001.

[35] L. E. Kidder, M. A. Scheel, S. A. Teukolsky, E. D. Carlson, and G. B. Cook. Black hole evolution by spectral methods. *Phys. Rev. D*, 62:084032, 2000.

[36] Heinz-Otto Kreiss and J. Lorenz. *Initial-Boundary Value Problems and the Navier-Stokes Equations*. Academic Press, New York, 1989.

[37] Luis Lehner. Numerical relativity: A review. *Class. Quantum Grav.*, 18:R25–R86, 2001.

[38] A. Lichnerowicz. L'intégration des équations de la gravitation relativiste et la problème des n corps. *J. Math Pures et Appl.*, 23:37, 1944.

[39] LIGO - http://www.ligo.caltech.edu/.

[40] R. L. Marsa and M. W. Choptuik. Black hole–scalar field interactions in spherical symmetry. *Phys Rev D*, 54:4929–4943, 1996.

[41] Richard A. Matzner, Mijan F. Huq, and Deirdre Shoemaker. Initial data and coordinates for multiple black hole systems. *Phys. Rev. D*, 59:024015, 1999.

[42] C. W. Misner, K. S. Thorne, and J. A. Wheeler. *Gravitation*. W. H. Freeman, San Francisco, 1973.

[43] A. R. Mitchell. *The finite element method in partial differential equations*. J. Wiley and Sons, U.S.A., 1977.

[44] Bernd Reimann and Bernd Brügmann. Maximal slicing for puncture evolutions of Schwarzschild and Reissner - Nordström black holes. *Phys. Rev. D*, 69:044006, 2004.

[45] R. D. Richtmyer and K.W. Morton. *Difference Methods for Initial Value Problems*. Interscience Publishers, New York, 1967.

[46] O. Sarbach, G. Calabrese, J. Pullin, and M. Tiglio. Hyperbolicity of the BSSN system of Einstein evolution equations. *Phys. Rev. D*, 66:064002, 2002.

[47] M. Scheel, T. Baumgarte, G. Cook, S. L. Shapiro, and S. Teukolsky. Numerical evolution of black holes with a hyperbolic formulation of general relativity. *Phys. Rev. D*, 56:6320–6335, 1997.

[48] M. A. Scheel, S. L. Shapiro, and S. A. Teukolsky. Collapse to black holes in Brans-Dicke theory: I. horizon boundary conditions for dynamical spacetimes. *Phys. Rev. D*, 51(8):4208–4235, 1995.

[49] Mark A. Scheel, Thomas W. Baumgarte, Gregory B. Cook, Stuart L. Shapiro, and Saul A. Teukolsky. Treating instabilities in a hyperbolic formulation of Einstein's equations. *Phys. Rev. D*, 58:044020, 1998.

[50] B. Schutz. *A First Course in General Relativity*. Cambridge University Press, 1985.

[51] E. Seidel and W.-M. Suen. Towards a singularity-proof scheme in numerical relativity. *Phys. Rev. Lett.*, 69(13):1845–1848, 1992.

[52] M. Shibata and Takashi Nakamura. Evolution of three-dimensional gravitational waves: Harmonic slicing case. *Phys. Rev. D*, 52:5428, 1995.

[53] Deirdre Shoemaker, Kenneth L. Smith, Ulrich Sperhake, Pablo Laguna, Erik Schnetter, and David Fiske. Moving black holes via singularity excision. *Class. Quantum Grav.*, 20:3729–3744, 2003. gr-qc/0301111.

[54] L. Smarr, A. Čadež, B. DeWitt, and K. Eppley. Collision of two black holes: Theoretical framework. *Phys. Rev. D*, 14(10):2443–2452, 1976.

[55] L. Smarr and J. York. Radiation gauge in general relativity. *Phys. Rev.D*, 17:1945, 1978.

[56] Larry Smarr and James W. York, Jr. Kinematical conditions in the construction of spacetime. *Phys. Rev. D*, 17(10):2529–2552, 15 May 1978.

[57] B. Szilágyi, Roberto Gómez, N. T. Bishop, and Jeffrey Winicour. Cauchy boundaries in linearized gravitational theory. *Phys. Rev. D*, 62:104006, 2000.

[58] B. Szilagyi, B. Schmidt, and Jeffrey Winicour. Boundary conditions in linearized harmonic gravity. *Phys. Rev. D*, 65:064015, 2002.

[59] TAMA - http://tamago.mtk.nao.ac.jp/.

[60] Jonathan Thornburg. Coordinates and boundary conditions for the general relativistic initial data problem. *Class. Quantum Grav.*, 4(5):1119–1131, September 1987.

[61] Jonathan Thornburg. *Numerical Relativity in Black Hole Spacetimes*. PhD thesis, University of British Columbia, Vancouver, British Columbia, 1993.

[62] VIRGO - http://www.virgo.infn.it/.

[63] R. M. Wald. *General Relativity*. The University of Chicago Press, Chicago, 1984.

[64] John A. Wheeler. *A journey into gravity and spacetime*. Scientific American Library, distributed by W. H. Freeman, New York, U.S.A., 1990.

[65] Jeffrey Winicour. Characteristic evolution and matching. *Living Rev. Rel.*, 1:5, 1998. [Online article].

[66] J. W. York. Conformal 'thin-sandwich' data for the initial-value problem of general relativity. *Phys. Rev. Lett.*, 82:1350–1353, 1999.

[67] James W. York. Kinematics and dynamics of general relativity. In Larry L. Smarr, editor, *Sources of Gravitational Radiation*, pages 83–126. Cambridge University Press, Cambridge, UK, 1979.