Lecture Notes on Numerical Relativity

– DRAFT –

Summer 2018

Thomas Baumgarte

This is a draft of the first two chapters of lectures notes on numerical relativity, intended for readers who have not necessarily studied general relativity in detail. This draft is meant for attendants of the summer school "Physics of Macronovae", who I ask to please not to share or circulate this draft. Feedback, comments and corrections, on the other hand, are highly appreciated.

Contents

Chapter 1

Newton's and Einstein's gravity

It is impossible to introduce general relativity in just one lecture, and we will not attempt that.[1] *Instead, we will review some properties of Newtonian gravity in this Chapter, and will then develop some of the key ideas, concepts and objects of general relativity by retracing the very same steps that we followed in our review of the Newtonian theory.*

1.1 A brief review of Newton's gravity

According to Sir Isaac Newton, space is flat, and gravity makes objects follow a curved trajectory. The origin of this curvature is a gravitational force $\mathbf{F}$ that acts on objects with nonzero mass. The gravitational force acting on an object is proportional to the mass m_G of this object; we may therefore write the force as

$$\mathbf{F} = m_G \mathbf{g}, \tag{1.1}$$

where $\mathbf{g}$ is the gravitational field created by all other objects. Note that we decorated the gravitational mass m_G with a subscript G in order to distinguish it from the inertial mass m_I that we will introduce in (1.6) below. From eq. (1.1) we see that the gravitational mass m_G describes how strongly an object couples to the gravitational field created by other masses. In electrodynamics, the equation equivalent to (1.1) would be $\mathbf{F} = q\mathbf{E}$, and we see that the electrodynamic cousin of the gravitational mass is the charge q.

We could verify that the gravitational field $\mathbf{g}$ is always irrotational, meaning that its curl vanishes. We refer to such fields as "conservative" and remember that we can always write such fields as the gradient of a scalar potential. For the gravitational field, in particular, we write

$$\mathbf{g} = -\mathbf{D}\Phi. \tag{1.2}$$

Here the negative sign follows convention, and we use the symbol $\mathbf{D}$ for the gradient operator because we will reserve the nabla operator ∇ for four-dimensional gradients. We will consider Φ the "fundamental quantity" in Newtonian theory, the quantity that encodes the gravitational interactions.

Throughout this text we will often use index notation; using an index i to denote the components in the vectors in eq. (1.2), for example, this equation would become

$$g_i = -D_i\Phi = -\frac{\partial}{\partial x^i}\Phi = -\partial_i\Phi, \tag{1.3}$$

[1] See, for example, Misner et al. (1973); Wald (1984); Hartle (2003); Carroll (2004); Moore (2013) for some excellent textbooks on general relativity.

where the components D_i of the gradient $\mathbf{D}$ are just partial derivatives. We also introduced the short-hand notation ∂_i to denote the partial derivative with respect to x^i. We could also ask whether it matters that the indices are "downstairs" rather than "upstairs" – it does, see Appendix A! – but we will ignore that subtlety for a little bit.

Newton's second law tells us that objects accelerate in response to a gravitational force $\mathbf{F}$ according to the Newtonian equations of motion

$$m_I \mathbf{a} = \mathbf{F} = -m_G \mathbf{D}\Phi \quad \text{or} \quad m_I a_i = F_i = -m_G D_i \Phi. \tag{1.4}$$

Here we have decorated the mass m_I on the left-hand side with the subscript I in order to emphasize that this is the inertial mass; it describes how strongly an object resists being accelerated. The gravitational mass m_G and the inertial mass m_I therefore describe two completely independent internal properties of objects, and a priori it is not clear at all why the two should be the same. Note that we can define the gravitational mass m_G without any involvement of acceleration; likewise, the inertial mass m_I measures an object's response to any force, not just a gravitational force. In fact, in the electrodynamic analogue the two related quantities m_I and q do not even have the same units. And yet, Newton assures us that the two masses are indeed equal,

$$m_G = m_I, \tag{1.5}$$

so that (1.4) reduces to

$$a_i = F_i = -D_i \Phi \tag{1.6}$$

in accordance with Galileo's remarkable observation that all objects fall at the same rate, independent of their mass. We now refer to this observation as the (weak) equivalence principle, and it will motivate our development of general relativity in Section 1.2.4 below.

It seems intuitive that we should be able to measure gravitational fields independent of reference frame, but the equivalence principle shows that this is not necessarily the case. Imagine we were in a freely-falling elevator. Newton would object, of course, that this is not an inertial reference frame, so that a new "fictitious" force accounting for the acceleration a_i^{elevator} of the elevator (measured in an inertial frame) would appear in (1.6), i.e.

$$a_i = F_i = -\partial_i \Phi - a_i^{\text{elevator}} \qquad \text{(in falling elevator)} \tag{1.7}$$

Now consider performing experiments – however short-lived they might be – in the falling elevator anyway. In particular, we could let an object drop, say a marshmallow. Since the elevator and the marshmallow drop at the same rate, though, we would not see the marshmallow fall at all. In (1.7), the two terms on the right-hand side exactly cancel each other, so that, as measured in the elevator, the marshmallow's acceleration vanishes. In a windowless elevator we would also be unable to determine the acceleration of the elevator.

Does that mean that, in the falling elevator, we could not measure gravitational fields at all? It turns out that we could, as long as the gravitational fields are inhomogeneous, i.e. not constant and independent of position. To see this, consider dropping not one but two marshmallows. Both marshmallows now drop towards the center of the Earth (assuming that we are falling in a terrestrial elevator), and their separation therefore decreases.

Say that the position vector of one marshmallow has components $x_i^{(1)}(t)$, while the other has components $x_i^{(2)}(t) = x_i^{(1)}(t) + \Delta x_i(t)$, where Δx_i measures the time-dependent deviation between the two marshmallows. The key idea is that the deviation Δx_i satisfies an equation that is independent of reference frame. We can evaluate

$$\frac{d^2\Delta x_i}{dt^2} = \frac{d^2}{dt^2}(x_i^{(2)} - x_i^{(1)}) = a_i^{(2)} - a_i^{(1)} = -(D_i\Phi)^{(2)} + (D_i\Phi)^{(1)}. \tag{1.8}$$

in either frame, i.e. either from (1.6) or (1.7), and obtain the same result, since the additional term $a_i^{\rm elevator}$ in (1.7) is the same for both marshmallows and hence drops out. The term $D_i\Phi$, on the other hand, has to be evaluated at $x_i^{(1)}$ for one marshmallow and at $x_i^{(2)}$ for the other. Let's assume that the two marshmallows are close to each other, so that the Δx_i are small (in comparison to length-scales over which Φ changes) and we may use a leading-order Taylor expression in Cartesian coordinates to express $(D_i\Phi)^{(2)} = (\partial_i\Phi)^{(2)}$ as

$$(\partial_i\Phi)^{(2)} = (\partial_i\Phi)^{(1)} + \sum_{j=1}^{3}\Delta x^j(\partial_j\partial_i\Phi)^{(1)} + \mathcal{O}(\Delta x^2) = (\partial_i\Phi)^{(1)} + \Delta x^j(\partial_j\partial_i\Phi)^{(1)} + \mathcal{O}(\Delta x^2). \tag{1.9}$$

In the middle term we have expressed the sum over the indices j explicitly; in the last term we have introduced the "Einstein summation convention", by which we automatically sum over all allowed values of any two repeated indices, one upstairs and one downstairs. Inserting (1.9) into (1.8) and ignoring higher-order terms we obtain

$$\frac{d^2\Delta x_i}{dt^2} = -\Delta x^j(\partial_j\partial_i\Phi)^{(1)}. \tag{1.10}$$

We conclude that, unlike the first derivatives $\partial_i\Phi$, we can indeed determine the second derivatives $\partial_j\partial_i\Phi$ independently of reference frame. The former describe the gravitational field, while the latter describe derivatives of the gravitational field, which are related to tidal forces – in fact, we will refer to the object

$$T_{ij} \equiv \partial_i\partial_j\Phi \tag{1.11}$$

as the "Newtonian tidal tensor". The word "tensor" is really short-hand for "rank-2 tensor". We properly define tensors in Appendix A; for our purposes here it is sufficient to say that a rank-n tensor carries n indices. Special cases are scalars, which are rank-0 tensors and have no indices, vectors, which are rank-1 tensors with one index, and the rank-2 tensors that we just met. We could display a rank-2 as a matrix, where rows correspond to one index and columns to the other.

With the help of the tidal tensor (1.11), eq. (1.10) takes the compact form

$$\frac{d^2\Delta x_i}{dt^2} = -T_{ij}\Delta x^j. \tag{1.12}$$

We have not yet discussed what equation governs the gravitational potential Φ itself. From Newton's universal law of gravitation, we can show that it satisfies the Poisson equation

$$D^2\Phi = 4\pi G\rho_0, \tag{1.13}$$

where ρ_0 is the mass density,[2] and D^2 the Laplace operator. The Poisson equation determines the fundamental quantity in Newtonian gravity, which, in turn, is a field – we should therefore think of the Poisson equation as the "Newtonian field equation".

[2] We use the subscript 0 to distinguish ρ_0 from the energy density ρ that we will encounter in the context of general relativity.

We now notice an interesting connection between the Newtonian field equation and the tidal tensor. Writing out the left-hand side of (1.13) we find

$$D^2\Phi = \frac{\partial^2\Phi}{\partial x^2} + \frac{\partial^2\Phi}{\partial y^2} + \frac{\partial^2\Phi}{\partial y^2} = T^i{}_i, \tag{1.14}$$

where we invoke the Einstein summation on the right-hand side. We are, admittedly, sloppy with indices being upstairs and downstairs, and will soon see how to clean this up, but in Cartesian coordinates the above is rigorous nevertheless. The important result is that we can write the Laplace operator acting on Φ as the so-called "trace" of the tidal tensor $T \equiv T^i{}_i$, i.e. the sum over its diagonal components (those for which both indices are the same). In terms of this trace we can then write the Newtonian field equation as

$$T = 4\pi G\rho_0. \tag{1.15}$$

Exercise 1.1 Consider a point mass M located at a position $\mathbf{r}_M$. At a position $\mathbf{r}$, the Newtonian potential Φ created by the point mass is

$$\Phi(\mathbf{r}) = -\frac{M}{s}, \tag{1.16}$$

where we have defined $\mathbf{s} = \mathbf{r} - \mathbf{r}_M$ and $s = |\mathbf{s}|$.

(a) Show that the tidal tensor can be written in the compact form

$$T_{ij} = \frac{M}{s^5}\left(\delta_{ij}s^2 - 3s_i s_j\right) \tag{1.17}$$

where δ_{ij} is the Kronecker delta: it takes the value one when both indices are equal, and zero otherwise.

(b) Choose a (Cartesian) coordinate system in which $\mathbf{r}_M = (0, 0, z_M)$ and find all non-zero components of the tidal tensor at the origin, $\mathbf{r} = (0, 0, 0)$.

(c) Consider two particles close to the origin that are separated by a distance $\Delta z \ll z_M$ along the direction towards M, i.e. $\Delta x^i = (0, 0, \Delta z)$, and find $d^2\Delta z/dt^2$. Then consider two particles close to the origin that are separated by a distance $\Delta x \ll z_M$ in a direction orthogonal to the direction towards M, e.g. $\Delta x^i = (\Delta x, 0, 0)$, and find $d^2\Delta x/dt^2$. If all went well, this should explain why objects in the gravitational field of a companion, e.g. the Earth in the gravitational field of the Moon, are elongated in the direction of the companion, and compressed in the orthogonal directions.

(d) Show that the trace $T = T^i{}_i$ of the tidal tensor (1.17) vanishes for all $s > 0$, as expected from the field equation (1.15).

This brings us to the following outline of our quick tour of Newtonian gravity:

1. The fundamental quantity in Newtonian gravity is the Newtonian potential Φ.

2. Trajectories of objects are, by virtue of the equations of motion (1.6), governed by first derivatives of the fundamental quantity.

3. Deviations between nearby trajectories are governed by second derivatives of the fundamental quantity, i.e. the tidal tensor, see (1.12).

4. The field equation (1.15) relates the trace of the tidal tensor to matter densities.

As it turns out, we can take a brief tour of Einstein's gravity by retracing these exact steps - in fact, every one of the items in the above list will have a corresponding subsection in our introduction to general relativity. As both a preview and summary we list all important gravitational quantities, and relate the Newtonian terms to their relativistic analogs, in Box 1.1.

Box 1.1: Cast of Gravitational Characters

	relation to fundamental quantity	Newton	Einstein
fundamental quantity	self	potential Φ	metric g_{ab}
equation of motion	derivative	$D_i\Phi$	Christoffel symbols Γ^a_{bc}
geodesic deviation	second derivatives	$T_{ij} = D_i D_j \Phi$	Riemann tensor $R^a{}_{bcd}$
field equation l.h.s.	trace of 2nd derivs.	$D^2 = N^i{}_i$	Einstein tensor G_{ab}
field equation		$D^2\Phi = 4\pi\rho_0$	$G_{ab} = 8\pi T_{ab}$

1.2 A first acquaintance with Einstein's gravity

Recall from our discussion above that, in order to reproduce Galileo's observation that all objects fall at the same rate, we had to assume the equivalence (1.5) of the gravitational and inertial masses m_G and m_I. Granted, anything else would probably seem very odd to most readers, but that is only because we are not used to even consider the possibility that they might be different – in fact, we rarely even use different symbols for the two intrinsic properties. As physicists, we "grew up" with the notion that they are the same! And yet, as we discussed above, the gravitational mass and the inertial mass measure two completely independent intrinsic properties, and there is no a priori reason at all for them to be the same. Newtonian gravity is therefore based on a remarkable coincidence: in order to reproduce a global observation, namely the trajectory's independence of the object's mass, all objects must share an individual property, namely that their two masses are equal.

We can either accept this as a coincidence, or, as Einstein did, we can think of this as a "smoking gun", suggesting that there is a deeper reason for this property of gravity. Einstein realized that, instead of relying on individual properties of individual objects, it seems much more natural to assume that the trajectory is not caused by the properties of the objects at all, but rather by something that is independent of the objects.

Consider dropping, next to each other, a marshmallow and a bowling ball, two objects with very different masses that nevertheless will fall at the same rate.[3] What these two objects share is the space through which they fall. And a property of this space that could affect the trajectory of the two objects is curvature. This is indeed the basic principle of Einstein's general relativity: instead of assuming that space is flat, as Newton did, Einstein assumed that space is curved, with the curvature representing the gravitational fields. Strictly speaking, it is actually spacetime that is curved, not space. And instead of assuming that gravitational forces make objects follow curved trajectories, Einstein assumed that they follows lines that are "as straight as possible" in these curved spacetimes – so-called "geodesics". As a result, the trajectories of the marshmallow and bowling ball do not depend on their masses at all. There is no need to make assumptions about the equivalence between gravitational and inertial mass; instead, the fact that they fall at the same rate is an immediate consequence of the properties of gravity. In the following subsections we will develop these concepts in some more detail, retracing our development of Newton's gravity in Section 1.1.

[3] Assuming that we can neglect friction, of course.

1.2.1 The metric

Our first question should be: how can we measure curvature of a space or even spacetime? In a nutshell, we can measure curvature by measuring lengths. For example, we could measure the length of the Earth's equator, say C, then measure the distance from a pole to the equator, say R. If the Earth were flat, the two numbers would be related by $C = 2\pi R$, but we would find $C < 2\pi R$ – a tell-tale sign that the surface of the Earth is curved.

Now we need a tool for measuring lengths. Imagine that our space, or spacetime, is covered by coordinates, so that every point in space, or every event in the spacetime, is labelled by certain values of these coordinates. A priori, these coordinates need not have any physical meaning. Street numbers, for example, are very useful coordinates, and yet the difference in street number between your home and your neighbor's has nothing to do with the physical distance to your neighbor. Translating a difference in coordinate value into a physical distance is what we need a *metric* for. In a one-dimensional space, the metric would simply be a factor that multiplies the "coordinate distance" into a physical distance. In higher dimensions we need to employ something similar to a Pythagorean theorem. In general, we write the *line element* between two points whose coordinate labels differ by dx^a, as

$$ds^2 = g_{ab} dx^a dx^b, \tag{1.18}$$

where g_{ab} is the *metric tensor*. We will consider the metric the fundamental quantity of general relativity; its role in Einstein's gravity is directly related to the role of the gravitational potential Φ in Newton's gravity. Note that the metric is a rank-2 tensor – it has two indices, and we could display the metric as a matrix (see below). We can already anticipate, then, that the relativistic cousins of all other objects that we encountered in Newton's gravity will likewise carry two more indices than their Newtonian counterparts. In particular, the relativistic generalization of the Newtonian field equation (1.15) should be a rank-2 tensorial equation.

Readers with knowledge of special relativity are already familiar with one particular metric, namely the *Minkowski metric*. The Minkowski metric describes a flat spacetime; we often denote it with η_{ab}, and in Cartesian coordinates it takes the form[4]

$$\eta_{ab} = \begin{pmatrix} -1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}. \tag{1.19}$$

We note that the indices a and b now run over four coordinates, e.g. t, x, y and z. We also note that we have assumed units here in which the speed of light is unity, $c = 1$. This may appear confusing, but, in fact, we colloquially do something very similar all the time. When asked, for example, how far Portland, Maine, is from Bowdoin College, the answer will often be "about half an hour", even though "hour" is not a unit of distance. The answer nevertheless makes sense, because there is a typical speed at which one travels between Bowdoin and Portland, namely the speed limit of 60 miles per hour or thereabout. We can therefore express distances in units of time, even though, admittedly, it is less common to express time in units of distance. In our context here we again have a typical speed, namely the speed of light, and we will express time and space in the same units so that $c = 1$. In fact, shortly we will express mass in the same unit also, thereby setting $G = 1$. We refer to this arrangement as "geometrized units".

[4] Different authors adopt different sign conventions. We use the convention by which space is positive and time negative, but among particle physicists, for example, the opposite convention is also common.

Inserting (1.19) together with the Cartesian displacement vector $dx^a = (dt, dx, dy, dz)$ into (1.18) we obtain

$$ds^2 = -dt^2 + dx^2 + dy^2 + dz^2, \tag{1.20}$$

which, as promised, is the generalization of the Pythagorean theorem for flat spacetimes. We can also transform the Cartesian form of the Minkowski metric into other coordinate systems. In spherical coordinates, with coordinates t, R, θ and ϕ, it takes the form

$$\eta_{ab} = \begin{pmatrix} -1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & R^2 & 0 \\ 0 & 0 & 0 & R^2 \sin^2\theta \end{pmatrix}, \tag{1.21}$$

and the line element (1.18) becomes

$$ds^2 = -dt^2 + dR^2 + R^2 d\theta^2 + R^2 \sin^2\theta d\phi^2 = -dt^2 + dR^2 + R^2 d\Omega^2, \tag{1.22}$$

where we have introduced the "solid angle" $d\Omega^2 \equiv d\theta^2 + \sin^2\theta d\phi^2$.

Since the "space" and "time" components of the spacetime metric enter with different signs, we can distinguish three different types of intervals between two nearby points in spacetimes, i.e. two nearby events. "Space-like" intervals are those with $ds^2 > 0$, for which the separation in space dominates; for such intervals we define the proper distance l by integrating over segments with $dl = (ds^2)^{1/2}$. "Time-like" intervals are those with $ds^2 < 0$, for which the separation in time dominates. Since objects with non-zero mass always travel at speeds less than the speed of light, two events on such an object's trajectory must always be separated by a time-like interval. For such intervals we then define the object's "proper time" τ from $d\tau = (-ds^2)^{1/2}$. Finally, "light-like" intervals are those for which $ds^2 = 0$. Light, and potentially other particles that travel at the speed of light, travel along trajectories whose events are separated by light-like intervals.

As an example, we can measure the proper distance between the origin of the Minkowski spacetime to a point with radial coordinate R. We will choose a path at constant time t and angles θ and ϕ, so that we have $dt = d\theta = d\phi = 0$. From (1.22) we then have $dl = (ds^2)^{1/2} = dR$, and evidently

$$l = \int_0^R dl = \int_0^R dR' = R, \tag{1.23}$$

not terribly surprising.

Exercise 1.2 Consider a circle in the equatorial plane (i.e. $\theta = \pi/2$) of the Minkowski spacetime, at constant time t and constant coordinate labels R. Confirm that this circle as a proper length $C = 2\pi R$.

Combining (1.23) we the result of exercise 1.2 shows that the proper distance l from the origin to a point with label R, and the circumference C of a circle connecting points with labels R, are indeed related by $C = 2\pi l$, in accordance with our flat-space intuition. Making measurements like this confirms that the Minkowski spacetime is indeed flat.

We observe that the metrics (1.19) and (1.21) describe the exact same (flat) spacetime, but clearly they look different. How can we then measure curvature in a coordinate-independent way? Doing that requires the tools of differential geometry. We will not be able to develop these tools in detail, but will introduce some of the most important concepts in the following Sections.

1.2.2 The geodesic equation

As we discussed above, Einstein proposed that freely-falling particles, both with and without mass, follow the shortest possible path through (potentially) curved spacetimes. For light we already now this to be true, since it is in accordance with Fermat's principle. Essentially, Einstein extended Fermat's principle to objects with mass. Such objects travel at speeds less than the speed of light, so that we can measure their proper time. For these objects, the trajectory of shortest possible path *maximizes* the advance of the object's proper time. This sounds confusing, but readers familiar with special relativity will recognize this fact from the so-called twin paradox: One twin stays on the Earth, while the other twin travels at high speed to a galaxy far, far away and then returns. Upon return the twin who stayed on the Earth aged more than the traveling twin. The asymmetry between the aging appears paradoxical until one recognizes that, in order to return to Earth, the traveling twin had turn around. This requires acceleration, meaning that that twin was not freely-falling at all times, while the home-bound one was (at least approximately). The twin on Earth ages more than any traveling twin, in accordance with the proposition that a freely-falling trajectory will maximize the advance of proper time.

When we draw the trajectory of an object through space, we draw different points, each of which represents the location of the object at one instance of time. We then connect these points to form a line. At each point we can also draw a tangent to the line; this tangent represents the object's velocity at the corresponding instance of time. In the absence of any forces an object travels in a straight line and does not accelerate; it therefore has a constant velocity vector that remains the same all along the object's trajectory. Mathematically we could express this by saying that the derivative of the velocity, taken in the direction of the velocity, vanishes.

In a spacetime diagram we include time as an additional axis. A point in a spacetime diagram therefore represents an "event"; its coordinates give us both space and time. Connecting the different events of an objects trajectory yields the object's "world-line". Tangents to this world-line now represent the "four-velocity", defined as

$$u^a = \frac{dx^a}{d\tau} \tag{1.24}$$

where τ is the object's proper time. We can now generalize the notation of "straight line" to curved spacetimes by constructing curves for which the derivative of the four-velocity u^a, taken in the direction of the four-velocity, vanishes,

$$u^b \nabla_b u^a = 0. \tag{1.25}$$

This is the *geodesic equation.* We again employ the Einstein summation convention on the left-hand side, and we also encounter a new notion of derivative, namely the *covariant derivative.* In four dimensions we will denote the covariant derivative with the nabla symbol ∇, and in three dimensions with $\mathbf{D}$, as before. The notion of the covariant derivative is needed for the following reason.

For a scalar function we can unambiguously compare values of the function at two different points and compute, from those values, the derivative – that is what we do when we compute a partial derivative. For a tensor, however, it is not good enough to compute the partial derivatives of the components. Remember that the components of a vector represent a linear combination of the basis vectors: when a vector has an x-component A^x, say, we can construct the vector $\mathbf{A}$ from the product $A^x\mathbf{e}_x$ plus terms for all other directions (see eq. (A.1) in Appendix A). Therefore, when we compute the derivative of a vector $\mathbf{A}$, we have to take into account not only changes in the components from one point to another, but also changes in the basis vectors $\mathbf{e}_a$. Since the metric g_{ab} is given by the dot product of the basis vectors (see (A.3)), we can compute these changes from derivatives of the metric. The resulting terms are encoded in the so-called Christoffel symbols Γ^a_{bc}.

A precise definition of the Christoffel symbols is given in eq. (B.5); for our purposes here it is sufficient to recognize that the Christoffel symbols contain first derivatives of the metric. It is therefore not surprising that the Christoffel symbols have three indices – two for the metric, and one for the first derivative. Extending the analogy between the metric g_{ab} and the Newtonian potential Φ as the fundamental quantities of Einstein's and Newton's theory, we therefore see that the Christoffel symbols play the same role in the former as the gravitational fields $-D_i\Phi$ in the latter.

For a scalar, the covariant derivative is the same as partial derivative – because there are no unit vectors to take care of. For a vector V^a we can express the covariant derivative as

$$\nabla_a V^b = \partial_a V^b + V^c \Gamma^b_{ac}. \tag{1.26}$$

Here the first term is responsible for the derivatives of the vector's components, and the second for those of the basis vectors. For higher-rank tensors there will be one addition Christoffel term for each additional index. Readers curious in seeing this formalism at work are encouraged to take a look at exercise A.1 and A.2.

The covariant derivative is closely related to the concept of *parallel transport*. Consider, for example, a vector V^a that is tangent to a sphere. Let's now move V^a a small distance in the direction of a second vector T^a that is also tangent to the sphere, and let's move V^a in such a way that it remains parallel to the original version of V^a. In general, V^a will no longer be tangent to the sphere at the new location. To compensate for that we will project the copy of V^a onto the sphere. By this process we "parallel transport" the vector V^a along the vector T^a. Now imagine that there is some given vector field V^a that is tangent to the sphere everywhere. If the covariant derivative of V^a in the direction T^a vanishes, then we know that V^a can be produced by parallel transport along T^a. Stated differently, the covariant derivative measures the deviation from parallel transport. The geodesic equation (1.25) then states that, for a geodesic, the velocity vector is parallel transported along itself.

Also, we no longer have an excuse for ignoring the differences between "upstairs" and "downstairs components. More technically, we refer to upstairs components as "contravariant" components, and downstairs as "covariant". Unfortunately the word "covariant" has two different meanings: we use it to describe a property that is independent of coordinate system, or to refer to downstairs components. While this is potentially confusing, the meaning should always be clear from the context. As we discuss in more detail in Appendix A, the two types of components differ in the way they transform in coordinate transformations. Using the chain rule for a displacement vector and a gradient, the "prototypes" of contravariant and covariant components, see see that they transform in "opposite" ways (see eqs. (A.24) and (A.25)). In a dot product, e.g. $A_a B^a$, we sum over one contravariant and one covariant component, so that the two different transformations cancel each other out and the result remains invariant – as it should for a scalar. We can always convert a contravariant component into a covariant component by "lowering" the index with the help of the metric, e.g. $A_a = g_{ab} A^b$. We similarly raise an index with the inverse metric, e.g. $A^a = g^{ab} A_b$. Expression (1.26) holds for the contravariant component of the covariant derivative; the covariant component of a covariant derivative is given by (B.3). Exercise B.1 shows that the covariant derivative of a scalar product $A_a B^a$ reduces to a partial derivative, as one would hope.

Returning to our story line, we insert (1.26) into (1.25) and notice the similarity of the relativistic equations of motion (1.25) with their Newtonian cousins (1.6): the derivative of the four-velocity u^a is an acceleration, and expanding the covariant derivative introduces the Christoffel symbols, which, in turn, are related to the Newtonian gravitational fields $-D_i\Phi$. In Section 1.1 we observed that the first derivatives $D_i\Phi$ do not provide a frame-independent measure of the gravitational fields, and the reader will not be shocked to hear that neither do the Christoffel symbols Γ^a_{bc}.

1.2.3 Geodesic deviation and the Riemann tensor

Following our Newtonian footsteps we now reach item 3 in the list at the end of Section 1.1, suggesting that we consider the deviation between two nearby, freely-falling objects in order to obtain a frame-independent measure of the gravitational fields, i.e. the curvature of spacetime. Conceptually we follow the same steps that we used in our Newtonian development: we evaluate the equations of motion at two nearby points, use a Taylor expansion, subtract the two results, and obtain an equation for the deviations Δx_i. The calculation is more complicated in this case, however, both because the relativistic equations of motion (1.25) is more complicated than its Newtonian counterpart (1.6), and also because we no longer want to restrict our analysis to Cartesian coordinates.

Instead of getting caught up in those details, we will observe the following. The geodesic equation (1.25) contains up to first derivatives of the our fundamental quantity, the metric g_{ab}, well hidden in the Christoffel symbols Γ^a_{bc}. Using a Taylor expansion in deriving the deviation between two geodesics will introduce a second derivative of g_{ab}, in exactly the same way as the Taylor expansion (1.9) introduced a second derivative of Φ in (1.12). And just like in the Newtonian case, where we defined the "tidal tensor" T_{ab} in (1.11) to absorb the second derivatives of Φ, we can again introduce a tensor that absorbs the second derivatives of g_{ab} together with a few non-linear terms. This tensor is the famous *Riemann tensor* $R^a{}_{bcd}$. With the help of this Riemann tensor the relativistic equation of geodesic deviation becomes

$$\frac{d^2\Delta x^a}{d\tau^2} = R^a{}_{bcd}u^b u^d \Delta x^c, \tag{1.27}$$

and we notice its similarity with its Newtonian counterpart (1.12).

We refer to Appendix B.1 for mathematical expressions and properties of the Riemann tensor. Its most important characteristic for our purposes here is that it contains second derivatives of the metric g_{ab}. Given that the latter, as a rank-2 tensor, already has two indices, it is not surprising that the Riemann tensor has four indices, making it a rank-4 tensor. As we anticipated, it has two more indices than its Newtonian counterpart, the tidal tensor T_{ab}, but plays the same role as the Newtonian tidal tensor: it is responsible for tides, and provides a frame-independent measure of the gravitational fields, which we now describe as curvature. For the Minkowski metric (1.19) or (1.21), for example, it vanishes identically independently of which coordinate system we evaluate it in.

We have introduced the Riemann tensor using the notion of geodesic deviation, but an alternative motivation is based on the concept of parallel transport. Consider parallel transporting a vector V^a around a closed loop. It is intuitive that, in flat spaces, we end up with the same vector that we started with. This is not true in curved spaces! As a concrete example, start with a vector V^a pointing north at some point on the equator of the Earth. Parallel transporting V^a along the equator half-way around the Earth it will still point north. Now we parallel transport V^a to the north pole, and continue to our original starting point – where it will arrive pointing south! The difference between a vector V^a and its copy after parallel transport along a closed loop therefore provides a measure of curvature, similar to the ratio between radius and circumferences that we have invoked before. Computing this difference for an infinitesimal loop leads to the definition (B.13) of the Riemann tensor.

1.2.4 Einstein's equations

Our Newtonian roadmap provides guidance for the next step as well, namely on how to find the field equations that determine the metric g_{ab}. As we saw in Section 1.1, the Newtonian field equation (1.15), a.k.a. the Poisson equation, relates the trace of the tidal tensor to mass densities; by analogy the relativistic field equation, a.k.a. Einstein's equations, should relate the a trace of the Riemann tensor to some measure of the energy density.

Forming a trace of a tensor means summing over a pair of its indices. The Riemann tensor has four indices, so it appears as if there were multiple choices for computing this trace. However, because of the symmetries of the Riemann tensor (see Appendix B.1), there is only one meaningful way of computing its trace, so that, up to a sign, there is no ambiguity. We call this trace the *Ricci tensor*,

$$R_{ab} \equiv R^c{}_{acb}, \tag{1.28}$$

where the index c is summed over. The result is a rank-2 tensor, which makes sense given that we should expect the relativistic field equation to be rank 2. While we are at it we also define the *Ricci scalar* as the trace of the Ricci tensor,

$$R \equiv R^a{}_a \tag{1.29}$$

(which should not be confused with the radius R). Here we have raised the first index of R_{ab} by contracting with the inverse metric, $R^a{}_c = g^{ab} R_{bc}$, see Appendix A.

Before proceeding we should discuss the right-hand side of the field equations. The Newtonian version, eq. (1.15), features the mass density ρ_0 on the right-hand side. We also anticipate that the relativistic generalization should be a rank-2 tensor. It turns out that the object that we are looking for is the *stress-energy* tensor T^{ab}. For different types of energy sources (fluids, electromagnetic fields, etc.) the stress-energy tensor depends on these energy sources in different functional form, but the time-time component T^{tt} is always the energy density ρ, the time-space components T^{ti} describe the momentum density, and the purely spatial components T^{ij} form the stress tensor. In vacuum we have, not surprisingly, $T^{ab} = 0$.

We know already that the Ricci tensor R_{ab} should make an appearance on the left-hand side of Einstein's equations, but other candidates for curvature-related, rank-2 tensors are Rg_{ab} and the metric g_{ab} itself. It turns out that R_{ab} and Rg_{ab} can only appear in a certain combination, since otherwise energy would not be conserved. We therefore absorb these two terms in the *Einstein tensor*

$$G_{ab} \equiv R_{ab} - \frac{1}{2} R\, g_{ab}. \tag{1.30}$$

Further requiring that our field equations reduce to the Newtonian field equations (1.15) in the Newtonian limit determines the relation between the Einstein tensor G_{ab} and the stress-energy tensor T_{ab}, leaving us with

$$\boxed{G_{ab} + \Lambda g_{ab} = 8\pi T_{ab}.} \tag{1.31}$$

This is our jewel, truly deserving a box – this is Einstein's field equation.[5]

The term Λ in Einstein's equations (1.31) is the famous cosmological constant. Einstein introduced this term originally in order to reproduce a static, time-independent universe, but quickly abandoned it after Hubble discovered the expansion of the universe.[6] Ironically, the much more recently discovered acceleration of the universe's expansion can be explained with such a cosmological constant, even though it is often explained as a "dark energy" component of the universe whose stress-energy tensor mimics this cosmological constant term. Regardless of what its true value is, even a non-zero cosmological constant would have effects only on large cosmological scales, and it is therefore irrelevant for most applications in numerical relativity. Accordingly we will disregard this term in the following.

Einstein's equations (1.31) form partial differential equations for the metric g_{ab}, just like its Newtonian counterpart, the Poisson equation (1.15), is a partial differential equation for the Newtonian potential Φ. Unfortunately – or fortunately from the perspective of numerical relativity – Einstein's equations are significantly more complicated than the Poisson equations, even though all the derivatives and all the non-linear terms are well hidden in the Einstein tensor G_{ab}. As a consequence, analytical solutions to Einstein's equations exist only under a number of simplifying assumptions, in particular symmetries. Finding more general solutions – for example calculating the inspiral and merger of two black holes, together with the emitted gravitational wave signal – requires solving the equations with approximate methods. In this volume we will discuss numerical relativity as one such approximate method – that will be the focus of the following two Chapters. Before we do that, though, we should review at least two important analytical results.

1.3 Two important analytical solutions

1.3.1 Schwarzschild black holes

Very shortly after Einstein put his final touches to his theory of general relativity, Karl Schwarzschild[7] famously derived an exact, static and spherically symmetric vacuum solution,

$$ds^2 = -\left(1 - \frac{2M}{R}\right) dt^2 + \left(1 - \frac{2M}{R}\right)^{-1} dR^2 + R^2 d\Omega^2. \tag{1.32}$$

Tragically, Schwarzschild died only months later, while fighting in World War I, and decades before this solution was recognized as describing non-rotating black holes, and appreciated for its monumental astrophysical importance.

The constant M in (1.32) turns out to be the black hole's mass. In particular we see that, for $M = 0$, we recover the flat Minkowski metric (1.21), which is reassuring. A priori the radial coordinate R need not have any particular physical meaning (think street numbers...), but it turns out that in this case it does. We could repeat exercise 1.2 for the line element (1.32) to find that the proper circumference C of a circle at radius R is $2\pi R$, as we might expect from our flat-space intuition. Similarly, a sphere of radius R has a proper area of πR^2, again in accordance with our intuition. We therefore refer to the radius R that appears in the Schwarzschild line element (1.32) as the *circumferential* or *areal* radius – as such, it does have in invariant physical meaning. The Schwarzschild spacetime is not flat, however. We could ascertain that in a frame-independent

[5]See Einstein (1915).

[6]It is often reported the Einstein considered the introduction of the cosmological constant his "biggest blunder".

[7]See Schwarzschild (1916).

way by computing the Riemann tensor R^a_{bcd}. This exercise would show that the curvature of the spacetime diverges as we approach $R \to 0$. The center of the Schwarzschild spacetime therefore features a *curvature singularity*. For large separations, $R \gg M$, on the other hand, we recover the flat Minkowski metric (1.21). We refer to such spacetimes as *asymptotically flat*.

It can be shown that the Schwarzschild geometry (1.32) is unique for spherically symmetric vacuum spacetimes; this is guaranteed by the so-called *Birkhoff* theorem. While the geometry is unique, the coordinate system in which it is described is not. The following two exercises provide examples of other useful coordinate systems for the Schwarzschild spacetime.

Exercise 1.3 Consider a new "isotropic" radius r that satisfies

$$R = r\left(1 + \frac{M}{2r}\right)^2 \tag{1.33}$$

(a) Show that under this transformation the Schwarzschild metric (1.32) takes the form

$$ds^2 = -\left(\frac{1 - M/(2r)}{1 + M/(2r)}\right)^2 dt^2 + \psi^4(dr^2 + r^2 d\Omega^2) \tag{1.34}$$

and find the "conformal factor" ψ. Note that the spatial part of this metric takes for form $\psi^4 \eta_{ij}$, where η_{ij} is the flat metric in any coordinate system (here in spherical polar coordinates). This form is called "isotropic", since it does not single out any direction.

(b) Find an expression for r in terms of R. Show that real solutions for r only exist $R \geq 2M$.

Exercise 1.4 (a) Show that under the coordinate transformation

$$dt = d\bar{t} - \frac{1}{f_0}\frac{M}{R - M}dR \tag{1.35}$$

the Schwarzschild metric (1.32) takes the new form

$$ds^2 = -f_0 d\bar{t}^2 + \frac{2M}{R - M}d\bar{t}dR + \left(\frac{R}{R - M}\right)^2 dR^2 + R^2 d\Omega^2. \tag{1.36}$$

(b) Now transform to a new spatial coordinate $r = R - M$ and show that the new metric takes the form

$$ds^2 = -\frac{r - M}{r + M}d\bar{t}^2 + 2\frac{M}{r}d\bar{t}dr + \psi^4(dr^2 + r^2 d\Omega^2) \tag{1.37}$$

with

$$\psi = \left(1 + \frac{M}{r}\right)^{1/2} \tag{1.38}$$

Note that the spatial part of the metric is again isotropic, so r is again an isotropic radius, as in exercise 1.3. Here, however, r is associated with a different time coordinate; it is therefore different from the isotropic radius r in exercise 1.3.

The defining property of black holes is the existence of a *horizon*, i.e. a surface through which light – and all other particles and objects – can fall into the black holes, but through which they cannot escape the black hole. Different notions of horizons make this statement precise in different ways. *Event horizons* separate those regions of a spacetime from which photons can escape to infinity from those from where they cannot. For the purposes of numerical relativity, the notion of an *apparent horizon* is often of greater practical importance. An apparent horizons is the outermost closed surface on which the expansion of an outgoing bundle of light is zero or negative. For static black holes, these different notions become identical, and for a Schwarzschild black hole we can simply locate the radius at which an outgoing radial photon no longer travels to larger radius.

To locate the horizon for the metric (1.37), we consider radial photons, i.e. photons along whose trajectory only dr and dt are non-zero. The trajectory must also be light-like, so that $ds^2 = 0$. From (1.37) we then have

$$\left(1 + \frac{M}{r}\right)^2 dr^2 + 2\frac{M}{r} d\bar{t} dr - \frac{r - M}{r + M} d\bar{t}^2 = 0 \tag{1.39}$$

or

$$\frac{dr}{d\bar{t}} = \left(\pm 1 - \frac{M}{r}\right) \frac{r^2}{(r + M)^2}. \tag{1.40}$$

Here the "+" sign yields the coordinate speed of outgoing photons, while the "-" sign yields that of ingoing photons. We should also emphasize that any local observer would still measure the light to propagate at the speed of light; the coordinate speed that we compute here has no immediate physical meaning. For $r \gg M$, the coordinate speeds (1.40) reduce to ± 1, consistent with the observation that, like the Schwarzschild metric (1.32), the metric (1.37) approaches the Minkowski metric (1.21) in that limit (recall that we are using units in which $c = 1$). At $r = M$, however, we see that even "outgoing" photons no longer propagate to larger radii; this location therefore marks the horizon. From exercise 1.4 we also know that this isotropic radius corresponds to an areal radius of $R = 2M$, which has an coordinate-independent meaning. We conclude that the horizon of a Schwarzschild black hole is at $R = 2M$.

Note that the Schwarzschild line element (1.32) becomes singular at $R = 2M$. We have seen already that we can remove this singularity by going to a different coordinate system; it is therefore just a coordinate singularity, akin to the singular behavior of longitude and latitude at the poles. For numerical purposes it is important that the metric remains non-singular on the horizon; this is a first hint, therefore, that a coordinate system with properties similar to that in exercise 1.4 might prove very useful in numerical simulations. Exercise 1.5 explores two other remarkable properties of this coordinate system, properties that we will later associated with so-called *trumpet* geometries of black holes.

Exercise 1.5 (a) Compute the proper length $\mathcal{L}$ of a circle at coordinate label r in the equatorial plane around the origin of the coordinate system (1.37).[8] Note that $\mathcal{L}$ remains greater than zero as $r \to 0$. This implies that the new coordinates of exercise 1.4 do not reach the curvature singularity at $R = 0$.

(b) Compute the proper distance $\mathcal{D}$ between the origin at $r = 0$ and a point at $r = \epsilon$. Show that $\mathcal{D}$ is infinite for any $\epsilon > 0$.

We conclude this section by noting that the Schwarzschild spacetime has been generalized in a number of different ways. Most importantly, the Kerr metric describes stationary rotating black holes.[9] Without any symmetry assumptions, however, black holes cannot be described analytically. Modeling the inspiral and merger of two black holes in binary orbit, for example, requires the tools of numerical relativity, which we will discuss in the following chapters.

1.3.2 Gravitational waves

Consider a small perturbation of a given "background" metric. Specifically, let us assume that the background metric is the flat Minkowski metric (1.19) in Cartesian coordinates. In that case

[8] Either carry our the integration, or simply recognize that, by definition of R, $\mathcal{L} = 2\pi R = 2\pi(r + M)$.

[9] See Kerr (1963).

we can write

$$g_{ab} = \eta_{ab} + h_{ab} \tag{1.41}$$

and assume the perturbations h_{ab} are small compared to unity, $|h_{ab}| \ll 1$. Given this metric we can compute the Christoffel symbols, the Riemann tensor, the Ricci tensor, and finally the Einstein tensor. At each step we can drop all terms that are higher than linear order in the h_{ab}, and we can also impose certain coordinate conditions to simplify our expressions. Inserting our result for the Einstein tensor into Einstein's equations (1.31) then yields the linearized equation

$$\left(-\partial_t^2 + D^2\right) h_{ab} = -16\pi T_{ab}. \tag{1.42}$$

Here ∂_t^2 denotes the second time derivative, and D^2 is the flat, spatial Laplace operator. Together, the two terms $(-\partial_t^2 + D^2)$ form a wave operator, meaning that Einstein's equations reduce to a wave equation for the perturbations h_{ab} in this limit. But where there is a wave equation, there will also be wave-like solutions: we see how gravitational waves emerge very naturally from Einstein's equations.

Even though Einstein recognized the existence of wave-like solutions very early on,[10] uncertainty over the physical reality of gravitational radiation remained for several decades (caused, in part, by Einstein himself). The issue was clarified in the late 1950's, however,[11] which then triggered the quest to detect gravitational waves from astrophysical sources. After decades of efforts, and almost exactly hundred years after Einstein first predicted the existence of this gravitational radiation, the Laser Interferometer Gravitational Wave Observatory (LIGO) detected gravitational waves for the first time on Sept. 14, 2015.[12] Named after the date, the event is called GW150914, and we now know that it was emitted by two black holes with masses of 36 and 29 solar masses that merged about a billion years ago.

Since (1.42) is a linear, flat wave equation, we can find solutions in terms of the Green function for the flat wave operator. The result can be written as

$$h_{ij} \simeq \frac{2}{r}\ddot{\mathcal{I}}_{ij}(t-r), \tag{1.43}$$

where r is the distance from the source, and where the double dot denotes a second time derivative. The so-called "reduced quadrupole moment"

$$\mathcal{I}_{ij} = I_{ij} - \frac{1}{3}\eta_{ij} I \tag{1.44}$$

can be computed from the second moment of mass distribution

$$I_{ij} = \int d^3x \rho x_i x_j \tag{1.45}$$

as well as its trace $I = I^i{}_i$.

Exercise 1.6 Consider a Newtonian binary consisting of two point masses m_1 and m_2 in circular orbit about their center of mass. Choose coordinates so that the binary orbits in the xy plane and is aligned with the x-axis at $t = 0$.

[10]See Einstein (1916).

[11]Of particular importance in these developments was the 1957 General Relativity Conference at Chapel Hill, see Bergmann (1957).

[12]See Abbott (2017).

(a) Show that the second moment of mass distribution can be written as

$$I_{ij} = \frac{1}{2}\mu R^2 \begin{pmatrix} 1+\cos 2\Omega t & \sin 2\Omega t & 0 \\ \sin 2\Omega t & 1-\cos 2\Omega t & 0 \\ 0 & 0 & 0 \end{pmatrix}, \tag{1.46}$$

where Ω is the orbital angular velocity, $\mu \equiv m_1 m_2/(m_1+m_2)$ the reduced mass, and R the binary separation.

(b) Show that the reduced quadrupole moment is

$$\mathcal{I}_{ij} = \frac{1}{2}\mu R^2 \begin{pmatrix} 1/3+\cos 2\Omega t & \sin 2\Omega t & 0 \\ \sin 2\Omega t & 1/3-\cos 2\Omega t & 0 \\ 0 & 0 & -2/3 \end{pmatrix} \tag{1.47}$$

(c) Show that the size of the metric perturbations h_{ij} is of order

$$h \simeq \frac{4}{r}\frac{\mu M}{R}, \tag{1.48}$$

where $M = m_1 + m_2$ is the total mass, and where we have used Kepler's law $\Omega^2 = M/R^3$.

(d) The gravitational wave source GW150914 consisted of two black holes of masses $m_1 = 36 M_\odot$ and $m_2 = 29 M_\odot$, at a distance of about 410 Mpc. In gravitational units a solar mass $M_\odot$ corresponds to about 1.4 km, and a parsec is about 3.1×10^{13} km. Estimate h for GW150914.

From (1.43) we can also estimate a source's rate of energy loss due to the emission of gravitational radiation,

$$\frac{dE}{dt} = -\frac{1}{5}\langle \dddot{\mathcal{I}}_{ij} \dddot{\mathcal{I}}^{ij} \rangle, \tag{1.49}$$

where $\langle\rangle$ denotes a time average. This is the famous *quadrupole formula*, similar to the Lamor formula in electrodynamics. A binary system, for example, loses energy in the form of gravitational radiation, which leads to a shrinking of the binary orbit, and ultimately to the merger of the two companions. The quadrupole formula (1.49) allows us to estimates for this inspiral.

Exercise 1.7 (a) Revisit the binary of exercise 1.6 and show that gravitational wave emission will result in the binary losing energy E at a rate

$$L_{\rm GW} = -\frac{dE}{dt} = \frac{32}{5}\frac{M^3\mu^2}{R^5}. \tag{1.50}$$

(b) Recall that the equilibrium energy of a binary in circular orbit is given by

$$E_{\rm eq} = -\frac{1}{2}\frac{M\mu}{R}. \tag{1.51}$$

In response to the energy loss of part (a) the binary separation R will therefore shrink. Assuming that we can consider this inspiral as "adiabatic", i.e. as a sequence of circular obits along which the radius changes on timescales much longer than the orbital period, we can compute the inspiral rate from

$$\frac{dR}{dt} = \frac{dE_{\rm eq}/dt}{dE_{\rm eq}/dR} = -\frac{L_{\rm GW}}{dE_{\rm eq}/dR} = -\frac{64}{5}\frac{M^2\mu}{R^3}. \tag{1.52}$$

Verify the last equality.

(c) Now integrate eq. (1.52) to find the binary separation R as a function of time

$$R(t) = 4\left(\frac{\mu M^2}{5}(T-t)\right)^{1/4}, \tag{1.53}$$

where T is the time of the coalescence when $R = 0$.

(d) Finally, combine your result from part (c) with Kepler's law to compute the orbital angular velocity Ω as a function of time.

While the quadrupole formula provides pretty good estimates, at least for large binary separations, it is not sufficiently accurate to predict gravitational wave signals from binary mergers for comparison with detected waves. This, again, requires the tools of numerical relativity, to which we will now turn.

Chapter 2

Foliations of Spacetime: Constraint and Evolution Equations

In Chapter 1 we motivated Einstein's equations, and saw how elegantly they relate the curvature of spacetime to its mass or energy content. From a numerical perspective, however, this elegance is somewhat of a hinderance. In Einstein's equations, all quantities are spacetime quantities, but numerically we would like to solve a so-called Cauchy *problem: we would like to start with a certain "state" of fields at some initial time, and then follow how these fields evolve forward in time. This requires splitting spacetime into space and time, i.e. introducing a so-called foliation of spacetime, or a "3+1" split. Leaning heavily on analogies with scalar waves and electrodynamics we will see that this results in split of the equations into so-called constraint and evolution equations: the former constrain the fields at each instance of time, while the later determine the evolution of the fields.*

2.1 A scalar wave equation

We will start our discussion with a simple massless scalar field, both in order to illustrate a few points, and to provide a reference point for our discussion later on. The field equation for a massless scalar field is the wave equation

$$\Box\Phi = \nabla^a \nabla_a \Phi = 4\pi\rho, \tag{2.1}$$

where we have allowed for a non-zero source term on the right hand side. Assuming that we solve this wave equation in a flat Minkowski spacetime, so that we can replace g_{ab} with η_{ab}, we can expand the wave operator on the left-hand side as

$$\Box\Phi \equiv g^{ab}\nabla_a\nabla_b\Phi = \eta^{ab}\nabla_a\nabla_b\Phi = \left(-\partial_t^2 + \gamma^{ij}D_iD_j\right)\Phi = \left(-\partial_t^2 + D^2\right)\Phi. \tag{2.2}$$

This requires a few words of explanation...First, we should remind the reader that g^{ab} is the inverse of the metric. It is defined by the requirement that $g^{ab}g_{bc} = \delta^a{}_c$, where $\delta^a{}_c$ is the Kronecker delta which we previously encountered in Exercise 1.1: it is zero when a is different from c, and one otherwise. In practice, however, we can compute the inverse of the metric exactly as one learns how to compute the inverse of a matrix in linear algebra. In the middle of the above equation we split the sum over the spacetime indices a and b into a term for time index t, and the remaining spatial terms. The latter only involve the spatial part of the metric, which we now denoted with

γ_{ij} and its inverse γ^{ij}, as well as the spatial covariant derivative D_i. Finally, we wrote the spatial Laplace operator as $D^2 = \gamma^{ij} D_i D_j$.

Note, by the way, that (2.1) reduces to the Newtonian field equation (1.13) if we can neglect the time derivatives. We also recognize that the field equation (2.1 for the scalar field again involves the trace of second derivatives, exactly as the Newtonian field equation (1.15) for Φ and Einstein's field equation (1.31) for the spacetime metric g_{ab}.

Eq. (2.1) is a second-order partial differential equation for Φ. Mathematicians classify this equation as a symmetric hyperbolic equation, and assure us of its very desirable properties: the problem is well-posed, meaning that solutions exist and are unique, and grow at most exponentially. Numerically, however, it is easier to integrate equations that are first-order in time. We could therefore define a new variable

$$\kappa \equiv -\partial_t \Phi, \tag{2.3}$$

where the inclusion of the minus sign will make sense later on, and rewrite the equation (2.1) as the pair of equations[1]

$$\begin{aligned} \partial_t \Phi &= -\kappa & (2.4) \\ \partial_t \kappa &= -D^2 \Phi + 4\pi\rho. & (2.5) \end{aligned}$$

The color coding will make comparisons with electrodynamics and general relativity easier, and we will apologize already to readers who are color-blind.

There are no *constraint equations* for the scalar wave. In particular, this means that we can freely choose initial data for Φ and κ; these initial data, given as functions of space only, then describe Φ and κ at some initial instance of time. Eqs. (2.4) and (2.5) are the *evolution equations* that determine how these variables evolve as time advances.

Solving the evolution equations (2.4) and (2.5) numerically is not very hard. In Appendix **??** we provide an example of how that could be done – assuming that our numerical code is correct, we should be able to solve these equations without any problems. Mathematicians might remind us that this a consequence of the equations being symmetric hyperbolic.

One could argue that eqs. (2.4) and (2.5) represent a *3+1 split* of the wave equation (2.2), since we started with an operator involving both time and space derivatives on the left-hand side of that equation, and ended up with a pair of equations in which we have explicitly separated the three space derivatives from the one time derivatives – hence "3+1". There was so little to do, however, that one easily misses the point. In particular, we only had to deal with the derivatives in the equation, but, since the variable here was a scalar, we did not have to separate its time components from the space components. What, however, do we need to do if the variables are tensors, so that we may want to split the tensors into time and space parts also? As a warm-up exercise we will consider electrodynamics in the Section, where the variables are rank-1 tensors.

2.2 Electrodynamics

2.2.1 Maxwell's equations

Maxwell's equations for the electric field $\mathbf{E}$ and the magnetic field $\mathbf{B}$ are

$$\begin{aligned} \mathbf{D} \cdot \mathbf{E} &= 4\pi\rho & \qquad \partial_t \mathbf{E} &= \mathbf{D} \times \mathbf{B} - 4\pi \mathbf{j} \\ \mathbf{D} \cdot \mathbf{B} &= 0 & \qquad \partial_t \mathbf{B} &= -\mathbf{D} \times \mathbf{B}, \end{aligned} \tag{2.6}$$

[1] In fact, many authors would introduce now variables for the first space derivatives also, thereby making the system first order in both space and time.

where we used vector notation, where $\mathbf{D}$ is the spatial nabla operator (rather than the more common symbol ∇, which we are reserving for four dimensions), ρ is the charge density, and $\mathbf{j}$ the charge current density. Since the magnetic field is divergence-free, we may write it as the curl of a vector potential $\mathbf{A}$, i.e.

$$\mathbf{B} = \mathbf{D} \times \mathbf{A}. \tag{2.7}$$

The divergence of $\mathbf{B}$ then vanishes identically,

$$\mathbf{D} \cdot \mathbf{B} = \mathbf{D} \cdot (\mathbf{D} \times \mathbf{A}) = 0, \tag{2.8}$$

so that the "bottom left" of Maxwell's equations (2.6) is satisfied automatically. The curl of the magnetic field in the "top right" equation now becomes

$$\mathbf{D} \times \mathbf{B} = \mathbf{D} \times (\mathbf{D} \times \mathbf{A}) = -D^2\mathbf{A} + \mathbf{D}(\mathbf{D} \cdot \mathbf{A}), \tag{2.9}$$

or, in words: the curl of $\mathbf{B}$ is now the negative Laplace operator acting on $\mathbf{A}$, plus the gradient of the divergence of $\mathbf{A}$. Finally, the bottom right of Maxwell's equations becomes

$$\partial_t(\mathbf{D} \times \mathbf{A}) = \mathbf{D} \times (\partial_t \mathbf{A}) = -\mathbf{D} \times \mathbf{E}. \tag{2.10}$$

Since the curl of a gradient vanishes identically, this implies

$$\partial_t \mathbf{A} = -\mathbf{E} - \mathbf{D}\Phi, \tag{2.11}$$

where Φ is an arbitrary gauge variable. Collecting our results, we see that Maxwell's equations (2.6) result in the two equations

$$\partial_t \mathbf{A} = -\mathbf{E} - \mathbf{D}\Phi \tag{2.12}$$
$$\partial_t \mathbf{E} = -D^2\mathbf{A} + \mathbf{D}(\mathbf{D} \cdot \mathbf{A}) - 4\pi\mathbf{j}, \tag{2.13}$$

or, returning to index notation,

$$\partial_t A_i = -E_i - D_i\Phi \tag{2.14}$$
$$\partial_t E_i = -D^2 A_i + D_i D^j A_j - 4\pi j_i. \tag{2.15}$$

We notice immediately the similarity with eqs. (2.4) and (2.5) for the scalar wave: in both cases the first equation relates the time derivative of the "green" variable to the "red" variable, and the second equation the time derivative of the red variable to the Laplace operator acting on the green variable, plus a "blue" matter term. In fact, we now appreciate the inclusion of the negative sign in the definition (2.3), which even makes all the signs consistent. We also notice important differences, in addition to the fact that eqs. (2.4) and (2.5) the dynamical variables were scalars, while in (2.14) and (2.15) they are vectors: in the electromagnetic case we encounter a new gauge variable in (2.14), high-lighted in gold, and (2.15) includes not only the Laplace operator of the green variable but also the gradient of the divergence. The latter will play an important role later on. Without this term, eqs. (2.14) and (2.15) could be combined to form a wave equation for the components E_i, but these new terms spoil this property. Mathematicians might be alarmed already...

As another difference, we also still have to satisfy the top left of Maxwell's equations (2.6),

$$D_i E^i = 4\pi\rho. \tag{2.16}$$

Unlike (2.14) and (2.15), this equation does not contain any time derivatives of the variables; instead, the electric field E_i has to satisfy this equation at all times. We refer to this equation as a *constraint equation*. In particular, any initial data that we choose for E_i have to satisfy the constraint equation (2.16) – we can no longer choose the initial data arbitrarily. Eqs. (2.14) and (2.15) do contain time derivatives of the fields; they therefore determine how the fields evolve forward in time, and they are therefore called *evolution equations*.

The above already illustrates the general procedure that is used to solve Einstein's equations numerically. As we will discuss in Section 2.3, Einstein's equations will also split into a set of constraint equations and a set of evolution equations, where the latter do involve time derivatives of the fields while the former do not. In fact, the equations will have a remarkable similarity with the set of equations above. We again have to solve the constraint equations first in order to construct valid initial data. These initial data will describe an instantaneous state of the gravitational fields at some instance of time – they might, for example, describe a snapshot of two black holes that are in the process of merging. Then we can solve the evolution equations to compute how the fields evolve in time – this would result in a "movie" of the subsequent coalescence, together with the emission of a gravitational wave signal.

The observant reader will notice, however, that the above example has not yet accomplished one goal: namely to provide an example of how we can split a spacetime tensor into space and time parts. Our example does not do that, because we started with Maxwell's equations already expressed in terms of the purely spatial electric and magnetic field vectors. Fortunately, it is easy to fix that problem – we could start with Maxwell's equations expressed in terms of the Faraday tensor instead of the electric and magnetic fields.

2.2.2 The Faraday tensor

Maxwell's equations show that electric and magnetic fields are intimately related. Lorentz transformations for electric and magnetic fields also show that one observer's electric field may manifest itself as a magnetic field to another observer, and vice versa. In particular, this shows that electric and magnetic fields do *not* transform like components of vectors. What kind of a relativistic object, then, do the fields form? Just counting the number of independent components gives us a good hint.

Evidently, there are six independent components in the electric and magnetic fields combined, more than a four-vector, i.e. a rank-1 tensor in four dimensions, could accommodate. A rank-2 tensor in four dimensions has $4 \times 4 = 16$ components – evidently that is too many. The metric tensor, for example, is a *symmetric* rank-2 tensor, with $g_{ab} = g_{ba}$. In four dimensions, such a tensor has $1 + 2 + 3 + 4 = 10$ components – still too many. An *antisymmetric* rank-2 tensor A_{ab} has the property $A^{ab} = -A^{ba}$ – in particular, this means that the diagonal components A_{aa} have to vanish. In four dimensions, such a tensor therefore has only $1+2+3 = 6$ independent components – exactly the number that we need to accommodate the electric and magnetic fields. That's a hint - and in fact, the covariant object describing electric and magnetic field is the so-called *Faraday* tensor, an antisymmetric rank-2 tensor.

An observer using Cartesian coordinates (in a local Lorentz frame) would identify the components of the Faraday tensor F^{ab} according to

$$F^{ab} = \begin{pmatrix} 0 & E^x & E^y & E^z \\ -E^x & 0 & B^z & -B^y \\ -E^y & -B^z & 0 & B^x \\ -E^z & B^y & -B^x & 0 \end{pmatrix}. \tag{2.17}$$

Since the Faraday tensor is antisymmetric, we now have to be careful with the identification of the components. By convention, we will let the first index in F^{ab} refer to the row, and the second to the column. For example, we identify $F^{tx} = E^x$.

For an observer with four-velocity u^a the Faraday tensor can also be written as

$$F^{ab} = u^a E^b - u^b E^a + \epsilon^{abcd} B_c u_d, \tag{2.18}$$

where E^a and B^a are the electric and magnetic fields as observed by this observer, with $u_a E^a = 0$ and $u_a B^a = 0$, and where ϵ^{abcd} is the *Levi-Civita* tensor. In a local Lorentz-frame, it is 1 for all even permutations $abcd$ of $txyz$, negative 1 for all odd permutations, and zero all other combinations of indices.

Exercise 2.1 Demonstrate that (2.18) reduces to (2.17) in the observer's local Lorentz-frame, where $u^a = (1, 0, 0, 0)$.

A tensor has well-defined properties under coordinate transformations (see Appendix A). In a new, primed, coordinate system, for example, the Faraday tensor can be found from (A.22). To find the transformation of the electric and magnetic fields under a Lorentz transformation, we identify the primed fields with the corresponding components of the primed Faraday tensor – for example $E^{x'} = F^{t'x'}$. Exercise A.2 demonstrates that electric and magnetic fields "mix" under a Lorentz transformation – what appears like an electric field in one reference frame, takes the form of a magnetic field in another, and vice versa.

In terms of the Faraday tensor, Maxwell's equations can be cast in the compact form

$$\nabla_b F^{ab} = 4\pi j^a \tag{2.19}$$

and

$$\nabla_{[a} F_{bc]} = 0. \tag{2.20}$$

In (2.19), the four-vector j^a has components

$$j^a = (\rho, j^i). \tag{2.21}$$

We usually express conservation of charge in terms of the continuity equation

$$\partial_t \rho = -D_i j^i; \tag{2.22}$$

in terms of the four-vector j^a this takes the more compact form

$$\nabla_a j^a = 0. \tag{2.23}$$

Previously we observed that the divergence of the magnetic field B^i vanishes, and we therefore wrote B^i as the curl of a vector potential A^i (see (2.7)). Similarly, equation (2.20) allows us to write the Faraday tensor in terms of a four-dimensional vector potential A^a,

$$F^{ab} = \nabla^a A^b - \nabla^b A^a. \tag{2.24}$$

Equation (2.20) is then satisfied identically, and the remaining Maxwell equation (2.19) takes the form

$$\nabla_b \nabla^a A^b - \nabla_b \nabla^b A^a = 4\pi j^a. \tag{2.25}$$

In fact, this form fits in well with our outline of Newton's and Einstein's gravity in Chapter 1. In the parlance of our discussion there, the four-dimensional vector potential A^a is now the fundamental quantity of our theory – the cousin of the Newtonian potential Φ in Newtonian gravity, and the spacetime metric g_{ab} in general relativity. The vector potential A^a is a rank-1 tensor, conveniently "half-way" between the rank-0 Newtonian potential Φ (or the scalar field Φ in wave equation (2.1)) and the rank-2 metric g_{ab} in general relativity. Electromagnetic forces on charges result from the electric and magnetic fields in the Lorentz force, and hence from derivatives of A^a. Finally, the field equations (2.25) contains the trace of second derivatives of A^a, in complete analogy to the Newtonian Poisson equation (1.15) (or its generalization, the wave equation (2.1)) and Einstein's equation (1.31). Equation (2.25) therefore provides a perfect "warm-up" problem for 3+1 decompositions.

We start with a 3+1 decomposition of the tensors appearing in (2.25). In (2.21) we have already split the four-dimensional current density vector j^a into the charge density ρ as the time component, and the spatial current density j^i as its spatial components. We similarly write the four-dimensional vector potential A^a as

$$A^a = (\phi, A^i). \tag{2.26}$$

This completes the decomposition of the tensors, and we next decompose the equations.

We first consider the time component of equation (2.25). Choosing $a = T$ in equation (2.25) then results in

$$\nabla_b \nabla^T A^b - \nabla_b \nabla^b A^T = \nabla_j \nabla^T A^j - \nabla_j \nabla^j \phi = 4\pi j^T = 4\pi\rho. \tag{2.27}$$

We have purposely denoted time with a capital T here, since we will generalize our approach in Section 2.2.2 below, where we will allow for more general time coordinates t. Note also that we were able to restrict the summation to spatial indices $b = j$, because for $b = T$ the two first terms cancel each other. In special relativity we may identify $\nabla_j = D_j$ and $\nabla_T = \partial_T$. We also have $\nabla^T = \eta^{Ta}\nabla_a = -\partial_T$, and therefore

$$D_j(-\partial_T A^j - D^j\phi) = 4\pi\rho. \tag{2.28}$$

We now identify

$$E^i = F^{Ti} = -\partial_T A^i - D^i\phi, \tag{2.29}$$

in complete consistency with (2.14), and obtain

$$D_i E^i = 4\pi\rho, \tag{2.30}$$

i.e. the constraint equation (2.16).

We next consider the spatial components $a = i$ in (2.25) and find

$$\nabla_T(\nabla^i A^T - \nabla^T A^i) = -\nabla_j \nabla^i A^j + \nabla_j \nabla^j A^i + 4\pi j^i, \tag{2.31}$$

where we have grouped the two terms with $b = T$ on the left-hand side, and the two terms with $b = j$ on the right-hand side. We restrict to special relativity again, identify $A^T = \phi$, insert (2.29) on the left-hand side, and obtain

$$\partial_T E_i = -D_j D^j A_i + D_i D^j A_j - 4\pi j_i. \tag{2.32}$$

Not surprisingly, this is identical to equation (2.15).

While, so far, we have only re-derived equations, we can already make the following observations: in the form (2.25), Maxwell's equations are written in terms of a four-dimensional equation for the four-dimensional tensor object A^a, our "fundamental quantity". We perform a 3+1 split by first splitting the tensors into their space and time parts; in particular, we observe that the time component of A^a plays the role of a gauge variable, and the spatial components are related to the time derivatives of what we identify as the electric field. We then saw that the time component of the equation resulted in the constraint equation, while the spatial components resulted in the evolution equation. The reader will not be shocked to hear that we will find a very similar structure in general relativity – before exploring that, however, we need to add one more layer of abstraction.

2.2.3 Arbitrary time slices: lapse, shift and all that...

In Section 2.2.2 we did introduce the notion of a 3+1 split of four-dimensional tensors and equations, but we did so by choosing given components T and j in the tensors and equations. By doing so, we have not yet taken advantage of the full coordinate freedom at our disposal. Essentially, we assumed that the spatial slices of our 3+1 decomposition have to line up with a given coordinate time T. We will now relax this assumption, and will allow for more general slices.

Specifically, let's assume that we are given a function $t = t(x^a)$ that depends, in some way, on our original coordinates. We have suggestively called this function t, because in a minute this will indeed become our coordinate time – but in principle they do not have to equal. We will assume that level surfaces of constant t are spacelike, and will therefore refer to these level surfaces as *spatial slices*. The *normal* n_a on these slices is proportional to the gradient of t, so that we can write it as

$$n_a = -\alpha \nabla_a t. \tag{2.33}$$

The normal on a spatial surface must be timelike; we therefore normalize the normal according to

$$n_a n^a = g^{ab} n_a n_b = -1. \tag{2.34}$$

This relation determines the constant of proportionality α, which we will give a physical meaning shortly. The choice of the negative sign in (2.33) makes the "up-stairs" contravariant time component of n^a in (2.38) below positive, meaning that this choice makes the normal vector future-oriented.

Note that the normalization (2.34) allows us to interpret the normal vector as a four-velocity u^a (which must be normalized in the same way). We therefore refer to a *normal observer* as an observer who always moves in the direction of the normal vector, and whose four-velocity is hence identical to n^a. More fancily we say that the normal observer's worldline is generated by the normal vector n^a.

We will now assume that our time coordinate is indeed the function t, so that surfaces of constant coordinate time coincide with our level surfaces. With this choice, the only non-zero component of $\nabla_a t = \partial_a t$ is the one for which $a = t$, i.e. the time component $\partial_t t = 1$. The normal (2.33) then reduces to

$$n_a = (-\alpha, 0, 0, 0). \tag{2.35}$$

Consider now two events separated by a proper time $d\tau$ along a normal observer's worldline. The two events are separated by the invariant four-vector $n^a d\tau$. Following (A.26) we can compute the change dt in the coordinate time t from

$$dt = \frac{\partial t}{\partial x^a}(n^a d\tau) = (\nabla_a t)(n^a d\tau) = -\frac{1}{\alpha} n_a n^a d\tau = \frac{d\tau}{\alpha}, \tag{2.36}$$

or

$$d\tau = \alpha dt. \tag{2.37}$$

We therefore see that the function α measures how much proper time elapses, according to a normal observer, as the coordinate time advances by dt, and we call α the *lapse function*.

We have identified our spatial slices, these give the normal n_a an invariant meaning, and we have used up our freedom to choose a time coordinate. We still have freedom to choose our spatial coordinates, however, and that choice affects the "up-stairs", contravariant components of the normal vector. Specifically, note that the normalization condition (2.34) is satisfied by

$$n^a = \frac{1}{\alpha}(1, -\beta^i), \tag{2.38}$$

for any spatial vector β^i. Viewing n^a as a four-velocity helps to interpret the function of β^i. For $\beta^i = 0$ the spatial components of the normal observer's four-velocity vanish, meaning that the spatial coordinates x^i of the normal observer do not change; we can think of the spatial coordinates being attached to the normal observer. For non-zero β^i, however, the normal observer moves with respect with to the spatial coordinates. Using an argument similar to (2.36) above we find that, during an advance of proper time of $d\tau$, the normal observer would measure the spatial coordinates change by

$$dx^i = \frac{\partial x^i}{\partial x^a}(n^a d\tau) = \delta^i{}_a(n^a d\tau) = n^i d\tau = -\beta^i dt. \tag{2.39}$$

Conversely, a *coordinate observer*, i.e. an observer who is attached to the spatial coordinates, appears to shift by $\beta^i dt$ in a proper time $d\tau$ as seen by the normal observer. We therefore call β^i the *shift vector*; it measures the rate at which spatial coordinates shift with respect to the normal observer.

In Section 2.2.2 we simply picked time and space components when we decomposed the tensors and equations. Now we need the projections of tensors either onto the spatial slices, or along the normal vector. For example, we now decompose the spacetime vector A^a according to

$$A^a = A^a_{||} + A^a_{\perp}, \tag{2.40}$$

where $A^a_{||}$ is parallel to the normal vector n^a, and $A^a_{\perp}$ is perpendicular on the normal vector, i.e. tangent to our spatial slice. The magnitude of $A^a_{||}$ is given by the projection of A^a along n^a

$$\phi \equiv -n_b A^b. \tag{2.41}$$

We can again interpret n_a as the four-velocity of the normal observer, and see that this ϕ is the time component of A^a, i.e. the gauge variable ϕ (see (2.26)), as seen by a normal observer. We then write

$$A^a_{||} = \phi\, n^a = -n^a n_b A^b. \tag{2.42}$$

Now we can solve (2.40) for $A^a_{\perp}$ to find

$$A^a_{\perp} = A^a - A^a_{||} = g^a{}_b A^b + n^a n_b A^b = (g^a{}_b + n^a n_b)A^b. \tag{2.43}$$

The term in parentheses is aptly called the *projection operator*

$$P^a_{\ b} \equiv g^a_{\ b} + n^a n_b, \tag{2.44}$$

since it projects a spacetime tensor A^a into a spatial slice. Acting on a general spacetime tensor it will result in a new tensor that is tangent to the spatial slice, and hence spatial. Collecting results we now have

$$A^a = \phi n^a + A^a_\perp. \tag{2.45}$$

Exercise 2.2 (a) Show that $A^a_\perp$ is indeed spatial by showing that its contraction with the normal vector n^a vanishes.

(b) Show that $A^a_\parallel$ is indeed normal by showing that its contraction with the projection operator $P^a_{\ b}$ vanishes.

Before proceeding we will consider projections of a few more tensors, since it will allow us to make an early acquaintance with objects that we will encounter again in the context of general relativity. Admittedly, though, this involves a few somewhat tedious calculations. To help with the process, here is roadmap to objects that we will meet along the way, and that will be important later on:

- we define the *extrinsic curvature* K_{ab} in (2.49); it measures how much a slice is "bent",
- we define the *spatial covariant derivative* D_a in (2.57); as the name suggests, it is the spatial cousin of the four-dimensional covariant derivative,
- we meet the *Lie derivative* in (2.63); it measures those changes in a tensor field that are not the result of a coordinate transformation,
- and finally we derive (2.71), which relates spacetime derivatives of A^a to spatial derivatives of $A^a_\perp$, plus terms to make up for the difference.

All the above will play a key role in our treatment of general relativity. Readers eager to move forward could just take a quick peak at the objects listed above, and then move on to general relativity in Section 2.3, but readers interested in learning some of the tricks of the trade may want to work through the remainder of this Section first.

We start with a decomposition of the charge density four-vector j^a,

$$j^a = g^a_{\ b} j^b = (P^a_{\ b} - n^a n_b) j^b = n^a(-n_b j^b) + P^a_{\ b} j^b, \tag{2.46}$$

where the metric $g^a_{\ b} = \delta^a_{\ b}$ after the first equality acts as a Kronecker delta. Identifying $\rho = -n_b j^b$ as the charge density as observed by the normal observer, and $j^a_\perp = P^a_{\ b} j^b$ as the charge density current, again as observed by the normal observer, we find

$$j^a = \rho n^a + j^a_\perp. \tag{2.47}$$

For rank-2 tensors we project each index individually, so that we can have completely normal, completely spatial, and mixed projections. As an example, consider the covariant derivative of the normal vector,

$$\begin{aligned} \nabla_a n_b &= g_a^{\ c} g_b^{\ d} \nabla_c n_d = (P_a^{\ c} - n_a n^c)(P_b^{\ d} - n_b n^d)\nabla_c n_d \\ &= P_a^{\ c} P_b^{\ d} \nabla_c n_d - P_b^{\ d} n_a n^c \nabla_c n_d - P_a^{\ c} n_b n^d \nabla_c n_d + n_a n_b n^c n^d \nabla_c n_d. \end{aligned} \tag{2.48}$$

The first term on the right-hand side of (2.48) is the complete spatial decomposition of $\nabla_a n_b$. It measures how much the normal vector changes as we move from point to point in our spatial slice, and hence how much our slice is "bent". This quantity plays a crucial role in the context of general relativity also; we therefore define the *extrinsic curvature* as

$$K_{ab} \equiv -P_a{}^c P_b{}^d \nabla_c n_d, \tag{2.49}$$

where the minus sign follows the convention that we will adopt.

Exercise 2.3 Insert the definition (2.33) into (2.49) to show that the extrinsic curvature is symmetric,

$$K_{ab} = K_{ba}, \tag{2.50}$$

or equivalently $K_{[ab]} = 0$.

The term

$$a_a \equiv n^b \nabla_b n_a \tag{2.51}$$

measures the *acceleration* of normal observer. For normal observers who follow geodesics, equation (1.25) implies $a_a = 0$, as expected. Finally, the normalization of the normal vector, $n_a n^a = -1$, implies that

$$n^a \nabla_b n_a = 0, \tag{2.52}$$

so that the last two terms on the right-hand side of (2.48) vanish. We can use the same argument to show that the acceleration a_a is purely spatial, $n^a a_a = 0$, so that $P_a{}^b a_b = a_a$. The decomposition of $\nabla_a n_b$ therefore reduces to

$$\nabla_a n_b = -K_{ab} - n_a a_b, \tag{2.53}$$

where the first term on the right-hand side is purely spatial, and the second term is a mixed spatial-normal projection.

We next consider projections of the Faraday tensor

$$\begin{aligned} F^{ab} &= g^a{}_c g^b{}_d F^{cd} = (P^a_{\ c} - n^a n_c)(P^b_{\ d} - n^b n_d) F^{cd} \\ &= P^a_{\ c} P^b_{\ d} F^{cd} - P^b_{\ d} n^a n_c F^{cd} - P^a_{\ c} n^b n_d F^{cd} + n^a n^b n_c n_d F^{cd} \end{aligned} \tag{2.54}$$

Because F^{ab} is antisymmetric, the contraction $n_c n_d F^{cd}$ in the last term must vanish identically. In order to simplify the middle two terms we evaluate the Faraday tensor (2.18) in the normal observer's frame, i.e. with $u^a = n^a$, to find

$$E^a = n_b F^{ab}. \tag{2.55}$$

The first term in (2.54) requires a little more work, but it also reduces to something very compact. Using the decomposition (2.45) we find

$$\begin{aligned} P^a_{\ c} P^b_{\ d} F^{cd} &= P^a_{\ c} P^b_{\ d} (\nabla^c A^d - \nabla^d A^c) = P^a_{\ c} P^b_{\ d} \left(\nabla^c (\phi n^d + A^d_\perp) - \nabla^d (\phi n^c + A^c_\perp) \right) \\ &= P^a_{\ c} P^b_{\ d} \left(n^d \nabla^c \phi - n^c \nabla^d \phi + \phi \nabla^c n^d - \phi \nabla^d n^c + \nabla^c A^d_\perp - \nabla^d A^c_\perp \right). \end{aligned} \tag{2.56}$$

The first two terms vanish because the spatial projection of the normal vector vanishes, $P^a_{\ c} n^c = 0$. The middle two terms can be rewritten in terms of the extrinsic curvature (2.49); we then recognize that the two terms cancel each other because the extrinsic curvature is symmetric (see (2.50)). That leaves only the last two terms, which present us with a new type of object, namely the spatial projection of the covariant derivative of a spatial tensor. We define this as the *spatial covariant derivative* and denote it with the operator D_i, e.g.

$$D_a A^b_\perp \equiv P_a^{\ c} P^b_{\ d} \nabla_c A^d_\perp. \tag{2.57}$$

In Section 2.3 below we will discuss this derivative in more detail, and will explain how it can be computed conveniently from spatial objects alone. Using the definition (2.57) in (2.56) we now have

$$P^a_{\ c} P^b_{\ d} F^{cd} = D^a A^b_\perp - D^b A^a_\perp. \tag{2.58}$$

Collecting terms, we see that the decomposition (2.54) of the Faraday tensor reduces to

$$F^{ab} = D^a A^b_\perp - D^b A^a_\perp + n^a E^b - n^b E^a. \tag{2.59}$$

Exercise 2.4 Show that the acceleration of a normal observer (2.51) is related to the lapse α according to

$$a_a = D_a \ln \alpha. \tag{2.60}$$

Hint: This requires several steps. Insert $n_a = -\alpha \nabla_a t$ into the definition of a_a, commute the second derivatives of t, use $n^b \nabla_a n_b = 0$, and finally recall the definition of the spatial covariant derivative.

In (2.59), we have used the electric field (2.55) to express a normal projection of the Faraday tensor. Expressing the same projection in terms of the vector potential A^a yields

$$\begin{aligned} n^c F_{cd} &= n^c(\nabla_c A_d - \nabla_d A_c) = n^c \left(\nabla_c(\phi n_d + A^\perp_d) - \nabla_d(\phi n_c + A^\perp_c)\right) \\ &= n^c \left(\phi \nabla_c n_d + n_d \nabla_c \phi + \nabla_c A^\perp_d - \phi \nabla_d n_c - n_c \nabla_d \phi - \nabla_d A^\perp_c\right) \\ &= \phi a_d + n^c n_d \nabla_c \phi + \nabla_d \phi + n^c \nabla_c A^\perp_d + A^\perp_c \nabla_d n^c. \end{aligned} \tag{2.61}$$

Here we used (2.52) to eliminate the fourth term in the second line, and the acceleration (2.51) to rewrite the first term. Using (2.60) we can now combine this term with the following two to find

$$\begin{aligned} \phi a_d + n^c n_d \nabla_c \phi + \nabla_d \phi &= \phi D_d \ln \alpha + (g^c_{\ d} + n^c n_d) \nabla_c \phi = \phi D_d \ln \alpha + P^c_{\ d} \nabla_c \phi \\ &= \frac{1}{\alpha}\left(\phi D_d \alpha + \alpha D_d \phi\right) = \frac{1}{\alpha} D_d(\alpha \phi). \end{aligned} \tag{2.62}$$

In the last term in (2.61) we used the fact that $n^c A^\perp_c = 0$, so that $-n^c \nabla_d A^\perp_c = A^\perp_c \nabla_d n^c$. The trained eye may now recognize the last two terms in (2.61) as the *Lie derivative* of $A^\perp_d$ along n^a,

$$\mathcal{L}_n A^\perp_d = n^c \nabla_c A^\perp_d + A^\perp_c \nabla_d n^c. \tag{2.63}$$

Whereas the covariant derivative of a vector A^a along another vector n^a vanishes if A^a is parallel transported along n^a, the Lie derivative vanishes if the changes in A^a merely result from an infinitesimal coordinate transformation generated by n^a. Accordingly, the Lie derivative measures those changes in the tensor field that are not produced by a coordinate transformation generated by n^a.[2] Writing the normal vector n^a as

$$\alpha n^a = t^a - \beta^a \tag{2.64}$$

[2]See, e.g., Appendix A in Baumgarte and Shapiro (2010) for a discussion of the Lie derivative.

with $t^a = (1, 0, 0, 0)$ we can rewrite the Lie derivative (2.63) as

$$\alpha \mathcal{L}_n A_d^{\perp} = \mathcal{L}_{\alpha n} A_d^{\perp} = \mathcal{L}_t A_d^{\perp} - \mathcal{L}_\beta A_d^{\perp} \tag{2.65}$$

It is one of the properties of Lie derivatives that the Lie derivative along a coordinate vector, like t^a, reduces to a partial derivative

$$\mathcal{L}_t A_d^{\perp} = \partial_t A_d^{\perp}. \tag{2.66}$$

Collecting terms we see that the normal projection of the Faraday tensor becomes

$$n^c F_{cd} = \frac{1}{\alpha} \left(D_d(\alpha\phi) + \partial_t A_d^{\perp} - \mathcal{L}_\beta A_d^{\perp} \right). \tag{2.67}$$

According to (2.55) the left-hand side must be equal to $-E_d$, and we therefore find

$$\partial_t A_a^{\perp} = -\alpha E_a - D_a(\alpha\phi) + \mathcal{L}_\beta A_a^{\perp}. \tag{2.68}$$

This equation is the generalization of (2.14) and (2.29) for arbitrary time slices, and reduces to the latter for $\alpha = 1$ and $\beta^i = 0$. Crudely speaking, we see that we may interpret E^a as the time derivative of $A_a^{\perp}$.

As the next step we decompose the divergence of the Faraday tensor. Using (2.59) we obtain

$$\nabla_a F^{ab} = P^b_{\ c} \left(\mathcal{L}_n E^c - E^c K + \nabla_a (D^a A^c_{\perp} - D^c A^a_{\perp}) \right) - n^b D_a E^a, \tag{2.69}$$

where $\mathcal{L}_n E^c = n^a \nabla_a E^c - E^a \nabla_a n^c$ and where $K = g^{ab} K_{ab} = \nabla_a n^a$ is the trace of the extrinsic curvature, also known as the *mean curvature.*

Exercise 2.5 Derive equation (2.71).

Finally, we can insert a Kronecker delta $\delta^a_{\ d} = g^a_{\ d} = P^a_{\ d} - n^a n_d$ into the third term in (2.71) in order to convert the ∇_a into a D_a,

$$P^b_{\ c} \nabla_a (D^a A^c_{\perp} - D^c A^a_{\perp}) = P^b_{\ c} (P^a_{\ d} - n^a n_d) \nabla_a (D^d A^c_{\perp} - D^c A^d_{\perp}) = \frac{1}{\alpha} D_a (\alpha D^a A^b_{\perp} - \alpha D^b A^a_{\perp}). \tag{2.70}$$

Collecting terms we may therefore rewrite (2.71) as

$$\alpha \nabla_a (\nabla^a A^b - \nabla^b A^a) = D_a (\alpha D^a A^b_{\perp} - \alpha D^b A^a_{\perp}) + +\alpha P^b_{\ a} \mathcal{L}_n E^a - \alpha E^b K - \alpha n^b D_a E^a \tag{2.71}$$

where the first four terms are purely spatial, and the last term is normal.

It is well worth to inspect equation (2.71) a little more carefully. On the left-hand side we have four-dimensional derivatives of our four-dimensional fundamental quantity, A^a. The first term on the right-hand side has a similar appearance, except that it involves spatial derivatives of the spatial projection of this fundamental quantity only. It is evident that this term alone cannot contain all the terms on the left hand side. For example, the left-hand side contains up to second time derivatives of the fundamental variable, which we cannot accommodate in the first term on the right hand side. These missing terms are accounted for by the remaining terms on the right-hand side – second time derivatives, for example, appear in the term $\partial_t E^b$ by virtue of (2.68).

We also note that we have not invoked Maxwell's equations in our derivation of (2.71) – instead, that equation results from the decomposition of the vector field A^a and its derivatives in terms of spatial objects. Stated differently, equation (2.71) is a consequence of differential geometry and independent of physical laws. Maxwell's equation in our 3+1 decomposition now follow immediately by inserting (2.71) into our field equation (2.25). In the spatial projection we can again express the Lie derivative in terms of the time derivative

$$\alpha P^b{}_c \mathcal{L}_n E^c = P^b{}_c \mathcal{L}_{\alpha n} E^c = \partial_t E^b - \mathcal{L}_\beta E^b \tag{2.72}$$

to find

$$\partial_t E^b = D_a(\alpha D^b A^a_\perp - \alpha D^a A^b_\perp) + \mathcal{L}_\beta E^b + \alpha E^b K - 4\pi\alpha j^b_\perp. \tag{2.73}$$

This generalizes our earlier result (2.15). The normal projection of (2.71) results in the constraint equation

$$D_a E^a = 4\pi\rho \tag{2.74}$$

which, in fact, takes a form identical to (2.16). In (2.74) we have also reintroduced our color-coding from the previous Sections. Doing the same for (2.68) and (2.73) we obtain

$$\partial_t A^\perp_a = -\alpha E_a - D_a(\alpha\phi) + \mathcal{L}_\beta A^\perp_a \tag{2.75}$$

$$\partial_t E^a = D_b(\alpha D^a A^b_\perp - \alpha D^b A^a_\perp) + \alpha E^a K - 4\pi\alpha j^a_\perp + \mathcal{L}_\beta E^a. \tag{2.76}$$

This pair of equations now generalizes (2.14) and (2.15) for general time slices.

2.3 General Relativity

Finally we are ready to discuss a 3+1 decomposition of Einstein's equations. We are ready indeed, since we have introduced almost all necessary objects and concepts in Section 2.2 already. There, the vector potential A^a was our fundamental four-dimensional quantity; now it is the spacetime metric g_{ab}. Following our treatment in Section 2.2 we decompose g_{ab} as

$$g_{ab} = g_a{}^c g_b{}^d g_{cd} = (P_a{}^c - n_a n^c)(P_b{}^d - n_b n^d) g_{cd} = P_a{}^c P_b{}^d g_{cd} - n_a n_b. \tag{2.77}$$

We now define the first term on the right-hand side as the *spatial metric*

$$\gamma_{ab} \equiv P_a{}^c P_b{}^d g_{cd} \tag{2.78}$$

and then have

$$\gamma_{ab} = g_{ab} + n_a n_b. \tag{2.79}$$

Drawing on the analogy with electromagnetism we note that the spatial metric γ_{ab} is the cousin of the spatial vector potential $A^a_\perp$.

Exercise 2.6 We can "count" the number of dimensions d of a space by taking the trace of the metric, i.e. $d = g^a{}_a$. Take the trace of the induced metric $\gamma_{ab} = g_{ab} + n_a n_b$ to show that the dimension of the spatial slice is one less than that of the spacetime.

The spatial metric plays the same role for spatial tensors as the spacetime metric does for spacetime tensors. For example, we can raise and lower indices of spatial tensors with the spatial metric instead of the spacetime metric. For example

$$\beta_a = g_{ab}\beta^b = (\gamma_{ab} - n_a n_b)\beta^b = \gamma_{ab}\beta^b, \tag{2.80}$$

since $n_b\beta^b = 0$. Since $n_a = (-\alpha, 0, 0, 0)$, this last equality also shows that $\beta^t = 0$. In fact, the contravariant ("upstairs") time component of spatial tensors must always vanish. Therefore, we may also restrict our indices to spatial indices when we are dealing with spatial tensors. In the following we will use the convention that indices $a, b, c, \ldots$ run over spacetime indices, and $i, j, k, \ldots$ over spatial indices only.

The four-dimensional line element can now be written as

$$ds^2 = g_{ab}dx^a dx^b = -\alpha^2 dt^2 + \gamma_{ij}(dx^i + \beta^i dt)(dx^j + \beta^j dt). \tag{2.81}$$

We can think of this as a "spacetime pythagorian": the first time on the right-hand side measures the proper distance from t to $t + dt$ along the normal vector (which we have already identified as αdt in (2.37)), while the remaining terms measure the proper distance inside the spatial slice.

Exercise 2.7 If all went well in Exercise 1.3, you should have found that the Schwarzschild metric in isotropic coordinates (on slices of constant Schwarzschild time) is given by

$$ds^2 = -\left(\frac{1 - M/(2r)}{1 + M/(2r)}\right)^2 dt^2 + \left(1 + \frac{M}{2r}\right)^4 (dr^2 + r^2 d\Omega^2). \tag{2.82}$$

Compare this line element with the "3+1" line element (2.81) to identify the lapse function α, the shift vector β^i and the spatial metric γ_{ij}.

In (2.57) we defined the spatial covariant derivative as the total spatial projection of the covariant derivative of a spatial tensor. It turns out, though, that we can compute the spatial covariant derivative using the exact same formalism as for the spacetime covariant derivative, e.g. (1.26), except that we replace all terms with the corresponding spatial terms. For example, we would find

$$D_i V^j = \partial_i V^j + V^k \Gamma^j_{ki}, \tag{2.83}$$

where V^j is a spatial vector, and where the Christoffel symbols Γ^j_{kj} are computed from derivatives of the spatial metric.

Next we'll ask the reader to show that we can express the extrinsic curvature (2.49) as the Lie derivative of γ_{ab} along the normal vector n^a.

Exercise 2.8 Evaluate the Lie derivative of $\gamma_{ab} = g_{ab} + n_a n_b$ along n^a,

$$\mathcal{L}_n \gamma_{ab} = n^c \nabla_c \gamma_{ab} + \gamma_{cb} \nabla_a n^c + \gamma_{ac} \nabla_b n^c, \tag{2.84}$$

to show that

$$K_{ab} = -\frac{1}{2}\mathcal{L}_n \gamma_{ab}. \tag{2.85}$$

Using (2.64) and the following discussion, we can rewrite (2.85) as

$$\partial_t \gamma_{ij} = -2\alpha K_{ij} + \mathcal{L}_\beta \gamma_{ij} = -2\alpha K_{ij} + D_i \beta_j + D_j \beta_i. \tag{2.86}$$

Evidently, we can think of the extrinsic curvature as measuring the time derivative of the spatial metric.

In full disclosure, two more properties of derivatives went into the derivation of the last equation. First, the covariant derivative of the associated metric always vanishes, i.e. $\nabla_a g_{bc} = 0$ (see Exercise B.2), and similarly $D_a \gamma_{bc} = 0$. Also, it turns out that the Lie derivative can be computed using *any* covariant derivative. In fact, evaluating the covariant derivatives in (2.84) shows that all Christoffel symbols drop out identically. We can therefore replace the covariant derivatives in (2.84) with partial derivatives, or, as we did in deriving (2.86), with the spatial covariant derivative. In that case the term $D_c\gamma_{ab}$ vanishes, and we are left with spatial covariant derivatives of the shift vector only.

Since we have previously identified γ_{ij} as the relativistic cousin of the $A^a_\perp$, we now compare (2.86) with (2.14) or (2.68) to see that the intrinsic curvature K_{ab} plays the same role here as the electric field E^a in electromagnetism.

Here is a point of potential confusion. In electromagnetism, we assume that a metric is given. In a four-dimensional description, the vector potential A^a then evolves according to (2.25) in this given spacetime. For example, if the spacetime is flat, we may take the spacetime metric g_{ab} to be given by the Minkowski metric η_{ab}. We then introduced spatial slices, which resulted in our definition (2.49) of the extrinsic curvature K_{ab}. Evidently, we could also define the spatial metric γ_{ab} as in (2.78) in the context of electromagnetism. In electromagnetism, however, neither γ_{ab} nor K_{ab} are dynamical variables; they are a result of the spacetime metric and the slicing only. Instead, the vector potential A^a can be considered the dynamical variable, or, in a 3+1 decomposition, the spatial vectors $A^i_\perp$ and E^i.

In general relativity, the spacetime metric g_{ab} itself becomes the dynamical variable – therefore it simultaneously plays the role of the metric and of the vector potential A^a in electrodynamics. By the same token, the spatial metric γ_{ij} and extrinsic curvature K_{ij} become dynamical variables in a 3+1 decomposition of general relativity. When comparing with electromagnetism, γ_{ij} therefore absorbs the dual roles of spatial metric and $A^i_\perp$, while the K_{ij} takes over the dual roles of extrinsic curvature and E^i.

The relativistic analog of the second derivatives of A^a in (2.25), or of ϕ in (2.1), or of Φ in (1.13), is the Einstein tensor (1.30), which involves the trace of the Riemann tensor. Just like we decomposed the second derivatives of A^a into normal and spatial parts in (2.71), we can decompose the Einstein tensor into normal and spatial parts. Usually this process starts with projections of the Riemann tensor, which leads to the famous equations by Gauss, Codazzi, Mainardi, and Ricci. Instead of deriving these equations here,[3] we will invoke the analogy with (2.71) to anticipate what kind of terms we should expect.

The first term on the right-hand side of (2.71) is the spatial cousin of the left-hand side – in general relativity, this becomes the spatial Riemann tensor, which is computed just like the spacetime Riemann tensor, but from spatial derivatives of the spatial metric only. The next term on the right-hand side of (2.71) is the Lie derivative of E^a along n^a – by analogy, we now expect the Lie derivative of the extrinsic curvature K_{ab} along n^a. As in electrodynamics, we can express this Lie derivative in terms of the time derivative of K_{ab} and the Lie derivative along β^i. The next term in (2.71) is the product E^aK. In general relativity, the extrinsic curvature takes over the role of both E^a and K, therefore we should not be surprised to find terms quadratic in K_{ab}. Finally, we expect to encounter the spatial divergence of K_{ab}, in analogy to the last term in (2.71).

Just like (2.71), the equations of Gauss, Codazzi, Mainardi and Ricci are based on differential geometry only, and do not yet invoke Einstein's equations. In order to complete our 3+1 decomposition of Einstein's equations, we now equate the projections of the Einstein tensor with projections of the stress-energy tensor on the right-hand side of (1.31). Projecting both indices along the normal results in the *Hamiltonian constraint*

$$R + K^2 - K_{ij}K^{ij} = 16\pi\rho, \tag{2.87}$$

[3]See, e.g., Section 2.4 in Baumgarte and Shapiro (2010) for a detailed treatment.

where

$$\rho \equiv n^a n^b T_{ab} \tag{2.88}$$

is the energy density as measured by a normal observer. A mixed spatial-normal projection now yields a second constraint equation, the *momentum constraint*

$$D_j(K^{ij} - \gamma^{ij} K) = 8\pi j^i, \tag{2.89}$$

where

$$j^i \equiv -\gamma^{ij} n^k T_{jk} \tag{2.90}$$

is the energy flux as seen by a normal observer. We have re-introduced our color-coding in order the emphasize the similarity with the electrodynamic constraint equation (2.74) – a divergence of the "red" variable is equal to a matter term.

Finally, a complete spatial projection results in the evolution equation for the extrinsic curvature. Combining this result with (2.86) we obtain the pair

$$\partial_t \gamma_{ij} = -2\alpha K_{ij} + \mathcal{L}_\beta \gamma_{ij} \tag{2.91}$$

$$\partial_t K_{ij} = \alpha(R_{ij} - 2K_{ik}K^k{}_j + KK_{ij}) - D_i D_j \alpha + 4\pi\alpha(2S_{ij} - \gamma_{ij}(S - \rho)) + \mathcal{L}_\beta K_{ij}, \tag{2.92}$$

where

$$S_{ij} = P_i{}^k P_j{}^l T_{kl} \tag{2.93}$$

is the stress according the a normal observer, and $S \equiv \gamma^{ij} S_{ij}$ its trace. The color coding again emphasizes the similarity with both the scalar wave of Section 2.1 – see equations (2.4) and (2.5) – and electrodynamics – compare either with the more common pair of equations (2.14) and (2.15) or the corresponding expressions for general time slices, equations (2.75) and (2.76). In all cases the first equation describes the time derivative of a first variable (the "green" variable) as the time derivative of s second variable (the "red" variable), while the second equation describes the time derivative of this second "red" variable in terms of second spatial derivatives of the "green" variable and "blue" matter sources. In electrodynamics and general relativity we find additional "golden" gauge terms on the right-hand sides; for general time slices, in particular, the Lie derivative along the shift vector β^i makes an appearance. Also note that, in general relativity, the spatial derivatives of the "green" variable are hidden in the Ricci tensor.

The constraint equations (2.87) and (2.89) together with the two evolution equations (2.91) and (2.92) form the famous "ADM" equations, named after Arnowitt, Deser and Misner.[4] In the following chapters we will discuss some strategies for how these solutions can be solved. Before proceeding to that discussion, though, we will comment on the role of the lapse function α and the shift vector β^i.

Note that the lapse and shift make appearances in the evolution equations (2.91) and (2.92) only, and not in the constraint equations (2.87) and (2.89). That is not entirely surprising – when we encountered the lapse and shift in Section 2.2.3, we saw that the lapse describes the advance of coordinate time, while the shift vector describes the evolution of the spatial coordinates. Accordingly, the lapse and shift make appearances in the evolution equations only. Note also, that Einstein's equations do not determine the lapse and shift. Instead, they encode the coordinate freedom inherent in general relativity. Part of our discussion in Chapter ??? will therefore be how to make suitable choices for the lapse and shift.

[4]See Arnowitt et al. (1962); also York, Jr. (1979).

Appendix A

A brief review of tensor properties

A tensor is a physical object and has an intrinsic meaning independently of coordinates or basis vectors. An example is a rank-1 tensor $\mathbf{A}$ that has a certain length and points in a certain direction independently of the coordinate system in which we express this tensor.

We can expand any tensor in terms of either basis vectors $\mathbf{e}_a$ or basis 1-forms $\tilde{\omega}^a$. Expanding a rank-1 tensor in terms of basis vectors, for example, yields

$$\mathbf{A} = A^a \mathbf{e}_a. \tag{A.1}$$

This expression deserves several comments. For starters, we have used the Einstein summation rule, meaning that we sum over repeated indices. The "up-stairs" index a on A^a refers to a *contravariant* component of $\mathbf{A}$ – meaning one that is used in an expansion of $\mathbf{A}$ in terms of basis vectors. The index a on the basis vector $\mathbf{e}_a$, on the other hand, does *not* refer to a component of the basis vector – instead it denotes the name of the basis vector (e.g. the basis vector pointing in the x direction). If we wanted to refer to the b-th component of the basis vector $\mathbf{e}_a$, say, we would write $(\mathbf{e}_a)^b$.

An immediate question is whether we always have $(\mathbf{e}_a)^b = \delta_a{}^b$, where $\delta_a{}^b$ is the Kronecker delta. The answer is no – this is true *only* if we are expressing normalized basis vectors in their own coordinate system, for example, if we are expressing Cartesian basis vectors in Cartesian coordinates. In general, however, this is not true. Think, for instance, about Cartesian basis vectors expressed in a spherical polar coordinate system.

We now write the dot product between two vectors as

$$\mathbf{A} \cdot \mathbf{B} = (A^a \mathbf{e}_a) \cdot (B^b \mathbf{e}_b) = A^a B^b \, \mathbf{e}_a \cdot \mathbf{e}_b. \tag{A.2}$$

Defining the metric as

$$g_{ab} \equiv \mathbf{e}_a \cdot \mathbf{e}_b \tag{A.3}$$

we obtain

$$\mathbf{A} \cdot \mathbf{B} = A^a B^b g_{ab}. \tag{A.4}$$

Expanding a rank-1 tensor in terms of 1-forms $\tilde{\omega}^a$ yields

$$\mathbf{B} = B_a \tilde{\omega}^a, \tag{A.5}$$

where the "down-stairs" index a refers to a *covariant* component. We call the basis 1-forms $\tilde{\omega}^a$ dual to the basis vectors $\mathbf{e}_b$ if

$$\tilde{\omega}^a \cdot \mathbf{e}_b = \delta^a{}_b. \tag{A.6}$$

This is what we will assume throughout. We can then compute the dot product between **B** and **A** as

$$\mathbf{A}\cdot\mathbf{B} = (A^a\mathbf{e}_a)\cdot(B_b\tilde{\omega}^b) = A^aB_b\,\mathbf{e}_a\cdot\tilde{\omega}^b = A^aB_b\,\delta_a{}^b = A^aB_a. \tag{A.7}$$

Since both this expression and (A.4) have to hold for any tensor **A**, we can compare the two and identify

$$B_a = g_{ab}B^b. \tag{A.8}$$

We refer to this operation as "lowering the index of B^a".

We define the inverse metric g^{ab} so that

$$g^{ac}g_{cb} = \delta^a{}_b. \tag{A.9}$$

We then "raise the index of B_a" using

$$B^a = g^{ab}B_b. \tag{A.10}$$

We can also show that

$$g^{ab} = \tilde{\omega}^a\cdot\tilde{\omega}^b. \tag{A.11}$$

Note that we can find the contravariant component of a rank-1 tensor **A** by computing the dot product with the corresponding 1-form,

$$A^a = \mathbf{A}\cdot\tilde{\omega}^a. \tag{A.12}$$

We can verify this expression by inserting the expansion (A.1) for **A**, and then using the duality relation (A.6). For covariant components we similarly take the dot product with basis vectors.

Under a change of basis, i.e. when we transform from one coordinate system x^a to another, say $x^{b'}$, basis vectors and basis 1-forms transform according to

$$\mathbf{e}_{b'} = M^a{}_{b'}\mathbf{e}_a \tag{A.13}$$
$$\tilde{\omega}^{b'} = M^{b'}{}_a\tilde{\omega}^a \tag{A.14}$$

where $M^{b'}{}_a$ is the transformation matrix and $M^a{}_{b'}$ its inverse, so that

$$M^{a'}{}_cM^c{}_{b'} = \delta^{a'}{}_{b'}. \tag{A.15}$$

We can then use the above rules to find the transformation rules for components of tensors, e.g.

$$A^{b'} = \mathbf{A}\cdot\tilde{\omega}^{b'} = \mathbf{A}\cdot M^{b'}{}_a\tilde{\omega}^a = M^{b'}{}_a\mathbf{A}\cdot\tilde{\omega}^a = M^{b'}{}_aA^a, \tag{A.16}$$

A specific example is a Lorentz transformation, for which the transformation matrix $M^{a'}_{\ b}$ is often denoted with $\Lambda^{a'}{}_b$. In this case we often refer to a transformation between two reference frames. For a boost with speed v in the z-direction, for example, this matrix takes the form

$$M^{a'}{}_b = \Lambda^{a'}{}_b = \begin{pmatrix} \gamma & 0 & 0 & -\gamma v \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ -\gamma v & 0 & 0 & \gamma \end{pmatrix} \tag{A.17}$$

where $\gamma \equiv (1-v^2)^{-1/2}$ is the Lorentz factor.

Exercise A.1 (a) A particle at rest in an unprimed coordinate system has a four-velocity $u^a = (1,0,0,0)$. Show that in a primed coordinate system, boosted with a speed of v in the negative z direction, the particle has a four-velocity $u^{a'} = (\gamma, 0, 0, \gamma v)$.

(b) A particle has a four-velocity $u^a = (\gamma_1, 0, 0, \gamma_1 v_1)$ in an unprimed coordinate system. Show that in a primed coordinate system, boosted with a speed of v_2 in the negative z direction, the particle has a four-velocity $u^{a'} = (\gamma', 0, 0, \gamma' v')$ with $v' = (v_1 + v_2)/(1 + v_1 v_2)$ – which is Einstein's famous rule for the addition of velocities. What is γ' in terms of γ_1, γ_2, and the two speeds v_1 and v_2?

Note that vectors and 1-forms transform in "inverse ways". This guarantees that the duality relation (A.6) also holds in the new coordinate system,

$$\tilde{\omega}^{a'} \cdot \mathbf{e}_{b'} = (M^{a'}{}_c \tilde{\omega}^c) \cdot (M^d{}_{b'} \mathbf{e}_d) = M^{a'}{}_c M^d{}_{b'} (\tilde{\omega}^c \cdot \mathbf{e}_d) = M^{a'}{}_c M^d{}_{b'} \delta^c{}_d = M^{a'}{}_c M^c{}_{b'} = \delta^{a'}{}_{b'}. \tag{A.18}$$

The components of a vector then transform according to

$$A^{b'} = \mathbf{A} \cdot \tilde{\omega}^{b'} = \mathbf{A} \cdot (M^{b'}{}_a \tilde{\omega}^a) = M^{b'}{}_a A^a \tag{A.19}$$

and similarly

$$B_{b'} = M^a{}_{b'} B_a. \tag{A.20}$$

The fact that contravariant and covariant components transform in "inverse" ways guarantees that the dot product (A.7) is invariant under coordinate transformations,

$$A^{b'} B_{b'} = M^{b'}{}_a A^a M^c{}_{b'} B_c = \delta^c{}_a A^a B_c = A^a B_a. \tag{A.21}$$

The dot product between two vectors therefore transforms like a scalar, as it is supposed to.

We can generalize all the above concepts to higher-rank tensors. A rank-n tensor can be expanded into n basis vectors or 1-forms, and we transform the components of a rank-n tensor with n copies of the transformation matrix or its inverse – for example

$$F^{a'b'} = M^{a'}{}_c M^{b'}{}_d F^{cd}. \tag{A.22}$$

Exercise A.2 The electric and magnet fields in an unprimed coordinate system are given by E^i and B^i, and the Faraday tensor by (2.17). An observer in a primed coordinate observer would identify the electric and magnetic fields from the Faraday tensor as observed in the primed coordinate system, e.g. $E^{i'} = F^{t'i'} = M^{t'}{}_c M^{x'}{}_d F^{cd}$. Find the electric and magnetic fields as observed in a reference frame that is boosted with speed v in the z-direction with respect to the unprimed reference frame. For the x'-component, for example, the answer is $E^{x'} = \gamma E^x - \gamma v B^y$.

Find the components of electric and magnetic fields in a primed coordinate system that is boosted with a speed v in the positive z direction.

For transformations between *coordinate bases*, for which the basis vectors are tangent to coordinate lines, we have

$$M^{b'}{}_a \equiv \frac{\partial x^{b'}}{\partial x^a} = \partial_a x^{b'}. \tag{A.23}$$

As an illustration of the above concepts, consider the components of a displacement vector dx^a, which measures the displacement between two points expressed in a coordinate system x^a. To compute the components of this vector in a different coordinate system, say a primed coordinate system $x^{b'}$, we use the chain rule to obtain

$$dx^{b'} = \frac{\partial x^{b'}}{\partial x^a} dx^a = M^{b'}{}_a dx^a \tag{A.24}$$

where we have used (A.23) in the last step. As expected, the components of dx^a transform like the vector components in (A.19).

As an example of a 1-form, consider the components of the gradient $\partial f/\partial x^a$ of a function f, again expressed in some coordinate system x^a. To transform to a new coordinate system $x^{b'}$ we again use the chain rule, but this time we obtain

$$\frac{\partial f}{\partial x^{b'}} = \frac{\partial x^a}{\partial x^{b'}}\frac{\partial f}{\partial x^a} = M^a_{\ b'}\frac{\partial f}{\partial x^a} \tag{A.25}$$

as in (A.20). We see that the "inverse" transformation of the components of a gradient are a result of the chain rule.

Finally, consider the difference *df* in the function values f at two (close) points. Clearly, this difference is an invariant, i.e. independent of coordinate choice. We can express this difference as the dot product between the vector displacement vector dx^a between the two points and the 1-form $\partial f/\partial x^a$,

$$df = \frac{\partial f}{\partial x^a}dx^a. \tag{A.26}$$

In equation (2.36), for example, we will use this relation to relate the advance of proper time to that of coordinate time as measured by a normal observer.

Exercise A.3 Apply the above transformation rules for the components of vectors and 1-forms to show that *df* is indeed invariant under a coordinate transformation.

We refer to a rank-2 tensor T_{ab} as *symmetric* if $T_{ab} = T_{ba}$, and as *anti-symmetric* if $T_{ab} = -T_{ba}$. We then introduce the notation

$$T_{(ab)} = \frac{1}{2}(T_{ab} + T_{ba}) \tag{A.27}$$

for the symmetric part of a general tensor T_{ab}, and

$$T_{[ab]} = \frac{1}{2}(T_{ab} - T_{ba}) \tag{A.28}$$

for the anti-symmetric part.

Exercise A.4 (a) Verify that the symmetric part of an anti-symmetric tensor vanishes, and likewise the other way around.

(b) Show that the complete contraction between a symmetric tensor S_{ab} and an antisymmetric tensor A^{ab} vanishes,

$$S_{ab}A^{ab} = 0. \tag{A.29}$$

Appendix B

Important Results

This appendix serves mostly as a reference; very few of these results will be used in the lectures.

B.1 Differential Geometry and General Relativity

- We denote the metric with g_{ab} and the inverse metric with g^{ab}
- Covariant derivative

  ... of a scalar

  $$\nabla_a \Phi = \partial_a \Phi \tag{B.1}$$

  ... of a rank-1 tensor in terms of contravariant component

  $$\nabla_a V^b = \partial_a V^b + V^c \Gamma^b_{ac} \tag{B.2}$$

  ... of a rank-1 tensor in terms of the covariant component

  $$\nabla_a V_b = \partial_a V_b - V_c \Gamma^c_{ab} \tag{B.3}$$

  ... for higher-rank tensors we add one more Christoffel symbol term for each index with the corresponding sign for contravariant or covariant indices , e.g.

  $$\nabla_a T_{bc} = \partial_a T_{bc} - T_{dc}\Gamma^d_{ba} - T_{bd}\Gamma^d_{cd} \tag{B.4}$$

- Christoffel symbols

  $$\Gamma^a_{bc} = \frac{1}{2} g^{ad}(\partial_c g_{db} + \partial_b g_{dc} - \partial_d g_{bc}) \tag{B.5}$$

  Symmetry

  $$\Gamma^a_{bc} = \Gamma^a_{cb} \tag{B.6}$$

Exercise B.1 Since the dot product $V^b W_b$ transforms like a vector (see (A.21)), the covariant derivative of $V^b W_b$ should be be the same as its partial derivative. Prove this explicitly by using the product rule $\nabla_a V^b W_b = V^b \nabla_a W_b + W_b \nabla_a V^b$ and inserting (B.2) and (B.3).

Exercise B.2 Show that

$$\nabla_a g_{bc} = 0. \tag{B.7}$$

We say that the covariant derivative ∇_a is *associated* with the metric g_{ab}.

Exercise B.3 Consider a flat, three-dimensional space in spherical polar coordinates, for which the metric and inverse metric are given by

$$g_{ij} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & r^2 & 0 \\ 0 & 0 & r^2 \sin^2\theta \end{pmatrix}, \qquad g^{ij} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & r^{-2} & 0 \\ 0 & 0 & r^{-2}\sin^{-2}\theta \end{pmatrix}. \tag{B.8}$$

(a) Show that the only non-vanishing Christoffel symbols are given by

$$\begin{array}{rclcrcl} \Gamma^r_{\theta\theta} & = & -r & & \Gamma^r_{\phi\phi} & = & -r\sin^2\theta \\ \Gamma^\theta_{\phi\phi} & = & -\sin\theta\cos\theta & & \Gamma^\theta_{r\theta} & = & r^{-1} \\ \Gamma^\phi_{r\phi} & = & r^{-1} & & \Gamma^\phi_{\phi\theta} & = & \cot\theta. \end{array} \tag{B.9}$$

and those that are related to the above by the symmetry $\Gamma^i_{jk} = \Gamma^i_{kj}$.

(b) Show that the Laplace operator acting on a scalar function f, defined as

$$\nabla^2 f \equiv g^{ij}\nabla_i\nabla_j f, \tag{B.10}$$

reduces to the well-known expression

$$\nabla^2 f = \frac{1}{r^2}\frac{\partial}{\partial r}\left(r^2\frac{\partial f}{\partial r}\right) + \frac{1}{r^2\sin\theta}\frac{\partial}{\partial\theta}\left(\sin\theta\frac{\partial f}{\partial\theta}\right) + \frac{1}{r^2\sin^2\theta}\frac{\partial^2 f}{\partial\phi^2}. \tag{B.11}$$

- Equation of geodesic deviation

$$\frac{d^2\Delta x^a}{d\tau^2} = R^a{}_{bcd}u^b u^d \Delta x^c \tag{B.12}$$

- Riemann tensor

Definition

$$\nabla_a\nabla_b v_c - \nabla_b\nabla_a v_c = R^d_{cba}v_d \tag{B.13}$$

Compute from

$$R^a{}_{bcd} = \partial_c\Gamma^a_{bd} - \partial_d\Gamma^a_{bc} + \Gamma^a_{ec}\Gamma^e_{bd} - \Gamma^a_{ed}\Gamma^e_{bc} \tag{B.14}$$

Symmetries

$$R_{abcd} = -R_{bacd} \qquad R_{abcd} = -R_{abdc} \qquad R_{abcd} = R_{cdab} \tag{B.15}$$

$$R_{abcd} + R_{adbc} + R_{acdb} = 0 \tag{B.16}$$

(also Bianchi identities.) In n dimensions have $n^2(n^2-1)/12$ independent components.

- Ricci tensor and Ricci scalar

$$R_{ab} \equiv R^c_{acb} \qquad R \equiv g^{ab}R_{ab} \tag{B.17}$$

Three ways to compute Ricci

$$R_{ab} = \partial_c\Gamma^c_{ab} - \partial_b\Gamma^c_{ac} + \Gamma^c_{ab}\Gamma^d_{cd} - \Gamma^c_{ad}\Gamma^d_{bc} \tag{B.18}$$

$$R_{ab} = \frac{1}{2}g^{cd}\left(\partial_a\partial_d g_{cb} + \partial_c\partial_b g_{ad} - \partial_a\partial_b g_{cd} - \partial_c\partial_d g_{ab}\right) + g^{cd}\left(\Gamma^e_{ad}\Gamma_{ecb} - \Gamma^e_{ab}\Gamma_{ecd}\right) \tag{B.19}$$

$$R_{ab} = -\frac{1}{2}g^{cd}\partial_d\partial_c g_{ab} + g_{c(a}\partial_{b)}\Gamma^c + \Gamma^c\Gamma_{(ab)c} + 2g^{ed}\Gamma^c_{e(a}\Gamma_{b)cd} + g^{cd}\Gamma^e_{ad}\Gamma_{ecb} \tag{B.20}$$

where $\Gamma^a \equiv g^{bc}\Gamma^a_{bc}$.

- Einstein tensor

$$G_{ab} \equiv R_{ab} - \frac{1}{2} g_{ab} R \tag{B.21}$$

- Einstein's equations

$$\boxed{G_{ab} = 8\pi T_{ab}} \tag{B.22}$$

where T_{ab} is stress-energy tensor, and where we assume $\Lambda = 0$.

B.2 The $3+1$ Decomposition

- normal vector

$$n^a = -\alpha g^{ab} \nabla_b t \tag{B.23}$$

- induced or spatial metric

$$\gamma_{ab} = g_{ab} + n_a n_b \tag{B.24}$$

- Extrinsic curvature

$$K_{ab} \equiv -\gamma_a{}^c \gamma_b{}^d \nabla_c n_d = -\nabla_a n_b - n_a a_b = -\frac{1}{2} \mathcal{L}_{\mathbf{n}} \gamma_{ab} \tag{B.25}$$

where $a_a \equiv n^b \nabla_b n_a$ is acceleration of normal observer.

- ADM equations
  - Constraint equations

$$R + K^2 + K_{ij} K^{ij} = 16\pi\rho \qquad \text{Hamiltonian constraint} \tag{B.26}$$

$$D_i (K^{ij} - \gamma^{ij} K) = 8\pi S^i \qquad \text{momentum constraint} \tag{B.27}$$

  - Evolution equations

$$\partial_t \gamma_{ij} = -2\alpha K_{ij} + D_i \beta_j + D_j \beta_i \tag{B.28}$$

$$\begin{aligned} \partial_t K_{ij} &= \alpha (R_{ij} - 2K_{ik} K^k{}_j + K K_{ij}) - D_i D_j \alpha - 8\pi\alpha \left(S_{ij} - \frac{1}{2} \gamma_{ij} (S - \rho) \right) \\ &\quad + \beta^k \partial_k K_{ij} + K_{ik} \partial_j \beta^k + K_{kj} \partial_i \beta^k \end{aligned} \tag{B.29}$$

(B.30)

B.3 Conformal decompositions

- metric

$$\gamma_{ij} = \psi^4 \bar{\gamma}_{ij} \tag{B.31}$$

- extrinsic curvature

$$K_{ij} = A_{ij} + \frac{1}{3} \gamma_{ij} K = \psi^{-2} \bar{A}_{ij} + \frac{1}{3} \psi^4 \bar{\gamma}_{ij} K \tag{B.32}$$

- Hamiltonian constraint

$$\bar{D}^2\psi - \frac{\psi}{8}\bar{R} + \frac{\psi^{-7}}{8}\bar{A}_{ij}\bar{A}^{ij} - \frac{1}{12}\psi^5 K^2 = -2\pi\psi^5\rho \tag{B.33}$$

- momentum constraint

$$\bar{D}_j\bar{A}^{ij} - \frac{2}{3}\psi^6\bar{\gamma}^{ij}\bar{D}_j K = 8\pi\psi^{10}S^i \tag{B.34}$$

Bibliography

Abbott, B. P. *et.al.*. (2017). GW170817: Observation of gravitational waves from a binary neutron star inspiral. *Phys. Rev. Lett. 119*, 161101.

Arnowitt, R., S. Deser, and C. W. Misner (1962). The dynamics of general relativity. In L. Witten (Ed.), *Gravitation: an Introduction to Current Research*, pp. 227–265. Wiley, New York. Also available at gr-qc/0405109.

Baumgarte, T. W. and S. L. Shapiro (2010). *Numerical relativity: Solving Einstein's equations on the computer*. Cambridge University Press, Cambridge.

Bergmann, P. G. (1957). Summary of the Chapel Hill Conference. *Rev. Mod. Phys. 29*, 352–354.

Carroll, S. (2004). *Spacetime and Geometry: An Introduction to General Relativity*. Addison-Wesley, San Francisco.

Einstein, A. (1915). Die Feldgleichungen der Gravitation. *Preuss. Akad. Wiss. Berlin, Sitzber.*, 844–847.

Einstein, A. (1916). Näherungsweise Integration der Feldgleichungen der Gravitation. *Sitzungsberichte der Königlich Preußischen Akademie der Wissenschaften (Berlin), Seite 688-696.*.

Hartle, J. B. (2003). *Gravity: an Introduction to Einstein's General Relavity*. Addison Wesley.

Kerr, R. P. (1963). Gravitational Field of a Spinning Mass as an Example of Algebraically Special Metrics. *Phys. Rev. Lett. 11*, 237–238.

Misner, C. W., K. S. Thorne, and J. A. Wheeler (1973). *Gravitation*. Freeman, New York.

Moore, T. A. (2013). *A General Relativity Workbook*. University Science Books.

Schwarzschild, K. (1916). Über das Gravitationsfeld eines Massenpunktes nach der Einsteinschen Theorie. *Sitzber. Deut. Akad. Wiss. Berlin, Kl. Math.-Phys. Tech.*, 189–196.

Wald, R. M. (1984). *General Relativity*. The University of Chicago Press, Chicago.

York, Jr., J. W. (1979). Kinematics and dynamics of general relativity. In L. L. Smarr (Ed.), *Sources of gravitational radiation*, pp. 83–126. Cambridge University Press, Cambridge.