GR meets astrophysics

Luis Lehner

July 20, 2016

1 Lecture 1

We want to make connection between GR and the plasma astrophysics. Why do we need this when we have gravitational waves now? There are a lot of limitations of gravitational waves. It is limited due to characteristics of signals and detectors. At higher frequency noise grows linearly, and sky location is very poor. Strong gravity effects in astrophysics are all tied to some analytical models, and are limited in scope. We also are interested in the EM connections to the gravitational wave signals.

Therefore we want to study mergers of compact objects within surrounding plasma. Two binary non-rotating black holes disturbs the spacetime enough that it launches jets from the surrounding plasma and allow the plasma to tap into the kinetic energy of the rotation.

We will be talking about the challenges and problems simulating and controlling dynamic space time. Let's go slowly from the beautiful theory of Einstein and put it into the computer to compute it dynamically.

1.1 Well-posedness

"Theorems are permanent, tricks are ephemeral."

We start with the theorem by Hadamard: any problem about a physical system should be well posed. By well posed (WP), we mean:

- Solution exists
- The solution to the problem is unique
- The solution must depend continuously on initial and boundary data

By the last statement we mean that the magnitude of solution in a general sense should

$$|u|_T \leq |u|_{t=0} K e^{\beta T} \tag{1.1}$$

where K and β should not depend on initial or boundary data. If any problem we are solving does not satisfy any of the statements we should throw it away immediately.

Lets talk about the sufficient and necessary conditions for WPness. If we can pose the problem in this form

$$u_{,t} = \sum A^i \partial_i u + \text{"Rest"} \tag{1.2}$$

then the sufficient condition is that the matrix A is diagonalizable, and the eigenvalues are real.

A simple example is $u_{,tt} = u_{,xx}$. If we define $f = u_{,x}$ and $g = u_{,t}$ then we can define the dynamic variables as u, f, g and it satisfy the sufficient condition.

Consider now this system

$$u_{,t} = u_{,x} + v_{,x}, \quad v_{,t} = u_{,x} \tag{1.3}$$

This system has A in a Jordan block form. It turns out that this system admits the solution in the form of

$$u = DF_1(t+x) \cdot t + F_2(t+x), \quad v = F_1(t+x) \tag{1.4}$$

the first part of u grows linearly with time. This solution is okay and satisfies our WPness condition, but the growth with time is alarming. This is called a weakly hyperbolic system, and we now need to care about the “Rest” of the problem. For example, if we couple this, and add u to $v_{,t}$, then the solution of v becomes

$$v = c_1 e^{c_1(t+x)} + c_2 e^{t\sqrt{c_2}} \tag{1.5}$$

and this solution is exponentially unstable to changes to initial condition, which is a badly posed problem. We especially want to avoid this in a numerical system because the initial conditions are never exact.

1.2 Crash Course in GR

How do we make sure our complex problem is well posed when such a simple one can be bad? Lets consider the Einstein equation

$$G_{ab} = 8\pi T_{ab} \tag{1.6}$$

The left hand side part has form of $\partial^2 g$, whereas the right hand side depends on the fluid quantities $\mathcal{F}$ and g. We also have the constraint equation of the fluid $\nabla_a T^{ab} = 0$, which concerns $\partial\mathcal{F}$ and ∂g. For our purposes they are decoupled, so we are allowed to treat T_{ab} as the “Rest”.

Lets say some words about gravity quickly. We have the metric tensor which describes the spacetime manifold g_{ab}. This determines a length

$$ds^2 = g_{ab} dx^a dx^b \tag{1.7}$$

Note that we use Einstein summation notation to add the indices a and b. Both indices take values $0, 1, 2, 3$ where 0 is for time part and others are for spatial part.

If we have a curved manifold, when we take derivatives we need to take into account the change of manifold in addition to the change of the field. We define the covariant derivative ∇_a which does exactly that. We have by definition

$$\nabla_a g_{bc} = 0 \tag{1.8}$$

The covariant derivative of a vector is defined as

$$\nabla_a v^b = \partial_a v^b + \Gamma^b{}_{ac} v^c \tag{1.9}$$

Some other definitions are

$$\nabla_a f = \partial_a f, \quad \nabla_a v_b = \partial_a v_b - \Gamma^c{}_{ab} v_c \tag{1.10}$$

Now we need to define what the Γ symbol is. It is called the Chrsitoffel symbol

$$\Gamma^a{}_{bc} = \frac{1}{2} g^{ae}(g_{be,c} + g_{ce,b} - g_{bc,e}) \tag{1.11}$$

Now we can define the so-called Riemann curvature tensor:

$$R^a{}_{bcd} = \partial_c \Gamma^a{}_{bd} - \partial_d \Gamma^a{}_{bc} + \Gamma^e{}_{bd}\Gamma^a{}_{ec} - \Gamma^e{}_{bc}\Gamma^a{}_{ed} \tag{1.12}$$

This curvature tensor describes the curvedness of the spacetime manifold.

If we have two world lines along the manifold, and the acceleration between them can be characterized by the Riemann tensor

$$a^a = -R^a{}_{cbd}\xi^c\xi^d L^b \tag{1.13}$$

where $\xi^c = dx^c/d\lambda\ L^b$ is the distance between the world line. A funny fact about the curved manifold is that derivatives depend on paths, and the order of taking them, and the Riemann tensor characterizes this:

$$[\nabla_b\nabla_c - \nabla_c\nabla_b]\,v_a = R^d{}_{abc}v_d \tag{1.14}$$

We finally define the Einstein tensor

$$G_{ab} = R_{ab} - \frac{1}{2}g_{ab}R, \quad R_{ab} = R^d{}_{adb}, \quad R = g^{ab}R_{ab} \tag{1.15}$$

1.3 Specializing into a Formalism

Now lets get into how to solve these equations in the computer. In a sense our problem is ill formed from the get go because everything depends on our choice of coordinates. So we need to find a formulation.

The one we introduce now is called the ADM formulation. Given a spacetime we can foliate it with space-like slices Σ_{t_i}. A way to understand space-like hypersurfaces is that its norm at any place n^a is time-like: $n_a n^a = -1$. Here we implicitly choose our signature to be $(-1, 1, 1, 1)$ so that the Minkowski spacetime is $ds^2 = -dt^2 + dx^2 + dy^2 + dz^2$. Time-like vectors are inside the lightcone and space-like vectors are outside.

We can now label these hypersurfaces with t equal to constant. Because the hypersurfaces have constant t, we can define $n_a = \nabla_a t = \partial_a t$. If we go from the hypersurface of t_1 to the next hypersurface of t_2. For a typical observer, he will not measure really time $t_2 - t_1$, but some time proportional to that:

$$\tau = \alpha(t_2 - t_1) \tag{1.16}$$

where α is called the lapse function. The vector that points from a point (t_1, x, y, z) to a point (t_2, x, y, z) is what we call ∂_t. However this may not coincide with the vector n_a we defined above, so we define a new vector

$$\beta^a = \partial_t^a - n^a \tag{1.17}$$

which is called the shift vector.

Now we can split up the metric tensor

$$ds^2 = -\alpha^2 dt^2 + h_{ij}\left(dx^i + \beta^i dt\right)\left(dx^j + \beta^j dt\right) \tag{1.18}$$

where h_{ij} will be the metric on the hypersurface.

We can find the relationship between the metric tensor of the individual hypersurfaces and the full metric:

$$h_{ab} = g_{ab} + n_a n_b \tag{1.19}$$

This metric has several properties:

- $h_a{}^b h_b{}^c = h_a{}^c$
- $h_{ab}n^a = 0$

- $h_{ab}S^b = S_a$

where S^b is a spacelike vector. So h_{ab} is like a projection operator onto the hypersurface.

We need one more concept so that we can foliate the spacetime. This is the so-called extrinsic curvature

$$K_{ab} = -\perp\perp \nabla_a n_b = -h_a{}^{\alpha} h_b{}^{\beta} \nabla_\alpha n_\beta \tag{1.20}$$

This is a useful quantity to measure the rate of change of the hypersurface when embedded into the whole manifold as one goes with the flow. It can be shown that

$$\partial_t h_{ab} = -2\alpha K_{ab} + \beta^l \partial_l h_{ab} + h_{al}\partial_b \beta^l + h_{bl}\partial_a \beta^l \tag{1.21}$$

Now we are equiped to look at numerical relativity. One takes the Einstein equations and take different projections. If we take a vector and dot it $v^a n_a = v_n$ to get the component along the direction of n. We can also dot it with h to get its projection to the hypersurface. If we dot the Einstein equation with $n^a n^b$ then we get

$$^{(3)}R + K^2 - K_{ij}K^{ij} = 16\pi\rho \tag{1.22}$$

Here ρ is the energy density that is measured by the observers that are moving normal to the spatial hypersurface, $\rho = T_{ab}n^a n^b$. $^{(3)}R$ is the intrinsic curvature of the hypersurface.

We can also imagine dotting the Einstein equation with $n_a h_c{}^b$:

$$D_b(K^{ab} - h^{ab}K) = 8\pi J^a \tag{1.23}$$

where D is the intrinsic covariant derivative along the hyper surface $D_b h_{ac} = 0$. Again J is the momentum measured by the same observers.

These two equations have no time derivatives and are called constraint equations. We have one final projection by $\perp\perp$

$$\begin{aligned}\partial_t K_{ab} &= \beta^l \partial_l K_{ab} + K_{al}\partial_b \beta^l + K_{bl}\partial_a \beta^l \\ &\quad - D_a D_b \alpha + \alpha\left[{}^{(3)}R_{ab} + KK_{ab} - 2K_{ad}K^d{}_b - 8\pi\left(S_{ab} - \frac{h_{ab}}{2}(S-\rho)\right)\right]\end{aligned} \tag{1.24}$$

We also have

$$\partial_t h_{ab} = \beta^l \partial_l h_{ab} + h_{al}\partial_b \beta^l + h_{bl}\partial_a \beta^l - 2\alpha K_{ab} \tag{1.25}$$

From these we know that the spatial curvatures have second order time derivatives, which makes it into a hyperbolic system, but the rest unfortunately turns the system into weakly hyperbolic.

1.4 Overcoming the Weakly Hyperbolic Problem

There is a whole bunch of choices of α and β, but lets first discuss the problem that this system is weakly hyperbolic. Lets take a step back and write down R_{ab}. Notice that taking the trace of the Einstein equation gives $R = -T$. So we can write

$$R_{ab} = T_{ab} - \frac{1}{2}g_{ab}T \tag{1.26}$$

The Ricci tensor has a very provocative form

$$R_{ab} \longrightarrow g^{cd}g_{ab,cd} - 2\nabla_{(a}\Gamma_{b)} + \Gamma\Gamma + \partial g \partial g + \dots \tag{1.27}$$

Since we are interested in the hyperbolic part, the first term is perfect which is like the wave equation, but the second term is nasty. Due to human nature we want to throw away things we don't like, so lets try to choose α and β such that $\Gamma_b = 0$ then the second term is zero. Does this do the job?

Suppose we have the equation

$$\Box\phi = \sum A\partial_J\phi\partial_R\phi \tag{1.28}$$

which is actually what happens with the above equation for R_{ab} and it turns out it has nice properties. Another problem people found is that $\Gamma_b = 0$ gives $\nabla_a\nabla^a x^b = 0$ which means coordinates fly away at speed of light.

Now do we need to give up our coordinates to solve the problem? What we do is to put a source $\Gamma_b = H_b$, so that $-\nabla_a\nabla^a x^b = H^b$. However now H_b shows up in the equation, but if we have a hyperbolic system for H where $\nabla_a\nabla^a H_b =$ "something" then we can close the system.

The remaining challenge is to satisfy the constraints, since a small departure from the constraint due to numerics then we might go completely away from the solution. Therefore we need to find a way to bring the solution back to the constraint hypersurface when things go wrong.

Imagine at some point we choose the coordinates as discussed above, and there is a deviation $C^a = H^a + \Gamma^a$, then our system becomes

$$R_{ab} - \nabla_{(a}C_{b)} = 0 \tag{1.29}$$

we can substitute the constraint equation $\nabla^b G_{ab} = 0$ into the above equation we get

$$0 = \nabla^a\nabla_a C_b + C^a\nabla_{(b}C_{a)} \tag{1.30}$$

so we have a wave equation for C. In theory if we start from zero then it will remain zero. But in reality we do see problems.

Consider instead

$$0 = \nabla^a\nabla_a C_b + C^a\nabla_{(b}C_{a)} + \gamma_0\left[t_{(a}C_{b)} - \frac{1}{2}g_{ab}t^c C_c\right] \tag{1.31}$$

where γ^0 is a numerical parameter and t a timelike vector field. If we do the same thing as above we have now

$$0 = \nabla^a\nabla_a C_b - 2\gamma_0\nabla^a\left[t_{(a}C_{b)}\right] \tag{1.32}$$

This is now a damped wave equation and will damp away deviation from the constraint on a time scale determined by γ_0.

2 Lecture 2

Why insist in hyperbolic? Why 1st order formulation? What is the difference between GR and hydro?

The reason we insist in strong/symmetric/strict hyperbolicity, is that we want local solutions to exist. However, we can't say anything about global solutions, since it is still an open problem. The reason we write every equation in first order formalism is that, if we have a set of eigenvectors, then we can diagonalize the system. We can write the equations into

$$U_{,t} = A\gamma_i U \longrightarrow T^{-1}U_{,t} = DT^{-1}\partial_i U \tag{2.1}$$

If the transformation matrices do not depend on position we can make $T^{-1}U$ our new unknowns, so now we can now easily do the propagation and evolution. Therefore this property of 1st order and diagonalizable is sufficient but not necessary for a WP problem.

On the problem of constraints. In MHD we have $\nabla \cdot \mathbf{B} = 0$, which has physical meaning: there is no monopoles. Therefore it is very important physically to respect this constraint, and failing that will probably lead to unphysical behavior. However in GR, we have energy and momentum conservation, but we don't have a very good physical motivation for the constraints apart from that. In GR we have ten degrees of freedom but 8 of the equations are some kind of constraints, only 2 of them physical evolution equations. This is kind of unfortunate.

Comparing hydro and GR, they actually both reduce to first order form. In this form there is almost no difference between the hydro equations and the GR equations. If we take the largest eigenvalue in the diagonalized matrix, it determines how fast things propagate and determines the dynamics of the system. However the one thing that is different between these two theories is that in hydro these eigenvalues may depend on the fluid property itself. This is the reason we can have shocks, etc. However in GR we have the form $g\partial^2 g$, and in mathematical terms it is called linearly degenerate. When solved GR has no modes with speed depending on the state of the system. Therefore the hydro equations are referred to as truly nonlinear equations, but GR equations are linearly degenerate nonlinear theories. The worse things can happen in GR are singularities.

The CFL condition is stricter in GR than in hydro since c is typically much larger than c_s. Therefore it is worth avoiding GR when it is possible.

We talked about constraint damping last time. It didn't mean that the problem was not well posed, because in the limit of infinite resolution, the right solution will be obtained, but for practical purposes the behavior is not good enough, therefore we need to constraint transport technique.

Also a word about coordinate conditions. Remember we said we want something like $\nabla_a \nabla^a H_b = -\xi$. This was proposed in 2005, where ξ has an α term and a damping term. Without this term the problem was not working. However two years later, it turned out that taking $H_b = 0$ was okay in evolving the system. The difference was just the resolution! 4x better refined grid and we found that $\Gamma_b = 0$ is okay. This is not to take away from the technique since it was way more robust, but we need to see the development as well.

Through the GH formulation, from $\Gamma_a = -H_a$ we expand to obtain

$$\partial_t \alpha = -\alpha^2 H_a n^a + \dots, \quad \partial_t \beta^i = \alpha^2 \perp H^i \tag{2.2}$$

Thus we can relate the choice of H to the choice of α and β.

Typically in astro people use Schwarzschild or Kerr in Boyl-Lindquist coordinates, but these are horrible coordinates because the metric blows up at the horizon. One should use the Eddington-Finkelstein coordinates instead to avoid issues.

We want to say a bit about how we treat black holes. We excise the region of the black holes so that we can avoid the singularity. In a dynamic simulation we can't know where is the horizon, since by definition it is a region which can't be accessed by infinite observers in finite time. We can instead use the idea of "trapped region" which is contained inside the event horizon. In the trapped region all null lines converge inwards. If we find this is the case, then we know that we are already inside the event horizon, and it is safe to excise this region.

The other strategy of avoiding singularity is to use a kind of "wormhole" solution. We can use such kind of solution to cut the region with singularity, patch it with a wormhole solution and map it to another region which we don't care that much. The excision method has some funny condition at the excision boundary with respect to the derivatives, but the avoidance method requires careful engineering of the coordinates.

Lets look at hyperbolic equations a bit more. We look at the equation $u_{,t} = u_{,x}$ again, because it is such a generic equation for hyperbolic systems. There is an analytic way of analyzing the stability for this equation. We can define the "energy" or the norm of the solution

$$E = \int u^2 \, dx \tag{2.3}$$

and now we can ask how this "energy" is going to evolve in time

$$\partial_t E = \int 2u\dot{u} \, dx = \int 2uu_{,x} \, dx = u^2\big|_L^R \tag{2.4}$$

Therefore $E_t = u_R^2 - u_L^2$, which means that the energy evolution is given by the energy flux at the left and right ends. This means also that energy has bounded norm.

Now if the space is discretized, we can redo the calculation

$$E_{,t} = \sum (u_i^2)_{,t} h = \sum 2u_i \dot{u}_i h = \sum 2u_i (Du)_i h \tag{2.5}$$

where Du is the finite derivative. If we have

$$\sum u_i (Du)_i h = \sum \left[D(u^2) \right]_i \tag{2.6}$$

then we say D satisfies "summation by parts" condition, and we have the same result as above. If we use e.g. RK3 then fully discrete norm will satisfy this equation. This is a general property of the hyperbolic system!

Now what do we add to this? One thing we want to explore is magnetic matter/plasma around compact objects like BHs and NSs. We could potentially look at the counterpart to GW events. Pre-merger the magnetic field is usually low enough that it can't be compared with gravitation, but when after a merger the magnetic field can potentially get the energy from the violent merging and crank up to even equipartition. This might impact the dynamics and can trigger strong energy outflows/bursts.

Other relevant systems are jets from BHs and pulsars. Both are traditional astrophysical systems and are intrinsically related to the systems of interest.

As an example of how things can relate to other fields, lets consider BZ effect and pulsars. We put a source at the center whether it is an NS or BH. In the case of BH the magnetic field could be anchored by some disk.

Already in the 60s it was realized that the region around compact objects is filled with low-density plasma. We assume that this low density plasma has no mass in any frame of reference, which is the force free approximation

$$T_{ab} = T_{ab}^{EM} + T_{ab}^{\text{matter}} \approx T_{ab}^{EM} \tag{2.7}$$

But since $\nabla_a T^{ab} = 0$ it means $F^{ab} I_a = 0$, which implies Lorentz force being zero. Thus the name "force-free". Some consequences of this approximation include

$$\mathbf{E} \cdot \mathbf{B} = 0, \quad \mathbf{J} = q \frac{\mathbf{E} \wedge \mathbf{B}}{B^2} + (\mathbf{J} \cdot \mathbf{B}) \frac{\mathbf{B}}{B^2} \tag{2.8}$$

If we have a current like this, we should be able to close the Maxwell equations. But one caveat is that in order for the approximation to hold, we need $B^2 > E^2$ otherwise the problem is ill-posed, because it means new physics is involved which is not contained in this approximation.

Implementing this approximation to NS gives the NS magnetosphere which has a light-cylinder, Y-point, and an equatorial current sheet. In BH case background field lines gets twisted in the ergosphere and form current sheets there. What happens when we have a dynamic situation? If we have a neutron star and suddenly it collapses on its own, the information only propagates out at finite speed. When the field lines find out this collapse they get twisted, and from the side it contracts and undergo reconnection, launching a magnetic loop carrying a large amount of energy.

Lets look at how we do this under the GR framework. We have the Maxwell equations

$$\nabla_b F^{ab} = I^a, \quad \nabla_b {}^*F^{ab} = 0 \tag{2.9}$$

However for a moment lets imagine we have instead

$$\nabla_b \left(F^{ab} + g^{ab}\psi \right) = I^a - \alpha\psi n^a, \quad \nabla_b \left({}^*F^{ab} + g^{ab}\phi \right) = -\alpha\phi n^a \tag{2.10}$$

Both ψ and ϕ fields satisfy the damped wave equation. This looks like the damped constraint equation, therefore these fields play the role of damping the deviation from the Maxwell equations. The reason we add fields here is necessity, because we don't have any more degrees of freedom to play with. When there is some curvature of the spacetime, we can also use this technique to evolve Maxwell equations in the GR framework.