Filling the gaps in geometry and relativity

Thomas F. C. Bastos

May 4, 2024

Abstract

The interplay between geometry and physics has been known for many decades, but it's only in fundamental physics where we can fully appreciate this relationship. In these notes, we will uncover one of these theories, General Relativity, and see how differential geometry plays an essential role. In the first part, we will strive to understand what curvature is, using ideas from Gauss, Cartan and Riemann. The second part is concerned with gravity: how it is an effect of the curvature of spacetime, and how such ideas came into being. It is meant to be an introduction to the introduction of general relativity, and for those who are already acquainted with the theory, it should fill some gaps that may have been left when you studied GR for the first time.

0 The physicist approach

If I asked people what they think a physicist would do to study geometry, they would probably say: "By throwing particles in it". That's precisely what we are going to do! No, seriously. Consider a free particle in space $\mathbb{R}^3$. It's lagrangian is just kinetic energy $L = m\frac{v^2}{2}$. Using cartesian coordinates the Euler-Lagrange equation gives us

$$\frac{d}{dt}\left(\frac{dx^i}{dt}\right) = 0$$

which is, unsurprisingly, a straight line in euclidean space. Consider now the case of a free particle constrained to the surface of a sphere. Using spherical coordinates we have:

$$v^2 = R^2\dot{\theta}^2 + R^2\dot{\phi}^2 \sin^2\theta$$

The Euler-Lagrange equations for θ and ϕ are:

$$\ddot{\theta} - \sin\theta\cos\theta\dot{\phi}^2 = 0$$
$$\ddot{\phi} + 2\cot\theta\dot{\theta}\dot{\phi} = 0$$

One solution to this problem is $\theta(t) = \frac{\pi}{2}$ and $\phi(t) = \omega t$, which is just a particle in the equator with constant velocity. In fact, from the symmetry we can rotate this solution in such a way that for every two points, the path the particle will take lies in a great circle[1] that passes through both points. Observe that in every case the free particle take the path of minimal length, i.e, take the path of geodesics, and it's easy to see that this is a general property of particles constrained to surfaces since the action is just $S = \frac{m}{2}\int v^2\, dt \sim \int ds$.

A point on a general surface may be parametrized by two degrees of freedom x_1, x_2

$$(x(x^1,x^2), y(x^1,x^2), z(x^1,x^2))$$

To measure the infinitesimal distance between neighboring points on surfaces, the distance must be something of the form:

$$ds^2 = dx^2 + dy^2 + dz^2 = \sum_{ij} g_{ij} dx^i dx^j$$

where we used the chain rule $dx = \sum_{i=1}^{2} \frac{\partial x}{\partial x_i} dx_i$ and defined the metric coefficients $g_{ij}(x)$ which are symmetric and depends on the coordinates we are using to describe the surface. For example, on 'flat' space the line element is

$$ds^2 = dx^2 + dy^2 + dz^2$$

and on the sphere

$$ds^2 = R^2 d\theta^2 + R^2 \sin\theta^2 d\phi^2$$

It's convenient to write $ds^2 = g_{ij} dx^i dx^j$ suppressing the sum while keeping in mind that every pair of repeated indices must be summed. Note that ds^2 is a distance so it does not depend on the choice of coordinates. In some sense, the information of the curvature on the surface must be encoded in the metric coefficients: the way we measure distances must give information about the geometry.

Now we seek for the equation of a geodesic in any surface, so we need to find the curve that connect two points [2] which makes the distance minimal. This is equivalent to finding the path a free particle will take on the surface, so the lagrangian of this problem is:

$$L = \frac{1}{2} g_{ij} \dot{x}^i \dot{x}^j = \frac{1}{2} v^2$$

where we consider a particle with $m = 1$ without loss of generality. Therefore

$$\frac{\partial L}{\partial x^l} = \frac{1}{2}\frac{\partial g_{ij}}{\partial x^l}\dot{x}^i\dot{x}^j = \frac{1}{2}\frac{d}{dt}\left(g_{ij}\frac{\partial \dot{x}^i}{\partial \dot{x}^l}\dot{x}^j + g_{ij}\frac{\partial \dot{x}^j}{\partial \dot{x}^l}\dot{x}^i\right) = \frac{1}{2}\frac{d}{dt}\left(g_{ij}\delta_{il}\dot{x}^j + g_{ij}\delta_{jl}\dot{x}^i\right) = \frac{d}{dt}\left(g_{il}\dot{x}^i\right)$$

where we use the fact that g_{ij} is symmetric and does not depend on the derivative of the coordinates. Hence

$$\frac{d}{dt}\left(g_{il}\dot{x}^i\right) = \frac{\partial g_{il}}{\partial x^j}\dot{x}^j\dot{x}^i + g_{il}\ddot{x}^i = \frac{1}{2}\left(\frac{\partial g_{il}}{\partial x^j}\dot{x}^j\dot{x}^i + \frac{\partial g_{jl}}{\partial x^i}\dot{x}^i\dot{x}^j\right) + g_{il}\ddot{x}^i$$

[1] A great circle is the circle that lies in the intersection of a sphere with a plane that passes through the center of the sphere

[2] Actually there are some issues in considering *any* two points because the extremal of the functional may not be a minimum, so in general we take two points in a neighborhood close enough to give a unique curve with minimal length.

Provided that the metric is non-degenerate and its inverse is given by g^{ij}, we can set

$$\Gamma^i_{jl} = \frac{1}{2} g^{il} \left(\frac{\partial g_{il}}{\partial x^j} + \frac{\partial g_{jl}}{\partial x^i} - \frac{\partial g_{ij}}{\partial x^l} \right)$$

the geodesic equations are given by

$$\ddot{x}^i + \Gamma^i_{jl} \dot{x}^j \dot{x}^l = 0 \tag{1}$$

Such quantities Γ^i_{jl} are called the Christoffel connection. For a physicist, equation 1 is just the acceleration of a particle in funny coordinates, or the acceleration of a particle on a surface.[3] We can make a slight change in equation 1

$$\dot{x}^l \left(\frac{\partial \dot{x}^i}{\partial x^l} + \Gamma^i_{jl} \dot{x}^j \right) = \dot{x}^l \nabla_l x^i = 0$$

where $\nabla_l = \frac{\partial}{\partial x^l} + \Gamma^i_{jl}$ is called the covariant derivative. Now this looks more like the equation of a 'straight line', but we had to change our notion derivative $\partial_k \to \nabla_k$ when geometry comes at play. In section 2 we will see that gravity can be completely incorporated by the change $\partial_\mu \to \nabla_\mu$ once we let spacetime be an arbitrary manifold of lorentzian metric. If you ever tried to do the quantum mechanics of a particle in a magnetic field, you may remember that you had to change the derivative $\partial_i \to \partial_k - iqA_k$ in Schrodinger's equations to accommodate the magnetic field. We just had our first hint of a gauge theory: turns out that electromagnetism is something like the geometry of a principal $U(1)$-bundle over spacetime. Partial derivatives commute $\partial_i \partial_j = \partial_j \partial_i$ while covariant derivatives don't

$$(\nabla_i \nabla_j - \nabla_j \nabla_i) x^k = R^k_{ijl} x^l$$

where you are invited to show that the Riemann tensor is

$$R^k_{ijl} = \partial_j \Gamma^k_{il} + \Gamma^{jr}_k \Gamma^{il}_r - (j \leftrightarrow l)$$

The failure of covariant derivatives to commute is something you can only find in curved geometries: parallel transport a vector in a loop and see what happens both on a piece of paper and a sphere. It turns out that the Riemann tensor contains all the information about the geometry of a surface. A physicist would be proud. If you want to skip ahead to the physics part in section 2, feel free to do it. But you are going to miss out some important concepts that is going to be crucial latter in order to really understand spacetime, black holes and singularities.

1 What is curvature?

In Book XI of the Confessions (397) Saint Augustine was trying to understand time. There he said something that struck me for a while:

[3]For a mathematician Γ^i_{jl} is the pull back of the Yang-Mills connection of a frame bundle. I guess physicists are winning this one.

"What then is time? If no one asks me, I know; if I want to explain it to a questioner, I do not know." (Augustine, p. 242)

I believe the same is true for many concepts, but it is specially true for curvature: you know exactly what it is, except when you have to compute it. Instead of looking at intuition as being a bug in our brains, I like to think it's actually a feature. It's time to roll up our sleeves and try to build curvature and geometry from our intuition. We will thank Augustine for this insight latter.

We start with one dimensional geometries, i.e. curves, and quickly exhaust everything there is to know about them using frame fields. It turns out that this method is extremely powerful and, provided some suitable modifications, it will carry over to any number of dimensions. In our path we will naturally encounter concepts such as manifolds, tangent spaces, metrics, connections and covariant derivatives. Hopefully this will make you see the whole picture and don't be bothered asking yourself questions like: why in the world is curvature a (1,3) tensor?

1.1 Curves and the power of Frenet Frames

We start with the simplest possible geometry: a parametrized curve $\gamma(t) : I \to \mathbb{R}^3$ in three dimensions. In figures 1 and 2, we can clearly see that 1 has no curvature while 2 has some curved sections. Curvature is a local property. Another basic fact is that a circle with a small radius is more curved than a circle with a large radius. This is easy to see: if you zoom in on any circle, it starts to look like a straight line. Therefore the curvature k of a circle should be inversely proportional to its radius R:

$$k \sim \frac{1}{R} \tag{2}$$

Now consider the helix in figure. It is intuitive that the helix not only have curvature but also twist as it moves, a feature that is not shared with the circle. Notice that all of these features are defined by how the curve changes it's directions through space. Let's formalize all of these statements.

Since we are only interested in the directions, we may only consider paths with unit speed $\|\gamma'\| = 1$ without loss of generality.

Definition 1. *A regular curve is a differentiable map $\gamma : I \to \mathbb{R}^3$ such that $\|\gamma'(s)\| = 1$*

There is a standard procedure to take any path $\gamma(t)$ without cusps and make it a regular curve. First calculate the integral

$$s = \int_0^t \|\gamma'\| \, dt$$

and then take the inverse of the function $s = s(t)$ so that $t = t(s)$ and $\gamma(s) = \gamma(t(s))$ where $\|\gamma'(s)\| = 1$. This is sometimes called arc length parametrization.[4] Making the rather confusing relabel $T \doteq \gamma'$ we are ready to define curvature:

[4] Forget about the parameters t, s as being time, it's just a real variable.

Definition 2. *The curvature $k(s)$ of a regular curve γ is given by:*

$$T' = kN \tag{3}$$

where $N(s)$ is a unit vector field $\|N\| = 1$.

At first this may look abstract but it's precisely what we wanted: k measures the amount of change of direction of the unit tangent T. We can use the inner product $\langle T, T\rangle = 1$ to show that the vector fields are orthogonal:

$$\langle T, T\rangle' = 2k\langle T, N\rangle = 0$$

Now define another vector field $B \doteq T \times N$ which is again unitary and orthogonal to both T and N. You should picture these vector fields T, N, B as moving frames attached to the curve.

Definition 3. *Given a regular curve γ its Frenet frame is the set of vector fields $\{T, N, B\}$.*

You can use the inner product to show (exercise) that $B' = -\tau N$ where $\tau(s) \in \mathbb{R}$ is called the torsion. The most important thing about the Frenet frames is that their derivatives T', N', B' are expressed in terms of themselves T, N, B. In this way we can keep track of all the changes in all directions of the curve.

Theorem 1.1. *If γ is a regular curve and T, N, B are its Frenet frame fields, then*

$$\begin{pmatrix} T' \\ N' \\ B' \end{pmatrix} = \begin{pmatrix} 0 & k & 0 \\ -k & 0 & \tau \\ 0 & -\tau & 0 \end{pmatrix} \begin{pmatrix} T \\ N \\ B \end{pmatrix} \tag{4}$$

Proof. The first and last equations of 4 are just definitions. Since the Frenet frame fields are orthonomal:

$$N' = \langle N', T\rangle T + \langle N', N\rangle N + \langle N', B\rangle B$$

From the identity $\langle N, T\rangle = 0$ it follows that $\langle N', T\rangle = -\langle N, T'\rangle = -k$ and by similar arguments $\langle N', B\rangle = \tau$ □

At this point we should check if this definition agrees with our expectations. A straight line $\gamma(s) = a + bs$ where $\|b\| = 1$ has curvature $k = \|b'\| = 0$. A circle can be parametrized by $\gamma(t) = (R\cos t, R\sin t, 0)$ but its arc length parametrization is different:

$$s = \int_0^t \|\gamma'(t)\|\, dt = Rt$$

so that $\gamma(s) = (R\cos\frac{s}{R}, R\sin\frac{s}{R}, 0)$. The curvature is then

$$k = \|T'\| = \|\gamma''(s)\| = \left\| -\frac{1}{R}(\cos\frac{s}{R}, \sin\frac{s}{R}, 0)\right\| = \frac{1}{R}$$

and the torsion is $\tau = 0$ since $B' = 0$. This is exactly what we expected from equation 2 and also from the lack of twisting of a circle. An arc lenght parametrization of the helix is (exercise):

$$\gamma = (a\cos\frac{s}{c}, a\sin\frac{s}{c}, \frac{b}{c}s)$$

where $c^2 = a^2 + b^2$. Then $k = \frac{a}{c^2}$ and $\tau = \frac{b}{c^2}$. Notice that when $b = 0$ the torsion goes to zero and the helix colapse into a circle.

These are some nice results, but there is a strong statement about curvature and torsion when you integrate the Frenet formulas 4:

Theorem 1.2. *Given two scalar fields $k, \tau : I \to \mathbb{R}$ there exists one regular curve (up to isometries) $\gamma : I \to \mathbb{R}^3$ such that its curvature and torsion are given by k, τ.*

In other words, k and τ completely determines the geometry of a curve. Well, that was kinda easy. The take away should be that if we want to understand curvature, we better have a way to construct frame fields attached to every point of our geometric object and express their derivatives in terms of themselves like in 4.

The thing is that there is nothing special about curves. We could choose a frame field that is defined on every point of $\mathbb{R}^3$. We would expect them to describe the curvature and torsion of the whole space $\mathbb{R}^3$, so it's interesting to generalize the formalism. But before we jump into that, we need to define extra objects that will greatly help us. These are covariant derivatives and forms. Let's start with some terminology:

Definition 4. *For each point $p \in \mathbb{R}^3$ the set of all vectors with base at p is denoted by $T_p(\mathbb{R}^3)$ and is called the tangent space at p. The set of all tangent spaces is called the tangent bundle $T\mathbb{R}^3$.*

Definition 5. *A smooth vector field is a smooth map $V : \mathbb{R}^3 \ni p \mapsto V(p) \in T_p(\mathbb{R}^3)$.*

Since we want to see how vector fields change in any direction, it's useful to define derivatives of these objects.

Definition 6. *Let W and V be smooth vector fields. The covariant derivative of W in the direction of V is another vector field $\nabla_V W$ such that:*

$$\nabla_V W(p) = W(p + V(p)t)'|_{t=0}$$

The definition of a covariant derivative looks like something you have never seen before, but you actually did. Take a point p, a vector $V(p)$ pointing somewhere with base on p and compute the directional derivative $W(p+V(p)t)'|_{t=0}$ of W in the direction $p+V(p)t$. For example, let $V = -y\hat{x} + x\hat{y}$ and $W = \cos(x)\hat{x} + \sin(y)\hat{y}$ be vector fields and $p = x\hat{x} + y\hat{y} + z\hat{z}$ be any point in $\mathbb{R}^3$. Then

$$p + V(p)t = (x - yt)\hat{x} + (y + xt)\hat{y} + z\hat{z}$$

$$\nabla_V W(p) = W(p + V(p)t)'|_{t=0} = y\sin(x)\hat{x} + x\cos(y)\hat{y}$$

Remember when your calculus teacher said that the differential

$$df = \frac{\partial f}{\partial x}dx + \frac{\partial f}{\partial y}dy + \frac{\partial f}{\partial z}dz$$

is not well defined? They lied to you (for a good reason). We can define them using forms, which turns out to be *extremely* useful in physics and mathematics.[5]

Definition 7. *A one-form is a linear map $\omega_p : T_p(\mathbb{R}^3) \to \mathbb{R}$. A one-form field is a linear map $\omega : T\mathbb{R}^3 \to \mathbb{R}$.*

Definition 8. *Let $f : \mathbb{R}^3 \to \mathbb{R}$ be a differentiable function. The differential df is a one-form such that:*

$$df(v) = v(f) = f(p + vt)'|_{t=0}$$

where $v \in T_p(\mathbb{R}^3)$.

Lemma 1.3. *The differentials dx, dy, dz form a basis[6] of the vector space $T_p^*\mathbb{R}^3$ of all one forms at a point $p \in \mathbb{R}^3$*

Proof. Notice that $dx(v) = v(x) = v_x$ and the same for the other differentials. For any one form $\omega_p \in T_p^*\mathbb{R}^3$ and vector $v = v_x\hat{x} + v_y\hat{y} + v_z\hat{z} \in T_p(\mathbb{R}^3)$ we have that:

$$\omega_p(v) = \omega_p(\sum_i v_i\hat{x}_i) = \sum_i v_i\omega_p(\hat{x}_i) = \left(\sum_i \omega_p(\hat{x}_i)dx_i\right)(v)$$

□

Corollary 1.3.1. *The differential can be written as $df = \frac{\partial f}{\partial x}dx + \frac{\partial f}{\partial y}dy + \frac{\partial f}{\partial z}dz$*

Proof. Trivial. □

Given two one-forms ω, ϕ we can make their product $\omega \wedge \phi$ which is a two-form. The simplest way to think about them is to consider their action on the differentials and the distributive property

Definition 9. *The wedge product $\wedge$ is a bilinear map that takes two one-forms and gives a two-form such that*

$$dx_i \wedge dx_j = -dx_j \wedge dx_i$$

Given any two one-forms $\omega = \sum_i \omega_i dx_i$ and $\phi = \sum_j \phi_j dx_j$ their product is

$$\omega \wedge \phi =$$

Needs a lot of attention. Mention the electromagnetic field as a one-form, and faraday tensor as a two-form $F = dA$.

Definition 10. *A frame field is a set of vector fields E_1, E_2, E_3 in $\mathbb{R}^3$ such that:*

$$\langle E_i, E_j \rangle = \delta_{ij}$$

where δ_{ij} is the Kronecker delta.

---

[5] That is an understatement.

[6] Recall that the dual V^* of a finite dimensional vector space V is again another vector space of the same dimension.

We know that something funny happens when we express derivatives of these frames in terms of themselves. Let V be any vector field and

$$\begin{aligned}
\nabla_V E_1(p) &= \omega_{11}E_1(p) + \omega_{12}E_2(p) + \omega_{13}E_3(p)\\
\nabla_V E_2(p) &= \omega_{21}E_1(p) + \omega_{22}E_2(p) + \omega_{23}E_3(p)\\
\nabla_V E_3(p) &= \omega_{31}E_1(p) + \omega_{32}E_2(p) + \omega_{33}E_3(p)
\end{aligned}$$

In a shorter notation using Einstein's convention of repeated indices $\nabla_V E_i(p) = \omega_{ij}E_j(p)$, the coefficients ω_{ij} clearly depends on the vector V and looks like a one-form field:

Lemma 1.4. *Let E_i be frames fields in $\mathbb{R}^3$, $v \in T_p(\mathbb{R}^3)$ any vector and*

$$\omega_{ij}(v) = \langle \nabla_v E_i(p), E_j(p) \rangle$$

Then ω_{ij} is a matrix valued one-form and $\omega_{ij} = -\omega_{ji}$. They are called connection coefficients.

Proof. You can use definition 6 and the chain rule to prove that $\nabla_{av+bu}E = a\nabla_v E + b\nabla_u E$ for any vectors $u, v \in T_p(\mathbb{R}^3)$ and numbers $a, b \in \mathbb{R}$. Therefore

$$\omega_{ij}(av + bu) = \langle \nabla_{av+bu}E_i(p), E_j \rangle = a\omega_{ij}(v) + b\omega_{ij}(u)$$

To prove the antisymmetry of ω notice that

$$\nabla_v \langle E_i, E_j \rangle = \langle \nabla_v E_i, E_j \rangle + \langle E_i, \nabla_v E_j \rangle = 0$$

□

Now we can use the dual 1-forms θ_i of the frame fields defined as:

$$\theta_i(v) = \langle v, E_i \rangle$$

to show a very interesting result that will clarify the geometry of $\mathbb{R}^3$

Theorem 1.5 (Cartan Structural Equations)**.** *Let E_i be frame fields, θ_i its duals and ω the connection coefficients. Then*

$$d\theta_i = \omega_{ij} \wedge \theta_j \tag{5}$$

$$d\omega_{ij} = \omega_{ik} \wedge \omega_{kj} \tag{6}$$

Proof. Note that any frame can be written in terms of the canonical vector field like $E_i = a_{ij}\hat{x}_j$, and so too the dual forms $\theta_i = a_{ij}dx_j$. On a more compact notation $\theta = Ad\xi$ where:

$$\begin{pmatrix}\theta_1\\ \theta_2\\ \theta_3\end{pmatrix} = \begin{pmatrix}a_{11} & a_{12} & a_{13}\\ a_{21} & a_{22} & a_{23}\\ a_{31} & a_{32} & a_{33}\end{pmatrix} \begin{pmatrix}dx_1\\ dx_2\\ dx_3\end{pmatrix} \tag{7}$$

and A is an orthogonal matrix. We have that

$$\omega_{ij}(v) = \langle \nabla_v E_i(p), E_j(p) \rangle = da_{ik}a_{jk}$$

or in matrix notation $\omega = dAA^T$. The Cartan equations then follows

$$d\theta = dAd\xi + Ad^2\xi = dAd\xi = dAA^T Ad\xi = \omega\theta$$

$$d\omega = d(dAA^T) = -dAdA^T = -dAA^T AdA^T = -\omega\omega^T = \omega\omega$$

□

We will see later that the second Cartan equation 6 means that $\mathbb{R}^3$ is flat. All the geometric properties are found in the connection coefficients! The next logical step in our investigation is to go up a dimension.

1.2 Surfaces and Gauss legacy

What exactly is a 2 dimensional surface? This is an important question that we will have to address in detail very soon. There we will encounter the notion of a manifold for the first time. But for now we are content with a simplified version of the answer. If you think about it, a 2-d surface is just a subset $M \subset \mathbb{R}^3$ where you can ascribe 2 coordinates. For a physicist its obvious that a surface has two degrees of freedom. In mathematical lingo we say that for each point $p \in M$ there exists an open set $V \subset M$ that contains p and a homeomorphism $x : U \subset \mathbb{R}^2 \to V \subset M$ [7] We call x a coordinate system, or chart, of the patch V. Have you noticed something odd? The space $\mathbb{R}^3$ didn't appear in the definition of a surface! We could choose S to be any set, not only a subset of $\mathbb{R}^3$. Of course we would have to establish a topology on S and so on and so fourth, but it's possible.

In figure we have an example of a surface.

Definition 11. *The set of all tangent vector to a point $p \in M$ on the surface is called the tangent space $T_p(M)$*

It's clear that $T_p(M)$ is really a 2 dimensional plane. This gives us a very natural way to attach a frame field to the surface: choose E_1, E_2 to be on the tangent plane and E_3 to be the unit normal vector field at each point of M. Notice that we can only hope for this construction to be true locally (hairy ball theorem). The structural equations are

Lemma 1.6. *Let M be a surface and E_i its attached frame fields. The structure equations are*

$$\theta_1 =$$

1.3 Higher dimensions, Riemann curvature and the modern stuff

Maybe not talk about the shape operator at all.. Now we only have to use Cartan structural equations and we are done! But before we do that, it is going to be useful to define the shape operator.

[7] See the appendix for topology.

Definition 12. *Let U be a unit normal to the surface M and $v \in T_p(M)$. The shape operator $S : T_p(M) \to T_p(M)$ is a linear operator defined as*

$$S_p(v) = -\nabla_v U$$

Its easy to see that the shape operator is linear by the linearity of the covariant derivative, and since $\nabla_v \langle U, U\rangle = -2\langle S_p(v), U\rangle = 0$ then S_p belongs to the tangent space.

Even though we had much success you can already tell something fishy is going on. It's odd that we are studying a two (one) dimensional geometry by embedding it in a 3d space, and this 3d space has a lot of structure which we take as given: a vector space (really a manifold) equipped with a very specific inner product $\langle ., .\rangle_E$ where E stands for Euclidean. In fact, $\langle ., .\rangle$ is the very definition for something to have a geometry! But we will get there latter.

1.4 Repère Mobile, Frame bundle and principal fiber bundles

---

2 Gravity enters the scene...

Now that we've delved deeply into the intricacies of geometry and curvature, we are in a position to confront the second big question: How exactly does gravity manifest as the curvature of spacetime? How did Einstein come up with this idea in the first place?

lots to be revised. change the order and add spacetime diagrams

2.1 It's simple once you know geometry

The idea that gravity is not a force is needed for Newtonian physics to be consistent! Consider what Newton's first law have to say:

(Law of Inertia): A body follows uniform straight motion unless acted on by a force.

Now imagine a universe with only a single particle. How can this particle tell that it is moving at all? Well, it can't tell. We need at least two particles, one the observer and other the observed. The observer will check with its coordinates and clocks if Newton's law holds for the other particle. But wait a second: both particles have mass, so there is a gravitational force that deviates the uniform straight motion. And if one includes more particles it only gets worse! It is clear that the force of gravity and Newton's first law cannot be both true since this leads to a contradiction. How can we resolve this problem?

It turns out that gravity is not a force, so it does not deviate particles from straight motion in spacetime. But for this to be true we must loose a bit our notion of *straight*, which is equivalent to consider a curved geometry. That is no problem for us! The straightest possible paths are geodesics in a more general geometry. If we can show that the effect of gravity is the same as a geodesic path in some sort of curved space, there are no more contradictions with Newton's first law since gravity is no longer a force.

Let us look at the simplest physical system where gravity is present: a free falling body in a gravitational potential Φ. The equation of motion is:

$$m\ddot{x} = -m\nabla\Phi$$

Notice that the mass appears on both sides. This is the difference of gravity to other forces like the electromagnetic one: all particles follows the same path regardless of the masses. A positively charged particle will follow a trajectory that is very different than a negative charged one when put on the same electromagnetic field. Rewriting the equation of motion

$$\ddot{x}^i + (\nabla\Phi)^i = 0 \tag{8}$$

We want this to look like a geodesic equation, but first derivatives are lacking on the second term. We are stuck: there is only three dimensions of space and none of them can be put together to build a geodesic out of equation (8). Turns out it's 21 century and there is another dimension we can play with: time. If we include the extra coordinate $x^0 = t$ of time on our recipe, it obviously obeys $\ddot{x}^0 = 0$ and we get a system of equations:

$$\begin{cases} \ddot{x}^0 = 0 \\ \ddot{x}^i + (\nabla\Phi)^i\dot{x}^0\dot{x}^0 = 0 \end{cases}$$

That is exactly a geodesic equation on *spacetime*, not just space! Notice that the law of inertia mentions space *and* time: a uniform straight motion. These are necessarily straight lines on spacetime diagrams, not just straight lines on space[8]. So it should not come as a surprise that we had to include time in the last step of the recipe.

With the geodesic equation we can obtain the connection $\Gamma^i_{00} = (\nabla\Phi)^i$, and $\Gamma i_{jk} = 0$ otherwise. With a connection at hand we find the Riemann curvature tensor $R^i_{0j0} = -\partial_j\partial^i\Phi$. Therefore we conclude that gravity is truly an effect of the curvature of spacetime.

A refinement of Newton's first law is:

(Enhanced Law of Inertia): In the absence of forces all particles follows geodesics in spacetime.

I should insist that in the new definition gravity is not a force. Its effect is encoded in the geodesic motion.

2.2 Einstein's happy thought

2.3 Generalize Special relativity

In the section above we gave some physical arguments to conclude that gravity must be the curvature of some geometry. But it may seem unnatural to mix space and time

[8]Think about some particle moving on a straight line with constant acceleration.

together, specially because we were dealing with a galilean structure of flat space and absolute time. A more natural arena is, of course, Minkowski spacetime of special relativity where the geometry is explicitly given to us by a metric $\eta_{\mu\nu}$. Our goal is to introduce gravity in special relativity. This will inevitably lead us to general relativity.

We know that in special relativity it is best to view space and time together, i.e. spacetime, as a set of points $X^\mu = (ct, x, y, z)$ that resembles the vector space $\mathbb{R}^4$ but has a very different notion of distance between points:

$$ds^2 = -c^2 dt^2 + dx^2 + dy^2 + dz^2 = \eta_{\mu\nu} dX^\mu dX^\nu \tag{9}$$

where the metric is defined as $\eta_{\mu\nu} = diag(-1, 1, 1, 1)$. When we look for coordinate transformations [9] $X'^\mu = \Lambda^\mu_\nu X^\nu$ that preserves (9) we get to the Lorentz transformations (given that it is proper orthocronus etc etc). They form a group $\Lambda^\mu_\nu \in SO(1, 3)$ where composition is given by the usual matrix multiplication.

The point is that all invariant quantities in SR, i.e. quantities that does not depend on a particular observer, should be made with the aid of the invariant metric (9). For example, the proper time:

$$d\tau = \frac{\sqrt{-ds^2}}{c}$$

is the time measured by a particle's clock in its rest frame $X'^\mu = (ct', 0, 0, 0)$. Different observers will experience different times t but they all agree on the value of τ. With these tools we can easily reconstruct the dynamics of particles in a relativistic invariant way. We define the 4-velocity and 4-momentum as:

$$U^\mu = \frac{dX^\mu}{d\tau}$$
$$P^\mu = mU^\mu$$

Let X^μ be a 4-vector in spacetime. We define $X_\mu X^\mu = X^2 \doteq \eta_{\mu\nu} X^\mu X^\nu$ as the contraction of X with itself. Then you can show that $U^2 = -c^2$ and $P^2 = -(mc)^2$. Newton's second law should be something like:

$$\frac{dP^\mu}{d\tau} = F^\mu$$

That is a nice way of getting to the dynamics of the theory, but it is not clear at all how to include some force or potential $V(x)$ in the equation above. We have look at the Lagrangian formulation of SR: What is the action of a free relativistic particle? Since Lorentz transformations are a symmetry of the system, we should look for a relativistic invariant action. The simplest thing to come up with is the following:

$$S = -m \int d\tau \tag{10}$$

[9] The set of coordinate transformations $X'^\mu = \Lambda^\mu_\nu X^\nu + a^\mu$ that preserves the metric (9) is called the Poincaré group $\mathcal{P} \simeq O(1, 3) \oplus R^{1,3}$. The reader should notice that this is the group of isometries of spacetime. The Lorentz transformations are the subgroup $SO(1, 3)$ of the Poincaré group connected to the identity.

From here on I'll use natural units where $c = 1$. To understand equation (10) first we notice that a particle describes a path in spacetime $X^\mu(\sigma) : \mathbf{R} \to \mathbf{R}^{1,3}$ parametrized by σ. The parameter does not have to be the proper time, we can always reparametrize the path just like we did when analysing curves in the Geometry part of the notes. Thus we can rewrite equation (10):

$$S = -m \int d\sigma \frac{d\tau}{d\sigma} = -m \int d\sigma \sqrt{-\eta_{\mu\nu} \frac{dX^\mu}{d\sigma} \frac{dX^\nu}{d\sigma}} \tag{11}$$

The action is fully relativistic, and more than that: it has reparametrization invariance. This is a form of gauge invariance: changing the parameter should not affect the underlying physics, i.e. the worldline of the particle. The Euler-Lagrange equations gives us (exercise):

$$\frac{d}{d\sigma} \left(m \frac{\partial \sqrt{-\eta_{\mu\nu} \dot{X}^\mu \dot{X}^\nu}}{\partial \dot{X}^\rho} \right) = m \frac{d^2 X_\rho}{d\tau^2} = 0$$

Which is exactly what we were expecting from our previous equation $\frac{dP^\mu}{d\tau} = F^\mu = 0$. In the Lagrangian formalism it is straightforward to insert a potential in the theory:

$$S = -m \int d\sigma \sqrt{-\eta_{\mu\nu} \frac{dX^\mu}{d\sigma} \frac{dX^\nu}{d\sigma}} - \int d\sigma \Phi$$

But this does not keep the reparametrization invariance! Any change in the parameter $\sigma \to \sigma'$ will be felt by the jacobian $\frac{\partial \sigma}{\partial \sigma'}$ in the second term. We can get around this by considering a four-potential A_μ instead of the scalar Φ, and contract it with the four-velocity:

$$S = -m \int d\sigma \sqrt{-\eta_{\mu\nu} \frac{dX^\mu}{d\sigma} \frac{dX^\nu}{d\sigma}} - \int d\sigma q A_\mu \dot{X}^\mu$$

where q is just a constant measuring the coupling with the potential. Notice that a reparametrization don't change the action because of the 4-velocity term! And everything keeps lorentz invariance as well. The suggestive notation is going to make more sense once we derive the Euler Lagrange equations for the system:

$$m \frac{d^2 X_\mu}{d\tau^2} = q \left(\frac{\partial A_\mu}{\partial X^\nu} - \frac{\partial A_\nu}{\partial X^\mu} \right) \dot{X}^\nu = q F_{\mu\nu} \dot{X}^\nu$$

This looks exactly like the (covariant) equation of a particle in an electromagnetic field, where $F_{\mu\nu} = \partial_\nu A_\mu - \partial_\mu A_\nu$ is the electromagnetic tensor! You can easily show that it satisfies $\partial_{[\rho} F_{\mu\nu]} = 0$ which is equivalent to the two homogeneous Maxwell's equations. Of course, it will only be Maxwell's EM when the other two inhomogeneous equations are provided $\partial^\mu F_{\mu\nu} = J_\nu$ where $J_\nu = (\rho, \mathbf{J})$ is the 4-current.

Does that mean that relativistic gravity is somehow a kind of gravitomagnetism? You could go in this route but eventually you would find serious inconsistencies in the solutions of your equations. Stuff like infinite energy, instability of simple closed orbits and worse.(reference here)

What now? We could go on with the previous idea and insert a relativistic tensor field of rank 2 $h_{\mu\nu}$ instead of the 4-potential A_μ and see where that leads us. We would actually get linearized general relativity in vacuum! But the procedure is much more involved and subtle. Since this is a field theory we would like to have positive kinetic terms (free of ghosts, if these words even make sense to you) and kill some degrees of freedom. This have to do with the latter quantization of the theory. In the appendix we show how you can achieve this, but for know we go through a much more simple route.

Instead of subtracting a potential, we change the metric:

$$S = -m \int d\sigma \sqrt{-g_{\mu\nu} \frac{dX^\mu}{d\sigma} \frac{dX^\nu}{d\sigma}} \tag{12}$$

where now $g_{\mu\nu} = \eta_{\mu\nu}$ except at the 00 component $g_{00} = \eta_{00} - 2\Phi$. This choice of lagrangian is inspired by what we did in the section above since this leads to the same connection $\Gamma^i_{00} = (\nabla\Phi)^i$ and we already have a clue of the geometric nature of gravity.

Exercise: Show that in the non relativistic limit $v << c$ the Lagrangian in (12) simplifies to $L \approx mv^2/2 - m\Phi$

The Lagrangian (12) not only reproduces all of the effects of gravity, but also introduces some new phenomena. Take a photon with a certain frequency and shoot it from the bottom of a high building. At the top, someone measures the same photon and realizes the frequency has changed! The gravitational redshift is predicted by the new metric $g_{\mu\nu}$. The photon 4-momentum is:

$$k^\mu = (E, 0, 0, E)$$

and the 4-velocities of the observers at the bottom and top of the building are $U^\mu_{\text{bottom}} = U^\mu_{\text{top}} = (1, 0, 0, 0)$. The ratio of the measured frequencies is:

$$\frac{\omega_2}{\omega_1} = \frac{k^\mu U_\mu(\text{top})}{k^\mu U_\mu(\text{bottom})} = \frac{g_{00} k^0 U^0(\text{top})}{g_{00} k^0 U^0(\text{bottom})} = \frac{1 - \frac{2GM}{r_2}}{1 - \frac{2GM}{r_1}}$$

Where it's clear that the frequency of the top ω_2 gets redshifted since $r_2 > r_1$. At this point there is nothing stopping us from thinking that the metric $g_{\mu\nu}$ may be anything, or at least anything that is consistent with the matter content of the physical system. Each $g_{\mu\nu}$ is interpreted as the geometry of spacetime or of just a portion of spacetime. The euler-lagrange equations of 12 gives us geodesics just like we did in the Geometry part. We thus arrive at the same conclusions drawn in the section above: gravity is the geometry of spacetime, and free particles move on geodesics. Furthermore, we have the necessary tools of differential geometry to make precise statements of these ideas in a much more general setting.

Definition: Spacetime is a four-dimensional smooth manifold equipped with a Lorentzian signature metric.

The four dimensions should be obvious. A smooth manifold structure is the least we can impose for something to look like spacetime without having any kind of weird topological phenomena.[10]. It also assures that everything is made without any reference to coordinate systems. A Lorentzian signature is necessary so that we maintain lorentz invariance locally (at a point), just like we showed that a Riemannian metric is trivial $g_{\mu\nu}(p) = \delta_{\mu\nu}$ at a point. It assures that when gravity (curvature) is not present, we go back to our beloved flat Minkowski spacetime full of twins and paradoxes.

How can we possible know the metric from the matter content in this general setting? Just like in electromagnetism where Maxwell's equations relate the electromagnetic field and the charge distribution, we need field equations relating $g_{\mu\nu}$ to the stress-energy-momentum tensor $T_{\mu\nu}$. You see, the energy momentum tensor is just a way of encoding the matter content in a tensorial, relativistic fashion. I will just say what it is: it's a matrix where the $T_{\mu\nu}$ component is the flux of P^μ momentum across the hypersurface perpendicular to the x^ν direction. For example, the 00 component is just the energy density where the 11 component is the pressure along the x direction

$$
\begin{aligned}
T_{00} &= \frac{P^0}{\Delta X \Delta Y \Delta Z} = \rho \\
T_{11} &= \frac{P^1}{\Delta T \Delta Y \Delta Z} = p_x
\end{aligned}
$$

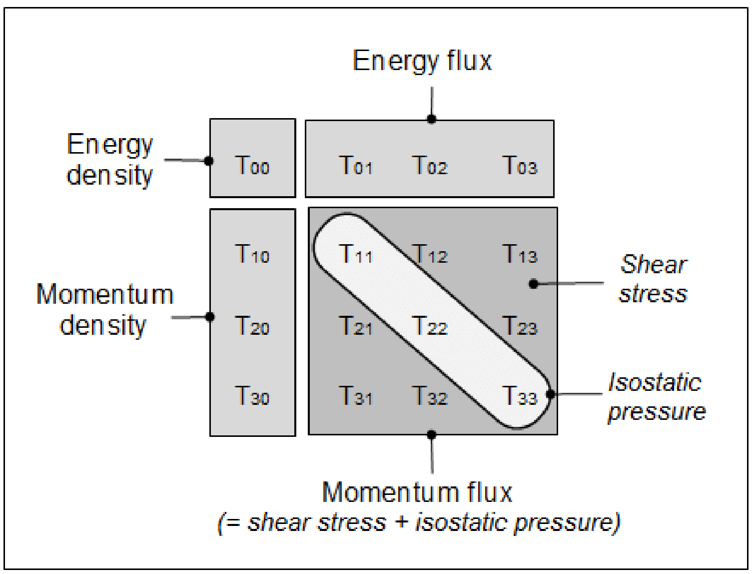

Figure 1: Energy momentum in all of its glory jumpscare.

If you know some field theory you may think that $T_{\mu\nu}$ is the Noether current associated to spacetime translations. Unfortunately this is not the same tensor we use in general relativity, since it's defined only in the context of special relativity. Hence it is not covariant under general coordinate transformations.

[10]By topological I mean continuity, paracompactness etc. The paracompactness of a manifold is a sufficient condition for the existence of a Riemannian metric. For a Lorentzian metric we have to impose one further condition: it has to admit a non-vanishing vector field. For a non-compact manifold that is no problem, but for compact manifolds this is equivalent to have zero Euler characteristic. Compact spacetimes could be 4-torus!

Recall that in classical mechanics Poisson's equation $\nabla^2\Phi = 4\pi G\rho$ determines the gravitation potential of a source. We know that Φ is just part of the 00 component of the metric tensor, so a natural generalization of Poisson's equation is

$$\nabla^2 g_{\mu\nu} = kT_{\mu\nu}$$

But the covariant derivative of the metric is zero by metric compatibility. Furthermore, we want the energy momentum tensor to be conserved $\nabla_\mu T^{\mu\nu} = 0$. On the left hand side we should have a symmetric tensor such that $\nabla_\mu G^{\mu\nu} = 0$ where G is made of second derivatives of the metric. We know just the guy to do it!

$$R_{\mu\nu} - \frac{R}{2} g_{\mu\nu} = kT_{\mu\nu}$$

finish section, black holes

2.4 It doesn't stop here: geometry in gauge theories

References