Numerical Relativity: A Brief Introduction

Brandon Li

March 3, 2025

1 Introduction

In general relativity, the goal is often to find solutions to the Einstein field equations and understand their behavior. Traditionally, people have either sought to find exact solutions to these equations or used perturbation theory to find approximate solutions around an exact background. In the past few decades, however, numerical relativity has emerged as an alternative to these two as it able to give approximate solutions in a wide variety of situations.

The problem with the traditional approaches is that analytic, closed-form solutions are very hard to come by. Furthermore, these spacetimes are usually highly symmetric and idealized. In fact, a lot of interesting physics happens when these symmetries are broken. The quintessential example is, perhaps, the merger of two black holes or neutron stars. In such an event, the period before the merger can be modelled using the post-Newtonian expansion and the period after can be described with perturbation theory [4]. What happens right at the point where the two bodies touch each other cannot be described with either method, but it can be computed using numerical relativity. We will now briefly touch upon the history of this field.

1.1 History

The reason numerical relativity has produced noteworthy results only relatively recently is partly due to its reliance on computing power, which has only become abundant in recent decades. The first numerical calculations were performed in the mid 70's and 80's, after which there were few developments up until the late 90's. After that, the field started rapidly growing in popularity with many new techniques being developed. This culminated in the first non-axisymmetric simulation of a black hole merger through the ringdown phase, performed by Pretorius in 2005. This breakthrough was a turning point for the field as it demonstrated the first realistic and complete calculation of a non-trivial system [5]. Now, before diving in to relativity, let us take a detour through the topic of Maxwell's equations.

1.2 Maxwell's equations and constraints

In any coordinate system, the Einstein equations simply become a set of coupled partial differential equations. We can explore some of the fundamental principles behind finding numerical solutions by studying another simpler set of PDEs - the Maxwell equations. In the absence of sources, these equations uniquely determine how the electromagnetic fields of a system evolve in time. In natural units they take the form

$$\nabla \cdot E = \rho \qquad \nabla \cdot B = 0 \tag{1}$$

$$\partial_t E = \nabla \times B - J \qquad \partial_t B = -\nabla \times E \tag{2}$$

To numerically solve these equations we first have to discretize spacetime. This means choosing a finite set of points (t_i, x_i) to approximate the continuum and approximating derivatives with their finite difference approximations. Let us suppose the t_i are equally spaced with spacing Δt. Using the most basic approximation of ∂_t for equations (2) gives

$$\frac{E(t_i + \Delta t, x_i) - E}{\Delta t} \approx \nabla \times B(t_i, x_i) - J(t_i, x_i) \tag{3}$$

$$\frac{B(t_i + \Delta t, x_i) - E}{\Delta t} \approx -\nabla \times E(t_i, x_i) \tag{4}$$

We see that knowing the E and B fields at t_i allows us to compute the fields at $t_{i+1} = t_i + \Delta t$. Solving for $E(t_i + \Delta t)$ and $B(t_i + \Delta t)$ gives us the following update rules:

$$E(t_i + \Delta t, x_i) = E(t_i, x_i) + \Delta t[\nabla \times B(t_i, x_i) - J(t_i, x_i)] \tag{5}$$

$$B(t_i + \Delta t, x_i) = B(t_i, x_i) - \Delta t[\nabla \times E(t_i, x_i)] \tag{6}$$

where $\nabla\times$ is now some finite difference approximation to the curl. This rule tells us if we know the fields at time t_0, then we can compute the fields at t_1, and then at t_2, and so on. It produces a full numerical solution for all times. Notice, however, that this somewhat crude solution only used the last two of Maxwell's equations. We are left with the two instantiations of Gauss's law that do seemingly nothing.

One way can explain this is by counting variables. If we break all the vectors down into their components, then it becomes apparent that there six variables - three electric field and three magnetic field components, and eight equations that must be obeyed. Since there are more equations than variables, the system is overconstrained. What that means is given initial conditions, just six of the equations is needed to determine the result of the solution. We will call these the *evolution equations* because they are sufficient to evolve the fields forward in time. The other two are what we call *constraint equations*. It can be shown that if E and B satisfy (2) and satisfy (1) at an initial time t_0, then it will also satisfy (1) for all of t. This is easily seen by taking the time derivative and integrating.

$$\partial_t \nabla \cdot E = \nabla \cdot \partial_t E = \nabla \cdot \nabla \times B - \nabla \cdot J = \partial_t \rho \tag{7}$$

$$\implies \int \partial_t \nabla \cdot E \, dt = \int \partial_t \rho \, dt \implies \nabla \cdot E(t) - \nabla \cdot E(t_0) = \rho(t) - \rho(t_0) \tag{8}$$

$$\implies \nabla \cdot E = \rho \tag{9}$$

A similar argument shows $\nabla \cdot B = 0$. This shows that if a constraint equation is satisfied at some time, then it will continue to hold if the evolution equations are obeyed. This is certainly true in theory, but in practice, we will see that in the case of GR, numerical errors will cause the constraint equations to be violated.

2 ADM Formalism

Moving on to general relativity, we will explore the details of the ADM formalism (Arnowitt, Deser, and Misner) which forms the basis of many of the popular formulations of numerical GR.

2.1 3+1 decomposition

We will start by briefly reviewing the 3+1 decomposition of spacetime which we also discussed in class. Let us make the assumption that the spacetime we wish to model is *globally hyperbolic*. There are many equivalent definitions of global hyperbolicity, but the definition we will use here is that the spacetime manifold is globally hyperbolic if it admits a special type of hypersurface called a *Cauchy surface*. A Cauchy surface is just a hypersurface that has exactly one intersection with every maximally extended timelike or null geodesic. Now, there is an important property of globally hyperbolic manifolds that we want to utilize. If M is such a manifold, then there exists a function $t : M \to \mathbb{R}$ such that the level sets $t = C$ are timelike hypersurfaces which foliate the spacetime. The utility of this is that we can use t as a "time" coordinate and put spacelike coordinates on the hypersurfaces, thus decomposing spacetime into space + time.

Let us go into a bit more detail on the coordinate system used in the 3+1 decomposition. Starting with the t coordinate, we can arbitrarily choose three additional coordinates $x^i : i = 1, 2, 3$ that along with t form the *adapted coordinate system*. Fixing t, the x^i coordinates parameterize each spatial slice Σ with coordinate vectors ∂_{x^i} residing in the tangent space of Σ. We define a unit normal vector field $n_\mu = \frac{\nabla_\mu t}{|\nabla_\mu t|}$, as well as the *lapse function* $\alpha = \frac{1}{|\nabla_\mu t|}$ and the *shift vector* $\beta^\mu = (\partial_t)^\mu - \alpha n^\mu$. Intuitively, the shift vector measures how much shearing there is in the coordinate system as t increases, in other words it is the difference between the unit normal and the coordinate t vector. This is depicted in the figure below.

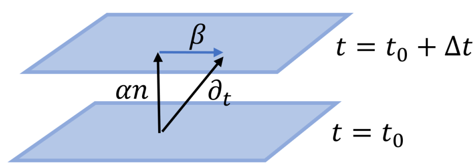

Figure 1: Lapse and shift vectors.

2.2 Math preliminaries

Before we derive the ADM formulation of the Einstein equations, we have to prove a few hypersurface identities. To begin, let us state the Gauss-Codazzi equations for spacelike hypersurfaces.

$$R_{abcd} = \tilde{R}_{abcd} - K_{ad}K_{bc} + K_{ac}K_{bd} \tag{10}$$

$$n^\mu R_{\mu abc} = D_c K_{ab} - D_b K_{ac} \tag{11}$$

Let us take equation (10) and contract the ac and bd indices. We have

$$R_{ab}{}^{ab} = R_{\alpha\beta\gamma\delta} P^\alpha_a P^\beta_b P^{\gamma a} P^{\delta b} = R_{\alpha\beta\gamma\delta}(g^{\alpha\gamma} + n^\alpha n^\gamma)(g^{\beta\delta} + n^\beta n^\delta) \tag{12}$$

$$= R + 2R_{\mu\nu} n^\mu n^\nu. \tag{13}$$

Note that the term $R_{\alpha\beta\gamma\delta} n^\alpha n^\gamma n^\beta n^\delta$ is zero because of the symmetries of the Riemann tensor. We find that the contracted Gauss equation becomes

$$R + 2R_{\mu\nu} n^\mu n^\nu = \tilde{R} + K^2 - K_{\mu\nu} K^{\mu\nu}. \tag{14}$$

Next we contract the ac indices in (11). This yields the contracted Codazzi equation

$$R_{b\mu} n^\mu \equiv P^\nu_b R_{\nu\mu} n^\mu = D_a K^a{}_b - D_b K. \tag{15}$$

Finally, the last equation we need requires a lengthy derivation. Here we will skip some details, but the full derivation can be found in [2]. We start with the Ricci identity along with a result which we will not prove but can be also found in [2]. They state

$$\nabla_\nu \nabla_\sigma n^\mu - \nabla_\sigma \nabla_\nu n^\mu = R^\mu{}_{\rho\nu\sigma} n^\rho \tag{16}$$

$$\nabla_\beta n_\alpha = -K_{\alpha\beta} - D_\alpha \ln \alpha \, n_\beta \tag{17}$$

(Note: The notation may be a bit confusing as α can refer to either the lapse function or an index. The context should distinguish the two). Using $n^\alpha \nabla_\beta n_\alpha = 0$ and $n^\alpha K_{\alpha\beta} = 0$, we find

$$h^\mu_\alpha n^\rho h^\nu{}_\beta n^\sigma R^\mu{}_{\rho\nu\sigma} = h_{\alpha\mu} n^\sigma h^\nu{}_\beta [-\nabla_\nu (K^\mu{}_\sigma + D^\mu \ln \alpha \, n_\sigma) + \nabla_\sigma (K^\mu{}_\nu + D^\mu \ln \alpha \, n_\nu))] \tag{18}$$

$$= h_{\alpha\mu} h^\nu{}_\beta [K^\mu{}_\sigma \nabla_\nu n^\sigma + \nabla_\nu D^\mu \ln \alpha \, + n^\sigma \nabla_\sigma K^\mu{}_\nu + D_\nu \ln \alpha \, D^\mu \ln \alpha] \tag{19}$$

$$= -K_{\alpha\sigma} K^\sigma{}_\beta + \frac{1}{\alpha} D_\beta D_\alpha \alpha + h^\mu{}_\alpha h^\nu{}_\beta n^\sigma \nabla_\sigma K_{\mu\nu} \tag{20}$$

Next, we define $m^\mu = \alpha n^\mu$ and consider the Lie derivative of K w.r.t. m. In coordinates, the definition of Lie derivative states

$$\mathcal{L}_m K_{\alpha\beta} = m^\mu \nabla_\mu K_{\alpha\beta} + K_{\mu\beta} \nabla_\alpha m^\mu + K_{\alpha\mu} \nabla_\beta m^\mu \tag{21}$$

Manipulating this expression a bit and projecting onto the hypersurface yields

$$\mathcal{L}_m K_{\alpha\beta} = \alpha \, h^\mu{}_\alpha h^\nu{}_\beta n^\sigma \nabla_\sigma K_{\mu\nu} - 2\alpha \, K_{\alpha\mu} K^\mu_\beta . \tag{22}$$

Plugging this in to (20) will give us the following equation:

$$h_{\alpha\mu} n^\rho h^\nu{}_\beta n^\sigma R^\mu_{\rho\nu\sigma} = \frac{1}{N} \mathcal{L}_m K_{\alpha\beta} + \frac{1}{\alpha} D_\alpha D_\beta \alpha + K_{\alpha\mu} K^\mu{}_\beta \tag{23}$$

Finally, we can combine this with the Gauss equation (10) to get the last necessary identity:

$$R_{ab} = -\frac{1}{\alpha} (\partial_t - \mathcal{L}_\beta) K_{ab} - \frac{1}{\alpha} D_a D_b \alpha + \tilde{R}_{ab} + K K_{ab} - 2 K_{ac} K^{c\,b} \tag{24}$$

2.3 Decomposition of Einstein equations

At this point we are ready to decompose the Einstein equations into separate components in accordance with the 3+1 decomposition. We start with the basic form of Einstein's equations without cosmological constant,

$$R_{\mu\nu} - \frac{1}{2} R g_{\mu\nu} = 8\pi T_{\mu\nu} . \tag{25}$$

The first step is to project along $n^\mu n^\nu$. Doing so gives

$$R_{\mu\nu} n^\mu n^\nu - \frac{1}{2} R g_{\mu\nu} n^\mu n^\nu = 8\pi T_{\mu\nu} n^\mu n^\nu \tag{26}$$

If we recall that $g_{\mu\nu} n^\mu n^\nu = -1$, we can apply equation (14). Let us also define $\rho \equiv T_{\mu\nu} n^\mu n^\nu$, a kind of "energy density". This equation becomes

$$\tilde{R} + K^2 - K_{ab} K^{ab} = 16\pi \rho , \tag{27}$$

also known as the *Hamiltonian constraint*. We will see in a moment that this is, indeed, a constraint equation. Next, we take the mixed projection of (25) onto $P^\mu_a n^\nu$. The $g_{\mu\nu}P^\mu_a n^\nu$ term vanishes since P and n are orthogonal. Let us define $j_a = P^\mu_a n^\nu T_{\mu\nu}$, the mixed projection of the stress-energy tensor. Using (15) this gives

$$D_a K^a_{\ b} - D_b K = 8\pi j_b. \tag{28}$$

This equation is known as the *momentum constraint*. Now let us take the spatial projection $P^\mu_a P^\nu_b$. Defining $S_{ab} \equiv P^\mu_a P^\nu_b T_{\mu\nu}$, we can apply (24) to get

$$-\frac{1}{\alpha}(\partial_t - \mathcal{L}_\beta)K_{ab} - \frac{1}{\alpha}D_a D_b \alpha + \tilde{R}_{ab} + KK_{ab} - 2K_{ac}K^{cb} - \frac{1}{2}Rh_{ab} = 8\pi S_{ab} \tag{29}$$

Now we wish to express R purely in terms of the dynamical quantities h, K, and/or S. This is done by taking the trace of the Einstein equations which gives $-R = 8\pi T^\mu_\mu = S^a_a - \rho \equiv S - \rho$. This leaves us with the third equation

$$(\partial_t - \mathcal{L}_\beta)K_{ab} = -D_a D_b \alpha + \alpha(\tilde{R}_{ab} + KK_{ab} - 2K_{ac}K^{cb}) + 8\pi\alpha(\frac{1}{2}(S-\rho)h_{ab} - S_{ab}). \tag{30}$$

Finally, the last equation is obtained starting with the identity $K_{ab} = \frac{1}{2}\mathcal{L}_n h_{ab}$ and substituting $n^\mu = \frac{1}{\alpha}[(\partial_t)^\mu - \beta^\mu]$. We obtain

$$(\partial_t - \mathcal{L}_\beta)h_{ab} = -2\alpha K_{ab}. \tag{31}$$

Putting all this together, we get the two constraint equations

$$\tilde{R} + K^2 - K_{ab}K^{ab} = 16\pi\rho, \tag{32}$$

$$D_a K^a_{\ b} - D_b K = 8\pi j_b. \tag{33}$$

These must be satisfied at all times for any solution of the Einstein equations. We also get the two evolution equations

$$(\partial_t - \mathcal{L}_\beta)h_{ab} = -2\alpha K_{ab}, \tag{34}$$

$$(\partial_t - \mathcal{L}_\beta)K_{ab} = -D_a D_b \alpha + \alpha(\tilde{R}_{ab} + KK_{ab} - 2K_{ac}K^{cb}) + 8\pi\alpha(\frac{1}{2}(S-\rho)h_{ab} - S_{ab}). \tag{35}$$

Let us get a sense for why we have categorized each of these as either constraint or evolution equations. Starting with the latter two, it can be shown that the Lie derivative $\mathcal{L}_\beta$ can be expressed only in terms of spatial derivatives of α and the hypersurface Christoffel symbols Γ^{jk}_i (which themselves only depend on spatial derivatives of h_{ab}) [2]. That means these equations are of the form $\partial_t h_{ab} = [\dots]$ and $\partial_t K_{ab} = [\dots]$ where $[\dots]$ depends only on the quantities K, h, α, and β at time t. So, like how we were able to evolve an EM field using two of Maxwell's equations in section 1.2, so too do the two constraint equations here allow us to compute an entire solution of Einstein's equations using just an initial slice of data. Analogously, the constraint equations contain no time derivatives, and if they are satisfied at some t, then they will continue to be satisfied forever afterwards.

We now have everything we need to solve the equations on a computer. It seems simple - provide initial data and update the fields using the evolution equations. When this was actually attempted, however, things went wrong quite quickly. It turns out that small numerical errors caused by rounding and truncation quickly cause the solution to blow up. The fundamental reason for this is that ADM is not a well posed system, a fact which was shown in the early 90's [3]. What this means is that there may not exist a unique solution that smoothly depends on the initial conditions. To get around this, people have tried different approaches. In the next two sections we will discuss two approaches to finding a functional numerical method.

3 BSSN Formalism

The BSSN (Baumgarte, Shapiro, Shibata and Nakamura) formalism is a modification of ADM developed in the mid 90's with the advantage of being a strongly hyperbolic system as opposed to the weakly hyperbolic nature of ADM. It is derived by first introducing three new variables: First we define $\tilde{h}_{ab} = \chi h_{ab}$ where χ is a conformal rescaling factor that makes $\det \tilde{h}_{ab} = 1$. Next, we define a conformally scaled, traceless version of the extrinsic curvature, $\tilde{A}_{ab} = \chi(K_{ab} - \frac{1}{3}h_{ab}K)$. Finally, define $\tilde{\Gamma}^i = \tilde{h}^{ab}\tilde{\Gamma}^i_{ab}$, a contraction of the Christoffel symbols. This comes with two new constraints. The first constraint is $\det \tilde{h}_{ab} = 1$, which obviously must be true by definition. The second constraint is $\tilde{h}^{ab}\tilde{A}_{ab} = 0$ which comes from $\tilde{A}$ being traceless. Working through the math leads to a modified set of evolution and constraint equations that dictate how the fields $\chi, \tilde{h}, K, \tilde{A}$ and $\tilde{\Gamma}$ evolve. We will not cover either the derivation nor the resulting equations since they are quite large and messy, the details can be found in [7]. We will, however, discuss the implications of this method.

The reason we included BSSN here is that it actually works. When implemented on a computer, numerical solutions do not blow up. Strong hyperbolicity means solutions are stable, and BSSN has been used to successfully model black hole mergers without the need for excision [4]. One downside of BSSN is that it exhibits constraint violation, meaning even if a solution satisfies the constraints at an initial time, numerical errors will cause gradual loss of constraint satisfaction. The next section describes a method that does not have this issue. Even so, BSSN is a still relatively popular technique used in research.

4 Generalized Harmonic Gauge

The generalized harmonic gauge (GHG) aims to solve some of the issues described previously. The foundations for this method were developed by Garfinkle in 2001 and developed further by Pretorius in 2004 [1][6].

4.1 Formulation

The way we generalize the standard harmonic gauge condition $\Box x^\alpha = 0$ is by adding a source term $\Box x^\alpha = H^\alpha$. Crucially, we will promote H^α to a dynamical variable which satisfies the constraint equation $C^\alpha = H^\alpha - \Box x^\alpha = 0$, and this comes simply from the definition of H. In this case, the Einstein equations can be rewritten as

$$\frac{1}{2}g^{\mu\nu}\partial_\mu\partial_\nu g_{\alpha\beta} = -\partial_\nu g_{\mu(\alpha}\partial_{\beta)}g^{\mu\nu} - \partial_{(\alpha}H_{\beta)} + H_\mu\Gamma^\mu_{\alpha\beta} - \Gamma^\mu_{\nu\alpha}\Gamma^\nu_{\mu\beta} - 8\pi(T_{\mu\nu} - \frac{1}{2}Tg_{\alpha\beta}) \tag{36}$$

We see that the left side of this equation is just the wave operator applied to $g_{\alpha\beta}$, therefore these equations are manifestly hyperbolic. This is promising, but in practice we still see the problem of constraint violation. Numerical errors means C^α may become non-zero even if they are supposed to be zero.

4.2 Constraint damping

In this case, the constraint violation problem was successfully solved by adding damping terms that make the violations go away. The damping terms take the form $\kappa[2n_{(\alpha}C_{\beta)} - \lambda g_{\alpha\beta}n^\mu C_\mu]$ where κ

and λ are constants which control the rate of damping, much like the diffusivity constant in the heat equation. Added to the Einstein equations, they look like

$$\frac{1}{2}g^{\mu\nu}\partial_\mu\partial_\nu g_{\alpha\beta} = -\partial_\nu g_{\mu(\alpha}\partial_{\beta)}g^{\mu\nu} - \partial_{(\alpha}H_{\beta)} + H_\mu\Gamma^\mu_{\alpha\beta} - \Gamma^\mu_{\nu\alpha}\Gamma^\nu_{\mu\beta} - 8\pi(T_{\mu\nu} - \frac{1}{2}Tg_{\alpha\beta}) \tag{37}$$

$$-\kappa[2n_{(\alpha}C_{\beta)} - \lambda g_{\alpha\beta}n^\mu C_\mu]. \tag{38}$$

With this modification, the GHG formulation has seen great success in practice. This method formed the numerical foundation for Pretorius's unprecedented calculation of a black hole merger in 2005. In addition, GHG with constraint damping has been used in the SXS (Simulating eXtreme Spacetimes) project partly developed here at Cornell.

5 Conclusion

Numerical relativity is an exciting field of research since it has the potential to produce so many unique and interesting insights. There is still much to be done, with many avenues of research that haven't been explored yet. In this paper, we have seen several techniques each with their own strengths and weaknesses, but we have only covered a fraction of what is out there. The future of this field seems quite bright. As computers continue to improve in power, it seems likely that we will discover new things about general relativity that no one has imagined yet.

References

[1] David Garfinkle. Harmonic coordinate method for simulating generic singularities. *Physical Review D*, 65(4), jan 2002.

[2] Eric Gourgoulhon. 3+1 formalism and bases of numerical relativity, 2007.

[3] Carsten Gundlach and José M. Martín García. Well-posedness of formulations of the einstein equations with dynamical lapse and shift conditions. *Physical Review D*, 74(2), jul 2006.

[4] Carlos Palenzuela. Introduction to numerical relativity. *Frontiers in Astronomy and Space Sciences*, 7, sep 2020.

[5] Frans Pretorius. Evolution of binary black-hole spacetimes. *Physical Review Letters*, 95(12), sep 2005.

[6] Frans Pretorius. Numerical relativity using a generalized harmonic decomposition. *Classical and Quantum Gravity*, 22(2):425–451, jan 2005.

[7] Ulrich Sperhake. Introduction to numerical relativity, 2016.