GAUSS - RIEMANN - EINSTEIN

or

What I Did on My Summer Vacation

Michael P. Windham

February 25, 2008

Preface

For the last forty years most every spring my fancy turned to differential geometry, among other things. I have always wanted to understand it and always struggled. I also wanted to know something about general relativity. Well, I think I have gotten a grip on the basics and that is what these notes are about.

I have organized the presentation around the key figures in the development of the subject, at least in the development of my understanding of the subject. They are Karl Friedrich Gauss, Bernhard Riemann, and Albert Einstein.

In 1827 Gauss produced a paper that fundamentally explained the geometry of surfaces in space. In 1854, Riemann generalized Gauss' ideas to something he called manifolds. In 1912 Einstein used manifolds to describe gravity. These contributions will provide the organization for these notes.

There is a lot to say here and I will not say it all. I hope only to follow a thread that will get me where I want to go with as few digressions as possible. Where I want to go is to general relativity. I will not be very complete in deriving or proving things. My plan is to give you a feel for the subject, not a rigorous development.

But, first a modest beginning with curves in space.

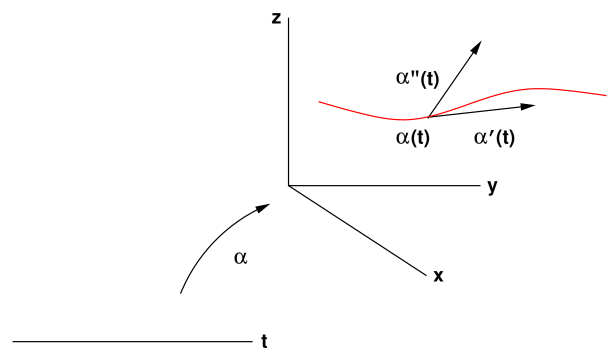
Michael P. Windham

Chapter 1

Curves

This little journey begins with a discussion of curves in space and how to investigate their geometry. The first step is to describe the curve in a quantitative way that allows you to apply the massive machinery of calculus to study it. The appropriate quantitative description is a *parameterization.* The curve is the image of a function $\alpha : R \to R^3$. In what follows I will implicitly assume any functions are sufficiently nice for what I say to be true. Usually this means the functions have as many derivatives as needed and the derivatives are not zero when it would be embarrassing for them to be so.

A parameterization $\alpha(t) = (x(t), y(t), z(t))$, describes the points on the curve as a function of the parameter t. Think of α as drawing the curve. The study of the curve begins with the derivative of α, $\alpha'(t) = \langle x'(t), y'(t), z'(t)\rangle$, where I am using $\langle x, y, z\rangle$ to denote a vector as opposed to (x, y, z) which denotes a point. A point describes a location and a vector, which describes a magnitude and direction, is the natural descriptor for change. In particular, as the curve is drawn, the speed with which it is drawn, how fast the "pen" is moving, and the direction it is being drawn are naturally described by a vector, and these are exactly what $\alpha'(t)$ quantifies. I may appear to be saying that the parameter t is time. That is not necessarily true, but it is seldom harmful to think so.



Perhaps the most basic geometric characteristic of a curve is its length. The derivative vector is the key to calculating length since, according to just about any first year calculus book

$$s(t) = \int_a^t |\alpha'(u)|\, du = \int_a^t \sqrt{\left(\frac{dx}{du}\right)^2 + \left(\frac{dy}{du}\right)^2 + \left(\frac{dz}{du}\right)^2}\, du \tag{1.1}$$

is the distance along the curve from $\alpha(a)$ to $\alpha(t)$. The length of the derivative vector is just the derivative of s with respect to t, so it measures how quickly the curve is being drawn relative to t.

If my goal was to study in detail the geometry of curves, the next step would be to look at the second derivative $\alpha''(t)$, since it describes how the drawing process changes. To do so, would lead us to curvature and torsion and to a complete characterization of curves in terms of them. As fascinating as this is, we will not go there. Let me just point out that if $\alpha''(t) = \mathbf{0}$ for all t, then the curve is a straight line. The converse is true for so called "constant speed curves", namely ds/dt is constant.

If we lived in the 17th century we would have no problem thinking of an infinitesimal piece of length along the curve, the "length of a point". Moreover, it would be natural to think of a point as an infinitesimal box with sides parallel to the axes of length dx, dy, and dz. The length of the point itself would simply be the infinitesimal length of the diagonal of the box, namely $ds = \sqrt{dx^2 + dy^2 + dz^2}$. This heuristic approach is frowned upon these days, but is wonderful to use to think about geometry and even physics, at least in the privacy of your own home.

It is also comfortable and even useful to think of a particle traveling through space and $\alpha(t)$ is its location at time t. The derivative $\alpha'(t)$ would then be the velocity of the particle, and $\alpha''(t)$ its acceleration. This interpretation will become more important when we leap into physics.

Chapter 2

GAUSS - Surfaces

Gauss takes the stage now because of his 1827 paper *Disquisitiones generales circa superficies curvas* (General investigations of curved surfaces). Michael Spivak has called this paper “the single most important work in the history of differential geometry.”

The problem the paper addresses is to study the geometry of surfaces in space. As with curves, a good way to represent them quantitatively is with a parameterization, a function $A : R^2 \to R^3$ whose image is the surface, or if you like $A(x_1, x_2) = (x(x_1, x_2), y(x_1, x_2), z(x_1, x_2))$ draws the surface using two parameters x_1 and x_2.

The geometric properties of the surface are quantified by the motion of its tangent planes and their normal vectors as you move about the surface. At a point $(x, y, z) = A(x_1, x_2)$ the tangent plane is the plane through (x, y, z) parallel to the vectors

$$A_{x_i} = \left\langle \frac{\partial x}{\partial x_i}, \frac{\partial y}{\partial x_i}, \frac{\partial z}{\partial x_i} \right\rangle$$

for $i = 1$ and 2. These two vectors are tangent to the surface because they are the tangent vectors of curves in the surface through (x, y, z) drawn by letting one of the parameters vary while the other is constant. The normal to the surface at the point is the cross product of the two tangent vectors. I will be assuming that the parameterization is sufficiently nice that the two tangent vectors are linearly independent and therefore the normal vector does not vanish. I will not go into details, but, for example, curvature of the surface can be quantified by derivatives of the normal vector, just as curvature of curves is quantified in terms of $\alpha''(t)$.

I do want to discuss the notions of *intrinsic* and *extrinsic* properties of the surface. To put it quaintly, intrinsic properties are those that little two dimensional people that lived "in" the surface could deduce about their world and extrinsic properties are those that you have to be outside the surface in space to see. The normal vector is extrinsic, but amazingly the tangent plane is intrinsic. To see how this is possible, you should think of the domain of the parameterization as a map of the surface. Each point on the map has coordinates (x_1, x_2) that correspond to the point $A(x_1, x_2)$ in the surface. Presumably the little people living in the surface would eventually produce such a map of their world. They would also eventually recognize the importance of drawing vectors on their map to describe directions and develop linear algebra to entertain themselves. Each vector $v = \langle v_1, v_2 \rangle$ drawn with its tail at a point (x_1, x_2) on the map corresponds to a vector in the tangent plane at the point $A(x_1, x_2)$ in the surface, namely $v_1 A_{x_1} + v_2 A_{x_2}$. Therefore, the little people could not see the tangent plane to their world, but they could do the same calculations with vectors on their map that we would with the tangent vectors to their surface in space. They would get the same information from their results, just as we get information about space itself with our vector calculations in it. More on this shortly, but first let us look at something that is easier to believe is intrinsic, length.

A curve in the surface is a curve in space. The length of a curve parameterized by $\alpha : R \to R^3$ whose image is in the surface can be calculated in the usual way by (1.1). However, the little people would see this curve and draw a corresponding curve on their map. Moreover the curve on the map would be parameterized by $a : R \to R^2$ satisfying $\alpha(t) = A(a(t))$. The tangent vectors to the curve on the surface would be related to corresponding tangent vectors $a'(t) = \langle a_1'(t), a_2'(t) \rangle$ to the curve on the map by

$$\alpha'(t) = a_1'(t) A_{x_1}(a(t)) + a_2'(t) A_{x_2}(a(t))$$

using the chain rule. Moreover, letting $\bullet$ denote the usual euclidean inner product, we have[1]

$$|\alpha'|^2 = \sum A_{x_i} \bullet A_{x_j} a_i' a_j'$$

The little people can then calculate the length of their tangent vector by

$$|a'|^2 = \sum g_{ij} a_i' a_j'$$

[1] I will not put any indices on summations. You can assume that any indices that do not appear outside the sum are summed over.

with $g_{ij}(x_1, x_2) = A_{x_i}(x_1, x_2) \bullet A_{x_j}(x_1, x_2)$ and integrate the length of their tangent vector to get the length of the curve. Of course, they are not aware of the A_{x_i}'s but being clever they would no doubt eventually derive the g_{ij}'s. Their linear algebra would have an inner product of vectors given by

$$v \bullet w = \sum g_{ij} v_i w_j$$

In fact, they could have infinitely many inner products, one for each point on their map, since the g_{ij}'s are functions of the location on the map. They could measure length and angles between vectors, but the measurements would depend on where they drew the vectors. This situation may appear disturbing, but we are faced with it all the time when we measure distances between points on a flat map of our world. After all, Greenland appears on a map of the world to be much larger than it really is on the world itself.

The 17th century folk would simply say

$$ds^2 = \sum g_{ij} dx_i dx_j$$

is the (infinitesimal) line element for the surface.

The next question to consider is what curves do the little people think are "straight lines". I already mentioned that for a curve in space, if α'' is zero along the curve then it is a straight line, that is, if the curve is drawn at a constant rate, without turning it is straight. If the curve is drawn in a surface, a tangent vector to the curve is tangent to the surface, so that the little people can perceive its effect. The vector $\alpha''(t)$ may not be tangent to the surface, so that the little people cannot see it - well, not all of it anyway. They can "see" the tangential component of $\alpha''(t)$. In particular, again using the chain rule,

$$\begin{aligned} \alpha'' &= \sum a_i'' A_{x_i} + a_i' a_j' A_{x_i x_j} \\ &= \sum (a_k'' + \sum \Gamma_{ij}^k a_i' a_j') A_{x_k} + \text{ the normal component} \end{aligned}$$

where the Γ_{ij}^k's are obtained from the tangential components of the $A_{x_i x_j}$'s and like the g_{ij}'s are functions of (x_1, x_2). Therefore, the little people would learn that for the curves on their maps

$$a'' = \left\langle a_1'' + \sum \Gamma_{ij}^1 a_i' a_j',\ a_2'' + \sum \Gamma_{ij}^2 a_i' a_j' \right\rangle$$

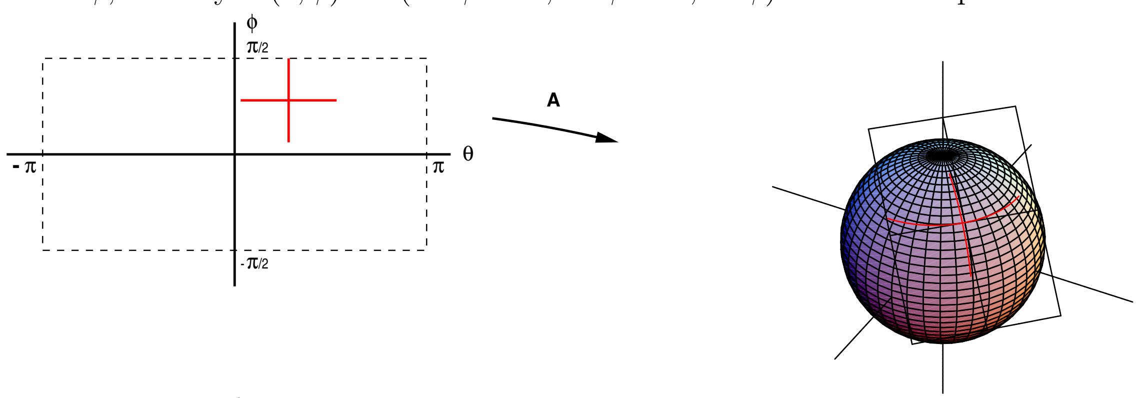
for some functions Γ_{ij}^k they would deduce by experience. They would think a line was "straight", if a'' were zero, that is, the components would be solutions to the differential equations

$$a_k'' + \sum \Gamma_{ij}^k a_i' a_j' = 0 \tag{2.1}$$

Such curves are called *geodesics*, they may curve in space, but not in the surface.[2] I have just been and probably will continue to be a little sloppy by calling curves on the map, $a(t)$, and curves on the surface, $\alpha(t) = A(a(t))$, geodesics, but you will get used to it.

I should point out that the shortest curve between two points is a geodesic. This fact takes a little calculus of variations to verify, so I will say no more. I should also point out that geodesics on the map may not look straight, that is, you may not be able to draw them with a ruler.

How about an example, one we can relate to, the unit sphere. The unit sphere centered at the origin can be parameterized by longitude θ and latitude ϕ, namely $A(\theta, \phi) = (\cos\phi\cos\theta, \cos\phi\sin\theta, \sin\phi)$ draws the sphere.



We have the following

$$ds^2 = \cos^2\phi \; d\theta^2 + d\phi^2$$

and

$$\begin{aligned} \Gamma_{11}^1 &= \Gamma_{22}^1 = \Gamma_{12}^2 = \Gamma_{21}^2 = \Gamma_{22}^2 = 0 \\ \Gamma_{12}^1 &= \Gamma_{21}^1 = -\tan\phi \\ \Gamma_{11}^2 &= \cos\phi\sin\phi \end{aligned}$$

Okay, these coefficients are not easy to compute. You can use the formula that appears below or you can do it the way I did, using *mathematica* (see Gray, 1999).

[2]With a little work and trickery one can show that any curve satisfying these equations is a constant speed curve, that is $\sum g_{ij} a_i' a_j'$ is constant. By rescaling the parameter you can, without loss of generality, assume the constant is one.

The geodesics can be obtained by solving $a'' = \mathbf{0}$ or

$$\begin{aligned} \theta'' - 2\tan\phi\ \theta'\phi' &= 0 \\ \phi'' + \cos\phi\sin\phi\ \theta'^{\,2} &= 0 \end{aligned}$$

to obtain great circles, curves on the sphere obtained by intersecting the sphere with a plane through the origin. Actually, the equations are quite difficult to solve, but at least it is clear that $a(t) = (t, 0)$ (the equator) and $a(t) = (c, t)$ for any constant c (lines of longitude) are solutions. You might find it interesting to show that these are the only geodesics that can be drawn on the map with a ruler.[3] On the other hand, there are many others. For example, $a(t) = \left(\sin^{-1}(\sin t/\sqrt{1+\cos^2 t}), \sin^{-1}(\sin t/\sqrt{2})\right)$ draws a geodesic going through $(0, 0)$ in the direction $\langle 1, 1\rangle/\sqrt{2}$. Check it out! In fact, pick any point and any initial direction, then standard existence results in differential equations say that there is a geodesic that goes through the point you picked in the direction you picked. Finally, that the equator is a geodesic makes it reasonable to believe that any great circle would be, since it could be moved around to coincide with the equator without distorting distances.

Before leaving surfaces, I want make one technical point that will have philosophical implications later on. The Γ^k_{ij}'s are called *Christoffel symbols* and the g_{ij}'s are the *metric coefficients*. Since the Christoffel symbols arose from orthogonally projecting onto the tangent plane it would seem reasonable that they would be related to the metric coefficients, which, in fact they are.

$$\Gamma^k_{ij} = \frac{1}{2}\sum g^{lk}\left(\frac{\partial g_{lj}}{\partial x_i} - \frac{\partial g_{ij}}{\partial x_l} + \frac{\partial g_{il}}{\partial x_j}\right)$$

where g^{ij}'s are the entries of the inverse of the matrix whose entries are the g_{ij}'s. The relationship is not pretty, but all you need to observe is that the Christoffel symbols can be calculated from the metric coefficients and their first derivatives.

[3] If we were the little people and the sphere were the earth, this mapping would be the *plate carrée* projection. There is a map called the *gnomonic* projection which allows you draw geodesics with a ruler.

Chapter 3

RIEMANN - Manifolds

By 1854 Bernhard Riemann had finished his doctoral dissertation in complex variables and his Habilitationsschrift on trigonometric series at the University of Göttingen and needed a job. He applied for the position of Privatdocent at the University (a lecturer who received no salary, but was merely forwarded fees paid by those students who chose to attend his classes - scary!). He was required to give a probationary inaugural lecture on a topic chosen by the faculty from a list of three that he was to propose. His first two choices were the topics in his dissertation and Habilitationsschrift, his third was something he called the foundations of geometry. Tradition had it that one of the two he was known for would be chosen. Gauss chose the third. The lecture, titled *Über die Hypothesen welche der Geometrie zu Grunde liegen* (On the Hypotheses that lie at the Foundations of Geometry) became the foundation of modern differential geometry and the tool that Einstein needed to deal with gravity.

Now, revolutionizing mathematics and physics was not what Riemann intended. He merely wanted to deal with a problem that had been around a while, but was a hot topic at the time, namely - is there something besides Euclid? Euclidean geometry was almost sacred. By the middle of the 19th century people began to wonder about its exalted position. The problem was the parallel postulate: given a line and a point not on the line there is exactly one line through the given point parallel to the given line. This axiom was the most complicated of Euclid's postulates and people began to wonder if it was really necessary to assume it. "Proofs" of the parallel postulate abounded. Gauss knew that the postulate was necessary, because the geometry of the sphere did not satisfy the parallel postulate. There are no parallel lines on a sphere. The analogs of straight lines are great circles, which always intersect. In 1829, Bolyai and Lobachevsky had independently constructed apparently consistent geometrical theories that assumed there could be more than one line through a given point parallel to a given line. Such a geometry can be realized on a surface that looks like a saddle. Gauss had this result even earlier, but told only a few people. He may have directly or indirectly inspired the work of Bolyai and Lobachevsky.

OK, so the Euclidean plane was not the only two-dimensional geometry, it was "flat" but there were others that were not. Of course, it was clear to everyone that space is flat, so Euclid was still good where it counts. Isaac Newton, for one said so. But, how can you be sure? Riemann in his lecture suggested a procedure for being sure and in a later paper worked out the details. I will describe essentially what Riemann proposed.

Riemann says we can view any "space" as an "n-fold extended manifold" where n is the number of independent directions we can go. We can lay out a coordinate system $(x_1, \ldots, x_n)$ in our space to quantify location. This notion has been made more precise with the modern definition of a differentiable manifold. The parameterizations I have described do the job for curves and surfaces. Riemann also suggests that we should be able to measure distances using an infinitesimal displacement or "line element" of the form

$$ds = \sqrt{\sum g_{ij}\, dx_i\, dx_j}$$

Looks familiar !

So, according to Riemann you have a set with a coordinate system (a manifold) and a metric (nowadays the two together are a Riemannian manifold) and away you go. Everything is intrinsic. A manifold does not sit in some larger place like surfaces sitting in space. It just is. He is now ready to show that with just this beginning you can tell if your manifold is flat - intrinsically.

The coefficients of the metric line element should tell you when your geometry is flat, but how? Certainly, if your metric coefficients are constant, such as $ds^2 = dx^2 + dy^2$ for R^2, then your world is flat. Unfortunately, the coefficients may not be constant and your world is still flat. For R^2 in polar coordinates we have $ds^2 = dr^2 + r^2 d\theta^2$. So, you can describe the same geometrical structure with different coordinate systems, and have different looking metrics.

On the other hand, if you have two different coordinate systems, then the coordinates themselves are related by a transformation, for example $x = r\cos\theta$ and $y = r\sin\theta$ relate rectangular and polar cordinates in R^2. The metric coefficients should also be related. For coordinate systems $(x_1, \ldots, x_n)$ and $(\bar{x}_1, \ldots, \bar{x}_n)$, the metric coefficients satisfy (more chain rule stuff)

$$g_{ij} = \sum \frac{\partial \bar{x}_k}{\partial x_i} \frac{\partial \bar{x}_l}{\partial x_j} \bar{g}_{kl} \tag{3.1}$$

Therefore the question of whether your world is flat or not, becomes: can you change your coordinate system $(x_1, \ldots, x_n)$ into a new system $(\bar{x}_1, \ldots, \bar{x}_n)$ so that your new metric coefficients are the euclidean coefficients, $\bar{g}_{ij} = 1$, if $i = j$ and zero otherwise. Which boils down to can the system of differential equations

$$g_{ij} = \sum \frac{\partial \bar{x}_k}{\partial x_i} \frac{\partial \bar{x}_k}{\partial x_j} \tag{3.2}$$

be solved for $\bar{x}_1, \ldots, \bar{x}_n$?

The system in (3.2) looks harmless enough, but in fact it is not in any standard form for which partial differential equations methods can be used. So, when in doubt differentiate and do a lot of algebra and maybe a miracle happens and indeed it does - you can get

$$\frac{\partial}{\partial x_i}\left(\frac{\partial \bar{x}_k}{\partial x_j}\right) = \sum \Gamma^l_{ij} \frac{\partial \bar{x}_k}{\partial x_l}$$

a system of linear equations with coefficients precisely our old friends the Christoffel symbols. There are a lot of equations, but each contains only one of the new coordinates, so that for each $\bar{x}_k$ you have a linear system for its partial derivatives. Solutions to these equations could then be integrated to get $\bar{x}_k$, with suitable initial conditions to ensure a non-trivial coordinate system. For solutions to exist certain "well known" integrability conditions on the coefficients must be satisfied[1], namely

$$\frac{\partial \Gamma^l_{ik}}{\partial x_j} - \frac{\partial \Gamma^l_{ij}}{\partial x_k} + \sum \left(\Gamma^p_{ik}\Gamma^l_{pj} - \Gamma^p_{ij}\Gamma^l_{pk}\right) = 0$$

[1]Integrability conditions are common for partial differential equations. Here is one you may know. Given P and Q defined on R^2, for there to be a function f so that $f_x = P$ and $f_y = Q$, the functions P and Q must satisfy the integrability condition $P_y = Q_x$.

(see Spivak, vol I p 187). Therefore, if your world is flat, then these combinations of the Christoffel symbols for your coordinate system are all zero. In fact, the condition is sufficient for local solutions, which in this case means that, if the condition is satisfied, then your world is flat.

The left side of this equation is denoted R^l_{ijk} and these numbers are the coefficients of the *Riemann curvature tensor*.[2] The miracle continues because the Riemann curvature tensor coefficients for two different coordinate systems are related by

$$\bar{R}^{\lambda}_{\alpha\beta\gamma} = \sum \frac{\partial \bar{x}_\lambda}{\partial x_l} \frac{\partial x_i}{\partial \bar{x}_\alpha} \frac{\partial x_j}{\partial \bar{x}_\beta} \frac{\partial x_k}{\partial \bar{x}_\gamma} R^l_{ijk} \tag{3.3}$$

which means that if the Riemann curvature tensor vanishes in one coordinate system, it vanishes in all. It does not matter what coordinate system you are using, if the Riemann curvature tensor is zero, your world is flat!

I mention in passing that since the Γ's are determined by the metric coefficients and their first derivatives, the R's are functions of the metric coefficients and their first and second derivatives. You can read the footnote again now, if you want to.

A couple of interesting simplifications of the Riemann curvature tensor have appeared. The first is the *Ricci curvature tensor*

$$R_{ij} = \sum R^l_{ilj}$$

which happens to be symmetric, $R_{ij} = R_{ji}$. The other is the *scalar curvature*

$$R = \sum g^{ij} R_{ij}$$

[2] I do not want to get into a big discussion of tensors, but will call something a tensor when it is one. So, what is it? Tensors are natural generalizations of vectors and matrices. They are fundamental objects for describing geometrical and even physical concepts. If you can pose a fact or law of nature in terms of a tensor's coefficients in one coordinate system the law will look the same in any other.

Let me just say that you could even define a tensor as a collection of numbers for each point in an n-dimensional coordinate system with subscripts and superscripts taking values from 1 to n that transform linearly when the coordinates are changed using the partial derivatives of one set of coordinates with respect to the other. Subscripts transform one way and superscripts another. Look carefully at (3.3), which comes up shortly and you will see what I mean. In fact, (3.1) says that the metric coefficients are the coefficients of a tensor. People, including me, often simply write the symbol for the coefficients of a tensor and refer to it as the tensor, for example I might say "the metric tensor g_{ij}."

This vague definition of tensor may be annoying or at least unmotivated, but it will do for our purposes and give you something to look into. Well, I guess this footnote has become a digression itself, but after all it is a footnote, you did not have to read it.

The two will become a good deal more interesting when we get to general relativity.

If you think about it, the Riemann curvature tensor can have a lot of coefficients, actually, n^4. Fortunately, there are symmetries and antisymmetries in the subscripts and superscripts, such as $R^l_{ijj} = 0$, so that there are not so many independent coefficients, $n^2(n^2 - 1)/12$ to be exact. So, for $n = 2$ there is only one and all you need to know is the scalar curvature, R. In three dimensions there are six, and all you need to know is the Ricci curvature tensor, R_{ij}. In dimension four, the interesting one for general relativity, there are twenty, ten of which can be summarized in the Ricci curvature tensor and the other ten, well, that is a long story.

I have used the word curvature a lot, but only to determine if things do not curve. Needless to say, there is more that could be said about how Riemann curvature is really curvature, but I will just leave it as more for you to look into. To stimulate your interest, let me make a couple of unsupported observations about curvature in surfaces. If you lived in a plane, you could determine it was flat because you would find that $R = 0$ at every point. On a sphere you would find that R was always positive and on a torus you would find that R was positive at some points, negative at some points, and even zero at some points. You might be surprised to know that little people living in the surface of a cylinder would think their world was flat. They would find that $R = 0$ at every point and, they would be correct. A cylinder is just a rolled up plane, rolled without stretching or folding. The length of curves, the geodesics, and all intrinsic aspects of the geometry are not changed by rolling up the plane. It only appears to be a cylinder to those looking at it from the outside - extrinsically.

Chapter 4

EINSTEIN - Gravity

Before looking at Einstein's gravity, I want to look at Newton's. According to Newton an object moves in a straight line (geodesic!) at a constant speed in the absence of forces. If you asked Newton what a force was he would say it was anything that causes you to deviate from travelling in a straight line at a constant speed.

Suppose we have an object with mass M sitting at the origin of a coordinate system in space, then, according to Newton the gravitational force exerted by the object on a passing particle will accelerate the particle by an amount inversely proportional to the square of the distance between them, that is, for a particle at a location $(x, y, z) = \alpha(t)$ at time t,

$$\alpha''(t) = -\frac{MG}{r^2}\mathbf{u}$$

where where G is Newton's gravitational constant of proportionality, $r = \sqrt{x^2 + y^2 + z^2} = |\alpha(t)|$ is the distance between the object and the particle and $\mathbf{u} = \alpha(t)/r$ is a unit vector in the direction from the object to the particle.

Gravitational force is conservative, that is, it is the gradient of a potential function, $\Phi(x, y, z) = -MG/r$ in this situation, so that

$$\alpha''(t) = -\nabla\Phi$$

Furthermore, the divergence of $\nabla\Phi$, that is, the Laplacian, $\nabla^2\Phi$, satisfies

$$\nabla^2\Phi = 0 \tag{4.1}$$

This equation is the *Newtonian gravitational field equation* and characterizes gravity in our situation. By "characterize" I mean that solutions to the equation tell you how gravity acts. That may seem a little far fetched, since a lot of functions satisfy the equation. Our situation has a single mass. If you add the physically reasonable assumption that gravity due to that mass is radially symmetric, then Φ is a function of r alone and (4.1) becomes a simple ordinary differential equation whose solutions are of the form $\Phi = a/r + b$. If you also add the physically reasonable assumption that the gravitational effect of the mass goes to zero as you get further away, then b must be zero and you are back to where we started.

In general, there is other mass about and the field equation becomes

$$\nabla^2\Phi = 4\pi G\rho$$

where ρ is the density of the mass at each point in the universe. You tell me ρ and I will tell you Φ, how gravity works for that distribution of mass.

That is gravity according to Newton.

Now, to Einstein, but not directly to gravity, first we need to talk about *spacetime*.

We live in spacetime, a four dimensional manifold where one coordinate measures time and the other three distance. The "points" in spacetime are called *events*. To Newton time and space were absolute, the same for everyone. To Einstein, the speed of light c was absolute. Suppose I think I am standing still, you are moving ahead with a velocity v, and a photon whizzes by both of us. According to Newton, if I measure the speed of the photon, I will get c, but if you measure it you will get $c - v$, assuming you were following the same line as the photon. According to Einstein we would both measure the speed of the photon to be c.

Many surprising conclusions follow from Einstein's statement that the speed of light is the same for all observers. Perhaps the most fundamental conclusion is that two different observers could measure different times and distances between the same two events. Time and space are relative, not absolute.[1]

[1] Actually, distance between events was not absolute, even for Newton. If a person riding a train going 120 km/hr drops a ball from a height of one meter, she would say that the ball traveled one meter straight down before it hit the floor of the train. Her mother-in-law, standing on the platform watching the train go by would say that the ball traveled about 3.5 meters along a diagonal path. The distance between events in Newtonian space depends on the observer. Who is correct? The mother-in-law, she always is.

It can be shown (don't you just hate that phrase) that there is one measurement that is the same for all observers, the so called *proper time*, $\Delta\tau$, given by

$$\Delta\tau^2 = \Delta t^2 - (\Delta x^2 + \Delta y^2 + \Delta z^2)/c^2$$

where Δt is the time and Δx, Δy, and Δz the distances between two events measured by an observer, any observer. Proper time is the arclength of spacetime. The Minkowski metric for spacetime is

$$d\tau^2 = dt^2 - (dx^2 + dy^2 + dz^2)/c^2$$

Actually that is the metric for "flat" spacetime, where there is no gravity and no forces.

According to Einstein there are forces, but *gravity is not one of them.* In the absence of forces, such as electromagnetism, free falling particles travel along geodesics, just like Newton's particles in the absence of forces except that for Einstein there is one less force. In other words, the path of a particle through spacetime $\alpha(\tau) = (t, x, y, z) = (\alpha_0(\tau), \alpha_1(\tau), \alpha_2(\tau), \alpha_3(\tau))$, satisfies the geodesic equations (2.1), namely

$$\alpha_k'' = -\sum \Gamma_{ij}^k \alpha_i' \alpha_j'$$

where the derivatives are with respect to proper time. But where do the Γ's come from? A metric tensor, of course. But where does the metric tensor come from? GRAVITY, of course. Mass changes the shape of the universe and the effect of the change in shape we call gravity.

The metric tensor g_{ij} that defines the geometry is essentially Einstein's gravitational potential function analogous to Newton's Φ. The Γ's are built from the metric coefficients and their first derivatives (I said that fact would become interesting) analogous to $\nabla\Phi$. All that remains is the *Einstein gravitational field equation* to characterize gravity analogous to Newton's $\nabla^2\Phi = 4\pi G\rho$, that is, to describe how the distribution of mass in the universe determines the metric tensor, hence the geometry of the universe.

At this point I have the impression that Einstein just started guessing and decided to come up with an equation that would be reasonable, not necessarily be based on some fundamental principle, but would survive if it agreed with experiment. Of course, his guesses were educated. In particular, the effect of gravity was due to mass, or mass-energy, really, since $E = mc^2$ said mass and energy were the same thing. There was already available a general description of mass-energy in the *energy-momentum tensor*, T_{ij}, a symmetric tensor that could describe the distribution of mass-energy in spacetime analogous to Newton's ρ. So, he wanted a symmetric tensor that was related to the geometry and should be built from the metric coefficients and their first and second derivatives, as was Newton's $\nabla^2\Phi$. Well folks, the Ricci curvature tensor, R_{ij} fits the bill. So, his first choice was simply to say that R_{ij} is proportional to T_{ij}, that is,

$$R_{ij} = kT_{ij}$$

for some constant k. Unfortunately, this equation only works when there is no mass except at a single point as in the example we used for Newtonian gravity where the equation becomes $R_{ij} = 0$. The reason it does not work in general is that there is a way of calculating something called the divergence of a tensor and the divergence of T_{ij} is zero, but the divergence of R_{ij} is not.

To fix the problem Einstein added more geometry. It can be shown that the divergence of

$$R_{ij} - \tfrac{1}{2}R\, g_{ij} + \Lambda\, g_{ij}$$

is zero where R is the scalar curvature and Λ is any constant. In fact, this tensor is the only tensor with zero divergence and the correct indices that can be built linearly from the Riemann curvature tensor. Einstein was able to determine that the proportionality constant k should be $8\pi G$ and could see no reason why Λ should be there at all, so he presented to the world the *Einstein field equation for gravitation*

$$R_{ij} - \tfrac{1}{2}R\, g_{ij} = 8\pi G\, T_{ij}$$

and that is where the journey almost ends.

The equation looks very simple and elegant, but it is, in fact, a nasty system of ten second order, non-linear differential equations. Solutions are not easy to come by, but several have been obtained in special circumstances. One of the first was presented to Einstein in 1916 by Karl Schwarzschild who was also calculating artillery trajectories on the Russian front at the time. Schwarzschild obtained the metric for the same situation I used for my Newtonian example, one mass M.

$$d\tau^2 = \left(1 - \frac{2GM}{r}\right) dt^2 - \left(1 - \frac{2GM}{r}\right)^{-1} dr^2 - r^2(\cos^2\phi\, d\theta^2 + d\phi^2)$$

Note the presence of Newton's potential function Φ. I also mention it because it predicts black holes! But, I will let you take that from here.

Not too long after the equation was published, Alexander Friedman told Einstein that his equation implied that the universe could be expanding. Einstein refused to believe that was possible. The sacred universe should be static. Friedman finally convinced Einstein, who then decided to change the equation by putting the Λg_{ij} term back in. Carefully choosing Λ would rule out an expanding or contracting universe. Λ became known as the *cosmological constant*. About five years later Edwin Hubble saw his red shift that showed that the universe was expanding. Einstein removed the cosmological constant once again and stated that putting it in was the worst mistake of his life. In recent years there have been indications that Λ should be put back in. I refer you to Brian Greene's *The Elegant Universe* for that story.

Some people have felt that at this point Einstein made a real mistake. He launched his ill-fated campaign for a *unified field theory* that incorporated all forces into the geometry of spacetime. He never found it, but I bet he could almost taste it, so near, yet so far. To be fair to him, you can almost taste it. At the beginning of the twentieth century there really was only one other force, electromagnetism. Quantum theory was in its infancy and its strong and weak nuclear forces had yet to come to prominence. Einstein did not think too much of quantum theory anyway. It was he who said about quantum theory that "I am convinced that He does not play dice." In his mind all he needed to do was to somehow incorporate electromagnetism into geometry. There was room. Gravity was covered by the Ricci tensor, which used up half of the available information in the curvature tensor, there were still ten independent coefficients left. There is also a tensor that characterizes the electromagnetic field, the Maxwell tensor, F_{ij}. It is antisymmetric, that is $F_{ji} = -F_{ij}$, so it had six degrees of freedom, plenty of room for it. There is also the current density tensor J_i with four degrees of freedom. It all seem to add up, at least, it seemed that it should. I can see Einstein saying to himself that it just has to fit together somehow, but he never could make it happen. In fact, Theodor Kaluza sent a paper to Einstein in 1919 showing how it might be possible to unify electromagnetism and gravity, but it required a five dimensional space, one time dimension and four spatial dimensions. Einstein and no one else, at least for a while, could conceive of such a thing - look around, there are obviously only three spatial dimensions.

It may yet be that both Einstein and Kaluza were on the right track. It may, in fact, be possible to find a unified "Theory of Everything" and it may live in a ten or more dimensional universe. But, that is where string theory comes in and Brian Greene tells that story very well. So, I think it is time for me to come to the end of my story.

Einstein's field equation must stand as one of the great intellectual achievements. I enjoyed very much trying to understand it.

References

Carroll, Sean M. (2004), *Spacetime and Geometry,* San Francisco: Addison Wesley.

Faber, Richard (1983), *Differential Geometry and Relativity Theory,* New York: Marcel Dekker.

Flanders, Harley (1989), *Differential Forms with Applications to the Physical Sciences,* New York: Dover.

Gray, Alfred (1999), *Modern Differential Geometry of Curves and Surfaces with Mathematica, 2nd Edition,* Boca Raton: CRC Press.

Greene, Brian (2003), *The Elegant Universe,* New York: Random House.

Greene, Brian (2004), *The Fabric of the Cosmos,* New York: Alfred A. Knopf.

Hartle, James B. (2003), *Gravity: an Introduction to Einstein's General Relativity,* San Francisco: Addison Wesley.

Lawden, Derek F. (1982), *Introduction to Tensor Calculus, Relativity and Cosmology,* New York: Dover.

Ludvigsen, Malcolm (1999), *General Relativity,* Cambridge: Cambridge Press.

Millman, Richard S. and Parker, George D. (1977), *Elements of Differential Geometry,* Englewood Cliffs: Prentice Hall.

Mlodinow, Leonard (2001), *Euclid's Window,* New York: The Free Press.

Spivak, Michael (1999), *A Comprehensive Introduction to Differential Geometry, 3rd Edition,* Houston: Publish or Perish.

Finally, I refer you to Wolfram's *Mathematica Information Center*, a veritable wealth of information and resources on many subjects, to be found at

`http://library.wolfram.com/infocenter/`