Curvature of Space and Time, with an Introduction to Geometric Analysis

Iva Stavrov Allen

PCMI, 2013

Contents

CHAPTER 1

Introduction to Riemannian geometry

1. Riemann's Habilitation lecture in examples

Introduction. German academic system recognizes a post-PhD qualification called Habilitation. Bernhard Riemann did his Habilitation work at University of Göttingen in early 1850's, while Carl Friedrich Gauss was the Chair of Mathematics. The practice at the time was to have the Habilitation candidate prepare several public lectures (not necessarily related to the Habilitation dissertation), and to have the Chair choose the one to be given. Gauss's choice was "Über die Hypothesen, welche der Geometrie zu Grunde liegen" ("On Hypotheses Which Lie at the Bases of Geometry"), which Riemann gave on June 10th 1854. The following quote[1] from Riemann's lecture summarizes the key ideas behind what is now called *Riemannian geometry*.

> Position-fixing being reduced to quantity-fixings, and the position of a point in the n-dimensional manifoldness being consequently expressed by means of n variables $x_1, x_2, x_3, ..., x_n$, the determination of a line comes to the giving of these quantities as functions of one variable. The problem consists then in establishing a mathematical expression for the length of a line, and to this end we must consider the quantities x as expressible in terms of certain units. I shall treat this problem only under certain restrictions, and I shall confine myself in the first place to lines in which the ratios of the increments dx of the respective variables vary continuously. We may then conceive these lines broken up into elements within which the ratios of quantities dx may be regarded as constant; and the problem is then reduced to establishing for each point a general expression for the linear element ds starting from that point, an expression which will thus contain the quantities x and quantities dx.

Our goal for today is to analyze this paragraph in detail. We will mainly do so on examples, and state the general principles at the end.

Example of Euclidean $\mathbb{R}^n$. Position of a point in $\mathbb{R}^n$ can be expressed in terms of n Cartesian coordinates[2]: $(x^1, x^2, ..., x^n)$. Paths / curves in $\mathbb{R}^n$ are best described parametrically

$$\left(x^1(t), x^2(t), ..., x^n(t)\right), \quad a \leq t \leq b,$$

[1] The translation used here is due to W. K. Clifford; see *Nature*, Vol. VIII. Nos. 183, 184, pp. 14 –17, 36, 37.

[2] We intentionally depart from the notation used in Riemann's original paper and adopt modern notational conventions according to which the indices on coordinate variables are superscripts.

although it should be noted that one and the same geometric path permits different parametrizations in t (which depend on the "traversing speed"). For each $1 \le j \le n$ the paths

$$x^i(t) = \begin{cases} t, & \text{if } \ i = j \\ \text{const}, & \text{if } \ i \neq j \end{cases}$$

define coordinate grid lines. The tangent vector field to the above indicates the direction in which the x^j-coordinate increase; such vector fields will be denoted by ∂_{x^j}, or ∂_j for short. At each point of $\mathbb{R}^n$ the vectors $\{\partial_j\}$ provide a basis for the space of vectors based at that point. In particular,

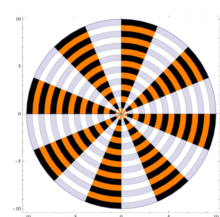
$$\sum_i \tfrac{\partial x^i}{\partial t} \partial_i = \sum_i (\partial_t x^i)\, \partial_i$$

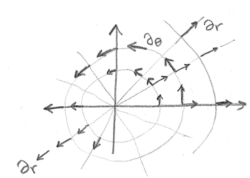
represents the tangent vector field to the curve given by $\big(x^1(t), x^2(t), \ldots.., x^n(t)\big)$. The length of the above curve can be conceptualized in two equivalent ways. As suggested by Riemann, one can compute the length of the curve by breaking it up into infinitesimal pieces (*line elements*) which are effectively straight (so that the ratios of infinitesimal coordinate increments dx_i *"may be regarded as constant"*), by computing the length ds of the resulting line element using Pythagorean Theorem

$$ds^2 = \sum_i (dx^i)^2 = \left(\sum_i (\partial_t x^i)^2 \right) dt^2$$

and by adding ds's up using definite integrals. Of course, this is the same as integrating the magnitude of the velocity vector $\sum_i (\partial_t x^i)\, \partial_i$ with respect to t.

It should be noted that there are many other ways of *coordinatizing* $\mathbb{R}^n$, or its subsets. For instance, in $\mathbb{R}^2$ one may want to utilize polar coordinates (r, θ). Their effect can be illustrated as follows. Note that ∂_r and ∂_θ are tangent vector fields to polar coordinate grid lines, and that they point in the direction in which r and θ increase.





It is important to have a means of relating expressions coming from different coordinate systems. In our situation one can use the formulae $x = r\cos\theta$, $y = r\sin\theta$ to relate Cartesian expressions to those given in terms of polar coordinates. For example, the expression for the Euclidean line element $ds^2 = dx^2 + dy^2$ becomes

$$ds^2 = dr^2 + r^2 d\theta^2.$$

The latter brings us to the following quote from Riemann's lecture:

> ".... the problem is then reduced to establishing **for each point** a general expression for the linear element ds starting from that point, an **expression which will thus contain the quantities** x and quantities dx."

As the illustration of the coordinate grids from the above shows, the curvilinear mapping $(r, \theta) \mapsto (r\cos\theta, r\sin\theta)$ distorts distances: segments $r = 2$ and $r = 6$ appear to be equally long on the coordinate grid on the left but the corresponding lengths in Euclidean $\mathbb{R}^2$ (middle diagram) are very different. The reason for this can be tracked back to the fact that the expression for ds^2 in polar coordinates *varies from point to point* and, more specifically, $d\theta^2$ contributes differently to ds^2 when $r = 2$ and when $r = 6$.

Generally speaking, when using curvilinear (e.g. polar, cylindrical, spherical) coordinates $(y^1, ..., y^n)$ for the Euclidean $\mathbb{R}^n$ we should expect the line elements to take the form of

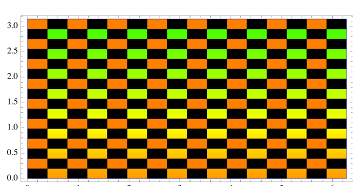
$$\boxed{ds^2 = \sum_{ij} g_{ij} dy^i dy^j,}$$

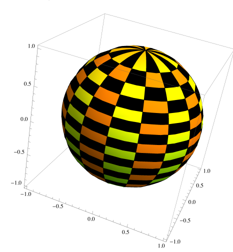
for some functions $g_{ij} = g_{ij}(y^1, y^2, ..., y^n)$. (In fact, the functions g_{ij} are in this case determined by the Euclidean inner-products of coordinate vector fields ∂_{y^i} and ∂_{y^j}; see exercise (8) following this lecture.) The differences in the sizes of different coordinate cells is all accounted for in the expression for the line element.

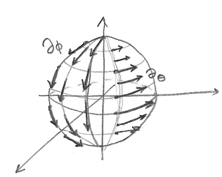
Spheres in Euclidean spaces. Map making process results in *coordinatization* of the surface of the Earth. In general, coordinates on a surface constitute a curvilinear mapping which transforms a map/rectangle in our atlas to the surface that is being coordinatized. One widely used example of a coordinatization comes from *spherical coordinates.* More specifically, the curvilinear mapping

$$(\theta, \phi) \mapsto (\varrho \sin\phi \cos\theta, \varrho \sin\phi \sin\theta, \varrho \cos\phi)$$

transforms a rectangle in our atlas (a piece of $\theta\phi$-plane[3] with $0 < \theta < 2\pi$ and $0 < \phi < \pi$) to the surface of the sphere $x^2 + y^2 + z^2 = \varrho^2$ of radius ϱ in $\mathbb{R}^3$. The nature of spherical coordinates is represented on the following diagrams, which show what is happening to coordinate grids and to coordinate vector fields ∂_θ, ∂_ϕ.





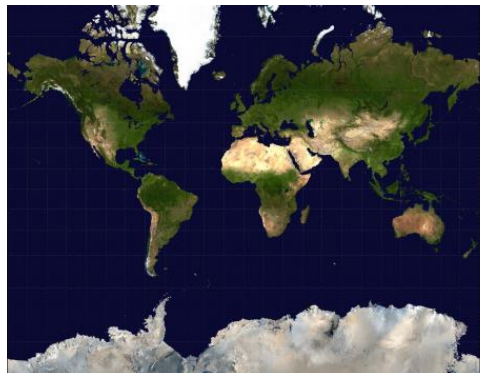


It is important to notice that since we need only two parameters (de-facto latitude and longitude) to describe the position of an object on its surface, the sphere is two dimensional, despite the fact that we tend to visualize the sphere as being a part of a larger three dimensional space. (This also explains why we use S^2 to denote the standard unit sphere in $\mathbb{R}^2$.) In the case when "sphere" is actually our planet Earth we are also interested in more detailed representations of certain regions of the Earth, and as a result we have a full atlas worth of maps (pun!!) from a rectangle (the sheet of paper on which the map is drawn) to the surface of

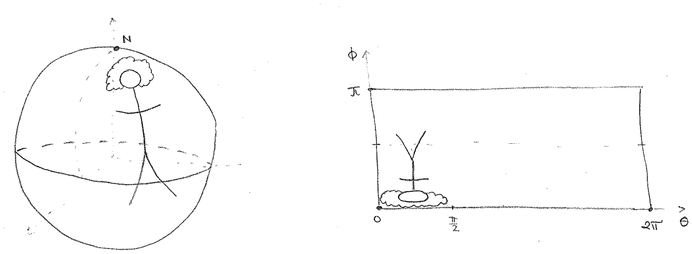
[3]Although in some sense the "North Pole" corresponds to $\phi = 0$, it is not described by a unique pair of coordinates (θ, ϕ); similar comments apply to the "South Pole" $\phi = \pi$ and the "Greenwich meridian" $\theta \in \{0, 2\pi\}$. For this reason it is advantageous to restrict our attention to the open rectangle $(0, 2\pi) \times (0, \pi)$.

the Earth. This atlas *coordinatizes* the Earth: every point is represented in some rectangle in a coordinate plane.

Once again one should observe the extent to which the the distances on the sphere differ from those in coordinate map. One can easily relate to the issue by looking at, say, the Mercator world map[4] in which Antarctica and Greenland look huge, especially in relation to countries in the tropics.



Here is what a map in our atlas would look like if it were to be based on spherical coordinates. Note how stretched the hair of stick-man appears in the map.



Now let us investigate this from phenomenon in spherical coordinates. Assuming the sphere is embedded in Euclidean geometry coordinatized by (x, y, z), the length of any path – so including those on the surface of the sphere – are to be computed using

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

Inserting $x = \varrho \cos\theta \sin\phi$, $y = \varrho \sin\theta \sin\phi$, $z = \varrho \cos\phi$ into the above we get

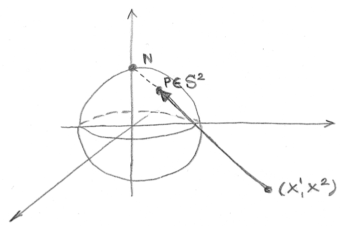
$$ds^2 = \varrho^2 \left[d\phi^2 + \sin^2(\phi)\, d\theta^2 \right].$$

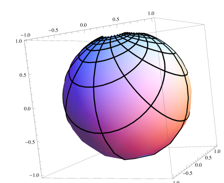
This is the formula for the *line element on spheres of radius* ϱ in Euclidean $\mathbb{R}^3$. Note that the expression for the line element is not the same everywhere on the sphere – one of the coefficients depends on ϕ. In other words, the expression for the line element depends on the North/South location. This type of dependance on the North/South location is exactly what we want in order to be able to account for the distorted distances on, say, the Mercator world map. The function $\sin^2(\phi)$ scrunches the contributions[5] of $d\theta^2$ as one approaches the North and the South Poles in such a way that the $\phi =$ const circles close up to form a dome.

[4]Retrieved from `http://www.learnnc.org/lp/multimedia/14596` on 5/3/13.

[5]There is a lot to be learned by comparing the situation here to that of polar coordinates in Euclidean $\mathbb{R}^2$ where the contributions of $d\theta$ were increasing proportionally to r.

There are other ways of coordinatizing spheres. One such coordinatization is based on something something called the *stereographic projection.* This transformation bijectively maps the surface of a sphere, excluding a point, onto $\mathbb{R}^2$. For instance, we may consider the excluded point to be "the North Pole" $(0,0,1)$ of the unit sphere and the target plane to be the equatorial plane $(x^1, x^2, 0)$; we require the North Pole, the point $P \in S^2$ and its projection $(x^1, x^2, 0)$ to be collinear. The inverse of the stereographic projection gives rise to a coordinatization of the unit sphere (without a point). To understand the visual aspects of this coordinatization observe the extent to which the "coordinate lines" on the surface of the sphere become denser the closer we get to the North Pole. Working out analytical details on this example, including the form of the spherical line element in such coordinates, is a worthwhile exercise.





Note that the entire sphere can be covered by two coordinate patches coming from the stereographic projection: one from the North Pole and one from the South Pole.

More generally, equations of the form $(x^1)^2+...+(x^n)^2 = \varrho^2$ represent spheres of radius ϱ in Euclidean $\mathbb{R}^n$. One can coordinatize such spheres using $n-1$ coordinates; this can be accomplished by means of stereographic projection. (See exercises.) As a result, spheres in $\mathbb{R}^n$ are $n-1$-dimensional. Unit sphere in Euclidean $\mathbb{R}^{n+1}$ is denoted by S^n. The line element it inherits from $\mathbb{R}^{n+1}$ is, in stereographic coordinates, given by

$$ds^2 = \frac{4}{(1+(x^1)^2+...+(x^n)^2)^2}\left((dx^1)^2+...+(dx^n)^2\right).$$

Once more, note how the multiplicative factor $\frac{4}{(1+(x^1)^2+...+(x^n)^2)^2}$ scrunches far away regions of $\mathbb{R}^n$. This, for example, explains why the coordinate grid lines are effectively denser near the North Pole of the above diagram.

Real projective plane. A much more abstract example is that of the real projective plane, denoted $\mathbb{R}P^2$. As a set $\mathbb{R}P^2$ consists of lines passing through the origin in $\mathbb{R}^3$. The study of such an object is motivated the fact that the set of lines through a painter's eye to some extent corresponds to the set of points on this painter's canvas. Each element of $\mathbb{R}P^2$ is a line spanned by some non-zero vector $\mathbf{u} = (u^1, u^2, u^3)^T \in \mathbb{R}^3$:

$$\ell_{\mathbf{u}} = \{\lambda \mathbf{u} \,|\, \lambda \in \mathbb{R}\}.$$

Note that $\ell_{\mathbf{u}} = \ell_{\mathbf{v}}$ if and only if $\mathbf{u}$ and $\mathbf{v}$ are collinear non-zero vectors. By rescaling if necessary we may always assume that at least one coordinate of $\mathbf{u} \neq \mathbf{0}$ is equal to 1. It now follows that the entire $\mathbb{R}P^2$ can be covered by 3 maps, the domain of which is $\mathbb{R}^2$:

$$(x^1,x^2) \mapsto \ell_{(1,x^1,x^2)^T}, \quad (x^1,x^2) \mapsto \ell_{(x^1,1,x^2)^T}, \quad (x^1,x^2) \mapsto \ell_{(x^1,x^2,1)^T}.$$

Certain regions of $\mathbb{R}P^2$ are represented in several different maps. For instance, the subset $\{\ell_{(u^1,u^2,u^3)^T} | u^1, u^2 \neq 0\} \subseteq \mathbb{R}P^2$ is covered by both $(x^1,x^2) \mapsto \ell_{(1,x^1,x^2)^T}$

and $(y^1, y^2) \mapsto \ell_{(y^1,1,y^2)^T}$. The two sets of coordinates for $\ell_{(1,x^1,x^2)^T} = \ell_{(y^1,1,y^2)^T}$ are related as follows:

$$y^1 = \frac{1}{x^1}, \quad y^2 = \frac{x^2}{x^1}; \quad x^1 = \frac{1}{y^1}, \quad x^2 = \frac{y^2}{y^1}.$$

The mapping which expresses how y-coordinates depend of x-coordinates is called a transition map and will be denoted by Φ_x^y; its inverse is Φ_y^x. Note that in our case Φ_x^y and Φ_y^x are differentiable bijections between $\{(x^1, x^2)|x^1 \neq 0\}$ and $\{(y^1, y^2)|y^1 \neq 0\}$.

What would be the line element on $\mathbb{R}P^2$? According to Riemann,

> ".... the problem is then reduced to establishing **for each point** a general expression for the linear element ds starting from that point, an **expression which will thus contain the quantities** x and quantities dx."

So in coordinates ds^2 can be just about anything of the form $\sum g_{ij}(dx^i)(dx^j)$. Of course, there are some limitations. First of all, we have to have $ds^2 > 0$. Perhaps more importantly, the coordinate expressions for ds^2 have to line up whenever two coordinate patches overlap. That is, the transition map Φ_x^y has to reduce the coordinate expression for ds^2 in y-coordinates to that in x-coordinates and – vice versa – Φ_y^x has to reduce the coordinate expression for ds^2 in x-coordinates to that in y-coordinates.

An example of such a line element is

$$ds^2 = \frac{(1+(x^2)^2)(dx^1)^2 - 2x^1x^2dx^1dx^2 + (1+(x^1)^2)(dx^2)^2}{(1+(x^1)^2+(x^2)^2)^2}.$$

More precisely, we claim that ds^2 can be taken to have this form in any of the three coordinate patches introduced above[6]. To see this first observe that $ds^2 > 0$ as its numerator can be expressed as a sum of squares. Secondly, consider the transition map Φ_y^x from the above. Inserting $x^1 = \frac{1}{y^1}, \;\; x^2 = \frac{y^2}{y^1}$ into the expression for ds^2 in x-coordinates and simplifying yields the claimed expression for ds^2 in y-coordinates:

$$ds^2 = \frac{(1+(y^2)^2)(dy^1)^2 - 2y^1y^2dy^1dy^2 + (1+(y^1)^2)(dy^2)^2}{(1+(y^1)^2+(y^2)^2)^2}.$$

Generalizations of $\mathbb{R}P^2$ include increasing the dimension (e.g. collection of lines in $\mathbb{R}^{n+1}$) or replacing $\mathbb{R}$ by some other normed division algebra such as $\mathbb{C}$; consult exercises following this lecture for further details.

Exercises for Lecture 1.

(1) (a) Find the explicit formulae relating the point $P \in S^2$ whose Cartesian coordinates are (x, y, z) to its stereographic projection $(x^1, x^2, 0)$, and vice versa. [Basic]

(b) Coordinatize the entire S^2 with two coordinate patches, one arising from the stereographic projection from the "North Pole" and the other arising from the stereographic projection from the "South Pole". Sketch a figure on the surface of S^2 and then sketch the corresponding figures in the two maps. [Basic]

[6] The fact that in this example ds^2 has the same form in each of the three coordinate patches is an artifact of the fact that this example is particularly natural. In general one should not expect this feature.

(c) Find the explicit formulae for the transition maps corresponding to the coordinate patches used in part (1b) above. Illustrate the effect of the transition maps on the polar coordinate grids in $\mathbb{R}^2 \smallsetminus \{(0,0)\}$, and convince yourself that the transition maps are smooth bijections of $\mathbb{R}^2 \smallsetminus \{(0,0)\}$. [Basic]

(2) (a) Based on the example of $\mathbb{R}P^2$ worked out in the lecture, roughly describe the coordinatization of the n-dimensional real projective space:

$$\mathbb{R}P^n = \{\ell_{\mathbf{v}} \,|\, \mathbf{v} \in \mathbb{R}^{n+1} \smallsetminus \{\mathbf{0}\}\}$$

for $\ell_{\mathbf{v}} = \{\lambda\mathbf{v} \,|\, \lambda \in \mathbb{R}\}$. [Basic]

(b) Let $\mathbb{C}P^n$ denote the set of all complex lines in $\mathbb{C}^{n+1}$ passing through the origin; in other words, let $\mathbb{C}P^n$ denote the set of all $\ell_{\mathbf{v}}$ where

$$\ell_{\mathbf{v}} = \{\lambda\mathbf{v} \,|\, \lambda \in \mathbb{C}\}$$

for some non-zero $\mathbf{v} = (v^1, v^2, ..., v^{n+1}) \in \mathbb{C}^{n+1}$. Show that $\mathbb{C}P^n$, called the complex projective space, can be coordinatized by using $2n$ variables. [Intermediate]

(c) Research the literature on projective spaces over quaternions and octonions. [Advanced]

(3) (a) Consider the coordinatization of $\mathbb{R}^2 \smallsetminus \{(0,0)\}$ using polar coordinates (r,θ). Express the vector fields ∂_r and ∂_θ in terms of Cartesian ∂_x and ∂_y. [Basic]

(b) Consider the coordinatization of $\mathbb{R}^3 \smallsetminus \{(0,0,0)\}$ using spherical coordinates (r,θ,ϕ). Express the vector fields ∂_r, ∂_θ and ∂_ϕ in terms of Cartesian ∂_x, ∂_y and ∂_z. [Basic]

(4) Provide a computation which shows that the (Euclidean) line element for $\mathbb{R}^2 \smallsetminus \{(0,0)\}$ in polar coordinates takes the form of $ds^2 = dr^2 + r^2 d\theta^2$. [Basic]

(5) Consider the polar coordinates (r,θ) in the Euclidean plane, and consider the line element $ds^2 = dr^2 + f(r)^2 d\theta^2$ for some positive function $f(r)$.

(a) Compute the length of $r(t) = t$, $\theta(t) = \theta_*$, $0 < t < \varrho$. [Basic]

(b) Compute the length of $r(t) = \varrho$, $\theta(t) = t$, $-\pi < t < \pi$ in terms of f. [Basic]

(c) Which geometry corresponds to the following choices of f? [Basic]

- $f(r) = r$;
- $f(r) = \sin(r)$, under the restriction $r \in (0,\pi)$;
- $f(r) = 1$;

(d) Based on the above provide a visual description of the geometries corresponding to the following choices of f. [Intermediate, Expected]

- $f(r) = \frac{1}{2}r$;
- $f(r) = 2r$;
- $f(r) = 1 + 2r$;

- $f(r) = 2 + \sin(r)$
- $f(r) = e^r$;
- $f(r) = \sinh(r) = \frac{1}{2}(e^r - e^{-r})$.

(6) (a) Develop formulas for a stereographic coordinatization of S^n. [INTERMEDIATE]

(b) Provide a computation which shows that the standard line element on S^n takes the form of

$$ds^2 = \frac{4}{(1 + (x^1)^2 + \cdots + (x^n)^2)^2} \left(d(x^1)^2 + \cdots + d(x^n)^2 \right)$$

in the coordinates arising from the stereographic projection. [INTERMEDIATE]

(7) (a) Provide a computation which shows that the Euclidean line element for $\mathbb{R}^3 \setminus \{(0,0,0)\}$ in spherical coordinates (r, θ, ϕ) takes the form of $ds^2 = dr^2 + r^2 d\Theta^2$, where $d\Theta^2$ denotes the spherical line element $d\Theta^2 = d\phi^2 + \sin^2 \phi \, d\theta^2$. [BASIC, EXPECTED]

(b) In analogy to the exercise 5, provide a visual description of the geometries on $\mathbb{R}^3 \setminus \{(0,0,0)\}$ corresponding to the following line elements. [INTERMEDIATE, EXPECTED]

- $ds^2 = dr^2 + \sin^2 r \, d\Theta^2$;
- $ds^2 = dr^2 + \sinh^2 r \, d\Theta^2$.

(8) Assume that $(x^1, \ldots, x^n)$ are Cartesian and assume that $(y^1, y^2, \ldots, y^n)$ are some (curvilinear) coordinates on $\mathbb{R}^n$. [INTERMEDIATE]

(a) Show that

(i) $dx^i = \sum_k \frac{\partial x^i}{\partial y^k} dy^k$;

(ii) $\partial_{y^i} = \sum_k \frac{\partial x^k}{\partial y^i} \partial_{x^k}$;

(b) Compute $\langle \partial_{y^i}, \partial_{y^j} \rangle$ where $\langle ., . \rangle$ signifies the Euclidean inner-product in $\mathbb{R}^n$;

(c) Compute g_{ij}, the coefficients of Euclidean line element expression in y-coordinates,

$$ds^2 = \sum_{ij} g_{ij} dy^i dy^j.$$

(d) Compare the answers to the last two questions.

(9) (a) It can easily be seen that

$$ds^2 = \frac{(1 + (x^2)^2)(dx^1)^2 - 2x^1 x^2 dx^1 dx^2 + (1 + (x^1)^2)(dx^2)^2}{(1 + (x^1)^2 + (x^2)^2)^2}$$

satisfies $ds^2 > 0$ and thus gives rise to a Riemannian metric on $\mathbb{R}^2$. Show that the schematic form of this line element is invariant under the change of coordinates $y^1 = \frac{1}{x^1}$, $y^2 = \frac{x^2}{x^1}$. [INTERMEDIATE]

(b) Show that the above defines a Riemannian metric on the entire $\mathbb{R}P^2$. [INTERMEDIATE]

(10) (a) Let $u, v \in \mathbb{C}$. Show that

$$ds^2 = \frac{(1+|v|^2)|du|^2 - 2\mathrm{Re}[(u\bar{v})(dv d\bar{u})] + (1+|u|^2)|dv|^2}{(1+|u|^2+|v|^2)^2}$$

satisfies $ds^2 > 0$ and thus gives rise to a 4-dimensional line element. [ADVANCED]

(b) Show that the above defines a Riemannian metric on the entire $\mathbb{C}P^2$. [ADVANCED]

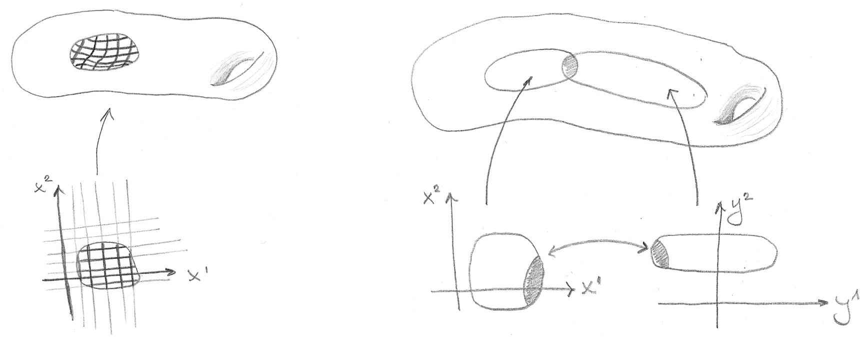
2. The Framework of Riemannian Geometry

Introduction. Gauss' book *"Disquisitiones generales circa superficies curvas"* ("General Investigations of Curved Surfaces") from 1828 investigates surfaces embedded in the Euclidean space $\mathbb{R}^3$, their associated line elements ds^2 and the concept of curvature. This work includes *Theorema Egregium* which identifies a formula (with more than a dozen terms) for the curvature based solely on the (derivatives of) line element coefficients. Thus, the concept of curvature does not depend as much on the ambient space, as it does on the expression for the line element. In other words, Gauss' work shifted the emphasis from the ambient space to the line element of the surface itself. In his Habilitation lecture Riemann discussed how the notions from Gauss' work, such as curvature, can be applied to *the space itself.* The idea is that, prima facie, space can be almost anything that is coordinatizable and line elements can be any expressions of the form

$$ds^2 = \sum g_{ij}(x^1, x^2, \ldots\ldots, x^n) dx^i dx^j,$$

as long as we have $ds^2 > 0$. (Actually, we can always in addition assume that $g_{ij} = g_{ji}$ due to $dx^i dx^j = dx^j dx^i$.) This line element, it turns out, determines the curvature of the space. We begin today's lecture by formulating this basic framework of Riemannian geometry somewhat more precisely, and end the lecture by investigating the concepts of distance and volume in such a framework.

The concept of an n-dimensional manifold. Riemannian geometry studies n-DIMENSIONAL MANIFOLDS. In a nutshell, the term n-dimensional manifold signifies a set M which is coordinatized (possibly with a big atlas of maps/charts) so that locations of points represented in the same map depend on n parameters $x^1, x^2, \ldots\ldots, x^n$. Certain domains within M may be covered with a number of different charts. In such situations we will always assume that the transition maps $\Phi_x^y : (x^1, \ldots, x^n) \mapsto (y^1, \ldots, y^n)$ and $\Phi_y^x : (y^1, \ldots, y^n) \mapsto (x^1, \ldots, x^n)$ are differentiable bijections between the corresponding domains in $\mathbb{R}^n$.

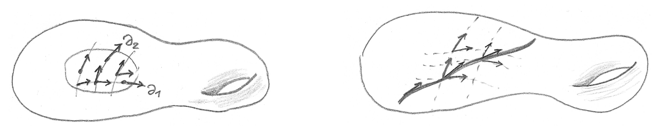


Disclaimer No.1. It is time to make a **major disclaimer**. To comply with the modern standards of rigor in mathematics, one first has to start by addressing the topological underpinnings of the manipulations done above. (E.g. how does one know that each map in an atlas requires the same number of parameters?) These lectures are meant to provide an introduction to the subject and are not designed to serve as a replacement for courses in differential and Riemannian geometry which address the very foundations of the subject. In particular, these lectures do not even attempt to answer questions such as the one raised above. Those interested in a more careful treatment of differential and Riemannian geometry are encouraged to consult resources such as the textbooks of John M. Lee, *Introduction to Topological Manifolds* and *Introduction to Smooth Manifolds*.

The concept of tangent vector (fields). A curve γ on a manifold can in coordinates be expressed as $\big(x^1(t), x^2(t), ..., x^n(t)\big)$, $a \leq t \leq b$. The curves

$$x^i(t) = \begin{cases} t, & \text{if } \ i = j \\ \text{const}, & \text{if } \ i \neq j \end{cases}$$

define coordinate grid lines. The idea rough here is that at each point P within the scope of x-coordinates the coordinate vector fields $\partial_i = \partial_{x^i}$ provide a basis for the vector space of directions based at P, also known as *the tangent space* at P. Thus each vector field V on M has coordinate representations of the form $\sum_i V^i \partial_i$; the functions V^i are often called *the components of* V. In particular, tangent vector field $\dot\gamma$ to a curve γ described above can be written as $\sum_i (\frac{\partial x^i}{\partial t})\partial_i = \sum_i (\partial_t x^i)\,\partial_i$.



Of course, different coordinate patches yield different coordinate vector fields. Considering the parametric equations for the x-grid lines in y-coordinates yields

$$\partial_{x^i} = \sum_j (\tfrac{\partial y^j}{\partial x^i})\partial_{y^j}.$$

More precisely, if $\Phi_x^y : (x^1, ..., x^n) \mapsto (y^1, ..., y^n)$ is a transition map then its linearization / Jacobi matrix $d\Phi_x^y$ (viewed as a linear transformation) provides a method of identifying tangent spaces in different coordinates.

$$d\Phi_x^y : \partial_i \mapsto \sum_j (\frac{\partial y^j}{\partial x^i})\partial_j.$$

Disclaimer No.2. The rough ideas described above are fairly intuitive in the cases when the manifold is manifestly embedded as a surface within some $\mathbb{R}^n$. However, when that is not the case the concepts of tangent space and vector (fields), are considerably more abstract and, as a result, somewhat difficult to define. One of the many approaches has us establish an equivalence relation between tangent spaces in different coordinates (such as $d\Phi_x^y$ above); the tangent space arises as the corresponding set of equivalence classes. In the interests of brevity we will shy away from the rigorous treatment of the subject. Those interested in details should consult textbooks such as those of John M. Lee referred to earlier.

Riemannian metric. From a more contemporary standpoint line elements are typically viewed as inner-product operations on the set of tangent vectors based at a particular point or, more globally, on the set of vector fields. Specifically, let

$$g(V, W) := \textstyle\sum_{ij} g_{ij} V^i W^j$$

where $V = \sum V^i \partial_i$ and $W = \sum W^j \partial_j$ are two tangent vector (fields). As discussed above we may, without loss of generality, assume $g_{ij} = g_{ji}$. Under this assumption g is symmetric in V and W. Also note that g is bilinear:

$$g(\alpha V_1 + \beta V_2, W) = \alpha g(V_1, W) + \beta g(V_2, W).$$

In fact, the above holds even when α and β are *functions* (i.e. scalars which change from point to point; *scalar fields*). Finally, the consideration that $ds^2 > 0$ gives us $g(V, V) > 0$ for all non-zero V. In other words, we have that g is a positive definite inner-product.

The operation g described above is called a *Riemannian metric*. Riemannian metric can be viewed as an operation on tangent vectors (the outcome of which is a number), or it can be viewed as an operation on vector fields (the outcome of which is a function / scalar field). The latter is not only bilinear with respect to constants, but also with respect to functions / scalar fields. In a sense one can think of Riemannian metric as an inner-product on $\mathbb{R}^n$ which changes from basepoint to basepoint.

Notation. Traditional notation for an inner-product, $\langle ., . \rangle$, is often used in place of and / or in parallel with g. (Occasionally, one even sees $\langle ., . \rangle_g$.) We use $|V|_g$ to denote the magnitude $\sqrt{g(V, V)}$ of V with respect to g. The quantities

$$g_{ij} = g(\partial_i, \partial_j)$$

are called the *metric components*. One often thinks of the metric components as entries of a symmetric $n \times n$ matrix. The condition that $g(V, V) > 0$ for nonzero V implies that the eigenvalues of the matrix $[g_{ij}]$ are all positive. In particular, $[g_{ij}]$ is invertible. The entries of its inverse are denoted by g^{ij}: $[g_{ij}]^{-1} = [g^{ij}]$. Note that by the definition of the inverse we have:

$$\sum_k g_{ik} g^{kj} = \begin{cases} 1, & \text{if } \ i = j, \\ 0, & \text{otherwise.} \end{cases}$$

Length and distance measurements. Riemannian metrics g and line elements ds^2 are related as follows: The (square of the) magnitude of a tangent vector to $(x^1(t), ..., x^n(t))$ is given by

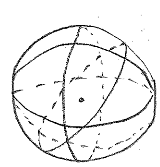
$$g(\sum_i (\partial_t x^i)\, \partial_i, \sum_j (\partial_t x^j)\, \partial_j) = \sum_{ij} g_{ij} (\partial_t x^i)(\partial_t x^j) = (\tfrac{ds}{dt})^2.$$

In other words, to measure the length of a curve γ one needs to integrate the magnitude of the tangent vector field $|\dot\gamma|_g$ with respect to the inner-product g:

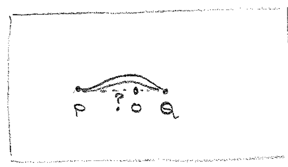
$$L(\gamma) = \int_a^b |\dot\gamma|_g \, dt.$$

One can show that the length of a curve does not depend on its parametrization. (See exercise (2).)

Equipped with a way of measuring lengths of paths, we are now able to determine distance between points. Naturally, *distance* between two points is the length of the shortest path between the two points, should such a thing exist. In such a case determining distance between points is basically an optimization problem which – like most optimization problems – can be solved by methods of calculus. We cannot go into the details of this now, but we will dedicate the entire next lecture to deriving and analyzing the ODE's which solve our optimization problem. At this stage we would like to at least announce the solution of the shortest path problem on a two dimensional sphere. A pair of points on the surface of the sphere is always contained in *great circle*, a circle which is the intersection of the sphere with a plane containing the center of the sphere. As we shall see in the next lecture, great circles serve as length minimizers on the sphere. In more practical terms, this is the reason why many intercontinental flights (e.g. between North America and Europe) take a route which goes over Artic.

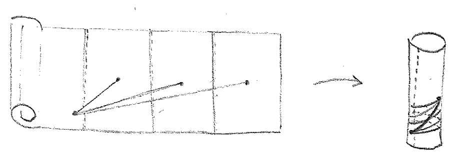


Note that it is not at all clear if for a given pair of points there even exists the path of shortest length joining them. Obviously, we want our manifolds to be *path connected* i.e. such that any pair of points can be joined by a path in the first place. Even then there are some easy examples where we have non-existence of shortest paths. Consider the Euclidean plane with a point O removed and consider two points P and Q which are collinear with and located on opposite sides of the excluded point O. Clearly, the shortest path between P and Q with respect to the Euclidean metric cannot be found within the plane while O is removed. Thus, in general, we cannot expect that there will always be the path of shortest length between two given points. For this reason, we *define the distance* $d(A, B)$ between two points A and B as the *infimum*, i.e. greatest lower bound, of the set of lengths of paths between A and B. In the our example the distance between P and Q would still have the same value as in the standard Euclidean plane.



A somewhat obvious remedy to the problem described above is to only consider (path connected) manifolds without punctures, holes, or otherwise missing points. Readers familiar with elementary analysis probably recognize that the desired property here is that of *completeness*. The technical meaning of this phrase will be addressed in more detail in the following lecture.

One should also notice that a path of shortest length between two points need not be unique. This is blatantly obvious on the example of a sphere where there are infinitely many paths of shortest length joining the North and South Pole: each and every meridian serves as a length minimizer. To make things even more complicated there may be many local length minimizers, in the sense that for a given pair of points there may exist paths which only have the shortest length amongst nearby paths. A good example of this comes from the circular cylinder in $\mathbb{R}^3$. As it can be obtained by a distance preserving folding of the Euclidean plane, the shortest paths between points come form straight line segments in the Euclidean plane. However, due to the folding effects there are many such line segments. As a result, each of the curves on the following diagram minimizes length amongst nearby curves joining the same points.



Volume measurements. A Riemannian metric also determines the *volume element*, the integral of which measures the volume of the domain of integration. To understand this let us revisit how we compute volumes of parallelotopes in Euclidean $\mathbb{R}^n$. A simple induction argument on n shows that the square of the volume of the parallelotope spanned by the vectors $\mathbf{a}_1, \ldots, \mathbf{a}_n$ is equal to the determinant of the matrix of Euclidean inner-products:

$$\mathrm{vol}^2 = \det([\langle \mathbf{a}_i, \mathbf{a}_j \rangle]).$$

Indeed, by performing column operations we can make the first column of the matrix on the right become $(\langle \mathbf{v}, \mathbf{a}_1 \rangle, 0, \ldots, 0)^T$ where $\mathbf{v}$ denotes the orthogonal component of $\mathbf{a}_1$ with respect to the subspace spanned by $\mathbf{a}_2, \ldots, \mathbf{a}_n$. Since $\langle \mathbf{v}, \mathbf{a}_1 \rangle = |\mathbf{v}|^2$, the inductive hypothesis gives us that $\det([\langle \mathbf{a}_i, \mathbf{a}_j \rangle]$ is equal to the product of $|\mathbf{v}|^2$ and the volume of the (lower dimensional) parallelotope spanned by $\mathbf{a}_2, \ldots, \mathbf{a}_n$. By definition of volume this product is the volume of the parallelotope spanned by $\mathbf{a}_1, \ldots, \mathbf{a}_n$.

The above identity is used to define volume of a parallelotope in general, with respect *any* inner-product $\langle ., . \rangle$ in $\mathbb{R}^n$. It can be checked that volume defined in this manner has the expected features: it is positive unless $\mathbf{a}_1, \ldots, \mathbf{a}_n$ are linearly dependent, it is additive, and it is preserved by isometries of $\langle ., . \rangle$. (Compare with exercise 5.) According to this definition, the square of the volume of the parallelotope spanned by $\partial_1, \ldots, \partial_n$ in the tangent space equipped with the Riemannian metric g, is simply

$$|g| := \det[g_{ij}].$$

Scaling to the infinitesimal level gives us the following formula for the volume element:

$$\mathrm{dvol} = \sqrt{|g|}\, dx^1 \ldots dx^n.$$

Finally, one checks (see exercise 8) that this expression for the volume element does not depend on the choice of coordinates.

Example. As an illustration, we compute the volume[7] of the standard unit three-dimensional sphere S^3. In stereographic coordinates we have

$$ds^2 = \tfrac{4}{(1+(x^1)^2+(x^2)^2+(x^3)^2)^2}\left[(dx^1)^2 + (dx^2)^2 + (dx^3)^2\right],$$

which can also be expressed as

$$ds^2_{S^3} = \tfrac{4}{(1+r^2)^2}\left[(dr^2) + r^2 ds^2_{S^2}\right],$$

where $0 < r < \infty$ and where $ds^2_{S^2}$ denotes the standard line element on S^2 in spherical coordinates. From here we get

$$\mathrm{dvol}_{S^3} = \tfrac{8r^2}{(1+r^2)^3}\mathrm{dvol}_{S^2} \quad \text{and} \quad \mathrm{Vol}(S^3) = \left(\int_0^\infty \frac{8r^2}{(1+r^2)^3}\, dr\right)\mathrm{Vol}(S^2).$$

The integral on the right can be easily computed using the trigonometric substitution $r = \tan s$; its value comes out to be $\frac{\pi}{2}$. Thus, the volume of the standard unit three-dimensional sphere S^3 is:

$$\mathrm{Vol}(S^3) = \tfrac{\pi}{2}\mathrm{Vol}(S^2) = 2\pi^2.$$

In general, *compact manifolds* have finite volume. We often think of compact manifolds as being closed and bounded (e.g. closed and bounded surfaces in $\mathbb{R}^3$), although this is not a part of their definition. Technically, we say that a Riemannian manifold (M, g) is compact if every bounded sequence[8] of points $\{P_m\}$ has a convergent subsequence $\{P_{m_i}\}$. One common feature that all compact manifolds share is that they can always be covered with finitely many coordinate patches. Another property that all compact manifolds have in common is that all continuous real-functions defined on compact manifolds reach their maximum and minimum. Proofs of these claims belong to elementary analysis and will not be done here.

Riemannian geometry and geometric analysis. *Riemannian geometry* is the general study of *Riemannian manifolds*, i.e. manifolds equipped with a Riemannian metric; as such does not refer to any specific geometry. One of the central notions in Riemannian geometry is that of *curvature*; many a theorem in Riemannian geometry addresses the kind of consequences that restrictions on the curvature (e.g. curvature bounds) have on geometric quantities such as distances and volumes or on topological qualities such as homeomorphism type or the fundamental group.

[7]Note: this does not refer to the volume enclosed by the sphere within $\mathbb{R}^4$ but to the surface "area".

[8]The phrase "bounded sequence" stands for a sequence for which there is some $C > 0$ and some $P \in M$ such that $d(P_m, P) \leq C$ for all m.

Imposing conditions on curvature often amounts to (a system of) differential equations on metric (components). Analysis of such differential equations is the subject of *geometric analysis*.

Exercises for Lecture 2.

(1) Consider the the line element $ds^2 = \frac{dx^2+dy^2}{y^2}$ on the upper half-plane $y > 0$. Use this line element to compute the lengths of the following paths joining $(-1, 1)$ and $(1, 1)$.

  (a) Let c_1 be the "straight" line segment $x(t) = t,\ \ y(t) = 1,\ \ -1 \leq t \leq 1$.

  (b) Let c_2 be the "circular arc" $x(t) = \sqrt{2}\cos(t),\quad y(t) = \sqrt{2}\sin(t)$ with $\frac{\pi}{4} \leq t \leq \frac{3\pi}{4}$.

  Which of the two paths is shorter? [BASIC]

(2) (a) Show that the length of a path does not change under smooth reparametrizations. (You may assume that the path is completely contained in a single coordinate patch.) [BASIC]

  (b) To avoid pathological situations one often assumes that curves γ are parametrized so that the tangent vector field $\dot{\gamma}$ is no-where vanishing. Show that under this assumption a curve on a Riemannian manifold can be smoothly reparametrized so that its tangent vector be of constant length. [INTERMEDIATE]

(3) This problem investigates how a change of coordinates

$$\Phi_x^y : (x^1, x^2, ..., x^n) \mapsto (y^1, y^2, ..., y^n)$$

  effects the coordinate expressions for vector fields and the components of a Riemannian metric. [INTERMEDIATE]

  (a) Let $\{\partial_i\}$ and $\{\partial_{i'}\}$ be the coordinate vector fields corresponding to the x and the y coordinates, respectively. Consider the decompositions $V = \sum_i V^i \partial_i = \sum_{i'} V^{i'} \partial_{i'}$ of a vector field V. Find a formula relating the functions V^i to the functions $V^{i'}$.

  (b) Let g_{ij} and $g_{i'j'}$ be the metric components with respect to the x and y coordinates, respectively. Find a formula relating $g_{i'j'}$ to g_{ij}.

  (c) Let a path γ be contained within the scope of both the x and the y coordinates. Show that the value of the length of γ, $L(\gamma)$, does not depend on the choice of the coordinate system.

(4) (a) The *diameter* of a compact Riemannian manifold (M, g) is the maximum value of $d(P, Q)$ for $P, Q \in M$. What is the diameter of the standard unit sphere S^2? What about S^n? [BASIC]

  (b) We define the circle of radius r centered at the point P of a 2-dimensional Riemannian manifold (M, g) as the set of all points Q with $d(P, Q) = r$. Describe and compute the circumference of the circle of radius r centered at the North Pole of the standard unit sphere S^2. [BASIC, EXPECTED]

(5) Let (M, g) be a compact Riemannian manifold, and let g_1 be another Riemannian metric on M. Show that there is a constant C such that for all points P and Q on M we have

$$C^{-1} d_g(P, Q) \leq d_{g_1}(P, Q) \leq C d_g(P, Q);$$

here the distances are computed with respect to metrics indicated in the subscripts. [ADVANCED]

(6) Let $\langle\langle ., . \rangle\rangle$ be some (not necessarily Euclidean) inner-product operation on $\mathbb{R}^n$. Define $a_{ij} := \langle\langle \partial_i, \partial_j \rangle\rangle$ and let A denote the corresponding matrix $[a_{ij}]$. [BASIC]

(a) Show that $\langle\langle \mathbf{v}, \mathbf{w} \rangle\rangle = \langle A\mathbf{v}, \mathbf{w} \rangle$, where $\langle ., . \rangle$ denotes the standard Euclidean inner-product.

(b) By a theorem from linear algebra, the symmetric A must have real eigenvalues. Show that the eigenvalues of A and the determinant of A are necessarily positive.

(c) By a theorem from linear algebra, there is an orthonormal basis for the Euclidean $\mathbb{R}^n$ consisting of eigenvectors of the symmetric A. Show that the same basis is orthogonal with respect to $\langle\langle ., . \rangle\rangle$.

(d) Consider the rectangular box spanned by the orthonormal set of eigenvectors of A in the Euclidean $\mathbb{R}^n$. What do you think should be the volume of this box from the perspective of $\langle\langle ., . \rangle\rangle$? How does your formula relate to our discussion of the formula for the volume element?

(7) Let $\langle\langle ., . \rangle\rangle$ be some (not necessarily Euclidean) inner-product operation on $\mathbb{R}^n$, let the matrix A be as above, and let L denote a $n \times n$ matrix and its corresponding linear transformation $L : \mathbb{R}^n \to \mathbb{R}^n$. We say that L is an *isometry* of $\langle\langle ., . \rangle\rangle$ if

$$\langle\langle L\mathbf{v}, L\mathbf{w} \rangle\rangle = \langle\langle \mathbf{v}, \mathbf{w} \rangle\rangle.$$

(a) Show that L preserves distances with respect to $\langle\langle ., . \rangle\rangle$. [BASIC]

(b) Show that $L^T A L = A$. [INTERMEDIATE]

(c) The above formula should be highly related to that in exercise 3b. Explain why this is not a coincidence. [INTERMEDIATE]

(8) (a) Let A be an invertible symmetric $n \times n$ matrix. Identify the set of $n \times n$ matrices with real entries with $\mathbb{R}^{n^2}$ and consider the map $\Psi : \mathbb{R}^{n^2} \to \mathbb{R}^{n^2}$ defined by

$$\Psi : X \mapsto X^T A X.$$

Compute the rank of the linearization / Jacobi matrix $d\Psi$ at $n \times n$ matrices L for which $L^T A L = A$. [ADVANCED]

(b) Review the statement of the Implicit and the Inverse function theorems. Then use exercise 7b to show that the set of isometries of an inner-product on $\mathbb{R}^n$ is itself a manifold. What is the dimension of this manifold? [ADVANCED]

(c) Describe the tangent vectors to this manifold. [ADVANCED]

(9) Show that the expression $\sqrt{|g|}\, dx^1 ... dx^n$ is invariant under the change of coordinates. [INTERMEDIATE]

(10) Compute:

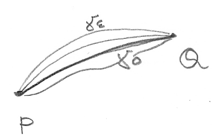
(a) The volume of S^4. [INTERMEDIATE]

(b) The volume of S^n. [ADVANCED]

3. Geodesics

Introduction. In the last lecture we saw that the distance between two points of a Riemannian manifold (M, g) is the infimum of the lengths of of paths joining the two points. In many circumstances this infimum is reached by a length minimizing path. The goal of today's lecture is to use methods of calculus to find such length minimizers. We end the lecture with an investigation of certain examples, one of which being of great historical significance.

The length functional. Suppose there is a path of minimum length γ_0 between two points P and Q, which for simplicity we assume lies within a single coordinate patch. Let γ_ε be a family of paths, indexed by a small parameter ε, between P and Q. One should think of this family as giving us *perturbations* or slight modifications of the path γ_0; in other words, $\gamma_\varepsilon \to \gamma_0$ as $\varepsilon \to 0$.



In coordinates the family γ_ε looks like

$$\gamma(\varepsilon, t) = \left(x^1(\varepsilon, t), \ldots, x^n(\varepsilon, t)\right).$$

Note that for a given geometric family of curves there are infinitely many representations of such a form: intuitively speaking they correspond to different speeds of traverse i.e different $|\dot\gamma(\varepsilon, t)|_g$. This kind of ambiguity can be resolved by requiring that $a \le t \le b$ for all ε and that all $|\dot\gamma(\varepsilon, t)|_g$ be constant:

$$|\dot\gamma(\varepsilon, t)|_g = \frac{L(\gamma_\varepsilon)}{b - a};$$

the reader is encouraged to compare this to the exercise 2b from the previous lecture. We make this assumption from now on. As γ_ε all join P to Q we know that $\gamma(\varepsilon, a)$ and $\gamma(\varepsilon, b)$ are each constant in ε. Therefore, we have

$$\frac{dx^i}{d\varepsilon}\Big|_{t=a} = \frac{dx^i}{d\varepsilon}\Big|_{t=b} = 0$$

for all i and at each ε.

The length $L(\gamma_\varepsilon)$ of each path γ_ε is given in coordinates by

$$L(\gamma_\varepsilon) = \int_a^b \left(\sum_{ij} g_{ij}(\gamma(\varepsilon, t)) \frac{dx^i}{dt}(\gamma(\varepsilon, t)) \frac{dx^j}{dt}(\gamma(\varepsilon, t)) \right)^{1/2} dt$$

If γ_0 indeed is the path of shortest length between the points P and Q, then the function $\varepsilon \mapsto L(\gamma_\varepsilon)$ reaches its global minimum at $\varepsilon = 0$. In particular, we must have

$$\frac{dL(\gamma_\varepsilon)}{d\varepsilon}\Big|_{\varepsilon=0} = 0.$$

In fact, if γ_0 is the shortest of all paths between P and Q then the above has to hold for _all_ possible perturbations γ_ε of γ_0.

The question arises: what is the best way of computing this derivative? We can try differentiating under the integral sign, but the presence of the square root promises computational complications. Instead, we will consider a related function which amongst its critical "points" includes the global minimum of the length functional $\gamma \mapsto L(\gamma)$.

(Optimizing the) energy functional. The *energy* of a curve $\gamma(t)$, $a \le t \le b$, is defined by

$$E(\gamma) = \frac{1}{2}\int_a^b |\dot\gamma|_g^2\, dt.$$

A big warning is in order: the value $E(\gamma)$ is not a geometric feature of γ as it depends on its parametrization! (Compare with exercise 2 following this lecture.) However, when $|\dot\gamma|_g$ is constant (as is our underlying assumption today) then clearly

$$E(\gamma) = \tfrac{L(\gamma)^2}{2(b-a)}.$$

In particular, minimizing length amongst paths of constant speed is equivalent to minimizing the corresponding energies. Going back to our family of curves γ_ε and its coordinate representations $\gamma(\varepsilon, t)$, we see that we should try to minimize

$$E(\gamma_\varepsilon) = \int_a^b \sum_{ij} g_{ij}(\gamma(\varepsilon,t))\frac{dx^i}{dt}(\gamma(\varepsilon,t))\frac{dx^j}{dt}(\gamma(\varepsilon,t))\, dt.$$

This can be done by imposing that the equation

$$\frac{dE(\gamma_\varepsilon)}{d\varepsilon}\Big|_{\varepsilon=0} = 0$$

hold for _all_ possible perturbations γ_ε of γ_0.

We proceed by investigating what this requirement has to say about γ_0. The computation we are about to do is highly related to the derivation of the Euler-Lagrange equations in the field of math called *the calculus of variations*. To simplify the notation *we adopt the notational conventional according to which* _*repeated indices*_ *are being* _*summed over*_.

Differentiating $E(\gamma_\varepsilon)$ under the integral sign produces:

$$\int_a^b \partial_l g_{ij}\frac{dx^l}{d\varepsilon}\frac{dx^i}{dt}\frac{dx^i}{dt}\, dt + \int_a^b g_{ij}\frac{d^2x^i}{dtd\varepsilon}\frac{dx^j}{dt}\, dt + \int_a^b g_{ij}\frac{dx^i}{dt}\frac{d^2x^j}{dtd\varepsilon}\, dt = 0,$$

where all the integrands are evaluated along γ_0. To extract a more useful information we apply integration by parts to the last two terms of this equation. For example, the middle term from the above can be manipulated as follows:

$$\begin{aligned}\int_a^b g_{ij}\frac{d^2x^i}{dt\,d\varepsilon}\frac{dx^j}{dt}\,dt =& g_{ij}\frac{dx^i}{d\varepsilon}\frac{dx^j}{dt}\Big|_{t=a}^{t=b} - \int_a^b \frac{d}{dt}\left(g_{ij}\frac{dx^j}{dt}\right)\frac{dx^i}{d\varepsilon}\,dt \\ =& -\int_a^b \partial_l g_{ij}\frac{dx^l}{dt}\frac{dx^j}{dt}\frac{dx^i}{d\varepsilon}\,dt - \int_a^b g_{ij}\frac{d^2x^j}{dt^2}\frac{dx^i}{d\varepsilon}\,dt;\end{aligned}$$

note that we used the fact that $\frac{dx^i}{d\varepsilon}$ vanish at $t = a$ and $t = b$. After gathering of like terms and relabeling indices we obtain

$$\sum_l \int_a^b \left(\sum_j g_{lj}\frac{d^2x^j}{dt^2} + \sum_{ij}\tfrac{1}{2}(\partial_i g_{jl} + \partial_j g_{li} - \partial_l g_{ij})\frac{dx^i}{dt}\frac{dx^j}{dt}\right)\frac{dx^l}{d\varepsilon}\,dt = 0.$$

Here is a crucial observation. The last equality is to hold for all perturbations γ_ε of γ_0. Thus, we have a great deal of freedom in how we choose our $\frac{dx^l}{d\varepsilon}$ – the only restriction is that these derivatives vanish at (ε, a) and (ε, b). In view of this (and exercise 3 following this lecture) we see that

$$\sum_j g_{lj}\frac{d^2x^j}{dt^2} + \sum_{ij}\tfrac{1}{2}(\partial_i g_{jl} + \partial_j g_{li} - \partial_l g_{ij})\frac{dx^i}{dt}\frac{dx^j}{dt} = 0$$

for each individual l. In other words, we have arrived at a system of ODEs which the coordinate expressions $(x^1(t), \dots, x^n(t))$ for any length minimizer have to satisfy! In fact, coordinate expressions for any of the critical "points" (critical parametrized paths, really) of the energy functional must solve the above system of ODEs.

Geodesics and the geodesic equations. In general, critical parametrized paths of the energy functional are called *geodesics*. They correspond to solutions of the above system of ODEs. Note that since energy functional is parametrization dependent the concept of a geodesic is as well. In particular, geodesics are necessarily parametrized so that their tangent vector fields have constant length. (Compare with exercises (4) and (5a) below.)

The system of ODEs for geodesics is most commonly expressed in a slightly simpler, equivalent form. We introduce the symbols

$$\Gamma^k_{ij} := \tfrac{1}{2}g^{kl}(\partial_i g_{jl} + \partial_j g_{li} - \partial_l g_{ij});$$

these symbols are known as *the Christoffel*[9] *symbols*. Multiplication by the inverse matrix $[g^{kl}]$ produces the *the geodesic equations*

$$\frac{d^2x^k}{dt^2} + \sum_{ij}\Gamma^k_{ij}\frac{dx^i}{dt}\frac{dx^j}{dt} = 0, \quad 1 \le k \le n.$$

Example of Euclidean $\mathbb{R}^n$*:* As a trivial example, consider the geodesic equations in the Euclidean plane, with respect to Cartesian coordinates. The metric components are constant, and as a result the Christoffel symbols vanish. The geodesic equations, therefore, boil down to saying that the second derivatives of all $x^k(t)$ vanish. In other words, the functions $x^k(t)$ must all be linear and the geometric curve described by them must be a (subset of) a line. This is to be expected as line segments are length minimizers in the Euclidean plane.

[9]They are named in honor of the German mathematician Elwin Bruno Christoffel.

Before we move on to more interesting examples we need to address existence and other properties (e.g uniqueness, time of existence, etc) of solutions of geodesic equations. This can be done through basic Existence, Uniqueness and Stability Theorems for initial value problems. One such theorem is stated below.

THEOREM 1. *Let* $F : \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}^n$ *be a smooth function. Then for all* $(\mathbf{y}_*, \mathbf{v}_*) \in \mathbb{R}^n \times \mathbb{R}^n$ *there exist:*

- *a neighborhood* $\mathcal{Y} \times \mathcal{V}$ *of* $(\mathbf{y}_*, \mathbf{v}_*)$ *in* $\mathbb{R}^n \times \mathbb{R}^n$*,*
- *a number* $\varepsilon > 0$ *and*
- *a smooth map* $\mathbf{Y} : \mathcal{Y} \times \mathcal{V} \times (-\varepsilon, \varepsilon) \to \mathbb{R}^n$

such that for each $(\mathbf{y}_0, \mathbf{v}_0) \in \mathcal{Y} \times \mathcal{V}$ *the function*

$$\mathbf{y} : t \mapsto \mathbf{Y}(\mathbf{y}_0, \mathbf{v}_0, t)$$

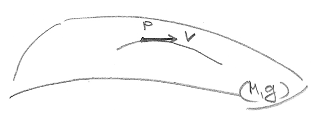
is the unique solution of the initial value problem

$$\begin{cases} \frac{d^2\mathbf{y}}{dt^2} = F(\mathbf{y}, \frac{\mathbf{y}}{dt}) \\ \mathbf{y}(0) = \mathbf{y}_0 \\ \frac{d\mathbf{v}}{dt}(0) = \mathbf{v}_0. \end{cases} \tag{1}$$

Thus, roughly speaking, solutions of initial value problems such as (1) exist, are unique, and vary smoothly with the initial data. It is important to notice that no claim is being made about the long-term existence of solutions; the only claim that is being made is that the solutions are defined on some small interval $(-\varepsilon, \varepsilon)$. When applied to geodesic equations the above theorem gives us the following result. (See also exercise 5 for further consequences.)

THEOREM 2. *Let* (M, g) *be a Riemannian manifold. For each* $P \in M$ *and each tangent vector* V *there is an* $\varepsilon > 0$ *and a unique geodesic* $\gamma(t)$ *with* $-\varepsilon < t < \varepsilon$ *which passes through* P *with the velocity vector* V*:*

$$\gamma(0) = P, \quad \dot{\gamma}(0) = V.$$



It is important to note that geodesics $\gamma(t)$ need not be defined for all $t \in \mathbb{R}$. We say that a manifold is *geodesically complete* if every geodesic (can be extended to be) defined on all $\mathbb{R}$. Geodesically complete are the subject of the HOPF-RINOW THEOREM[10]. We only offer the statement of the theorem here; the proof is too difficult for us at this point.

THEOREM 3. *If a path connected Riemannian manifold* (M, g) *is geodesically complete, then any two points of* M *can be joined by a length minimizing geodesic.*

Remark. For the benefit of the readers familiar with elementary analysis we point out to a consequence of the Hopf-Rinow theorem, which states that a Riemannian manifold (M, g) is geodesically complete if and only if it is complete as a metric space. For this reason it is appropriate to use the phrase *complete manifold*

[10]The theorem is attributed to Heinz Hopf and his student Willi Rinow.

whenever the manifold is geodesically complete. *Unless specifically stated otherwise,* we will always assume our manifolds are complete and path connected. It is useful to know that every compact manifold without boundary is automatically complete.

Examples.

Example of the sphere $(x^1)^2 + \cdots + (x^n)^2 = 1$ *in* $\mathbb{R}^n$*.* We now show that great circles are geodesics. Due to symmetries of the sphere it suffices to argue that the great circle in the subspace $x^3 = \cdots = x^n = 0$ is a geodesic through the point $P(1, 0, \ldots, 0) \in S^{n-1}$. This can be done by verifying that this parametrization of the circle solves the geodesic equations (see also exercise 6 following this lecture). An easier way to this is the following. The map

$$\mathcal{I} : (x^1, x^2, x^3, \ldots, x^n) \mapsto (x^1, x^2, -x^3, \ldots, -x^n)$$

preserves distances in $\mathbb{R}^n$ and, consequently, the metric on S^{n-1}. (One says that $\mathcal{I}$ is an *isometry.*) Consequently, $\mathcal{I}$ has to take geodesics to geodesics. Since both P and the tangent vector $V = (0, 1, 0 \ldots, 0)^T$ is mapped to itself under $\mathcal{I}$, the *unique* geodesic passing through P with velocity vector V must also be mapped to itself under $\mathcal{I}$. It follows that this geodesic is (contained within) the great circle in the subspace $x^3 = \cdots = x^n = 0$.

Examples of hyperbolic spaces. This example examines the n-dimensional upper half-space $\mathbb{H}^n$, given by $x^n > 0$, with the Riemannian metric

$$g = (x^n)^{-2}(d(x^1)^2 + \cdots + d(x^n)^2).$$

On an intuitive qualitative level, the factor of $(x^n)^{-2}$ enlarges the distances near the boundary plane $x^n = 0$ and diminishes the distances far away from $x^n = 0$. In some sense, there is a whole lot more "space" in $\mathbb{H}^n$ near $x^n = 0$ than it might appear to our Euclidean eyes.

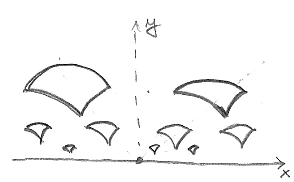
Here are some more concrete observations which can help us visualize this new geometry. For example, the dilation of the form

$$\mathcal{I} : (x^1, \ldots, x^n) \mapsto (\lambda x^1, \ldots, \lambda x^n), \quad \lambda > 0$$

preserves the line element g:

$$(\lambda x^n)^{-2}(d(\lambda x^1)^2 + \cdots + d(\lambda x^n)^2) = (x^n)^{-2}(d(x^1)^2 + \cdots + d(x^n)^2).$$

Thus, the map $\mathcal{I}$ preserves distances and should be thought of as an *isometry.* In particular, it maps shapes to congruent shapes. The same is true for "horizontal" translations, $(x^1, ..., x^{n-1}, x^n) \mapsto (x^1 + a^1, ..., x^{n-1} + a^{n-1}, x^n)$, boundary preserving reflections $(..., x^i, ..., x^n) \mapsto (..., -x^i, ..., x^n)$ and boundary preserving rotations $(..., x^i, ..., x^j, ...) \mapsto (..., \cos\alpha\, x^i + \sin\alpha\, x^j, ..., -\sin\alpha\, x^i + \cos\alpha\, x^j, ...)$. Pictorially, this is to say that all of the following shapes are congruent.



The question we would like to address right now is that of geodesics on $\mathbb{H}^n$. Utilizing the above symmetries it suffices to analyze the geodesic problem in $\mathbb{H}^2$:

the upper half-plane $y > 0$ with the metric

$$g = y^{-2}(dx^2 + dy^2).$$

To find the geodesic equation we first compute the Christoffel symbols. The only non-vanishing Christoffel symbols here are:

$$\Gamma^y_{xx} = \tfrac{1}{y}, \quad \Gamma^x_{xy} = \Gamma^x_{yx} = \Gamma^y_{yy} = -\tfrac{1}{y}.$$

Thus, the geodesic equations are:

$$\begin{cases} \frac{d^2x}{dt^2} - \frac{2}{y}\frac{dx}{dt}\frac{dy}{dt} = 0 \\ \frac{dy^2}{dt^2} + \frac{1}{y}\left(\frac{dx}{dt}\right)^2 - \frac{1}{y}\left(\frac{dy}{dt}\right)^2 = 0. \end{cases}$$

A consequence of the first equation is that $\frac{d}{dt}\left(y^{-2}\frac{dx}{dt}\right)$ vanishes. Consequently, we have $\frac{dx}{dt} = Cy^2$ for some constant C. We could insert this information into the second geodesic equation but we can also make use of the fact that geodesics have constant speed parametrizations; without loss of generality we may assume that geodesics are of unit speed. (Compare with exercises 4 and 5a below.) Therefore, we have

$$(\tfrac{dx}{dt})^2 + (\tfrac{dy}{dt})^2 = y^2 \quad \text{and} \quad \tfrac{dy}{dt} = y\sqrt{1-(Cy)^2}.$$

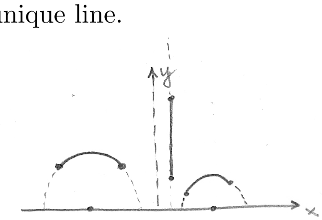
The latter can be solved by the method of separation of variables. In case when $C = 0$ we obtain $y = y_0 e^t$ and $x = x_0$. In other words, vertical rays are geodesics! In the case when $C \neq 0$ we get[11]

$$y = \frac{1}{C\cosh(t+t_0)} \quad \text{and consequently} \quad x = \frac{1}{C}\tanh(t+t_0) + D,$$

for some constants t_0 and D. To visualize these geodesics observe that

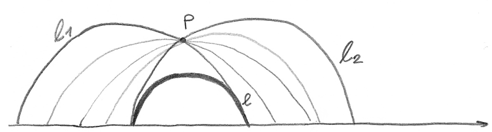
$$(x-D)^2 + y^2 = \left(\tfrac{1}{C}\right)^2;$$

thus these geodesics are upper semi-circles centered at the boundary $y = 0$. Overall, geodesics in $\mathbb{H}^2$ are either (contained in) vertical rays or upper semi-circles centered at the boundary $y = 0$. Although each pair of distinct points determines such a geodesic uniquely, it is not at this point clear that such a geodesic is necessarily length minimizing. However, our explicit formulae show that each geodesic on $\mathbb{H}^2$ is defined on all of $\mathbb{R}$, and that therefore $\mathbb{H}^2$ is geodesically complete. It now follows from the Hopf-Rinow Theorem each pair of points in $\mathbb{H}^2$ can be joined by a unique length minimizing geodesic: it is either a vertical line segment or an arc of an upper semi-circle centered at the boundary $y = 0$. The same holds in $\mathbb{H}^n$, due to symmetries of $\mathbb{H}^n$ discussed above. Since in Euclidean geometry lines are the length minimizers, we can think of the last statement as saying that in $\mathbb{H}^n$ each pair of points can be joined by a unique line.



[11]The hyperbolic trigonometric functions are defined through $\cosh(x) = \frac{1}{2}(e^x + e^{-x})$ and $\sinh(x) = \frac{1}{2}(e^x - e^{-x})$.

The broader context. The last example has a big historical significance. After nearly 2 000 years of attempts to clarify the status of Euclid's Fifth Postulate several mathematicians managed to develop a geometry based on the negations of Euclid's Fifth Postulate. As a reminder, Euclid's Fifth Postulate has a somewhat complicated statement which was quickly recognized to be equivalent to the statement[12] that through a point not on a line ℓ one can draw exactly one line which does not intersect ℓ. For centuries nobody questioned the validity of Euclid's Fifth Postulate but at the same time nobody managed to successfully prove the conjecture that this postulate is not logically independent from the remaining postulates. The intellectual climate of the late 18th and early 19th century surrounding mathematics was to a great extent influenced by the work of Prussian philosopher Immanuel Kant who in his CRITIQUE OF PURE REASON (1781) addressed the nature of mathematical statements and knowledge. This work was understood as saying that the laws of Euclidean space are already in human mind and that without these laws no consistent reasoning is possible. Thus, coming out with a geometrical theory based on a negation of Euclid's Fifth Postulate was highly controversial. The two mathematicians to do so (independently of one another) were the Russian mathematician Nikolai Ivanovich Lobachevsky (1826) and the Hungarian mathematician Janos Bolyai (1829). It took a long time for their work to reach public acceptance (e.g Gauss' lukewarm reception of Janos Bolyai's work in which he suggested that he already made the same discoveries but has been unwilling to share them publically). Eventually, models of axiomatic geometries developed by Bolyai and Lobachevsky were discovered and the logical validity of their *hyperbolic geometries* was established. The first (partial) model was due to the Italian mathematician Eugenio Beltrami (1868). The example we saw above is the *Poincaré upper half-space model* of hyperbolic geometry, in which the term "line" is interpreted as a geodesic. The following diagram shows that the opposite of Playfair's axiom holds:



$$\ell_1 \cap \ell = \ell_2 \cap \ell = \emptyset$$

This model is named in honor of the French mathematician Henri Poincaré who in his 1882 papers on the theory of functions of complex variables pointed out to a remarkable connection between, on one hand, the class of functions of complex variables he was studying and, on the other hand, the line element $ds^2 = \frac{dx^2+dy^2}{y^2}$ and hyperbolic geometries. His comments on the matter can be found in §2 of *"Théorie des groupes Fuchsiens"*, Acta Math. 1 (1882).

We end by noting that there is another (isomorphic) model of hyperbolic geometry which is attributed to Poincaré, the so-called *disk model.* In this model we consider the interior of the Euclidean unit disk

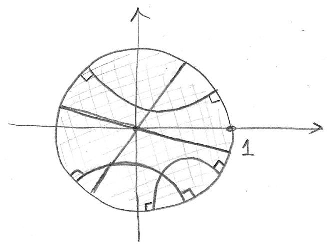
$$\{\mathbf{x} \in \mathbb{R}^n \big| |\mathbf{x}| < 1\}$$

[12]This statement is known as Playfair's Axiom, in honor of the late 18th and early 19th century Scottish mathematician John Playfair who advocated that the Fifth Postulate be replaced by this axiom.

and the line element

$$ds^2 = \tfrac{4}{(1-|\mathbf{x}|^2)^2}|d\mathbf{x}|^2.$$

The geodesics in this model are line segments through the origin and the circles orthogonal to the unit sphere $|\mathbf{x}| = 1$.



As we shall see, it is no coincidence that the form of this hyperbolic line element is so similar to the spherical line element in stereographic coordinates. Interestingly enough, this general form of the line element was foreseen by Riemann who in his Riemann's Habilitation lecture were discusses line elements of the form

$$ds^2 = \frac{1}{(1+\frac{\alpha}{4}|\mathbf{x}|^2)}|d\mathbf{x}|^2.$$

Exercises for Lecture 3.

(1) (a) Starting from the fact that $\int_a^b (f(t)+sg(t))^2\,dt \geq 0$ for all s, conclude that

$$\left(\int_a^b f(t)g(t)\,dt\right)^2 \leq \left(\int_a^b f(t)^2\,dt\right)\left(\int_a^b g(t)^2\,dt\right);$$

this is the Cauchy-Schwarz inequality for integrals. [Basic]

(b) Show that $\left(\int_a^b f(t)g(t)\,dt\right)^2 = \left(\int_a^b f(t)^2\,dt\right)\left(\int_a^b g(t)^2\,dt\right)$ only when the functions f and g are scalar multiples of one another. [Basic]

(c) Show that for any (thus not necessarily constant speed) path $\gamma(t)$, $a \leq t \leq b$ we have

$$L(\gamma)^2 \leq 2(b-a)E(\gamma).$$

Furthermore, show that the equality holds if and only if γ is parametrized by constant length. [Basic]

(2) (a) Show that energy of a curve depends on a parametrization. [Basic]

(b) Show that for a given geometric curve energy is minimized by parametrizations of constant speed. [Basic]

(3) (a) Let $f(t)$, $a \leq t \leq b$ be a continuous function with $\int_a^b f(t)\chi(t)\,dt = 0$ for all continuous functions $\chi(t)$ with $\chi(a) = \chi(b) = 0$. Show that $f(t) = 0$. [Intermediate]

(b) Complete the details of the derivation of the geodesic equations. [Intermediate, Expected]

(4) Show that any solution $\gamma(t) = (x^1(t), \ldots, x^n(t))$ of the geodesic equation satisfies

$$\tfrac{d}{dt}(|\dot\gamma|_g^2) = 0.$$

In particular, show that geodesics are necessarily parametrized in such a way that their tangent vectors are of constant length. [Intermediate]

(5) (a) Let c be a non-zero constant. Show that a curve $\gamma(t)$ is a geodesic if and only if $\gamma(ct)$ is. [Basic]

(b) Let P be a point of the Riemannian manifold (M, g) and let B_ε denote the set of all tangential vectors V at P of magnitude less than ε: $|V|_g < \varepsilon$. Show that there is $\varepsilon > 0$ and a smooth map $\gamma : B_\varepsilon \times (-2, 2) \to M$ such that for each $V \in B_\varepsilon$ the parametrized curve

$$\gamma_V(t) := \gamma(V, t)$$

is a geodesic passing though P with velocity vector V. [Intermediate]

(6) (a) Compute the Christoffel symbols for the standard unit 2-dimensional sphere with respect to the standard spherical coordinates (θ, ϕ). [Basic, Expected]

(b) Express the geodesic equations on the standard unit sphere, and verify that a particular great circle is a geodesic. [Basic]

(7) Find the distance between the following set of points of the Poincaré upper half-plane model $\mathbb{H}^2$ of hyperbolic geometry. [Basic]

(a) $(0, e^a)$ and $(0, e^b)$;

(b) $(-1, 1)$ and $(1, 1)$;

(c) $(0, 2)$ and $(-1, 1)$.

(8) Investigate circles in hyperbolic plane geometry using the following guidelines. [Intermediate]

(a) Express the metric of the 2-dimensional Poincaré disk model in polar coordinates.

(b) Find the hyperbolic distance between $(0, 0)$ and $(r, 0)$.

(c) Describe circles of radius ϱ centered at $(0, 0)$.

(d) What are circumferences of these circles?

(9) This problem addresses surfaces which for a given boundary minimize the enclosed surface area; such surfaces are higher dimensional analogues of length minimizers. Critical surfaces for area functionals $S \mapsto \mathrm{Area}(S)$ are called *minimal surfaces*. [Advanced]

(a) Consider $x^{n+1} = f(x^1, \ldots, x^n)$ for $(x^1, \ldots, x^n) \in \Omega$; this defines an n-dimensional surface S in Euclidean $\mathbb{R}^{n+1}$. Show that the area functional is given by

$$\int_\Omega \sqrt{1 + |\mathrm{grad} f|^2}\, dx^1 \ldots x^n.$$

(b) Show that S is a minimal surface if and only if

$$\operatorname{div}\left(\frac{1}{\sqrt{1+|\operatorname{grad} f|^2}}\operatorname{grad} f\right) = 0.$$

(10) This problem specifically addresses minimal surfaces in $\mathbb{R}^2$. (Refer to the previous problem for background material on minimal surfaces.) [ADVANCED]

(a) Find and visualize all rotationally symmetric minimal surfaces in $\mathbb{R}^3$. Hint: re-write the minimal surface equation from the previous problem in polar coordinates, and rely on the fact that the function f depends only on r.

(b) Find all minimal surfaces which are graphs of functions of the form $f(x,y) = f_1(x) + f_2(y)$. (If you need help visualizing your solution look up Scherk's first minimal surface on the internet.)

4. Introduction to Differential Calculus

Introduction. The goal of today's lecture is to develop an understanding of how to measure the rates of change of a function or a vector field on a Riemannian manifold. This is typically done in terms of directional derivatives ∇_V where V is a vector field. While directional differentiation of functions turns out to be relatively straightforward, a surprise is waiting for us in the case of vector fields. It turns out that when differentiating vector fields one cannot work with coordinate functions alone: coordinate systems change with respect to themselves and thus contribute to the overall rate of change of a vector field.

Directional derivatives of functions. One of the basic concepts in multivariable calculus is that of a *directional derivative*. The directional derivative of $f(\mathbf{x})$, a scalar function of n variables $\mathbf{x} = (x^1, \ldots, x^n) \in \mathbb{R}^n$, with respect to a vector field V on $\mathbb{R}^n$ is the function $\nabla_V f$ given by $(\nabla_V f)(\mathbf{x}) := \lim\limits_{t\to 0} \frac{1}{t}(f(\mathbf{x}+tV) - f(\mathbf{x}))$. Thus intuitively $\nabla_V f$ measures the rate of change of f in the direction of V. In practice one computes this directional derivative as follows: if $V = \sum_i V^i \partial_i$ then

$$\nabla_V f = \sum_i V^i \partial_i f. \tag{2}$$

While the earlier definition of $\nabla_V f$ does not easily generalize to manifolds due to the fact that $\mathbf{x} + tV$ is undefined, the latter identity does carry over to manifolds. One first proves (see exercise (1) following this lecture) that for a given function f and a given vector field V the value of $\sum_i V^i \partial_i f$ does not depend of the choice of coordinates. Consequently, the local coordinate expressions (2) for $\nabla_V f$ do seamlessly overlap to give a globally defined $\nabla_V f$.

If $V = \dot\gamma$ along some curve $\gamma(t) = (x^1(t), \ldots, x^n(t))$ then along that curve we have:

$$\nabla_V f = \sum_i \frac{dx^i}{dt}\partial_i f = \frac{d}{dt} f(\gamma(t)).$$

As the intuition suggests, the value of $\nabla_V f$ along γ depends only on V and f along γ. The derivative $\frac{d}{dt} f(\gamma(t))$ where f is a function defined along γ is often denoted by $\nabla_{\dot\gamma} f$.

We would like to emphasize the following features of $\nabla_V f$; note that the last of the properties serves as a de-facto Product Rule.

- $\nabla_{\alpha V+\beta W} f = \alpha\nabla_V f + \beta\nabla_W f$ for all functions α, β;
- $\nabla_V(f+g) = \nabla_V f + \nabla_V g$;
- $\nabla_V(\alpha f) = (\nabla_V \alpha) f + \alpha(\nabla_V f)$ for all (differentiable) functions α, β.

Although partial derivatives of a (continuously differentiable) function f always commute, $\partial_i\partial_j f = \partial_j\partial_i f$, one should be careful not to commute ∇_V and ∇_W:

$$\begin{aligned}\nabla_V\nabla_W f - \nabla_W\nabla_V f &= \sum_{ij} V^i\partial_i(W^j\partial_j f) - \sum_{ij} W^i\partial_i(V^j\partial_j f)\\ &= \sum_{ij}(V^i\partial_i W^j - W^i\partial_i V^j)\partial_j f\end{aligned}$$

The structure of the last formula suggests that the commutator $\nabla_V\nabla_W - \nabla_W\nabla_V$ behaves as directional differentiation with respect to the vector field

$$\sum_{ij}(V^i\partial_i W^j - W^i\partial_i V^j)\partial_j.$$

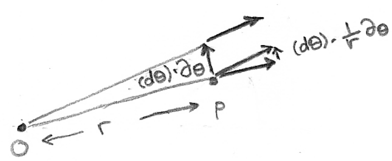
The latter is just a local formula which, prima facie, need not correspond to a globally defined vector field. (Of course, there are no such issues if we are working over $\mathbb{R}^n$.) Thankfully, it can be shown that the last expression is invariant under changes of coordinates and so it does indeed give rise to a globally defined vector field. (See also exercise 1 below.) This new vector field is denoted by $[V, W]$ and called *the commutator of V and W*. It is sometimes also called the *Lie bracket* in honor of the Norwegian mathematician Sophus Lie, who in the second half of the 19th century developed a whole field of mathematics (*Lie algebras*) based on objects which behave like the commutators $[V, W]$.

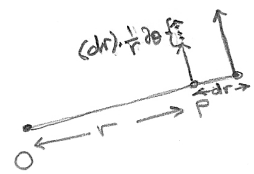
The naive approach to differentiating vector fields does not work. Measuring the rate of change of one vector field with respect to another can be complicated. In $\mathbb{R}^n$ we can do this differentiation component-wise, provided we use the standard Cartesian coordinates: If V and $W = \sum W^i\partial_i$ on $\mathbb{R}^n$ then

$$\nabla_V W = (\nabla_V W^i)\partial_i.$$

Note that in particular $\nabla_{\partial_i}\partial_j = \vec{0}$. On the other hand, the next example shows that this no longer holds if "curvilinear" coordinate systems are used. In essence, the reason for this is that, generally speaking, coordinate systems change with respect to themselves! For this reason one should somehow first understand how a coordinate system changes with respect to itself, and then apply linearity and some sort of product rule.

Example. We now pictorially investigate how coordinate systems change with respect to themselves. Arguments we are about to make are informal, and could be backed up by an explicit computation, e.g. exercise 3. The hope is that pictures will aid in developing an intuitive rather than pure computational understanding of what is going on.





Consider polar coordinates in the Euclidean plane $\mathbb{R}^2 \smallsetminus \{(0,0)\}$, and the coordinate vector fields ∂_r and ∂_θ. Let us study the rate of change of ∂_r in the direction of ∂_θ, i.e. $\nabla_{\partial_\theta}\partial_r$. Starting with ∂_r based at a point P with coordinates (r,θ) and moving infinitesimally in the ∂_θ direction produces ∂_r which is rotated by $d\theta$. Consider the infinitesimal triangle with two sides of length $|\partial_r| = 1$, enclosing the angle of $d\theta$ at P. The change in ∂_r can be seen as the side of this triangle across from P. There is a similar (pun intended!) triangle with the infinitesimal angle of $d\theta$ at the origin O and with sides of size $|\vec{OP}| = r$; the remaining side is the infinitesimal $(d\theta)$ multiple of ∂_θ. Using the properties of similar triangles we are lead to

$$\nabla_{\partial_\theta}\partial_r = \tfrac{1}{r}\partial_\theta.$$

Likewise, to find $\nabla_{\partial_r}\partial_\theta$ move infinitesimally (dr) away from P in the ∂_r direction. The two vectors ∂_θ are parallel but of different length: one is of length r and the other is of length $r + dr$. In particular, the change in ∂_θ is a multiple of dr and the unit vector $\frac{1}{r}\partial_\theta$. It follows that the rate of change of ∂_θ in the direction of ∂_r is

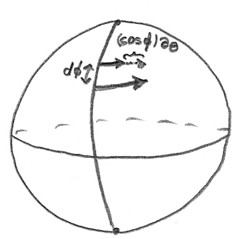
$$\nabla_{\partial_r}\partial_\theta = \tfrac{1}{r}\partial_\theta.$$

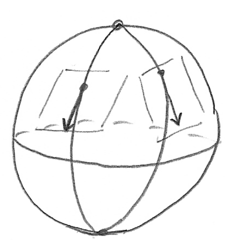
Similar arguments show that $\nabla_{\partial_r}\partial_r = \vec{0}$ and $\nabla_{\partial_\theta}\partial_\theta = -r\partial_r$. The four directional derivatives we found show how *the polar coordinate system changes with respect to itself.*

In principle, analogous "computations" can be executed on the sphere S^2. For instance, to find $\nabla_{\partial_\phi}\partial_\theta$ we observe that the coordinate vectors ∂_θ at two nearby points with coordinates (θ,ϕ) and $(\theta,\phi + d\phi)$ are parallel and pointing in the same direction while their magnitudes are $\sin(\phi)$ and $\sin(\phi + d\phi)$. Since ∂_ϕ is unit it follows that

$$\nabla_{\partial_\phi}\partial_\theta = \cos(\phi)\partial_\theta.$$

However, finding some of the remaining directional derivatives (such as $\nabla_{\partial_\theta}\partial_\phi$) is much harder to do primarily because it is hard to compare vectors which lie in two different tangent spaces and especially if we want to stay on the surface of the sphere alone.





Throughout the above discussion we used the word "parallel" and our primary tool for comparing vectors with different base points has been *parallel transport* along a coordinate line. This approach can become very confusing very quickly.

For example, it is natural for "a Flatlander" on S^2 to perceive the vector field ∂_ϕ along a meridian to be parallel to itself (thus giving us $\nabla_{\partial_\phi}\partial_\phi = \vec{0}$), since ∂_ϕ is always unit and always points "due South". However, from the perspective of $\mathbb{R}^3$ in which S^2 lives the vector field ∂_ϕ changes along the meridian as it bends towards the South Pole. The following question arises: Given two tangent vectors with different base points, how does one know if they are to be treated as "parallel"?

Ultimately, one comes to the conclusion that parallelism and differentiation are little bit like a chicken and an egg. If one knows how to parallel-transport a coordinate vector along a coordinate "axis", then one can deduce the formulae which describe the rates of change of a coordinate system with respect to itself and – by means of linearity and some sort of product rule – the derivative of any vector field with respect to any other vector field. Conversely, if one knows how to differentiate vector fields then one can parallel transport a vector along a curve by requiring that its derivative along the way vanish. It looks like that in either case one has to be open to the possibility that an additional piece of mathematical structure, parallelism or the differentiation operation for vector fields, is to be imposed on Riemannian manifolds; it may be the case that this new piece of mathematical structure is completely independent of the Riemannian metric.

Connections. Differentiation operation should have two inputs: one for the direction in which we are differentiating and one for the vector field which is being differentiated. We expect the map $\nabla : (V, W) \mapsto \nabla_V W$ to have the following properties:

- $\nabla_{\alpha U + \beta V} W = \alpha \nabla_U W + \beta \nabla_V W$ for all functions α, β;
- $\nabla_V (W_1 + W_2) = \nabla_V W_1 + \nabla_V W_2$;
- $\nabla_V(\alpha W) = (\nabla_V \alpha) W + \alpha(\nabla_V W)$ for all (differentiable) functions α, β.

Any operation on vector fields $\nabla : (V, W) \mapsto \nabla_V W$ which satisfies the three properties stated above is called a *connection*. The word connection is used because differentiation (through the notion of parallel transport which we briefly discussed above) provides a way of relating tangent vectors based at different points. Just like a given manifold can carry many different metrics, a given Riemannian manifold can carry many different connections and consequently: many different ways of comparing tangent spaces at different base points.

One consequence of the defining properties of connections is that

$$\nabla_V W = \sum_{ij} V^i \partial_i (W^j) \partial_j + \sum_{ij} V^i W^j \nabla_{\partial_i} \partial_j$$

for all vector fields $V = \sum V^i \partial_i$ and $W = \sum W^i \partial_i$. This identity clearly shows that an understanding of $\nabla_{\partial_i}\partial_j$ i.e. how coordinate systems change with respect to themselves, together with linearity and the Product Rule, determine $\nabla_V W$ in general. The coordinates A^k_{ij} of $\nabla_{\partial_i}\partial_j$,

$$\nabla_{\partial_i}\partial_j = \sum_k A^k_{ij} \partial_k,$$

are called the *connection coefficients*. They can be thought of as some sort of derivatives of a coordinate system with respect to itself.

If along a curve $\gamma(t) = (x^1(t), \ldots, x^n(t))$ we have $V = \dot\gamma$, then along γ the vector field $\nabla_V W$ depends only on V and W along γ:

$$\nabla_V W = \sum_i (\nabla_{\dot\gamma} W^i)\partial_i + \sum_{ijk} (W^i \tfrac{dx^j}{dt} A^k_{ij})\partial_k.$$

This fact allows us to define the derivative $\nabla_{\dot\gamma} W$ in the direction of $\dot\gamma$ of any vector field W which is defined along γ.

Levi-Civita connection / Covariant differentiation. From a geometric standpoint it is natural to expect/require our differentiation/connection/parallelism structure to be such that tangent vector fields to any geodesic remain parallel to one another: $\nabla_{\dot\gamma}\dot\gamma = \vec{0}$. This requirement is equivalent to

$$\sum_i \tfrac{d^2x^i}{dt^2}\partial_i + \sum_{ijk} (\tfrac{dx^i}{dt}\tfrac{dx^j}{dt} A^k_{ij})\partial_k = \vec{0}$$

or, equivalently, to the system of equations

$$\tfrac{d^2x^k}{dt^2} + \sum_{ij} A^k_{ij}\tfrac{dx^i}{dt}\tfrac{dx^j}{dt} = 0, \quad 1 \le k \le n.$$

We expect/require that all geodesics $\gamma(t) = (x^1(t), \ldots, x^n(t))$ satisfy this system of equations. By comparing with the geodesic equations we conclude that the *geometrically natural connections have Christoffel symbols as the connection coefficients:*

$$A^k_{ij} = \Gamma^k_{ij} = \tfrac{1}{2}\sum_l g^{kl}(\partial_j g_{il} + \partial_i g_{lj} - \partial_l g_{ij}).$$

Of course, this formula does not prove that such a geometrically natural connection exists – it only proves that such a connection has to be unique. The good news is that one can exhibit an almost explicit formula for a connection ∇ whose coefficients are equal to Γ^k_{ij}:

$$\begin{aligned} g(\nabla_U V, W) = \tfrac{1}{2}\{&\nabla_U(g(V,W)) + \nabla_V(g(W,U)) - \nabla_W(g(U,V)) \\ &+ g([U,V],W) - g([V,W],U) + g([W,U],V)\} \end{aligned}$$

The verification details is left as an exercise: see 4 below.

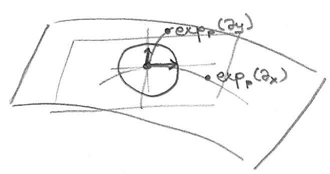
This geometrically natural connection will be assumed throughout our lectures. It is called the *Levi-Civita connection* in honor of the Italian mathematician Tullio Levi-Civita who, at the turn of the 20th century, participated in the discovery of *tensor calculus* and the development of *covariant differentiation*. We will study both of these concepts in the next several lectures. For now you may consider the phrase "covariant differentiation" to be synonymous to the phrase "Levi-Civita connection" although as we encounter more examples of tensors we will see that they can be covariantly differentiated as well.

Properties of Christoffel symbols worthy of notice. Christoffel symbols are not as much of a geometric invariant of a Riemannian manifold, as much as a reflection of our choice of coordinates. In fact, given a point P on a manifold one can always choose coordinate system which is orthonormal at that point and with respect to which the Christoffel symbols at P vanish.

$$g_{ij}(P) = \begin{cases} 1 & \text{if } \ i = j \\ 0 & \text{if } \ i \ne j \end{cases} \quad \text{and} \quad \Gamma^k_{ij}(P) = 0.$$

As we shall see in the following in the following lectures, there are geometric obstructions to having Christoffel symbols vanish everywhere.

We now describe one coordinate system with respect to which Christoffel symbols vanish (at a point). Consider a point P and the ball B_r of tangent vectors V based at P for which $|V|_g < r$. If r is sufficiently small then for all $V \in B_r$ the geodesic γ_V passing through P with velocity vector V is defined at least on the interval $(-2, 2)$. (Compare with exercise 5 from yesterday.) One can also show that for sufficiently small r the geodesic γ_V remains a minimizing geodesic over the entire interval $(-2, 2)$. Overall, there is a well-defined map $V \mapsto \gamma_V(1)$. This map is called *the exponential map*, and it plays a huge role in Riemannian geometry. It is denoted by $\exp_P$. Unfortunately, we do not have the time in these lectures to go deeply into its properties. The point which we would like to emphasize today is that the exponential map we just described is a map from $\exp_P : B_r \subseteq \mathbb{R}^n$ to M, i.e. a coordinatization of a neighborhood of P. It is with respect to these coordinates that the Christoffel symbols vanish at P. For more details on this and the related subjects please consult exercises 5 and 6 following this lecture.



Another feature one should notice is that Christoffel symbols are symmetric:

$$\Gamma^k_{ij} = \Gamma^k_{ji}.$$

While this is saying that $\nabla_{\partial_i}\partial_j = \nabla_{\partial_j}\partial_i$, one should not rush into concluding that $\nabla_V W$ and $\nabla_W V$ are the necessarily the same. Similarly to what happens in the case of $\nabla_V \nabla_W f$ and $\nabla_W \nabla_V f$, the difference of the two is related to the commutator $[V, W]$:

$$\nabla_V W - \nabla_W V = [V, W].$$

The proof is left as an exercise. In general, connections which satisfy this property are said to be *torsion-free.*

Parallel transport. Finally, we go back to the idea of parallel transport mentioned earlier in the lecture. We say that a vector field V which is defined along a curve γ is *parallel along* γ if $\nabla_{\dot\gamma} V = 0$. Since

$$\nabla_{\dot\gamma} V = \sum_k \left(\tfrac{d}{dt} V^k + \sum_{ij} (\Gamma^k_{ij} \tfrac{dx^j}{dt}) V^i \right) \partial_k$$

the condition that a vector field be parallel along a curve is just a system of first order ODE's:

$$\tfrac{d}{dt} V^k + \sum_{ij} (\Gamma^k_{ij} \tfrac{dx^j}{dt}) V^i = 0, \quad 1 \le k \le n.$$

This is a linear system, and its (long-term) solutions are well understood: for a given initial condition $V\big|_P = V_0$ there is exactly one long-term solution of the system. (See also exercise 8 below.) In other words, we have the following theorem.

THEOREM 4. *Let $\gamma : I \to M$ be a curve passing through $\gamma(t_0) = P$. For any tangent vector V_0 at P there exists a unique vector field V which is parallel along γ and satisfies $V\big|_P = V_0$.*

In particular, if γ is a curve joining points P and Q then the tangent vector $V\big|_Q$ is said to be *the parallel transport* of $V\big|_P$ along γ. Parallel transport can be viewed as a map $V\big|_P \mapsto V\big|_Q$ between tangent spaces at P and Q. Since the parallel transport equations are linear, so is this map. In fact even more is true:

THEOREM 5. *Parallel transport between tangent spaces is an isometry.*

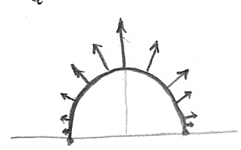
As a reminder, isometries are linear transformations which preserve the inner-product – in our case g – on tangent spaces. In other words, our claim is that parallel transports of vectors $V\big|_P$ and $W\big|_P$ satisfy:

$$g(V\big|_Q, W\big|_Q) = g(V\big|_P, W\big|_P).$$

The proof of theorem is a nice little exercise which we leave for the reader. (See exercise 9 below.) We end the lecture with a quick example which illustrate parallel transport.

Example. Consider the Poincaré upper half-space model of the hyperbolic plane, $ds^2 = \frac{dx^2+dy^2}{y^2}$ with $y > 0$. Consider the upper semi-circle γ of $x^2 + y^2 = 1$ and the vector $V_0 = \partial_y$ at $P(0, 1)$. To find parallel transport of V along γ note that γ is a geodesic and that V_0 is a unit vector which is orthogonal to the (tangent vector to the) geodesic. Since parallel transport serves as an isometry, and since the tangent field to a geodesic is necessarily parallel, we see that parallel transport of V_0 remains unit and orthogonal to γ. Vectors orthogonal to γ at a point $Q(x, y)$ are multiples of $x\partial_x + y\partial_y$. By normalizing we see that the parallel transport of V_0 at $Q(x, y)$ is

$$V\big|_Q = y\,(x\partial_x + y\partial_y)\,.$$



Exercises for Lecture 4.

(1) (a) Show that for a given function f and a given vector field V the value of $\sum_i V^i \partial_i f$ does not depend of the choice of coordinates. [INTERMEDIATE]

(b) Let $\{\partial_i\}$ and $\{\partial_{i'}\}$ be the coordinate vector fields corresponding to the x and the y coordinates, respectively, and let V and W be two vector fields. Prove the following equality. [INTERMEDIATE]

$$\sum_{ij}(V^i\partial_i W^j - W^i\partial_i V^j)\partial_j = \sum_{i'j'}(V^{i'}\partial_{i'} W^{j'} - W^{i'}\partial_{i'} V^{j'})\partial_{j'}.$$

(2) Prove the following properties of the Lie bracket. You should assume α is some differentiable function. [BASIC, EXPECTED]

(a) $[V, W] = -[W, V]$;

(b) $[U,[V,W]]+[V,[W,U]]+[W,[U,V]]=0$;

(c) $[V,\alpha W]=(\nabla_V\alpha)W+\alpha[V,W]$.

(3) Consider the standard polar coordinates (r,θ) in the Euclidean plane. Recall the expressions for ∂_r and ∂_θ in Cartesian coordinates from exercises following Lecture 1. Compute $\nabla_{\partial_r}\partial_r$, $\nabla_{\partial_\theta}\partial_r$, $\nabla_{\partial_r}\partial_\theta$ and $\nabla_{\partial_\theta}\partial_\theta$ directly, with no use of Christoffel symbols. Express your answers in terms of ∂_r and ∂_θ. [Basic]

(4) (a) Let $\langle .,. \rangle$ be an inner-product on a finite dimensional vector space $\mathcal{V}$ and let $L:\mathcal{V}\to\mathbb{R}$ be a linear map. Show that there is a unique vector $V\in\mathcal{V}$ such that $\langle V,W\rangle = L(W)$ for all $W\in\mathcal{V}$. [Intermediate]

(b) Let

$$\begin{aligned}\mathcal{L}(U,V,W) := \tfrac{1}{2}\{&\nabla_U(g(V,W))+\nabla_V(g(W,U))-\nabla_W(g(U,V))\\ &+g([U,V],W)-g([V,W],U)+g([W,U],V)\}.\end{aligned}$$

Prove the following. You should assume α and β are some differentiable functions. [Intermediate]

(i) $\mathcal{L}(U,V,\alpha W_1+\beta W_2)=\alpha\mathcal{L}(U,V,W_1)+\beta\mathcal{L}(U,V,W_2)$;

(ii) $\mathcal{L}(\alpha U_1+\beta U_2,V,W)=\alpha\mathcal{L}(U_1,V,W)+\beta\mathcal{L}(U_2,V,W)$;

(iii) $\mathcal{L}(U,V_1+V_2,W)=\mathcal{L}(U,V_1,W)+\mathcal{L}(U,V_2,W)$;

(iv) $\mathcal{L}(U,\alpha V,W)=(\nabla_U\alpha)g(V,W)+\alpha\mathcal{L}(U,V,W)$.

(c) Show that there exists a unique connection ∇ for which

$$g(\nabla_U V,W)=\mathcal{L}(U,V,W).$$

Furthermore, show that the corresponding connection coefficients are equal to the Christoffel symbols Γ^k_{ij}. [Intermediate]

(5) Let $(x^1,\ldots,x^n)$ be coordinates on a neighborhood of $P\in M$, with $(0,\ldots,0)$ corresponding to P.

(a) Show that one can make a linear change of coordinates $y^i := \sum_j L^i_j x^j$ so that the coordinate vectors $\{\partial_{y^1},\ldots,\partial_{y^n}\}$ are orthonormal at P. [Advanced]

(b) Assume coordinates $(y^1,\ldots,y^n)$ as above, and let $C^k_{ij}=C^k_{ji}$ be a family of constants. Use the Inverse Function Theorem to show that $(z^1,\ldots,z^n)$ for

$$z^k := y^k + \sum_{ij} C^k_{ij}y^iy^j$$

is a coordinatization of a (potentially smaller) neighborhood of P. [Advanced]

(c) Assume coordinates $(z^1,\ldots,z^n)$ as above. Show that the coordinate vectors $\{\partial_{z^1},\ldots,\partial_{z^n}\}$ are orthonormal at P and that the constants C^k_{ij} can be chosen so that all the partial derivatives $\partial_k g_{ij}$ of the metric components with respect to the z-coordinates, and thus all the Christoffel symbols, vanish at P. [Advanced]

(6) Review exercise 5 following yesterday's lecture, and the definition of the exponential map from this lecture. Coordinates on (a neighborhood of) $P \in M$ defined by $\exp_P$ are called *geodesic normal coordinates centered at* P. [INTERMEDIATE]

(a) Show that $t \mapsto \exp_P(tV)$ is the geodesic passing though P with the velocity vector V.

(b) Show that in geodesic normal coordinates centered at P geodesics passing through P take the form of

$$(x^1(t), \dots, x^n(t)) = (tV^1, \dots, tV^n)$$

where $V = \sum_i V^i \partial_i$ serves as the tangent vector to the geodesic as it passes through P.

(c) Show that the Christoffel symbols with respect to geodesic normal coordinates centered at P satisfy

$$\sum_{ij} \Gamma_{ij}^k(P) V^i V^j = 0$$

for all $V = \sum_i V^i \partial_i$.

(d) Show that $\Gamma_{ij}^k(P) = 0$ for all i, j, k. Hint: use the above with judicious choices of V such as $V = \partial_i + \lambda \partial_j$ for fixed i, j but variable λ.

(7) Prove that Levi-Civita connection is torsion-free: $\nabla_V W - \nabla_W V = [V, W]$. [BASIC]

(8) In coordinates, parallel transport equations take the form of

$$\tfrac{d}{dt} \mathbf{V} = A(t) \mathbf{V}$$

where $\mathbf{V} : I \to \mathbb{R}^n$ is a vector-valued unknown function of single variable t, and where $A(t)$ is $n \times n$-matrix-valued. Show that solutions of the above take the form of

$$\mathbf{V}(t) = \exp\left(\int_0^t A(s)\, ds\right) \mathbf{V}_0$$

for some constant $\mathbf{V}_0 = \mathbf{V}(0) \in \mathbb{R}^n$. Here $\int A(s)\, ds$ stands for entry-wise integration ($\int [a_{ij}(s)]\, ds = [\int a_{ij}(s)\, ds]$) and $\exp(M) = \sum_{m=0}^{\infty} \frac{1}{m!} M^m$ denotes the matrix exponential. Hint: Use the integrating factor of $\exp\left(-\int_0^t A(s)\, ds\right)$. [INTERMEDIATE]

(9) Show that parallel transport is

(a) a linear transformation;

(b) an isometry.

Hint: Show that $\frac{d}{dt} g(V, W) = \frac{d}{dt}\left(\sum_{ij} g_{ij} V^i W^j\right) = 0$ whenever vector fields $V = \sum V^i \partial_i$ and $W = \sum W^j \partial_j$ are parallel along $\gamma(t)$. [INTERMEDIATE]

(10) Find parallel transport of V_0 along the curve γ, for the following choices of V_0 and γ. In each case, illustrate what parallel transport does along the curve. [INTERMEDIATE]

(a) γ is the (piece-wise smooth) counterclockwise contour of a geodesic triangle on the surface of S^2 with one vertex at the North Pole P; V_0 is the unit tangent to the side of the triangle leaving P;

(b) γ is the "line" segment $x(t) = t$, $y(t) = 1$ with $0 \le t \le 1$ in the Poincaré upper half-space, and $V_0 = \partial_y$ at the point $P(0, 1)$.

5. Vector Calculus

Introduction. The main goal of today's lecture is to develop an understanding of the gradient, divergence and Laplace operators compatible with a Riemannian metric. Theorem-wise, we are working towards the Divergence Theorem and the Maximum Principles – particularly in the context of compact manifolds (without boundary).

The linear algebra framework. In the previous lecture, and throughout this course, we view differentiation as an operation with two inputs: one for the object being differentiated (function, vector field, etc...) and one for the direction of differentiation. If the object which we are differentiating is fixed and known – a function f, for example – then we are dealing with ∇f, a linear map defined by

$$\nabla f : V \mapsto \nabla_V f.$$

Although one might be tempted to think of ∇f as the gradient vector field of f this would be incorrect: ∇f is *not* a vector field. Instead, it is what is known as a covector[13] field: a linear map which takes vector fields to scalar functions.

A more thorough understanding of the underlying linear algebraic framework pays dividends here! Vector fields form a structure similar to that of a vector space in which the role of scalars is played by functions. Namely, one can add vector fields $(V + W)$ and multiply them with scalar functions (αV); these operations result in vector fields and obey all the usual vector space rules. We use $\mathfrak{X}$ to denote this "vector space"[14] and $\mathfrak{F}$ to denote this set of scalars. Unless specifically stated otherwise one may assume that the coordinate representation of any element in $\mathfrak{X}$ and $\mathfrak{F}$ is smooth.

Covector fields are $\mathfrak{F}$-linear maps from $\mathfrak{X}$ to $\mathfrak{F}$; the set of covector fields will be denoted by $\mathfrak{X}^*$. Since

$$\nabla_{\alpha V + \beta W} f = \alpha \nabla_V f + \beta \nabla_W f$$

for all $\alpha, \beta \in \mathfrak{F}$, covariant derivatives ∇f are examples of covector fields. On the other hand, covariant derivatives of vector fields should be viewed as $\mathfrak{F}$-linear maps which take vector fields to vector fields:

$$\nabla X : \mathfrak{X} \to \mathfrak{X}, \quad \nabla X : V \mapsto \nabla_V X.$$

Terminology applicable to such maps will be discussed in the next lecture.

---

[13]It is not a coincidence that the phrases co-variant differentiation and co-vector field have the same prefix!

[14]From a formal standpoint there are two issues here. One is that $\mathfrak{F}$ is not a field but a ring; thus the more appropriate word to use here is that of *module*. The other issue is that we have made no commitments about the domains of our functions and vector fields. (Are these defined on all of M, or just on portions of M?) In the interests of brevity we avoid dealing with this issue in detail. The upshot is that for the most part we are able to treat locally defined fields/functions in essentially the same manner as the fields/functions which are globally defined.

Implications of $\mathfrak{F}$-linearity. If a map $\mathcal{L}(\dots,V,\dots)$ is $\mathfrak{F}$-linear in the V-entry then the value of $L(\dots,V,\dots)$ at a particular point $P \in M$ depends only on the value of V at P, and *not* on any nearby values of V. For instance, if $V, W \in \mathfrak{X}$ are the same at P then in coordinates $V - W = \sum_i \alpha^i \partial_i$ with each of the functions α^i vanishing at P. Since L is $\mathfrak{F}$-linear in the V-entry we have:

$$\begin{aligned} L(\dots,V,\dots) =& L(\dots,W,\dots) + L(\dots,V-W,\dots) \\ =& L(\dots,W,\dots) + \sum_i \alpha^i L(\dots,\partial_i,\dots). \end{aligned}$$

The last term vanishes at P due to the fact that each α^i vanishes at P. Thus, the values of $L(\dots,V,\dots)$ and $L(\dots,W,\dots)$ are the same at P.

The above is important for the following reason. Although technically covector fields act on vector fields, they restrict to a single point and give a well-defined operation on tangential directions/vectors based at that single point. For example, restricting ∇f to a point produces a linear map $\mathbb{R}^n \to \mathbb{R}$ whose rule can be described by "give me a direction based at P and we can find the rate of change of f in that particular direction at P"); likewise, restricting ∇X to a point produces a linear map $\mathbb{R}^n \to \mathbb{R}^n$. In contrast, differential operators which act on vector fields do not restrict to a single point in the same sense of the word: knowledge of a vector X at a point alone is insufficient for understanding the outcome of the differentiation $\nabla_V X$ at that point.

Duality between $\mathfrak{X}$ and $\mathfrak{X}^*$. Those familiar with the concept of a vector space dual will recognize that in some sense covectors fields are dual to vector fields. In fact, Riemannian metric gives rise to a natural correspondence between $\mathfrak{X}$ and $\mathfrak{X}^*$. Let us address this phenomenon in more detail.

If $V \in \mathfrak{X}$, then the map $W \mapsto \langle V, W \rangle$ is $\mathfrak{F}$-linear and defines a covector field. This covector field is commonly denoted by $V_\flat$, and should be thought of as a dual to V. Allowing V to vary as well, we arrive at

$$_\flat : \mathfrak{X} \to \mathfrak{X}^*, \quad \text{defined by} \quad V \mapsto \{W \mapsto \langle V, W \rangle\}.$$

It can be easily checked that $_\flat$ is $1-1$; see exercise 1. More interestingly, the map $_\flat$ is onto. That is, for each $\omega \in X^*$ there is a vector field V such that $\langle V, W \rangle = \omega(W)$ for all $W \in \mathfrak{X}$. To exhibit the existence of this vector field we utilize coordinates. It can easily be checked that

$$V = \sum_{ij} (g^{ij} \omega(\partial_j)) \partial_i$$

satisfies that relationship $g(V, W) = \omega(W)$, at least within the domain a particular coordinate grid. One can also verify that the last expression is independent of the choice of coordinates; see exercise (2). Consequently, if ω extends beyond a single coordinate grid then the vector field $\sum_{ij} (g^{ij} \omega(\partial_j)) \partial_i$ extends seamlessly to domains covered by other coordinates. The vector field V corresponding to the covector field ω in the manner described above is commonly denoted by $\omega^\sharp$. The fields ω and $\omega^\sharp$ are to be thought of a being duals of one another.

The gradient vector field. Motivated by the properties of the gradient vector field for functions on $\mathbb{R}^n$ we define $\mathrm{grad}(f) \in \mathfrak{X}(M)$ to be the unique vector field for which

$$\langle \mathrm{grad}(f), V\rangle = \nabla_V f.$$

In other words, we define $\mathrm{grad}(f)$ to be the vector field dual to ∇f. As discussed above, the coordinate expression for the gradient is

$$\mathrm{grad}(f) = \sum_{ij} (g^{ij}\partial_j f)\partial_i.$$

Divergence. Trace is a geometric invariant of a linear operator $L : \mathcal{V} \to \mathcal{V}$ commonly defined as $\mathrm{Tr}(L) := \sum \langle Le_i, e_i\rangle$, where $\{e_i\}$ is an orthonormal basis for an inner-product space $\mathcal{V}$. The value of $\mathrm{Tr}(L)$ can be proven to be independent of the choice of $\{e_i\}$. In fact, one can show that $\mathrm{Tr}(L)$ is equal to the sum of the diagonal entries of <u>any</u> of the matrix representations of L (not necessarily with respect to an orthonormal basis)

$$\mathrm{Tr}(L) = \sum L^i_i,$$

where L^i_j denotes the entry on the i-th row and j-th column of a matrix representation of L. (See exercise 3 for details.) It is often helpful to view trace as some sort of averaging. For instance, if V is the Euclidean $\mathbb{R}^n$ one can prove

$$\mathrm{Tr}(L) := \frac{n}{\mathrm{vol}(S^{n-1})} \int_{V\in S^{n-1}} \langle LV, V\rangle \, \mathrm{dvol}_{S^{n-1}}$$

for the standard volume element $\mathrm{dvol}_{S^{n-1}}$.

Recall that restricting ∇X (for a given vector field X) to a point produces a linear operator on $\mathbb{R}^n$. This means that ∇X can be traced point-wise, ultimately producing a scalar function. This process yields the *divergence* of the vector field X:

$$\mathrm{div}(X) := \mathrm{Tr}(\nabla X) = \sum (\nabla X)^i_i = \sum \left(\nabla_{\partial_i} X\right)^i = \sum \left(\partial_i(X^i) + \Gamma^i_{ij}X^j\right),$$

where we have used the notational convention $\nabla_{\partial_i} X = \left(\nabla_{\partial_i} X\right)^j \partial_j$. It is remarkable that the quantity $\sum_i \Gamma^i_{ij}$ can be simplified further. Namely, it follows from the definition of the Christoffel symbols that $\sum_i \Gamma^i_{ij} = \sum_{il} \frac{1}{2} g^{il}\partial_j g_{il}$. Replacing g^{il} by its expression in terms of determinants and cofactors produces $\sum_i \Gamma^i_{ij} = \frac{1}{2|g|}\partial_j(|g|)$; see exercises 5 and 6 for further details. Overall, we arrive at an alternate formula for the divergence:

$$\mathrm{div}(X) = \sum_i \left(\partial_i(X^i) + \frac{1}{2|g|}\partial_i(|g|)X^i\right) = \frac{1}{\sqrt{|g|}} \sum_i \partial_i(X^i\sqrt{|g|}).$$

The nature of divergence as the net outward flux per unit of enclosed volume is best revealed by the Divergence Theorem.

THEOREM 6 (DIVERGENCE THEOREM). *Let Ω be a region of a Riemannian manifold M with piece-wise smooth boundary $\partial\Omega$ and outward pointing unit normal ν. We have*

$$\int_\Omega \mathrm{div}(X)\, \mathrm{dvol}_\Omega = \int_{\partial\Omega} \langle X, \nu\rangle \, \mathrm{dvol}_{\partial\Omega}.$$

In particular, if M is compact and without boundary we have:

$$\int_M \operatorname{div}(X)\,\mathrm{dvol}_M = 0.$$

We will work out the proof of the Divergence Theorem in the case when the region Ω is a coordinate box. The case of general Ω is handled by breaking Ω into box-like regions and addition; note that addition of boundary integrals contributed by two adjacent boxes leaves only the outmost boundary integral. In the case when the region of integration is compact and without boundary all the boundary integrals cancel, leading to the second statement of our theorem.

Sketch of a proof: Assume that in some coordinates the region Ω takes the form of $B := [-c_1, c_1] \times ... \times [-c_n, c_n]$. We start by examining what the statement of the Divergence Theorem looks like in these coordinates. First observe that the unit normal to the boundary component given by $x_a = \pm c_a$ has the coordinate expression $\pm \sum_j \frac{g^{aj}}{\sqrt{g^{aa}}}\partial_j$. Therefore, $\langle X, \nu\rangle = \pm\frac{X^a}{\sqrt{g^{aa}}}$. The volume element for this component of the boundary is determined (pun intended!) through the submatrix of $[g_{ij}]$ obtained by removing the a-th row and the a-th column; according to the formula for the inverse matrix the absolute value of the determinant of this submatrix is equal to $g^{aa}|g|$. Overall, we need to argue that:

$$\begin{aligned}&\int_B \sum_i \partial_i(X^i\sqrt{|g|})dx^1...dx^n\\ &=\sum_a \left(\int_{x_a=c_a} X^a\sqrt{|g|}\,dx^1..\widehat{dx^a}...dx^n - \int_{x_a=-c_a} X^a\sqrt{|g|}\,dx^1..\widehat{dx^a}...dx^n\right),\end{aligned}$$

where $\widehat{dx^a}$ signifies that dx^a has been omitted. The latter follows immediately from the Divergence Theorem in $\mathbb{R}^n$. □

The following is an immediate consequence of the Divergence Theorem and the identity $\operatorname{div}(fX) = \nabla_X f + f\operatorname{div}(X)$; see exercise 8.

Theorem 7 (Integration by Parts). *If M is compact without boundary, then*

$$\int_M f\operatorname{div}(X)\mathrm{dvol}_M = -\int_M \langle \operatorname{grad} f, X\rangle \mathrm{dvol}_M = -\int_M (\nabla_X f)\mathrm{dvol}_M.$$

The Laplace operator. We define the Laplace operator Δ by

$$\Delta f = \operatorname{div}(\operatorname{grad}(f)).$$

Coordinate expression for Δf can easily be found from the coordinate expression for gradient and divergence we found earlier:

$$\Delta f = \frac{1}{\sqrt{|g|}}\sum_{ij}\partial_i(g^{ij}\sqrt{|g|}\partial_j f).$$

The following is an immediate consequence of the Divergence Theorem and the integration by parts (Theorem 7):

Theorem 8. *If M is compact without boundary, then*

(1) $\int_M \Delta\alpha\,\mathrm{dvol}_M = 0$;
(2) $\int_M \alpha(\Delta\beta)\mathrm{dvol}_M = -\int_M\langle\operatorname{grad}\alpha, \operatorname{grad}\beta\rangle\mathrm{dvol}_M = \int_M(\Delta\alpha)\beta\mathrm{dvol}_M$.

Maximum Principles. The Laplacian Δf carries a certain amount of information about the local minima and maxima of f. To understand this suppose that f reaches its local extremum at a point P. In coordinates where $g_{ij}(P) = \begin{cases} 1 & i = j \\ 0 & i \neq j \end{cases}$ and $\partial_k g_{ij}(P) = 0$ the coordinate expression for the Laplacian Δf at P takes the form of $\sum_i \partial^2_{ii} f$, since at P the components of the metric tensor take the Euclidean form while their derivatives vanish. If f reaches its local maximum at P then we must have $\Delta f \leq 0$ at P and if f reaches its local minimum at P then $\Delta f \geq 0$ at P. This proves the following:

THEOREM 9 (WEAK MAXIMUM PRINCIPLE). *If $\Delta f > 0$ over a region Ω then f cannot attain a local maximum inside Ω. In particular, the maximum of f can only be attained on the boundary of Ω.*

By replacing f by $-f$ one obtains the corresponding WEAK MINIMUM PRINCIPLE. Often times Maximum Principle is used in one of the following forms.

THEOREM 10 (WEAK MAXIMUM PRINCIPLE ON COMPACT MANIFOLDS WITHOUT BOUNDARY). *Let M be compact without boundary, let c be a positive function, and let $Lf := \Delta f - cf$.*

(1) *If $Lf \geq 0$ on M, then $f \leq 0$.*
(2) *If $Lf \leq 0$ on M, then $f \geq 0$.*
(3) *If $Lf = 0$ on M, then $f = 0$ everywhere.*

PROOF. It remains to prove the first claim. Suppose the opposite; then f achieves a positive interior maximum. Then at that point $\Delta f - cf \leq 0$, which contradicts our assumptions. $\square$

Maximum Principles are one of the most often used tools in Geometric Analysis. There is a stronger version of the principle we addressed above, but its proof of goes beyond the scope of these lectures.

THEOREM 11 (STRONG MAXIMUM PRINCIPLE). *Let M be compact without boundary, let c be a non-negative function, and let $Lf := \Delta f - cf$. If $Lf = 0$ then f must be constant. If in addition c is not identically zero, then we must have $f = 0$.*

Exercises for Lecture 5.

(1) Show that the map $\flat$ introduced in this lecture is $1-1$. [BASIC]
(2) Let ω be a covector field. Show that the vector field $\sum_{ij}(g^{ij}\omega(\partial_j))\partial_i$ is invariant under the change of coordinates. [BASIC]
(3) Let $L : \mathbb{R}^n \to \mathbb{R}^n$ be a linear map, let $\{e_i\}$ be a basis for $\mathbb{R}^n$ and let $[L^i_j]$ be the matrix representation of L with respect to the basis $\{e_i\}$ where L^i_j denoties the entry in the i-th row and the j-th column.
    (a) Show that the quantity $\sum_i L^i_i$ is independent of the choice of $\{e_i\}$. [BASIC]
    (b) Show that if $\{e_i\}$ is orthonormal then $\sum_i L^i_i = \sum_i \langle Le_i, e_i\rangle$. [BASIC]

(4) Let $L : \mathbb{R}^n \to \mathbb{R}^n$ be a linear map. Show that
$$\mathrm{Tr}(L) := \frac{n}{\mathrm{vol}(S^{n-1})} \int_{V \in S^{n-1}} \langle LV, V\rangle \,\mathrm{dvol}_{S^{n-1}}$$
for the standard volume element $\mathrm{dvol}_{S^{n-1}}$. *Hints:* pick an orthonormal basis $\{e_i\}$ and coordinatize S^{n-1} with $(\phi, v) \in [0, \pi] \times S^{n-2}$ by setting $V = \cos\phi e_1 + \sin\phi v$; use the fact that $\mathrm{dvol}_{S^{n-1}} = \mathrm{d}\phi^2 + \sin^2(\phi)\mathrm{dvol}_{S^{n-2}}$ and use induction. [Advanced]

(5) Let the entries of the invertible matrix $A(t) = [a_{ij}(t)]$ be differentiable functions of variable t. Recall that:
$$\det(A) = \sum_\sigma (-1)^{\mathrm{sgn}(\sigma)} a_{1\sigma(1)} a_{2\sigma(2)} \ldots a_{n\sigma(n)},$$
where the summation goes over all permutations σ of indices $\{1, 2, \ldots, n\}$.
   (a) Prove that $\frac{d}{dt}\det(A(t)) = \sum_{ij} (-1)^{i+j} \frac{da_{ij}}{dt} M^{ij}$ where M^{ij} is the determinant of the matrix obtained when the i-th row and the j-th column of $A(t)$ are removed. [Intermediate]
   (b) Prove that $\frac{d}{dt}\det(A(t)) = \det(A(t)) \sum_{ij} \frac{d\,a_{ij}}{dt} a^{ji}$, where $[a^{ji}]$ denotes the inverse matrix $A(t)^{-1}$. [Intermediate]

(6) Complete the details behind the formula $\mathrm{div}(X) = \frac{1}{\sqrt{|g|}} \sum_i \partial_i (X^i \sqrt{|g|})$. [Intermediate]

(7) Verify the claim regarding $\langle X, \nu\rangle$ from the proof of the Divergence Theorem. [Basic]

(8) Verify the formula $\mathrm{div}(fX) = \nabla_X f + f\mathrm{div}(X)$. [Basic]

(9) Express $\Delta(f^\alpha)$ and $\Delta(fg)$ in terms of (covariant) derivatives of functions f and g individually. [Intermediate]

(10) Let f be a spherically symmetric function on $\mathbb{R}^n$, i.e. let f be a function on $\mathbb{R}^n$ which depends only on $r = |x|$.
   (a) Show that Euclidean Laplacian satisfies $\Delta f = \frac{\partial^2 f}{\partial r^2} + \frac{n-1}{r}\frac{\partial f}{\partial r}$. [Basic]
   (b) Use the above to find all spherically symmetric solutions of $\Delta f = 0$ in the Euclidean space. [Basic]

6. Differential Calculus with Tensors; Introduction to Curvature

Introduction. One of our tasks for today is to make sense out of second and higher derivatives. Although in the case of second derivatives of functions we have been able to get away with $\nabla\mathrm{grad}(f)$, the natural approach in the general case is inductive: $\nabla^k f = \nabla(\nabla^{k-1} f)$. Similar comments apply to higher derivatives of vector fields, $\nabla^k X$. The first step in this inductive process is to understand $\nabla(\nabla f)$ or, rather, $\nabla\omega$ for covector fields ω. We expect $\nabla\omega$ to be a linear map taking a vector field V to a covector field $\nabla_V \omega$, which eventually amounts to a map taking a pair of vector fields to a scalar function. Similarly, $\nabla^2 X$ amounts to a map taking a pair of vector fields to a vector field. Based on this it is natural to expect higher covariant derivatives $\nabla^k f$ and $\nabla^k X$ to behave as maps taking k vector fields to a scalar or vector field. In fact, even more is true: these maps are $\mathfrak{F}$-multilinear! So, the first task for today is to learn how to differentiate $\mathfrak{F}$-multilinear maps. As we are about to see, there is a somewhat unexpected feature of the second covariant derivative. Ultimately, this feature leads to the notion of curvature.

Tensors. In linear algebra one studies the concept of a tensor product, which essentially identifies the set of multilinear maps with the set of linear maps on an appropriately modified domain. In view of this, multilinear maps such as the ones we encountered above are called *tensor fields*, or *tensors* for short. We will use the phrase $(k,0)$-*tensor field* (resp. $(k,1)$-*tensor field*) for a $\mathfrak{F}$-linear map taking k vector fields to a scalar function (resp. vector field). Alternatively, $(k,0)$-tensor fields are called *covariant k-tensor fields.* Note that in this terminology vector fields are the same as $(0,1)$-tensor fields and Riemannian metrics are positive definite symmetric covariant 2-tensors. We will stop short of discussing the more general (k,l)-tensors.

Maps $\mathcal{L}(\dots,V,\dots)$ which are $\mathfrak{F}$-linear in the V-entry are said to be *tensorial in the V-entry.* For example, $\nabla_V X$ is tensorial in the V-entry but it is *not* tensorial in the X-entry. If $\mathcal{L}$ is tensorial in V then, as we discussed in the previous lecture, the value of $\mathcal{L}(\dots,V,\dots)$ at a particular point $P \in M$ depends only on the value of V at P, and *not* on any nearby values of V. Consequently, all tensors can be restricted to a point! At each point $P \in M$ a covariant tensor T restricts to $\mathbb{R}$-multilinear map which takes tangent vectors at P to real numbers, that is, $T\big|_P : \mathbb{R}^n \times \dots \times \mathbb{R}^n \to \mathbb{R}$.

Another simple consequence of $\mathfrak{F}$-multilinearity which is of great interest is the fact that a tensor T is completely determined by the set of values $T(\partial_{i_1}, \partial_{i_2}, ..., \partial_{i_k})$ for different choices of $i_1, ..., i_k \in \{1, ..., n\}$. For reasons of notational efficiency we denote these components by $T_{i_1,\dots,i_k} := T(\partial_{i_1}, ..., \partial_{i_k})$ in the case of $(k,0)$-tensors, and by $T^i_{i_1,\dots,i_k}$ in the case of $(k,1)$-tensors:

$$T(\partial_{i_1}, ..., \partial_{i_k}) = \sum_i T^i_{i_1,\dots,i_k} \partial_i.$$

For example, for $(2,1)$-tensors T we have $T(V,W) = \sum_{ijk} T^k_{ij} V^i W^j \partial_k$. It is worth emphasizing that despite the notational similarities the Christoffel symbols Γ^k_{ij} are *not* components of a tensor.

Covariant differentiation of tensors. Basic examples of linear maps seen in any introductory linear algebra course involve matrix multiplication. Thus, for a covector field ω and a vector field V it is reasonable to view $\omega(V)$, which is a scalar function, as some sort of multiplication. This line of thinking motivates a formal product rule relationship

$$\nabla_W(\omega(V)) = (\nabla_W \omega)(V) + \omega(\nabla_W V)$$

between the rates of change in the direction of W. Guided by this we define $\nabla\omega$ by

$$\nabla\omega(V;W) := \nabla_W(\omega(V)) - \omega(\nabla_W V).$$

(Since symmetry of $\nabla\omega$ is not be taken for granted we choose to use the semicolon to distinguish the directional derivative entry whenever confusion is possible.) While it is clear from the definition that $\nabla\omega(V;W)$ is tensorial in the W-entry, the presence of $\nabla_W V$ in the formula for $\nabla\omega(V;W)$ makes it is much less clear that $\nabla\omega$

is tensorial in the V-entry. This is something that should be checked directly:

$$\begin{aligned}\nabla\omega(\alpha V;W) =&\nabla_W(\alpha\omega(V)) - \omega(\nabla_W(\alpha V))\\ =&(\nabla_W\alpha)\omega(V) + \alpha\nabla_W(\omega(V)) - (\nabla_W\alpha)\omega(V) - \alpha\omega(\nabla_W V)\\ =&\alpha\left(\nabla_W(\omega(V)) - \omega(\nabla_W V)\right) = \alpha\nabla\omega(V;W).\end{aligned}$$

Thus, as suggested in the introduction, $\nabla\omega$ is a $(2,0)$-tensor.

In general, identities such as $T(U,V) = \sum_{ij} T_{ij}U^iV^j$ suggest an extended product rule formula:

$$\nabla_W\left(T(U,V)\right) = (\nabla_W T)(U,V) + T(\nabla_W U, V) + T(U, \nabla_W V),$$

which is turn motivates us to <u>define</u> the covariant derivative of a (k,i)-tensor field T, with $i = 0,1$, by:

$$\begin{aligned}\nabla T(V_1,\ldots,V_k;W) :=&\nabla_W(T(V_1,\ldots,V_k))\\ &-T(\nabla_W V_1,\ldots,V_k) - \cdots - T(V_1,\ldots,\nabla_W V_k).\end{aligned}$$

One important point of the story here is that *as soon as we know how to differentiate scalar functions and vector fields we also know how to differentiate covector and more general tensor fields.* As above, one can verify directly that ∇T for a (k,i)-tensor T is a $(k+1,i)$-tensor.

Examples.

- Computing in coordinates. It is important to gain some experience in computing coordinate expressions for covariant derivatives. Here we include the example of the covariant derivative of a covector field ω, but the reader should make sure to do at least one more example on their own (e.g exercise 1).

$$\begin{aligned}(\nabla\omega)_{i;j} =&(\nabla_{\partial_j}\omega)_i = \partial_j(\omega(\partial_i)) - \omega(\nabla_{\partial_j}\partial_i)\\ =&\partial_j(\omega_i) - \omega(\textstyle\sum_k \Gamma^k_{ji}\partial_k) = \partial_j(\omega_i) - \sum_k \Gamma^k_{ij}\omega_k.\end{aligned}$$

- Differentiating the metric tensor. To compute the covariant derivative of the metric tensor it suffices to find the components of ∇g in coordinates. We have:

$$\begin{aligned}(\nabla g)_{ij;k} &= \partial_k g_{ij} - g(\nabla_{\partial_k}\partial_i, \partial_j) - g(\partial_i, \nabla_{\partial_k}\partial_j)\\ &= \partial_k g_{ij} - \textstyle\sum_l \Gamma^l_{ki} g_{lj} - \sum_l \Gamma^l_{kj} g_{li}\\ &= \partial_k g_{ij} - \tfrac{1}{2}(\partial_k g_{ij} + \partial_i g_{jk} - \partial_j g_{ik}) - \tfrac{1}{2}(\partial_k g_{ij} + \partial_j g_{ik} - \partial_i g_{jk}) = 0.\end{aligned}$$

The last computation proves the following important result.

Theorem 12. *We have:*

(1) $\nabla g = 0$.

(2) $\nabla_W(\langle U,V\rangle) = \langle \nabla_W U, V\rangle + \langle U, \nabla_W V\rangle$.

See exercise 10 for one of the many reasons why this result is so significant.

Hessian and higher covariant derivatives. The idea here is quite simple: k-th covariant derivative of a tensor T is defined inductively by $\nabla^k T = \nabla(\nabla^{k-1}T)$. The second covariant derivative is often called the Hessian. For example, the Hessian of a scalar function f satisfies

$$\nabla^2 f(V,W) = \nabla_W(\nabla_V f) - \nabla_{\nabla_W V} f$$

and its coordinate expression is given by:

$$\nabla^2 f(\partial_i, \partial_j) = \partial^2_{ij} f - \textstyle\sum_k \Gamma^k_{ij}\partial_k f.$$

Since the latter is symmetric in i and j we see that $\nabla^2 f$ is itself symmetric:

$$\nabla^2 f(V,W) = \nabla^2 f(W,V).$$

(See exercise 2 for an alternative proof of this symmetry.) The symmetry result is perhaps expected as ordinary partial derivatives of a function commute. The correspondence between $\nabla\mathrm{grad}(f)$ and $\nabla^2 f$ is investigated in exercise 3.

In general, however, the Hessian is not symmetric! The following computation examines the symmetry (or the lack there of) of the Hessian of a vector field X:

$$\begin{aligned}\nabla^2 X(V,W) - \nabla^2 X(W,V) =& (\nabla_W(\nabla_V X) - \nabla_{\nabla_W V} X) - (\nabla_V(\nabla_W X) - \nabla_{\nabla_V W} X)\\ =& \nabla_W(\nabla_V X) - \nabla_V(\nabla_W X) - \nabla_{[W,V]} X.\end{aligned}$$

It is not at all clear if the latter simplifies, and so it may be a good idea to retreat to a familiar example and investigate the situation there.

Example: Consider the coordinatization of the 2-dimensional sphere of radius Λ involving longitude ($0 \le \theta \le 2\pi$) and latitude ($0 \le \phi \le \pi$). In these coordinates the standard Riemannian metric on the unit sphere takes the form of $\mathrm{d}\phi^2 + \sin^2(\phi)\mathrm{d}\theta^2$. Using the expression for the Christoffel symbols in terms of the metric components we easily find that $\nabla_{\partial_\theta}\partial_\phi = \nabla_{\partial_\phi}\partial_\theta = \frac{\cos(\phi)}{\sin(\phi)}\partial_\theta$. Thus, we have:

$$\begin{aligned}(\nabla^2\partial_\phi)(\partial_\phi, \partial_\theta) - (\nabla^2\partial_\phi)(\partial_\theta, \partial_\phi) &= \nabla_{\partial_\theta}(\nabla_{\partial_\phi}\partial_\phi) - \nabla_{\partial_\phi}(\nabla_{\partial_\theta}\partial_\phi) - \nabla_{[\partial_\theta,\partial_\phi]}\partial_\phi\\ &= -\nabla_{\partial_\phi}\left(\tfrac{\cos(\phi)}{\sin(\phi)}\partial_\theta\right) = \tfrac{1}{\sin^2\phi}\partial_\theta - \tfrac{\cos^2(\phi)}{\sin^2(\phi)}\partial_\theta = \partial_\theta. \quad \square\end{aligned}$$

The above example shows very clearly that we should not (in general) expect the Hessian of a vector field to be symmetric. Consequently, one needs to be very careful in situations where one would like to exchange $\nabla_V\nabla_W X$ for $\nabla_W\nabla_V X$, even when V and W are coordinate vector fields:

$$\nabla_{\partial_i}(\nabla_{\partial_j} X) \ne \nabla_{\partial_j}(\nabla_{\partial_i} X).$$

In this lecture, and throughout the course, we will use the notation $R(V,W)X$ to denote $\nabla_V(\nabla_W X) - \nabla_W(\nabla_V X) - \nabla_{[V,W]}X$; this gives us the formula

$$\nabla_V(\nabla_W X) = \nabla_W(\nabla_V X) + \nabla_{[V,W]}X + R(V,W)X.$$

A natural task at this stage is to investigate the properties of $R(V,W)X$. It follows from $R(V,W)X = \nabla^2 X(W;V) - \nabla^2 X(V;W)$ that R is tensorial and antisymmetric in V and W. Moreover, <u>*R is tensorial in X*</u>! This is a hugely important property that is well worthy of a verification by hand:

$$\begin{aligned} R(V,W)(\alpha X) =& \nabla_V(\nabla_W(\alpha X)) - \nabla_W(\nabla_V(\alpha X)) - \nabla_{[V,W]}(\alpha X) \\ =& \nabla_V\left((\nabla_W\alpha)X + \alpha\nabla_W X\right) - \nabla_W\left((\nabla_V\alpha)X + \alpha\nabla_V X\right) - \nabla_{[V,W]}(\alpha X) \\ =& \left\{\nabla_V(\nabla_W(\alpha)) - \nabla_W(\nabla_V(\alpha)) - \nabla_{[V,W]}\alpha\right\}X + \alpha R(V,W)X \\ =& \alpha R(V,W)X. \end{aligned}$$

Therefore, we should view $R(V,W)X$ as a $(3,1)$-tensor.

The definition of the curvature tensor. The $(3,1)$-tensor defined by

$$R(V,W)X := \nabla_V(\nabla_W X) - \nabla_W(\nabla_V X) - \nabla_{[V,W]}X$$

is called the *Riemann curvature tensor*. The "dual" covariant 4-tensor defined by

$$R(V,W,X,Y) := \langle R(V,W)X, Y\rangle$$

is also referred to as the Riemann curvature tensor. If/when notational confusion is possible we will use $\mathrm{Riem}(V,W,X,Y)$ in place of $R(V,W,X,Y)$.

It should be emphasized that there are two equally common sign conventions involving curvature. A number of geometry resources have the curvature tensor defined as the negative of our curvature tensor, so one should always familiarize themselves with curvature conventions of the text before citing any results.

We will spend the rest of this week studying the more geometric/visual/physical interpretations of the curvature tensor. For the remainder of the day we study its formal properties.

Curvature in Coordinates. For practice reasons, more than anything else, let us analyze the curvature in coordinates. By definition we have:

$$\begin{aligned} R(\partial_i,\partial_j)\partial_k =& \nabla_{\partial_i}(\nabla_{\partial_j}\partial_k) - \nabla_{\partial_j}(\nabla_{\partial_i}\partial_k) = \nabla_{\partial_i}(\textstyle\sum_l \Gamma^l_{jk}\partial_l) - \nabla_{\partial_j}(\textstyle\sum_l \Gamma^l_{ik}\partial_l) \\ =& \textstyle\sum_l \left(\partial_i(\Gamma^l_{jk})\partial_l + \Gamma^l_{jk}\nabla_{\partial_i}\partial_l - \partial_j(\Gamma^l_{ik})\partial_l - \Gamma^l_{ik}\nabla_{\partial_j}\partial_l\right) \\ =& \textstyle\sum_l \left(\partial_i(\Gamma^l_{jk})\partial_l + \sum_m \Gamma^l_{jk}\Gamma^m_{il}\partial_m - \partial_j(\Gamma^l_{ik})\partial_l - \sum_m \Gamma^l_{ik}\Gamma^m_{jl}\partial_m\right) \\ =& \textstyle\sum_l \left(\partial_i(\Gamma^l_{jk}) - \partial_j(\Gamma^l_{ik}) + \sum_m \Gamma^m_{jk}\Gamma^l_{im} - \sum_m \Gamma^m_{ik}\Gamma^l_{jm}\right)\partial_l. \end{aligned}$$

The exact nature of the above formula for R^l_{ijk} is not as important as much as its schematic form:

$$\partial\Gamma - \partial\Gamma + \Gamma\Gamma - \Gamma\Gamma.$$

Since the Christoffel symbols are made out of terms which look like $g^{\cdot\cdot}\partial_{\cdot}g_{\cdot\cdot}$, we see that the curvature components behave as a *non-linear second order differential operator of the components of the metric tensor*. Therefore, any sort of requirement placed on curvature becomes a de-facto non-linear second order differential equation on g. Many a problem in geometric analysis boils down to an analysis of (solutions of) such differential equations.

The schematic formula for R^l_{ijk} shows that the curvature tensor of the Euclidean $\mathbb{R}^n$ vanishes. It also shows that the presence of a non-zero curvature component (such as $R(\partial_\theta,\partial_\phi)\partial_\phi$ in the examples of the 2-dimensional sphere; see above) serves as an obstruction to finding coordinates in which the metric tensor takes the Euclidean form everywhere. In particular, we now have a proof of the fact that one can never find a to-scale accurate map of any – no matter how small – portion of the Earth! Manifolds are often depicted as spaces which locally look like $\mathbb{R}^n$. It is important to understand that this is true only on the level of coordinatizations and not on the level of geoMETRY.

Symmetries. Most of the time one is able to avoid dealing with the formula for the curvature components in full detail, but if absolutely necessary one can execute curvature computations at a fixed point P of M (which is OK due to the tensorial nature of Riem) in the coordinates with respect to which the metric components take the Euclidean form at P while their first derivatives vanish. *In such coordinates* we have

$$\begin{aligned} R_{ijkl} =& \partial_i(\Gamma^m_{jk})g_{ml} - \partial_j(\Gamma^m_{ik})g_{ml} \\ =& \tfrac{1}{2}(\partial^2_{ij}g_{kl} + \partial^2_{ik}g_{jl} - \partial^2_{il}g_{jk}) - \tfrac{1}{2}(\partial^2_{ji}g_{kl} + \partial^2_{jk}g_{il} - \partial^2_{jl}g_{ik}) \\ =& \tfrac{1}{2}(\partial^2_{ik}g_{jl} + \partial^2_{jl}g_{ik} - \partial^2_{jk}g_{il} - \partial^2_{il}g_{jk}). \end{aligned}$$

One nice aspect of the last formula is that it exhibits certain (anti-)symmetries of the curvature tensor. Identities such as $R_{ijkl} = -R_{jikl}$, $R_{ijkl} = R_{klij}$ and $R_{ijkl} + R_{jkil} + R_{kijl} = 0$ imply the following.

THEOREM 13.

(1) $R(V, W, X, Y) = -R(W, V, X, Y)$;

(2) $R(V, W, X, Y) = R(X, Y, V, W)$;

(3) $R(V, W, X, Y) + R(W, X, V, Y) + R(X, V, W, Y) = 0$.

These symmetries show that there is only one "interesting" curvature component for the standard (unit) 2-sphere: $R(\partial_\theta, \partial_\phi, \partial_\phi, \partial_\theta) = \langle \partial_\theta, \partial_\theta \rangle = \sin^2(\phi)$. Using the above we can express all the other curvature components in terms of $R(\partial_\theta, \partial_\phi, \partial_\phi, \partial_\theta)$. The presence of $\sin^2(\phi)$ is directly related to the fact that $\|\partial_\theta\|^2 = \sin^2(\phi)$. Since the sphere is rotationally symmetric we may normalize the above to conclude that

$$R(e_1, e_2, e_2, e_1) = 1$$

for any orthonormal set $\{e_1, e_2\}$ of tangential directions at any point of the sphere.

The last of the three identities in Theorem 13 is sometimes called The First Bianchi Identity. There is also something called the Second Bianchi Identity; its proof is left for an exercise.

THEOREM 14 (The Second Bianchi Identity).

$$\nabla R(V, W, X, Y; Z) + \nabla R(V, W, Y, Z; X) + \nabla R(V, W, Z, X; Y) = 0.$$

Exercises for Lecture 6.

(1) Let T be a $(2, 1)$-tensor. Prove the following formula. [BASIC, EXPECTED]

$$(\nabla_{\partial_i} T)^l_{jk} = \partial_i(T^l_{jk}) + \textstyle\sum_m \Gamma^l_{im} T^m_{jk} - \sum_m \Gamma^m_{ij} T^l_{mk} - \sum_m \Gamma^m_{ik} T^l_{jm}.$$

(2) Use the property $\nabla_V W - \nabla_W V = [V, W]$ and the definition of the bracket to show that $\nabla^2 f$ is symmetric. [BASIC]

(3) (a) Recall the definition of the map $\flat$ from the previous lecture. Then explain how Riemannian metrics can be used to make a correspondence between $(2, 0)$-tensors and $(1, 1)$-tensors. [INTERMEDIATE]

(b) Show that under the correspondence established above in (3a) $\nabla \mathrm{grad} f$ and $\nabla^2 f$ correspond to one another. More specifically, use Theorem 12 to show that $\langle \nabla_W \mathrm{grad}(f), V\rangle = \nabla^2 f(V, W)$ for all V and W. [INTERMEDIATE]

(4) How does $\nabla_V \nabla_W T$ relate to $\nabla_W \nabla_V T$ if T is

(a) a covector field? [BASIC, EXPECTED]

(b) a covariant 2-tensor field? [BASIC, EXPECTED]

(c) a $(1, 1)$-tensor field? [BASIC, EXPECTED]

(5) We proved Theorem 13 by means of a coordinate computation. To gain an insight into what proving this theorem without resorting to coordinates might look like, stick to the following guidelines. [ADVANCED]

(a) Consider the function $f = g(X, Y)$, where X and Y are some fixed vector fields. Compute $\nabla^2 f(V, W)$.

(b) Use the expression from the above to show that:

$$\nabla^2 f(V, W) - \nabla^2 f(W, V) = R(W, V, X, Y) + R(W, V, Y, X).$$

(c) Conclude that $R(W, V, X, Y) = -R(W, V, Y, X)$.

(6) Compute the non-vanishing curvature components of

(a) the 2-dimensional sphere of radius ϱ in $\mathbb{R}^3$; [BASIC, EXPECTED]

(b) the upper half-plane model of the hyperbolic plane. [BASIC, EXPECTED]

(7) Use curvature symmetries to show the following identity. The derivatives are to be evaluated at $(t, s) = (0, 0)$. [INTERMEDIATE]

$$\begin{aligned} R(W, V, X, Y) = &\frac{1}{6}\partial^2_{ts} R(V + tX, W + sY, W + sY, V + tX) \\ &-\frac{1}{6}\partial^2_{ts} R(V + tY, W + sX, W + sX, V + tY). \end{aligned}$$

(8) (a) Prove the following schematic formula for the components of ∇R. [ADVANCED]

$$R_{ijkl;m} = \left(\partial^2_{km}\Gamma^p_{il} - \partial^2_{lm}\Gamma^p_{ik}\right) g_{pj} + \text{ terms in } \Gamma\partial\Gamma \text{ and } \Gamma\Gamma\Gamma.$$

(b) Use the above to prove the second Bianchi Identity, Theorem 14. [ADVANCED]

(9) Revisit the claim that *as soon as we know how to differentiate scalar functions and vector fields we also know how to differentiate covector and more general tensor fields.* Apply it to the following example, which relates to something called the *Lie derivative*. The Lie derivative with respect to the vector field V, denoted $\mathcal{L}_V$, acts on functions and vector fields in the following manner:

- $\mathcal{L}_V f = \nabla_V f$;
- $\mathcal{L}_V X = [V, X]$.

Note that $\mathcal{L}_V$ is not tensorial in V. Based on the above, define the Lie derivatives $\mathcal{L}_V$ of a general (k, i)-tensor and show that the outcome is still a (k, i)-tensor. [INTERMEDIATE]

(10) Suppose the operation $D_V X$ on vector fields V and X satisfies the following:
- $D_V(c_1 X + c_2 Y) = c_1 D_V X + c_2 D_V Y$ for all constants c_1, c_2;
- $D_{\alpha V + \beta W} X = \alpha D_V X + \beta D_W X$ for all $\alpha, \beta \in \mathfrak{F}$;
- $D_V X - D_X V = [V, X]$;
- $\langle D_V X, Y\rangle + \langle X, D_V Y\rangle = \nabla_V(\langle X, Y\rangle)$.

(a) Prove the following. [Intermediate]

$$\begin{aligned} 2\langle D_X Y, Z\rangle =& \nabla_X(\langle Y, Z\rangle) + \nabla_Y(\langle Z, X\rangle) - \nabla_Z(\langle X, Y\rangle) \\ &+ \langle [X, Y], Z\rangle - \langle [Y, Z], X\rangle + \langle [Z, X], Y\rangle. \end{aligned}$$

(b) Show that $D = \nabla$. [Intermediate]

CHAPTER 2

Curvature of space and time

7. Intuiting curvature via Jacobi equation

Introduction. The concept of curvature was formally introduced in the previous lecture. The goal for today's lecture is to develop an intuitive understanding of this concept. We will do so by means of a differential equation (Jacobi equation) which controls the behavior of nearby geodesics. This differential equation features a particular curvature operator whose sign (or, rather, the sign of its eigenvalues) determines whether nearby geodesics are ultimately attracted to (e.g sphere) or forced away from one another (e.g. hyperbolic space). By the end of the lecture we will be able to prove our first *comparison theorems*, which take the schematic form of "if a manifold M curves at least as much as sphere S_ϱ of a particular radius ϱ then M is at most as big as S_ϱ." In particular, we will discuss Bonnet-Myers Theorems.

The Jacobi equation on examples. Assume $\gamma_0(t)$ is a geodesic belonging to a family of geodesics

$$\gamma_s(t) = \gamma(s,t) = (x^1(s,t), x^2(s,t), \dots, x^n(s,t))$$

which is smooth in the parameter s. The vector field $Y := \partial_s\gamma = \sum_i(\partial_s x^i)\partial_i$, defined along γ_0, provides a measure of separation between geodesics in our family which are infinitesimally close to γ_0. The following examples provide further insight into the nature of the vector field Y.

Example of $\mathbb{R}^n$: Let γ_0 be a straight line through a point with Cartesian coordinates $\mathbf{x}$, in the direction of the unit vector $\mathbf{u}$. Let $\mathbf{v}$ be a unit vector which is orthogonal to $\mathbf{u}$, and consider the following family of lines through $\mathbf{x}$:

$$\gamma(s,t) = \mathbf{x} + t(\cos s\,\mathbf{u} + \sin s\,\mathbf{v}).$$

We then have $Y = t\mathbf{v}$. For future purposes we note that $\ddot{Y} = 0$.

Example of S^n: Let γ_0 be a great circle through the point $\mathbf{x} \in S^n \subseteq \mathbb{R}^{n+1}$ of the standard unit sphere, initially going in the direction of the unit tangent vector $\mathbf{u}$ at $\mathbf{x}$. Taking advantage of the ambient $\mathbb{R}^{n+1}$ we may express γ_0 as

$$\gamma_0(t) = \cos t\,\mathbf{x} + \sin t\,\mathbf{u}.$$

Let $\mathbf{v}$ be a tangent vector at $\mathbf{x}$ which is orthogonal to $\mathbf{u}$. (This makes $\mathbf{x}$, $\mathbf{u}$ and $\mathbf{v}$ pairwise orthogonal.) Consider the following family of great circles through $\mathbf{x}$:

$$\gamma(s,t) = \cos t\,\mathbf{x} + \sin t\,(\cos s\,\mathbf{u} + \sin s\,\mathbf{v}).$$

We then have $Y = \sin t\,\mathbf{v}$; note that Y is tangent to S^n along γ_0. This vector field describes the rate at the which the geodesics through $\mathbf{x}$ are spreading apart. At first, due to $\sin(t) \approx t$, the geodesics spread at the rate similar to that in the Euclidean space. As t increases the rate of spreading out diminishes, and the geodesics are spread out the most at $t = \frac{\pi}{2}$. (Intuitively speaking: Meridians are spread out the most at the equator.) After $t = \frac{\pi}{2}$ our geodesics start approaching one another. The whole thing culminates at $t = \pi$ which is when all our geodesics meet at a point. Another way to look at all of this is through the differential equation

$$\ddot{Y} + Y = 0,$$

which our deviation vector field satisfies. There is an obvious structural similarity with the equation for a harmonic oscillator; the role of the spring constant is played by $+1$. In contrast, the corresponding equation in $\mathbb{R}^n$ is $\ddot{Y} = 0$ and its spring constant is 0. Relative to $\mathbb{R}^n$, therefore, the geometry of S^n is such that spreading geodesics are being pulled back by a spring-like mechanism.

Example of $\mathbb{H}^n$: Here we invoke the Poincaré disk model of the hyperbolic space. Recall that this model entails the unit disk $|\mathbf{x}| < 1$ and the metric $ds^2 = \frac{4}{(1-|\mathbf{x}|^2)^2}|d\mathbf{x}|^2$. In this model geodesics passing through the origin are usual (Euclidean) line segments. The parametrization of these geodesics takes the form of $t \mapsto f(t)\mathbf{u}$ where $\mathbf{u}$ denotes the tangent vector at the origin, but one should be cautious not to assume that these geodesics are parametrized as $t \mapsto t\mathbf{u}$. Since geodesics necessarily have constant (without loss of generality unit) speed we see that the function f is a solution to the initial value problem

$$\begin{cases} 2\frac{\dot{f}}{1-f^2} = 1, \\ f(0) = 0. \end{cases}$$

The unique solution of this problem (verify!) is $f(t) = \tanh(t/2)$. Overall, the geodesics discussed above take the form of $\gamma_0(t) = \tanh(t/2)\mathbf{u}$, with the Euclidean magnitude of $\mathbf{u}$ equal to 1. The vector field $\dot{\gamma_0} = \frac{1}{2\cosh^2(t/2)}\mathbf{u}$ is parallel and unit along γ_0. Next, consider the family of geodesics

$$\gamma(s,t) = \tanh(t/2)(\cos s\,\mathbf{u} + \sin s\,\mathbf{v}),$$

where $\mathbf{v}$ is orthogonal to $\mathbf{u}$ and also of Euclidean magnitude 1. The corresponding deviation vector field along γ_0 is

$$Y = \tanh(t/2)\mathbf{v}.$$

To be able to make a fair comparison with the previous two examples we should re-write Y as a multiple of a parallel unit vector field, in this case $\frac{1}{2\cosh^2(t/2)}\mathbf{v}$. (Check this vector field is indeed parallel along γ_0!) Using the identity $\sinh(t) = 2\sinh(t/2)\cosh(t/2)$ we get:

$$Y = \sinh(t)\left(\frac{1}{2\cosh^2(t/2)}\mathbf{v}\right).$$

As $\sinh(t)$ is increasing and asymptotic to $\frac{1}{2}e^t$ we see that geodesics on $\mathbb{H}^n$ repel away from each other. In fact, since $\frac{d^2}{dt^2}\sinh(t) = \sinh(t)$ the vector field Y satisfies

$$\ddot{Y} - Y = 0.$$

This again resembles the harmonic oscillator equation, but its "spring constant" is negative. One way to intuit this is to think of a hypothetical scenario is which the "spring" in a "mass-spring system" is not pulling back when the "spring" is stretched out of equilibrium but is instead pushing further away with the force proportional to the displacement from the equilibrium. Compounding such influences yields the displacement which, generically speaking, increases exponentially. Geodesics of hyperbolic space are subjected to such a mechanism. In view of this it should not be too surprising that circumferences of circles, surface areas of spheres and volumes of balls in hyperbolic space increase exponentially with the radius. We will make these claims more precise later on.

The analogy with the harmonic oscillator exists in general. It is provided by the *Jacobi equation*. Solutions of the Jacobi equation are called *Jacobi vector fields*.

THEOREM 15 (Jacobi Equation).

$$\ddot{Y} + R(Y, \dot{\gamma}_0)\dot{\gamma}_0 = 0.$$

PROOF. On the basis of $[\partial_t, \partial_s] = 0$ we have:

$$\ddot{Y} = \nabla_{\partial_t}(\nabla_{\partial_t}\partial_s) = \nabla_{\partial_t}(\nabla_{\partial_s}\partial_t) = \nabla_{\partial_s}(\nabla_{\partial_t}\partial_t) + R(\partial_t, \partial_s)\partial_t.$$

Since each ∂_t is a geodesic, we have $\nabla_{\partial_t}\partial_t = 0$ and thus $\ddot{Y} = R(\partial_t, Y)\partial_t$. The antisymmetry of the curvature tensor in the first two entries now proves the Jacobi equation. □

In view of the above examples, we are lead to the following interpretation of the operator $Y \mapsto R(Y, \dot{\gamma}_0)\dot{\gamma}_0$. Very loosely speaking, this operator is analogous to the "spring constant" of a harmonic oscillator and it describes the mechanism by which a geometry forces its geodesics to converge towards or repel away from one another. The more "positive" $Y \mapsto R(Y, \dot{\gamma}_0)\dot{\gamma}_0$ is, the more the geodesics are attracted to one another creating a more tightly curved geometry. (By the end of the lecture will be able to make this statement very precise.) On the flip side, the more "negative" $Y \mapsto R(Y, \dot{\gamma}_0)\dot{\gamma}_0$ becomes the more spread apart are the geodesics.

The definition and the examples of sectional curvature. At the end of the last lecture we saw that, due to multi-linearity and (anti-)symmetry properties of the curvature tensor, 2-dimensional manifolds really only have one independent, non-vanishing curvature component:

$$R(\partial_1, \partial_2, \partial_2, \partial_1).$$

Restricting to a tangent space at a point, this leads us to studying the values of $R(X, Y, Y, X)$ where $\{X, Y\}$ is any basis for the said tangent space. One can show (see exercise 2 below) that the value of

$$\mathscr{K} = \frac{R(X, Y, Y, X)}{|X|_g^2 |Y|_g^2 - g(X, Y)^2}$$

is independent of the choice of $\{X, Y\}$. When working with orthonormal bases $\{U, V\}$, the expression for $\mathscr{K}$ simply becomes $\mathscr{K} = R(U, V, V, U)$. The value of $\mathscr{K}$ identifies everything there is to know about the curvature tensor of a 2-dimensional manifold. In the case of the standard 2-dimensional sphere we identified this component to be identically equal to 1. (See the end of the last lecture.) The "1" we discovered in the last lecture is the same "1" which appears as a "spring constant" of the Jacobi equation.

When working with manifolds of general dimension each 2-dimensional subspace π of a tangent space generates its own value of $\mathscr{K}$. More specifically, the value of

$$\mathscr{K}(\pi) = \frac{R(X,Y,Y,X)}{|X|_g^2|Y|_g^2 - g(X,Y)^2}$$

is independent of the choice of a basis $\{X, Y\}$ of π. The function

$$\pi \mapsto \mathscr{K}(\pi)$$

is called *the sectional curvature* (M, g). It is very common to abuse the terminology and simply refer to the values of $\mathscr{K}(\pi)$ as "sectional curvature". We will do the same.

As was explored in exercise 7 following the last lecture, *sectional curvature determines the curvature tensor*. Thus having a complete knowledge of sectional curvature should enable us to get an explicit formula for the curvature tensor. We explore this claim on the example of S^n and $\mathbb{H}^n$.

In the examples from the beginning of this lecture we were free to chose vectors $\mathbf{u}$ and $\mathbf{v}$ as long as they were mutually orthogonal and unit. Through comparison with the Jacobi equation (as stated in the previous theorem), we conclude that:

$$R(\mathbf{v},\mathbf{u})\mathbf{u} = \begin{cases} \mathbf{v} & \text{in the case of } \ S^n \\ -\mathbf{v} & \text{in the case of } \ \mathbb{H}^n. \end{cases}$$

It follows that $R(V,U,U,V) = 1$ on S^n and $R(V,U,U,V) = -1$ on $\mathbb{H}^n$ for all orthonormal $\{U, V\}$. In other words, both S^n and $\mathbb{H}^n$ have constant sectional curvatures; their values are $\mathscr{K} = 1$ and $\mathscr{K} = -1$, respectively.

In the case of constant sectional curvature $\mathscr{K}$ we have that all (not necessarily linearly independent) tangent vectors X and Y satisfy

$$R(X,Y,Y,X) = \mathscr{K}\left(|X|_g^2|Y|_g^2 - g(X,Y)^2\right).$$

Combining this identity with the formula

$$\begin{aligned} R(W,V,X,Y) =& \frac{1}{6}\partial^2_{ts} R(V+tX, W+sY, W+sY, V+tX) \\ &- \frac{1}{6}\partial^2_{ts} R(V+tY, W+sX, W+sX, V+tY), \end{aligned}$$

with all the derivatives evaluated at $(t,s) = (0,0)$ (see yesterday's exercises) produces the following result.

THEOREM 16. *The standard Riemannian manifolds $\mathbb{R}^n$, S^n and $\mathbb{H}^n$ have constant sectional curvature $\mathscr{K} = 0$, $\mathscr{K} = 1$ and $\mathscr{K} = -1$, respectively. Their curvature tensors take the form*

$$R(V,W,X,Y) = \mathscr{K}\left(g(V,Y)g(W,X) - g(V,X)g(W,Y)\right).$$

It is important to note that the sectional curvature of a sphere of radius ϱ is

$$\mathscr{K} = \varrho^{-2}.$$

Consequently, the corresponding curvature tensor takes the form discussed in the previous theorem. We leave the proof as an exercise.

Manifolds of strictly positive sectional curvature. Loosely speaking, positive sectional curvature "forces" geodesics to converge towards one another. We now investigate one precise statement of this sort. It claims that a manifold which curves at least as much as a sphere of radius ϱ has to close up at least as fast as this sphere. In particular, distances between points of such a manifold cannot be bigger than $\pi\varrho$, i.e. cannot be bigger than the maximum distance between points of a sphere of radius ϱ.

THEOREM 17. *Suppose that the sectional curvature of a complete manifold (M, g) satisfies*

$$\mathscr{K} \geq \varrho^{-2}.$$

Then for all points $P, Q \in M$ we have that

$$d_g(P, Q) \leq \pi\varrho$$

and that the manifold M is compact.

PROOF. Let $P, Q \in M$. Since (M, g) is complete there is a minimizing geodesic $\gamma_0 : [0, L] \to M$ joining P and Q; without loss of generality we may assume that γ_0 is of unit speed. Our task now is to prove $L \leq \pi\varrho$.

To this end consider a vector field X which is parallel, unit and orthogonal to γ_0; such vector fields exists in abundance as they are all parallel transports of $X\big|_P$. Next, let

$$Y = \sin\tfrac{\pi t}{L} X.$$

The consideration of Y is inspired by the form of the Jacobi fields on spheres; note that Y vanishes at both P and Q. Finally, let γ_s be any (not necessarily geodesic) perturbation of γ_0 for which $\partial_s\big|_{s=0} = Y$. Since minimizing geodesics also minimize the energy functional[1], we see that

$$\frac{d^2E(\gamma_s)}{ds^2}\Big|_{s=0} \geq 0.$$

A general computation which is structurally very similar to the one in the proof of the Jacobi equation for geodesic deviations, as is thus left as an exercise, shows that

$$\frac{d^2E(\gamma_s)}{ds^2}\Big|_{s=0} = -\int_0^L \langle \ddot{Y} + R(Y, \dot{\gamma}_0)\dot{\gamma}_0, Y\rangle \, dt.$$

In our case we have that

$$\ddot{Y} = \tfrac{d^2}{dt^2}\left(\sin\tfrac{\pi t}{L}\right) X = -(\tfrac{\pi}{L})^2 Y,$$

because X is parallel along γ_0. Furthermore, since $\mathscr{K} \geq \varrho^{-2}$ we have

$$R(Y, \dot{\gamma}_0, \dot{\gamma}_0, Y) \geq \varrho^{-2}|Y|_g^2.$$

Inserting these two facts into the expression for the second derivative of the energy functional produces

$$\frac{d^2E(\gamma_s)}{ds^2}\Big|_{s=0} = -\int_0^L \langle -(\tfrac{\pi}{L})^2 Y + R(Y, \dot{\gamma}_0)\dot{\gamma}_0, Y\rangle \, dt \leq \left((\tfrac{\pi}{L})^2 - \varrho^{-2}\right)\int_0^L |Y|_g^2.$$

Since $\frac{d^2E(\gamma_s)}{ds^2}\big|_{s=0}$ is non-negative, we must have $(\frac{\pi}{L})^2 \geq \varrho^{-2}$ i.e. $L \leq \pi\varrho$. □

[1] $E(\gamma) = \frac{1}{2}\int_a^b |\dot{\gamma}|_g^2 \, dt$, see Lecture 3 for details.

Remark. Those familiar with the concept of the fundamental group will be interested to know that the fundamental group of a manifold with strictly positive curvature is necessarily finite. This is basically a consequence of the fact that the above argument applies to the universal cover of the manifold and that – as a result – the universal cover is compact.

The theorem we just went through was originally due to the 19th century French mathematician Pierre Ossian Bonnet; it was subsequently improved by Hopf and Rinow, and by an American mathematician Sumner Byron Myers. In fact, in early 1940's Myers published a much stronger result which is now known as *Myers' Theorem*. We proceed by investigating this theorem. In paragraphs below we adopt the notation from our proof of Theorem 17.

We begin by noting that instead of considering one single perturbation γ_s in the proof of Theorem 17, one may want to consider $n-1$ perturbations $\gamma_s^{(i)}$ each going in a different direction orthogonal to $\dot\gamma_0$. To construct such perturbations one should start with an orthonormal basis for the orthogonal complement $\dot\gamma_0^\perp$ within the tangent space at an endpoint of γ_0 and consider its parallel transport $\{X^{(1)}, \ldots, X^{(n-1)}\}$ along γ_0. We set

$$Y^{(i)} := \sin(\tfrac{\pi t}{L})X^{(i)}$$

and consider any perturbation $\gamma_s^{(i)}$ in the direction of $Y^{(i)}$. Reasoning as in the proof of Theorem 17 we obtain

$$\frac{d^2}{ds^2}E(\gamma_s^{(i)}) = -\int_0^L \langle -(\tfrac{\pi}{L})^2 Y^{(i)} + R(Y^{(i)}, \dot\gamma_0)\dot\gamma_0, Y^{(i)}\rangle\, dt \geq 0 \quad \text{for all} \quad i.$$

In terms of the orthonormal set $\{X^{(i)} | 1 \leq i \leq n-1\}$ the last inequality states:

$$\int_0^L \sin^2(\tfrac{\pi t}{L})\left((\tfrac{\pi}{L})^2 - R(X^{(i)}, \dot\gamma_0, \dot\gamma_0, X^{(i)})\right) \geq 0 \quad \text{for all} \quad i.$$

Adding these up produces

$$\int_0^L \sin^2(\tfrac{\pi t}{L})\left((n-1)(\tfrac{\pi}{L})^2 - \sum_i R(X^{(i)}, \dot\gamma_0, \dot\gamma_0, X^{(i)})\right) \geq 0 \quad \text{for all} \quad i.$$

From the last inequality we learn that the conclusions of Theorem 17 still hold if one *only* assumes that

$$\sum_i R(X^{(i)}, \dot\gamma_0, \dot\gamma_0, X^{(i)}) \geq (n-1)\varrho^{-2}.$$

Note that this condition permits some of the sectional curvatures to be negative.

The definition of Ricci curvature. The last observation points to the importance of the average sectional curvature $\mathscr{K}(\pi)$ over the set of π's which contain a particular direction, e.g. $\dot\gamma_0$. It is time to develop a linear-algebraic object which captures such an average. The basic idea here is to take a trace[2] of the full Riemann curvature tensor.

[2] Consult Lecture 5 for details on taking the trace.

An easy linear algebra exercise (see exercises following this lecture) shows that for two given tangent vectors V, W the value of $\sum_i R(e_i, V, W, e_i)$ is independent of the choice of orthonormal basis $\{e_i\}$ of tangent vectors. Consequently,

$$\mathrm{Ricci}(V, W) = \sum_i R(e_i, V, W, e_i)$$

defines a covariant 2-tensor. It is called *Ricci curvature tensor*, in honor of an Italian mathematician Gregorio Ricci-Curbastro who (together with Levi-Civita) developed tensor calculus. Based on exercises following Lecture 5 it should be easy to see that

$$\mathrm{Ricci}(V, W) = \sum_{ij} g^{ij} R(\partial_i, V, W, \partial_j) = \frac{n}{\mathrm{vol}(S^{n-1})} \int_{X \in S^{n-1}} R(X, V, W, X)\, \mathrm{dvol}_{S^{n-1}}$$

for the standard volume element $\mathrm{dvol}_{S^{n-1}}$. The latter formula makes it particularly clear that $\mathrm{Ricci}(V, V)$ is to be thought of as some sort of average of all sectional curvatures $\mathscr{K}(\pi)$ where π contains V.

Based on Theorem 16, Ricci curvature of an n-dimensional sphere of radius ϱ is found to be

$$\mathrm{Ricci}_{S^n_\varrho} = (n-1)\varrho^{-2} g$$

where $g = \varrho^2 g_{S^n}$ denotes the metric on the sphere of radius ϱ. Likewise, the Ricci curvature of the hyperbolic space $\mathbb{H}^n$ is

$$\mathrm{Ricci}_{\mathbb{H}^n} = -(n-1) g_{\mathbb{H}^n}.$$

The computational details are left as an exercise.

Theorems of S. B. Myers and S. Y. Cheng. The work presented above proves the following improvement of Theorem 17. Note that the remark we made after Theorem 17 regarding the fundamental group applies here as well.

THEOREM 18 (Myers' Theorem). *Suppose that the Ricci curvature of a complete n-dimensional manifold (M, g) satisfies*

$$\mathrm{Ricci} \geq (n-1)\varrho^{-2} g$$

in the sense that $\mathrm{Ricci}(V, V) \geq (n-1)\varrho^{-2} g(V, V)$ *for all tangent vectors V. Then for all points $P, Q \in M$ we have that*

$$d_g(P, Q) \leq \pi\varrho.$$

Furthermore, the manifold M is compact.

While we are discussing this subject we should mention a related *rigidity result*. It was proven by Shiu-Yuen Cheng in the 1970's. The proof of this theorem is too advanced for us right now.

THEOREM 19. *If (M, g) is a complete n-dimensional Riemannian manifold with*

$$\mathrm{Ricci} \geq (n-1)\varrho^{-2} g \quad \textit{and} \quad \max_{P,Q \in M} d_g(P, Q) = \pi\varrho$$

then M is (isometric to) the standard sphere S^n_ϱ of radius ϱ.

Exercises for Lecture 7.

(1) Work through all the details of the opening example of the hyperbolic space. [Basic]

(2) Let π be a 2-dimensional subspace of the tangent space at a point of an n-dimensional manifold (M, g). Show that the value of

$$\frac{R(X,Y,Y,X)}{|X|_g^2|Y|_g^2 - g(X,Y)^2}$$

is independent of the choice of a basis $\{X, Y\}$ of π. [Basic, Expected]

(3) Compute the following.

(a) The curvature tensor, the sectional curvature and the Ricci curavture of the sphere of radius ϱ in $\mathbb{R}^{n+1}$. [Basic, Expected]

(b) The curvature tensor, the sectional curvature and the Ricci curavture of the unit disk $|\mathbf{x}| < 1$ with the metric $ds^2 = \frac{4K^2}{(1-|\mathbf{x}|^2)^2}|d\mathbf{x}|^2$. [Basic, Expected]

(4) The covariant derivative ∇R for Riemannian manifolds of constant sectional curvature. [Intermediate]

(5) This problem investigates the space of Jacobi vector fields from the perspective of initial and/or boundary value problems. [Intermediate]

(a) Apply the basic existence and uniqueness theorems for initial value problems to show that the space of Jacobi vector fields along a geodesic of an n-dimensional manifold is $2n$-dimensional.

(b) Suppose a Jacobi vector field Y along a geodesic γ_0 is such that both Y and $\dot{Y}$ vanish at one of the endpoints of γ_0. Show that $Y = 0$.

(c) Is it possible to have a non-trivial ($Y \neq 0$) Jacobi vector field which vanishes at two distinct points?

(6) This problem investigates the decomposition of the space of Jacobi vector fields into vector fields which are orthogonal and vector fields which are tangential to γ_0. [Intermediate]

(a) Let Y be Jacobi field along a geodesic γ_0. Compute $\frac{d^2}{dt^2}g(Y, \dot{\gamma}_0)$.

(b) Show that a Jacobi field Y along γ_0 is everywhere orthogonal to γ_0 if and only if both Y and $\dot{Y}$ are orthogonal to γ_0 at one of its endpoints.

(c) Show that a Jacobi field along γ_0 is everywhere orthogonal to γ_0 if and only if it is orthogonal to γ_0 at two distinct points.

(d) Show that Jacobi fields along a geodesic γ_0 which are everywhere orthogonal to γ_0 form a $2n-2$-dimensional subspace of the space of all Jacobi fields along γ_0.

(e) Show that the space of Jacobi fields which are tangential to γ_0 is 2-dimensional. Visualize one particular set of basis elements for this space.

(7) (a) At the beginning of our course we introduced geodesics as critical points of the energy functional $E(\gamma) = \frac{1}{2}\int_a^b |\dot\gamma|_g^2\, dt$; geodesics γ_0 have the property that $\frac{dE(\gamma_s)}{ds}\big|_{s=0}$ vanishes for all small perturbations γ_s of γ_0 with $\gamma_s(a) = \gamma_0(a)$ and $\gamma_s(b) = \gamma_0(b)$ for all small s. In this process we left one very natural question unanswered: Is it possible to detect if a geodesic is a minimizing geodesic by studying the Hessian

$$\frac{d^2E(\gamma_s)}{ds^2}\Big|_{s=0}$$

of the energy functional? What is the main reason why we had to postpone this discussion? [Basic]

(b) Adopt the framework outlined above, and let $Y = \partial_s\big|_{s=0}$. Verify the formula

$$\frac{d^2E(\gamma_s)}{ds^2}\Big|_{s=0} = -\int_a^b \langle \ddot Y + R(Y,\dot\gamma_0)\dot\gamma_0, Y\rangle\, dt$$

for the Hessian of the energy functional. [Intermediate]

(c) Continue with the above framework. Compute the corresponding formula for the Hessian $\frac{d^2L(\gamma_s)}{ds^2}\Big|_{s=0}$ of the length functional. [Intermediate]

(8) Let $\gamma_0 : [a,b] \to M$ be a geodesic, and let $\mathfrak{X}_0$ denote the space of vector fields along γ_0 which vanish at the endpoints $\gamma_0(a)$ and $\gamma_0(b)$.

(a) Show that $I(X,Y) = -\int_a^b \langle X, \ddot Y + R(Y,\gamma_0)\gamma_0\rangle\, dt$ is symmetric and $\mathbb{R}$-bilinear map on $\mathfrak{X}_0$. [Basic]

(b) Show that if γ_0 is a minimizing geodesic then $I(Y,Y) \geq 0$ for all $Y \in \mathfrak{X}_0$. [Basic]

(9) (a) Let V and W be two tangent vectors. Prove that the value of $\sum_i R(e_i, V, W, e_i)$ is independent of the choice of orthonormal basis $\{e_i\}$ of tangent vectors. [Basic]

(b) Prove the coordinate formula $\mathrm{Ricci}(V,W) = \sum_{ij} g^{ij} R(\partial_i, V, W, \partial_j)$. [Intermediate]

(10) Show that on 3-dimensional manifolds Ricci curvature determines sectional curvature. Hint: For orthonormal bases $\{e_1, e_2, e_3\}$ we have that $\mathrm{Ricci}(e_1,e_1) = \mathscr{K}(\mathrm{Span}(e_1,e_2)) + \mathscr{K}(\mathrm{Span}(e_1,e_3))$. What can you say about $\mathrm{Ricci}(e_2,e_2)$ and $\mathrm{Ricci}(e_3,e_3)$? [Intermediate]

8. Comparison Theorems

Introduction. The word "curvature" can stand for many things. The most general of all is the Riemann curvature tensor, a covariant 4-tensor which measures the lack of symmetry of Hessian on vector fields. The word "curvature" may also refer to sectional curvature

$$\mathscr{K}\,(\mathrm{Span}(X,Y)) = \frac{R(X,Y,Y,X)}{|X|_g^2|Y|_g^2 - g(X,Y)^2}.$$

In the last lecture we explored how sectional curvature relates to the Jacobi equation, and the fact that sectional curvature determines the full curvature tensor. At times the word "curvature" is used in the context of the Ricci curvature tensor

$$\mathrm{Ricci}(V, W) = \frac{n}{\mathrm{vol}(S^{n-1})} \int_{X \in S^{n-1}} R(X, V, W, X)\, \mathrm{dvol}_{S^{n-1}}$$

or even its associated quadratic form $\mathrm{Ricci}(V, V)$. The latter can be thought of as an average of sectional curvatures of 2-dimensional planes containing the vector V. Finally, there is also something called *scalar curvature*. It is the trace of the Ricci tensor. In other words, we have

$$\mathrm{Scal} := \sum_i \mathrm{Ricci}(e_i, e_i) = \sum_{ij} g^{ij} \mathrm{Ricci}_{ij} = \frac{n}{\mathrm{vol}(S^{n-1})} \int_{X \in S^{n-1}} \mathrm{Ricci}(X, X)\, \mathrm{dvol}_{S^{n-1}},$$

where $\{e_i\}$ denotes any orthonormal basis of tangent vectors. One should think of Scal as being an average of all sectional curvatures. It is easy to see that the scalar curvature of the standard n-dimensional sphere of radius ϱ is $n(n-1)\varrho^{-2}$, while the scalar curvature of the hyperbolic space $\mathbb{H}^n$ is $-n(n-1)$. The goal for today's lecture is to gain a more thorough understanding of all these concepts of curvature, mostly by going through a number of geometric comparison theorems.

More detailed analysis of Jacobi fields for radial families of geodesics. Study of geodesics emanating from a point P is most easily done in coordinates arising from the exponential map $\exp_P$. These coordinates were very briefly mentioned is Lecture 4. For reader's convenience we now summarize their key features[3].

- At the very least these coordinates are defined on a small ball $B^n_{\mathbf{R}}$ centered at the origin of $\mathbb{R}^n$. In the case of unit spheres one can take $\mathbf{R} = \pi$, while in the case of the hyperbolic space it turns out one may take $R = \infty$, i.e. $B^n_{\mathbf{R}} = \mathbb{R}^n$.
- In these coordinates P corresponds to $\mathbf{0}$. We will abuse the notation and identify P with $\mathbf{0}$ from now on.
- Assuming an employment of Cartesian coordinates $(x^1, \ldots, x^n)$ within the coordinate ball $B^n_{\mathbf{R}}$, the metric components and the Christoffel symbols at P satisfy

$$g_{ij}(P) = \begin{cases} 1 & \text{if } \ i = j \\ 0 & \text{if } \ i \neq j \end{cases} \quad \text{and} \quad \Gamma^k_{ij}(P) = 0.$$

- In these coordinates unit speed geodesics passing through P take the form $\gamma(t) = t\mathbf{u}$ where $\mathbf{u}$ is some unit vector based at the origin. In accordance to this, and in accordance with the standard notation used in the context of spherical coordinates, we often use r in place of t.
- The Euclidean sphere $S^{n-1}_{\mathbf{r}} \subseteq B^n_{\mathbf{R}}$ of radius r centered at the origin quite literally corresponds to the sphere on (M, g) with radius $\mathbf{r}$ and center P. (The latter refers to the set of all points $Q \in M$ such that $d_g(P, Q) = \mathbf{r}$.) Similar statement can be made about the ball $B^n_{\mathbf{r}} = \{\mathbf{x} \in \mathbb{R}^n \big| |\mathbf{x}| < \mathbf{r}\}$.

[3]Rigorous proofs for the most part require a bit more technical skill than one can obtain from these lecture notes.

- Assuming an employment of spherical coordinates $(r, \mathbf{u}) \in (0, \infty) \times S^{n-1}$ on $B^n_{\mathbf{R}}$, the metric g can be expressed as
$$g = dr^2 + h(r),$$
where $h(r)$ is a smooth family of metrics on S^{n-1}. Trivial examples of this include the expressions for the Euclidean metric in polar and spherical coordinates:
$$ds^2 = dr^2 + r^2 d\theta^2, \quad ds^2 = dr^2 + r^2 ds^2_{S^{n-1}},$$
where $ds^2_{S^{n-1}}$ denotes the line element on the standard unit sphere. More generally, in dimension 2 the metric g can always be expressed as
$$ds^2 = dr^2 + f(r, \theta)^2 d\theta^2.$$
Note that there can be no cross-terms (e.g $dr\, d\theta$) because ∂_r has to be orthogonal to the level set of the function r. The particular formula for the metric (components) in *geodesic polar coordinates* discussed here is often called the *Gauss' Lemma.*

Consider a unit speed geodesic $\gamma_0(r) = r\mathbf{u}$ through P[4], and its variation
$$\gamma_s(r) = r(\cos s\, \mathbf{u} + \sin s\, \mathbf{v})$$
where $\mathbf{v}$ is a unit vector at P which is orthogonal to $\mathbf{u}$. The variation vector field $Y = \partial_s\big|_{s=0}$ can be identified as
$$Y(r) = r\mathbf{v}.$$
Furthermore, we have $Y(0) = \mathbf{0}$ and $\dot{Y}(0) = \mathbf{v}$.

The constant vector field $U = \mathbf{u}$, being the velocity field of the geodesic γ_0, is parallel along γ_0. However, we need to be very careful not to assume that the constant vector field $\mathbf{v}$ is parallel along γ_0. Instead, let us introduce the parallel transport V of $\mathbf{v}$ along γ_0. Basic Mean Value/Taylor Theorems give us the asymptotic formula
$$Y = rV + O(r^2),$$
where $O(r^2)$ signifies that the error of the approximation $Y \approx rV$ grows no faster than some constant multiple of r^2.

Our next task is to push this Taylor-like approximation further. Temporarily set $Y \approx rV + \frac{r^2}{2}V_2 + \frac{r^3}{3!}V_3 + \dots$ where $V_2, V_3, \dots$ are some parallel vector fields[5] along γ_0. Inserting this expansion into the Jacobi equation $\ddot{Y} + R(Y, U)U = 0$ produces
$$W_2 + rW_3 + \dots = -rR(V, U)U + \dots.$$
It follows that $W_2(0)$, and thus W_2 itself, vanishes while
$$W_3 = -R(V, U)U + O(r).$$
Overall, we have

$$Y = rV - \frac{r^3}{3!}R(V, U)U + O(r^4). \tag{3}$$

---

[4]Note that the variable r plays the role of the variable t from the previous lecture(s).

[5]Every vector field along γ_0 permits such an expansion. To see this consider an orthonormal basis at an endpoint of γ_0 and its parallel transport along γ_0. Decomposing the given vector field into a linear combination of the described basis fields, followed by an application of the Taylor Theorem to the resulting coefficients, proves our claim.

Applications to 2-dimensional manifolds. We now specialize the above framework to 2-dimensional manifolds. Along $\gamma_0(r) = r\mathbf{u}$ the constant vector field $\mathbf{v}$ can be identified with $\frac{1}{r}\partial_\theta$. Recall that, by the Gauss Lemma, $|\partial_\theta|_g = f(r,\theta)$. In dimension 2 the vector fields $\mathbf{v}$ and V are necessarily proportional and thus:

$$Y = fV.$$

The asymptotic formula (3) shows that

$$f(r) = g(Y,V) = r - \tfrac{r^3}{3!}R(U,V,V,U) + O(r^4).$$

By taking advantage of $R(U,V,V,U) = R(\mathbf{u},\mathbf{v},\mathbf{v},\mathbf{u}) + O(r)$ we obtain

$$ds^2 = dr^2 + \left(r - \mathscr{K}_P\frac{r^3}{3!} + O(r^4)\right)^2 d\theta^2,$$

where $\mathscr{K}_P$ denotes the value of sectional curvature at P.

The most immediate application of this result yields an asymptotic formula for the circumferences of small circles and areas of small disks centered at P.

THEOREM 20. *The circumference $C_{\mathbf{r},P}$ of $S^1_{\mathbf{r}}(P) = \{Q \in M \big| d_g(P,Q) = \mathbf{r}\}$ and the surface area $A_{\mathbf{r},P}$ of $B^2_{\mathbf{r}}(P) = \{Q \in M \big| d_g(P,Q) \leq \mathbf{r}\}$ for $\mathbf{r} \ll 1$ satisfy*

$$\begin{aligned} C_{\mathbf{r},P} =& 2\pi\mathbf{r} - \mathscr{K}_P\frac{\pi\mathbf{r}^3}{3} + O(\mathbf{r}^4), \\ A_{\mathbf{r},P} =& \pi\mathbf{r}^2 - \mathscr{K}_P\frac{\pi\mathbf{r}^4}{12} + O(\mathbf{r}^5). \end{aligned}$$

For proof one just needs to integrate $C_{\mathbf{r},P} = \int_0^{2\pi} \mathbf{r} - \mathscr{K}_P\frac{\mathbf{r}^3}{3!} + O(\mathbf{r}^4)\, d\theta$ and $A_{\mathbf{r},P} = \int_0^{\mathbf{r}}\int_0^{2\pi} r - \mathscr{K}_P\frac{r^3}{3!} + O(r^4)\, d\theta\, dr$. It follows that in positive sectional curvature small circles/disks have consistently smaller circumference/surface area than the corresponding circles/disks in Euclidean space. On the other hand, negative sectional curvature has the opposite effect: circumferences and surface areas are bigger than what one would expects from the Euclidean experience. To fully appreciate what this means from the visual standpoint the reader is encouraged to do exercise 5 below, in which one constructs a paper version of a "discretized" hyperbolic plane.

The above theorem shows that sectional curvature is related to higher order corrections to $C_{\mathbf{r},P} \approx 2\pi\mathbf{r}$ and $A_{\mathbf{r},P} \approx \pi\mathbf{r}^2$ for small $\mathbf{r}$. More precisely, we have

$$\mathscr{K}_P = \lim_{\mathbf{r}\to 0}\frac{2\pi\mathbf{r} - C_{\mathbf{r},P}}{\frac{\pi\mathbf{r}^3}{3}} = \lim_{\mathbf{r}\to 0}\frac{\pi\mathbf{r}^2 - A_{\mathbf{r},P}}{\frac{\pi\mathbf{r}^4}{12}}.$$

It is in this sense of the word that the sectional curvature of a 2-dimensional manifold can be viewed as a relative failure of the circumference/surface area of small circles/disks to be what they are in Euclidean space. It is important to understand that, as such, sectional curvature can be measured by a "Flatlander" − from a purely intrinsic perspective. Any result of this type in dimension 3 would allow us to detect if our 3-dimensional space is curved!

Ideas of this sort go back to at least Gauss. Elementary Riemannian geometry texts do not fail to provide an interpretation of sectional curvature at a point of a 2-dimensional manifold as

$$\mathscr{K}_P = \lim_{Q,R\to P}\frac{\angle P + \angle Q + \angle R - \pi}{\mathrm{Area}(\Delta PQR)},$$

where ΔPQR refers to a triangle whose sides are geodesics (geodesic triangle). Anyone taking a more thorough Riemannian geometry course will most definitely see a rigorous proof of this result. For this reason, we only discuss some heuristic reasons which might bring one to the result in the first place. For simplicity we focus on the case when $\mathscr{K}_P > 0$. In that case we have

$$\begin{aligned} ds^2 = dr^2 + \left(r - (\mathscr{K}_P)\frac{r^3}{3!} + O(r^4)\right)^2 d\theta^2 =& dr^2 + \frac{\sin^2((\mathscr{K}_P)^{1/2}r)}{\mathscr{K}_P}d\theta^2 + O(r^4), \\ =& \frac{1}{\mathscr{K}_P}\left(d\rho^2 + \sin^2(\rho)d\theta^2\right) + O(\rho^4), \end{aligned}$$

for $\rho = (\mathscr{K}_P)^{1/2}r$. The point is that locally the metric g is very well approximated by the metric on the sphere of radius $(\mathscr{K}_P)^{-1/2}$. On this sphere the formula for the area of geodesic triangles states that

$$\text{Area}(\Delta PQR) = (\mathscr{K}_P)^{-1}(\angle P + \angle Q + \angle R - \pi), \quad \text{i.e} \quad \mathscr{K}_P = \frac{\angle P + \angle Q + \angle R - \pi}{\text{Area}(\Delta PQR)}$$

To a very high order this formula is also accurate for small geodesic triangles on (M, g) with one vertex at P.

The result we just discussed is better known in its integral form, and under the name of Gauss' Theorem.

Theorem 21 (Gauss' Theorem). *If ΔPQR is a geodesic triangle on a Riemannian manifold (M, g) then*

$$\angle P + \angle Q + \angle R - \pi = \int_{\Delta PQR} \mathscr{K}\ \text{dvol}_g.$$

We should mention that there are some important generalizations of this theorem. In today's exercises you will be guided through *Gauss-Bonnet Theorem*, a result which relates total sectional curvature $\int_M \mathscr{K}$ dvol of a 2-dimensional compact manifold without boundary to its overall shape (topology). One of the consequences of the Gauss-Bonnet Theorem is that the total sectional curvature of compact Riemannian manifold is the same for *all* metrics. In the case of the sphere this common value is always 4π!

Asymptotic formula for volumes of small balls in arbitrary dimension. The asymptotic formula (3) can be used to prove analogues of Theorem 20 in general dimension. To make this possible we need to relate the volume element on (M, g) to its Jacobi vector fields.

Adopt the notation used in the discussion leading to the asymptotic formula (3), and choose Cartesian coordinates on $B^n_{\mathbf{R}}$ so that $U = \partial_n$ along γ_0. Letting the role of $\mathbf{v}$ be played by each and every ∂_i with $1 \leq i \leq n-1$ leads us to $n-1$ Jacobi vector fields

$$Y_i = r\partial_i, \quad 1 \leq i \leq n-1.$$

If V_i denote the parallel transports of $\partial_i\big|_P$ along γ_0, then

$$Y_i = rV_i - \tfrac{r^3}{3!}R(V_i, U)U + O(r^4).$$

Being tangential to the coordinate spheres the vector fields $\partial_1, \ldots, \partial_{n-1}$ are orthogonal to ∂_n along γ_0. (Compare with the decomposition $g = dr^2 + h(r)$

claimed by the Gauss' Lemma.) It follows that along γ_0 the volume element can be expressed as:

$$
\begin{aligned}
\mathrm{dvol}_g\big|_{(r,\mathbf{u})} &= \sqrt{\det([\langle\partial_i,\partial_j\rangle]_{1\le i,j\le n})}\,\mathrm{d}x^1\ldots\mathrm{d}x^n \\
&= \sqrt{\det([\langle\partial_i,\partial_j\rangle]_{1\le i,j\le n-1})}\, r^{n-1}\mathrm{d}r(\mathrm{dvol}_{S^{n-1}})\big|_{\mathbf{u}} \\
&= \sqrt{\det[\langle Y_i, Y_j\rangle]}\,\mathrm{d}r(\mathrm{dvol}_{S^{n-1}})\big|_{\mathbf{u}}.
\end{aligned}
\tag{4}
$$

It should be noted that this particular formula for the volume element is of interest in its own right. For example, our understanding of Jacobi vector fields on the hyperbolic space ($Y = \sinh(r)V$, with V parallel, unit and orthogonal to γ_0) leads to $Y_i = \sinh(r)V_i$ and

$$\mathrm{dvol}_g\big|_{(r,\mathbf{u})} = \sinh^{n-1}(r)\,\mathrm{d}r(\mathrm{dvol}_{S^{n-1}})\big|_{\mathbf{u}}.$$

Thus the volume of a ball $B_{\mathbf{r}}^{\mathbb{H}^n}$ of radius $\mathbf{r}$ in hyperbolic space is given by

$$\mathrm{vol}(B_{\mathbf{r}}^{\mathbb{H}^n}) = \int_{S^{n-1}}\int_0^{\mathbf{r}} \sinh^{n-1}(r)\,\mathrm{d}r\,\mathrm{dvol}_{S^{n-1}}.$$

For example, in dimension 2 we have

$$\mathrm{vol}(B_{\mathbf{r}}^{\mathbb{H}^2}) = \int_0^{2\pi}\int_0^{\mathbf{r}} \sinh(r)\,dr d\theta = 2\pi(\cosh(\mathbf{r})-1).$$

Note that for large $\mathbf{r}$ we have

$$\mathrm{vol}(B_{\mathbf{r}}^{\mathbb{H}^2}) \approx \pi e^{\mathbf{r}}.$$

In other words, the volume grows exponentially! Although the exact expression for the volume is a bit more complicated, the same holds in general dimension. There is a sequence of constant c_n such that

$$\mathrm{vol}(B_{\mathbf{r}}^{\mathbb{H}^n}) \approx c_n e^{(n-1)\mathbf{r}} \quad \text{if} \quad \mathbf{r} \gg 1.$$

We now go back to our task of studying the small geodesic spheres/balls; our next step is to compute the asymptotic expansion of $\sqrt{\det[\langle Y_i, Y_j\rangle]}$. It follows from $Y_i = rV_i - \frac{r^3}{3!}R(V_i,U)U + O(r^4)$, and the fact that $\{V_i\}$ are everywhere orthonormal, that

$$[\langle Y_i, Y_j\rangle] = r^2\mathrm{Id} - \tfrac{r^4}{3}[R(V_i,U,U,V_j)] + O(r^5).$$

In general, one has $\det(\mathrm{Id} + r^2 A) = 1 + r^2\mathrm{Tr}(A) + O(r^4)$. Applying this identity to our situation (e.g. using the $(n-1)\times(n-1)$ matrix $A = -\frac{1}{3}[R(V_i,U,U,V_j)]$) yields

$$\det[\langle Y_i, Y_j\rangle] = r^{2(n-1)}\left(1 - \tfrac{r^2}{3}\mathrm{Ricci}(U,U) + O(r^4)\right).$$

In view of $\mathrm{Ricci}(U,U) = \mathrm{Ricci}(\mathbf{u},\mathbf{u}) + O(r)$, this means that

$$\sqrt{\det[\langle Y_i, Y_j\rangle]} = r^{n-1} - \mathrm{Ricci}(\mathbf{u},\mathbf{u})\tfrac{r^{n+1}}{6} + O(r^{n+2}).$$

Overall, we have

$$\mathrm{dvol}_g\big|_{(r,\mathbf{u})} = \left(r^{n-1} - \mathrm{Ricci}(\mathbf{u},\mathbf{u})\tfrac{r^{n+1}}{6} + O(r^{n+2})\right)\,\mathrm{d}r(\mathrm{dvol}_{S^{n-1}})\big|_{\mathbf{u}}.$$

Integrating this, and using the formula $\int_{\mathbf{u}\in S^{n-1}}\mathrm{Ricci}(\mathbf{u},\mathbf{u})\,\mathrm{dvol}_{S^{n-1}} = \frac{\mathrm{vol}(S^{n-1})}{n}\mathrm{Scal}$, proves the following theorem:

THEOREM 22. *Let* $\text{vol}(B^M_{\mathbf{r}}(P))$ *denote the volume of the geodesic ball*

$$B^M_{\mathbf{r}}(P) = \{Q \in M \big| d_g(P,Q) \le \mathbf{r}\} \subseteq M$$

of radius $\mathbf{r}$ *centered at* P. *For sufficiently small* $\mathbf{r} \ll 1$ *we have*

$$\text{vol}(B^M_{\mathbf{r}}(P)) = \left(\frac{\mathbf{r}^n}{n} - \frac{\text{Scal}\,(P)}{6n(n+2)}\mathbf{r}^{n+2} + O(\mathbf{r}^{n+3})\right)\text{vol}(S^{n-1}).$$

In particular, if $\text{vol}(B^{\mathbb{R}^n}_{\mathbf{r}}) = \frac{\mathbf{r}^n}{n}\text{vol}(S^{n-1})$ *denotes the volume of the standard ball of radius* $\mathbf{r}$ *in* $\mathbb{R}^n$ *then*

$$\frac{\text{vol}(B^M_{\mathbf{r}}(P))}{\text{vol}(B^{\mathbb{R}^n}_{\mathbf{r}})} = 1 - \frac{\text{Scal}\,(P)}{6(n+2)}\mathbf{r}^2 + O(\mathbf{r}^3).$$

Bottom line? This theorem provides a very geometric understanding of scalar curvature: the scalar curvature at P determines higher order corrections to

$$\text{vol}(B^M_{\mathbf{r}}(P)) \approx \text{vol}(B^{\mathbb{R}^n}_{\mathbf{r}}), \quad \mathbf{r} \ll 1.$$

Exercises for Lecture 8.

(1) Prepare a short essay in which you are explaining different concept(s) of curvature (sectional, Ricci, scalar) to somebody with basic understanding of mathematics (e.g somebody at the level of basic calculus). Illustrate the subject of Riemannian geometry on the example of comparison theorems. Take my word for this: developing the ability to explain your math to a lay person will pay huge dividends. [EXPECTED]

(2) Verify the formula for the circumference of small circles and the surface area of small disks on 2-dimensional spheres through a direct computation. [BASIC]

(3) (a) Complete all the asymptotic computations behind Theorem 22. [INTERMEDIATE]

(b) Compute the asymptotic formula for the "surface area" of the geodesic sphere $S^M_{\mathbf{r}}(P)$ of small radius $\mathbf{r}$ centered at the point $P \in M$. [BASIC]

(4) Prove that

$$\text{vol}(S^{2n}) = \frac{(4\pi)^n(n-1)!}{(2n-1)!}, \quad \text{vol}(S^{2n+1}) = 2\frac{\pi^{n+1}}{n!}$$

by using the volume element formula (4). [INTERMEDIATE]

(5) You will need a ruler, lots of paper, scissors and a fair amount of tape for this exercise. [BASIC, EXPECTED]

- Tile a sheet of paper with regular triangles.
- Make several copies (four or five should suffice) of the above.
- Cut some of your sheets into individual triangles. Note that taping the triangles together so that at each vertex exactly six of them meet reproduces our (Euclidean) sheet of paper.
- Now tape the triangles together so that at each vertex exactly five of the triangles meet. What surface are you reproducing? Comment on how this relates to Theorem 20.

- Now tape the triangles together so that at each vertex exactly seven of the triangles meet. The extremely ruffly surface you are creating is a discretized version of a particular geometry. Which geometry is that?

(6) Divergence of a covariant 2-tensor T is defined as the trace of ∇T in the following sense:
$$(\mathrm{div}T)_i = \sum_{jk} g^{jk} T_{ij;k}.$$
Use the second Bianchi identity to show that tensor $G = \mathrm{Ricci} - \frac{1}{2}\mathrm{Scal}\, g$ is divergence free. (The tensor G is called the Einstein tensor; it is used in general relativity.) [Advanced]

(7) Riemannian manifolds whose Ricci tensor is proportional to the metric tensor, $\mathrm{Ricci} = \lambda g$, are called *Einstein manifolds*.

(a) Show that in the case of n-dimensional Einstein manifolds we necessarily have $\mathrm{Ricci} = \frac{\mathrm{Scal}}{n} g$ i.e $\lambda = \frac{\mathrm{Scal}}{n}$. [Basic]

(b) Use the second Bianchi identity to prove that
$$\nabla \mathrm{Scal} = 2\nabla\lambda$$
on Einstein manifolds; conclude that in dimensions $n \geq 3$ Einstein manifolds necessarily have constant scalar curvature. [Advanced]

(8) Show that 3-dimensional Einstein manifolds (see the previous problem for definition) necessarily have constant sectional curvature. Hint: review yesterday's exercises. [Intermediate]

(9) There is another curvature tensor of interest. It is called the *Weyl tensor*. It is defined for manifolds of dimension $n \geq 3$.
$$W := R - \frac{\mathrm{Scal}}{2n(n-1)}\, g \cdot g - \frac{1}{n-2}\left(\mathrm{Ricci} - \frac{\mathrm{Scal}}{n}\, g\right) \cdot g,$$
where the operation $\cdot$ stands for the Kulkarni-Nomizu product $h \cdot k$ of two symmetric covariant 2-tensors:
$$\begin{aligned} h \cdot k\,(V_1, V_2, V_3, V_4) = {} & h(V_1, V_3)k(V_2, V_4) + h(V_2, V_4)k(V_1, V_3) \\ & - h(V_1, V_4)k(V_2, V_3) - h(V_2, V_3)k(V_1, V_4). \end{aligned}$$

(a) Show that Kulkarni-Nomizu product of two symmetric 2-tensors necessarily satisfies the curvature symmetries of Theorem 13. Conclude that the Weyl tensor satisfies the curvature symmetries of Theorem 13. [Basic]

(b) Compute the Weyl tensor for S^n and $\mathbb{H}^n$. [Basic]

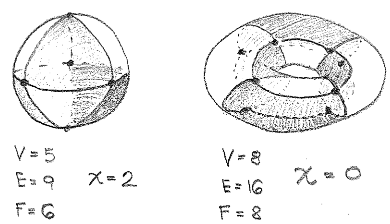
(c) It can be shown that if $\tilde{g} = e^u g$ then $W_{\tilde{g}} = e^u W_g$. Taking this as a fact, review your answer to the previous question. [Intermediate]

(d) Very roughly speaking, the Riemann curvature tensor measures how badly a certain manifold fails to be flat. In view of the above, how would you describe what the meaning of the Weyl curvature tensor? [Intermediate]

(10) This problem guides you towards the Gauss-Bonnet Theorem. The theorem involves a topological invariant of a manifold, which for 2-dimensional compact manifolds can be defined in terms of its (geodesic) polygonal decompositions. More precisely, for a given 2-dimensional compact manifold M the value

$$\chi = V - E + F$$

where V denotes the number of vertices, E denotes the number of edges and F denotes the number of faces of a polygonal decomposition turns out to be independent of the choice of the polygonal decomposition. Thus, this value constitutes an *invariant of M*. This invariant is called *the Euler*[6] *characteristic of* M and is denoted by $\chi(M)$. It is a topological invariant in the sense that it remains the same under "elastic deformations" and is not affected by the change of metric. [Basic, Expected]



(a) Consider a triangulation of a 2-dimensional compact manifold M without boundary using small geodesic triangles. Apply Gauss' Theorem (Theorem 21) to each geodesic triangle, and add up the equations you got. Convince yourself that

$$\int_M \mathscr{K}\,\mathrm{dvol} = (2\pi)V - \pi F.$$

(b) Argue that $F = \frac{3}{2}E$ and consequently $V - \frac{1}{2}F = V - E + F$.

(c) Show that $\int_M \mathscr{K}\mathrm{dvol} = 2\pi\chi(M)$. This is the celebrated Gauss-Bonnet Theorem.

(d) What is the total sectional curvature of any metric on S^2?

(e) Does there exist a metric of everywhere non-positive sectional curvature on S^2?

(f) What is the total sectional curvature of any metric on a torus (donut)? What about tori with multiple holes?

(g) If a torus carries a constant curvature metric, what sign does it have to be? What about tori with multiple holes?

9. The framework of general relativity

Introduction. From a physics standpoint, there are two fundamental principles of special relativity: *the Principle of Relativity* (which dates back to Galileo) and *the Universality of the Speed of Light* (which is what makes special relativity "special"). The Principle of Relativity encompasses the idea that *no one is*

---

[6] Leonhard Euler's discovery of the formula $V - E + F = 2$ for convex polyhedra marked the beginning of the investigations of this invariant.

at absolute rest and that all unaccelerated (also known as inertial) observers are "created equal". A typical statement of the Principle of Relativity goes along the lines of "all inertial observers observe the same laws of physics", but a more careful investigation reveals that such a statement relies on some further (often implicit) assumptions. For example, it is assumed that each inertial observer has their own way of assigning spatial (x, y, z) and temporal t-coordinates to an event, that the spatial geometry is always Euclidean and independent of t, etc. A more careful statement of such implicit assumptions can, for example, be found in Bernard Schutz's *"A First Course in General Relativity"*.

The principle of the Universality of the Speed of Light was postulated by Einstein based on earlier experimental evidence. It states that all inertial observers measure the same speed of light[7]. Experimental evidence suggests that this speed is about 299 792 458 meters per second, although in mathematical relativity it is common to choose the units of space measurement based on units of time measurement so that the speed of light is always equal to 1. (Note that in particular this makes speed dimension-less.)

From the perspective of inertial observers the set of events constitutes a space-time $\{(t, x, y, z)\}$, in which t-axis represents observer's own trajectory. Light, on the other hand, travels along *light cones*

$$(\Delta x)^2 + (\Delta y)^2 + (\Delta z)^2 = (\Delta t)^2 \quad \text{i.e} \quad -(\Delta t)^2 + (\Delta x)^2 + (\Delta y)^2 + (\Delta z)^2 = 0.$$

This in particular means that, for two different inertial observers, the nullsets of $-(\Delta t)^2 + (\Delta x)^2 + (\Delta y)^2 + (\Delta z)^2$ and $-(\Delta \tilde{t})^2 + (\Delta \tilde{x})^2 + (\Delta \tilde{y})^2 + (\Delta \tilde{z})^2$ are identical. Based on this equality an argument can be made (see Schutz's book cited above) that two inertial observers have to observe the same value of

$$-(\Delta t)^2 + (\Delta x)^2 + (\Delta y)^2 + (\Delta z)^2 = -(\Delta \tilde{t})^2 + (\Delta \tilde{x})^2 + (\Delta \tilde{y})^2 + (\Delta \tilde{z})^2. \tag{5}$$

This *invariant* of special relativity is called the *interval*. It motivates the use of inner-product

$$\mathfrak{m}(\mathbf{u}_1, \mathbf{u}_2) = -t_1 t_2 + x_1 x_2 + y_1 y_2 + z_1 z_2 \tag{6}$$

on the vector space $\mathbb{R}^{1+3} = \{\mathbf{u} = (t, x, y, z)^T\}$, and an articulation of laws of special relativity in terms of geometry of this inner-product space. To a large extent this idea goes back to the German mathematician Hermann Minkowski who in 1907/8 re-formulated Einstein's 1905 work on special relativity (combined with an even earlier work of the Dutch physicist Hendrik Antoon Lorentz) in this new geometric language. In his honor the inner-product space defined by (6) is called *Minkowski space-time*. Deducing the special relativistic effect known as the time dilation from the geometry of Minkowski space-time is a cute and easy exercise. We leave it to the reader; see exercise(2) below.

Here we choose to invest our time into addressing the broader mathematical setting to which Minkowski space-time belongs, and the ways in which one can incorporate gravity into special relativity. Formulating a theory which combines gravity and special relativity was achieved by late 1915, through work of many mathematicians and physicists including Einstein and David Hilbert. The theory is known under the name *general relativity*.

[7]Built into this is the assumption that light moves along straight paths with time-independent speed which is equal in each direction.

Minkowski space-time. Minkowski inner-product $\mathfrak{m}$ is an example of a *non-degenerate inner-product*, i.e $\mathbb{R}$-valued symmetric bilinear map $\langle .,. \rangle$ for which

$$(\forall \mathbf{v}) \, \langle \mathbf{u}, \mathbf{v} \rangle = 0 \ \text{ implies } \ \mathbf{u} = 0.$$

The concepts of magnitude, angle and distance are at least somewhat problematic in this framework, since non-degenerate inner-products permit vectors $\mathbf{u} \neq \mathbf{0}$ with $\langle \mathbf{u}, \mathbf{u} \rangle < 0$ or even $\langle \mathbf{u}, \mathbf{u} \rangle = 0$. In analogy with the special-relativistic interpretations the former are called time-like and the letter are called light-like (or null) vectors; the phrase space-like vectors is reserved for vectors $\mathbf{u}$ with $\langle \mathbf{u}, \mathbf{u} \rangle > 0$. It can be shown (see exercise (4) below) that each non-degenerate inner-product permits what is sometimes called a pseudo-orthonormal basis: a basis $\{\mathbf{u}_i\}$ of vectors for which

$$\langle \mathbf{u}_i, \mathbf{u}_j \rangle = \begin{cases} \pm 1 & \text{if } \ i = j \\ 0 & \text{if } \ i \neq j. \end{cases}$$

Clearly, $\{\partial_t, \partial_x, \partial_y, \partial_z\}$ forms a pseudo-orthonormal basis for the Minkowski space-time. Furthermore, one shows that the number r of time-like vectors of a pseudo-orthonormal basis does not depend on the choice of the pseudo-orthonormal basis. Assuming the underlying vector space is n-dimensional, the ordered pair $(r, n-r)$ is called *the signature* of the inner-product. The signature of the Minkowski inner-product $\mathfrak{m}$ is $(1,3)$. In the literature signature $(1,3)$ is commonly referred to as the *Lorentz signature.*

For reasons we do not have time to go into, massive particles move at speeds slower than the speed of light. This means that the velocity vector fields to their space-time trajectories

$$\gamma(\tau) = (t(\tau), x(\tau), y(\tau), z(\tau))$$

are everywhere time-like. The meaning of the parameter τ is maybe a bit mysterious as it cannot simply be thought of as "time"[8]. Its meaning is pinned down by requiring that $\dot{\gamma}$ be unit time-like:

$$\langle \dot{\gamma}, \dot{\gamma} \rangle = -1.$$

This requirement is very natural given that inertial observers move along t-axis $\gamma(\tau) = (\tau, 0, 0, 0)$. The overall point here is that τ can be though of as "time" as measured by an observer going along with the particle. The parameter τ is often called *proper time.*

Review of Newtonian gravity. According to Newton's Law of Gravity the magnitude of the gravitational force between two point objects with masses M and m is inversely proportional to the square of their distance r:

$$G \frac{mM}{r^2}.$$

Per unit of mass, therefore, the point mass M makes a gravitational contribution of $GM\frac{1}{r^2}$. More specifically, if the point mass M is located at $\mathbf{x} \in \mathbb{R}^3$ then its *gravitational field*, i.e force per unit mass which point mass M exerts at the location $\mathbf{y} \in \mathbb{R}^3$, is:

$$-GM \frac{\mathbf{y} - \mathbf{x}}{|\mathbf{y} - \mathbf{x}|^3}.$$

[8] As a consequence of effects such as time dilation, there is concept of "time" which all observers can agree on.

What if we are not dealing with point masses but, say, with a cloud of dust particles described by mass density $\rho(\mathbf{x})$? Since the gravitational contribution of an infinitesimal body of volume $\mathrm{dvol}_{\mathbf{x}}$ and mass density $\rho(\mathbf{x})$ is

$$-G\rho(\mathbf{x})\frac{\mathbf{y}-\mathbf{x}}{|\mathbf{y}-\mathbf{x}|^3}\mathrm{dvol}_{\mathbf{x}}$$

the *gravitational field* generated by the said dust cloud is given by

$$\mathbf{F}(\mathbf{y}) = -G\int_{\mathbb{R}^3}\rho(\mathbf{x})\frac{\mathbf{y}-\mathbf{x}}{|\mathbf{y}-\mathbf{x}|^3}\mathrm{dvol}_{\mathbf{x}} = -G\int_{\mathbb{R}^3}\rho(\mathbf{y}-\mathbf{x})\frac{\mathbf{x}}{|\mathbf{x}|^3}\mathrm{dvol}_{\mathbf{x}}.$$

One might want to be at least a little bit concerned about the last two expressions, as their integrands are unbounded. Through an employment of spherical coordinates on $\mathbb{R}^3$ one sees that the integrals in question are absolutely convergent as long as $\int_{\mathbb{R}^3}\rho(\mathbf{x})\mathrm{dvol}_{\mathbf{x}}$ is. Not to open too many cans of worms, we will from this point on assume that *the dust cloud is bounded at that its mass density is smooth* so that this integral condition on ρ is fulfilled. One should note that under such assumptions the gravitational field F is smooth and its partial derivatives are given as follows:

$$\partial^k\mathbf{F}(\mathbf{y}) = -G\int_{\mathbb{R}^3}(\partial^k\rho)(\mathbf{y}-\mathbf{x})\frac{\mathbf{x}}{|\mathbf{x}|^3}\mathrm{dvol}_{\mathbf{x}}.$$

The following is known as the Gauss' Law.

THEOREM 23. $\mathrm{div}\mathbf{F} = -4\pi G\rho$.

PROOF. Differentiation under the integral sign produces

$$\mathrm{div}\mathbf{F}(\mathbf{y}) = -G\int_{\mathbb{R}^3}\frac{1}{|\mathbf{x}|^3}\langle\mathrm{grad}\big|_{\mathbf{y}-\mathbf{x}}\rho,\mathbf{x}\rangle\mathrm{dvol}_{\mathbf{x}}.$$

The key observation for us now is that $\mathrm{div}\left(\frac{\mathbf{x}}{|\mathbf{x}|^3}\right) = 0$ and that therefore

$$\tfrac{1}{|\mathbf{x}|^3}\left\langle\mathrm{grad}\big|_{\mathbf{y}-\mathbf{x}}\rho,\mathbf{x}\right\rangle = -\mathrm{div}\left(\rho(\mathbf{y}-\mathbf{x})\tfrac{\mathbf{x}}{|\mathbf{x}|^3}\right)$$

away from the origin. In what follows we would like to apply the Divergence Theorem, but we first need to deal with the singularity at the origin. To that end, we excise a small ball B_ε of radius ε centered at the origin. Observe that, due to Cauchy-Schwartz inequality, we have

$$\left|\int_{B_\varepsilon}\frac{1}{|\mathbf{x}|^3}\left\langle\mathrm{grad}\big|_{\mathbf{y}-\mathbf{x}}\rho,\mathbf{x}\right\rangle\mathrm{dvol}_{\mathbf{x}}\right| \leq C\int_{B_\varepsilon}\tfrac{1}{|\mathbf{x}|^2}\mathrm{dvol}_{\mathbf{x}}$$

where C is some upper bound on the size of the gradient of ρ. By using spherical coordinates the remaining integral is easily found to be some multiple of ε. In particular, we now have:

$$\begin{aligned}
\operatorname{div}\mathbf{F}(\mathbf{y}) = & -G\lim_{\varepsilon\to 0}\int_{\mathbb{R}^3\smallsetminus B_\varepsilon}\tfrac{1}{|\mathbf{x}|^3}\left\langle \operatorname{grad}\big|_{\mathbf{y}-\mathbf{x}}\rho,\mathbf{x}\right\rangle \mathrm{dvol}_{\mathbf{x}}\\
=& G\lim_{\varepsilon\to 0}\int_{\mathbb{R}^3\smallsetminus B_\varepsilon}\operatorname{div}\left(\rho(\mathbf{y}-\mathbf{x})\tfrac{\mathbf{x}}{|\mathbf{x}|^3}\right)\mathrm{dvol}_{\mathbf{x}}\\
=& -G\lim_{\varepsilon\to 0}\int_{\partial B_\varepsilon}\rho(\mathbf{y}-\mathbf{x})\frac{\langle\mathbf{x},\nu\rangle}{|\mathbf{x}|^3}\mathrm{dvol}_{\mathbf{x}}\\
=& -G\lim_{\varepsilon\to 0}\frac{1}{\varepsilon^2}\int_{\partial B_\varepsilon}\rho(\mathbf{y}-\mathbf{x})\mathrm{dvol}_{\mathbf{x}}\\
=& -4\pi G\lim_{\varepsilon\to 0}\frac{1}{4\pi\varepsilon^2}\int_{|\mathbf{z}-\mathbf{y}|=\varepsilon}\rho(\mathbf{z})\mathrm{dvol}_{\mathbf{z}},
\end{aligned}$$

where we used ν to denote the outward pointing unit normal to the small sphere ∂B_ε of radius ε centered at the origin. (Note that we used the fact that ρ vanishes in far way regions.) The last expression is the average of the mass density over a small sphere centered at $\mathbf{y}$; as the radius of the sphere approaches 0 this average is more and more like the value of the mass density at the center of the sphere $\mathbf{y}$. Thus, $\operatorname{div}\mathbf{F}(\mathbf{y}) = -4\pi\rho(\mathbf{y})$. □

It is important to notice that under our assumptions $\mathbf{F}(\mathbf{y})$ is the negative gradient

$$\mathbf{F} = -\operatorname{grad}\Phi$$

of the scalar function

$$\Phi(\mathbf{y}) = -G\int_{\mathbb{R}^3}\frac{\rho(\mathbf{x})}{|\mathbf{x}-\mathbf{y}|}\mathrm{dvol}_{\mathbf{x}} = -G\int_{\mathbb{R}^3}\rho(\mathbf{y}-\mathbf{x})\frac{1}{|\mathbf{x}|}\mathrm{dvol}_{\mathbf{x}}.$$

In the case of the point mass M located at $\mathbf{x}$ this amounts to

$$\Phi(\mathbf{y}) = -GM\frac{1}{|\mathbf{y}-\mathbf{x}|}.$$

This underlying scalar function is called *the gravitational potential*. The following is immediate from Gauss' Law:

THEOREM 24 (Poisson equation).

$$\Delta\Phi = 4\pi G\rho.$$

The acceleration of a particle in free fall is determined by the gravitational field $\mathbf{F} = -\operatorname{grad}\Phi$. More precisely, if the trajectory of a particle in free fall is given by $\mathbf{y}(t)$ then

$$\tfrac{d^2}{dt^2}\mathbf{y} = \mathbf{F}\big|_{\mathbf{y}(t)}.$$

In the basic example of the gravitational field of a point mass M centered at the origin we have:

$$\tfrac{d^2}{dt^2}\mathbf{y} = -GM\tfrac{\mathbf{y}}{|\mathbf{y}|^3}.$$

Analysis of this differential equation eventually yields Kepler's laws. (We will do certain aspects of this in the next lecture.)

If $\mathbf{y_1}(t)$ and $\mathbf{y_2}(t)$ are two nearby particles in free fall then by Taylor's Theorem we have

$$\tfrac{d^2}{dt^2}(\mathbf{y}_2-\mathbf{y}_1) = -\operatorname{grad}\Phi\big|_{\mathbf{y}_2} + \operatorname{grad}\Phi\big|_{\mathbf{y}_1} \approx -\operatorname{Hess}(\Phi)\big|_{\mathbf{y}_1}(\mathbf{y}_2-\mathbf{y}_1),$$

where $\mathrm{Hess}(\Phi)\big|_{\mathbf{y}_1}$ denotes the Hessian of Φ viewed as a matrix acting on the separation vector $\mathbf{y}_2 - \mathbf{y}_1$. This in particular means that the separation vector field $\mathbf{Y}(t)$ for infinitesimally close particles in free fall (or, rather, their relative acceleration $\frac{d^2}{dt^2}\mathbf{Y}$) satisfies

$$\frac{d^2}{dt^2}\mathbf{Y} = -\mathrm{Hess}(\Phi)\mathbf{Y}.$$

General Relativity. Acceleration due to Earth's gravity is independent of the nature of the falling object; this was first shown by Galileo at the Leaning Tower of Pisa and has since been confirmed to a very high accuracy by modern experiments. Abstracted into a general principle, this is to say that (at least when working on a small scale with respect to which any variation in the gravitation field is negligible) there is no observable difference between effects of gravity and acceleration. There is no reason to believe that there are "gravitationally neutral observers" and, thus, in presence of gravity no observer is truly unaccelerated.

The theory of special relativity relies on the idea that there is no absolute rest and no such thing as an absolute velocity of an observer. All inertial (unaccelerated) observers are "created equal", and each one of them is at rest with respect to their own inertial coordinates. Moreover, the relativistic effects due to a small scale relative velocity $\mathbf{v}$ (i.e. $|\mathbf{v}| \ll 1$) between inertial observers are negligible and in that sense of the word Newtonian physics serves as an approximate to special relativity.

If we want to include gravity in a relativistic framework then everything said in the previous paragraph needs to be taken one step beyond. There is no such thing as absolute acceleration; all *observers in free-fall* (that is, observers which are exposed to no forces other than gravity) are "created equal"; each one of them can view themselves as unaccelerated with respect to some appropriate coordinate system; the effects of a small scale relative acceleration between observers in free-fall are negligible and in such a framework laws of special relativity apply (at least approximately).

The emphasis placed on inertial (unaccelerated) observers in special relativity needs to be shifted to observers in free-fall, who with respect to themselves appear to be like inertial observers of special relativity. Thus we expect that on small scale space-time can be coordinatized and be given a metric like that of Minkowski space-time. From the perspective of observers in free-fall laws of special relativity apply over small distances and short times. The perfect backdrop for uniting relativity and gravity is that of a 4-dimensional manifold whose tangent spaces (think "linearization") are equipped with Minkowski geometry and theory of special relativity.

The most basic premise of general relativity is that space-time is a 4-dimensional manifold endowed with a *pseudo-Riemannian metric of Lorentz signature.* By this we mean a symmetric covariant 2-tensor g which on each tangent space induces a non-degenerate inner-product of Lorentz signature.

The geometry of pseudo-Riemannian manifolds can be developed in parallel to what was done thus far in the lectures. The geodesics, the Christoffel symbols, the Levi-Civita connection and the curvature tensor can be introduced in the exact same manner; the Jacobi equation for geodesic deviation still holds. To be entirely honest, one does need to be cautious. Due to the sign difference the volume element in Lorentzian framework is

$$\mathrm{dvol}_g = \sqrt{-\det[g_{ij}]}\, dx^1 \dots dx^n,$$

the Laplace operator changes its nature and becomes more of wave equation operator, the spheres all of a sudden are not compact, comparison theorems no longer apply as stated. Nevertheless, the framework of pseudo-Riemannian (or semi-Riemannian) geometry is remarkably similar to that of Riemannian geometry. Readers interested in seeing the details should consult Barrett O'Neill's *"Semi-Riemannian Geometry With Applications to Relativity"*.

In accordance with the idea that inertial observers move in straight lines, in general relativity we think of *free-falling observers as moving along geodesics*. As in special relativity we assume the geodesics γ are parametrized by proper time, $\langle \dot\gamma, \dot\gamma \rangle = -1$.

From Poisson to Einstein equation. In the presence of a non-constant gravitational field (and when working on a larger scale) relative accelerations of free-falling observers do come into play. The equation which describes the relative acceleration of nearby particles in free-fall in Newtonian gravity is

$$\frac{d^2}{dt^2}\mathbf{Y} = -\mathrm{Hess}(\Phi)\mathbf{Y}.$$

The equation which describes the relative acceleration of nearby geodesics in pseudo-Riemannain geometry is the Jacobi equation

$$\tfrac{d^2}{dt^2} Y = -R(Y, \dot\gamma)\dot\gamma.$$

The structural similarity is very striking! The general relativistic version of the role of the Hessian of the gravitational potential

$$\mathbf{Y} \mapsto \mathrm{Hess}(\Phi)\big|_{\dot\gamma}\mathbf{Y}$$

is played by the curvature-based *Jacobi operator*

$$Y \mapsto R(Y, \dot\gamma)\dot\gamma.$$

The gravitational potential Φ is related to the matter distribution by the Poisson equation $\Delta\Phi = 4\pi\rho$. The Laplace operator is the trace of the Hessian, and in general relativity the Hessian of the gravitational potential corresponds to the Jacobi operator. Thus, we expect there to be a *relativistic version of the Poisson equation* which relates the trace of the Jacobi operator to the matter distribution. The trace of the Jacobi operator, of course, is the Ricci curvature. So, schematically the relativistic version of the Poisson equation takes the form of

$$\mathrm{Ricci}(\dot\gamma, \dot\gamma) = \text{“matter” along } \dot\gamma.$$

The most straight-forward case of the above occurs in vacuum regions of spacetime where the matter content can be set to zero. Despite how this may sound it does not mean that there is absolutely nothing going on in those vacuum regions. To convince yourself of this consider the fact that the gravitational potential Φ of a point mass M in Newtonian gravity is $\Phi(\mathbf{y}) = -GM\frac{1}{|\mathbf{y}-\mathbf{x}|}$ and that it satisfies the Poisson equation $\Delta\Phi = 0 = 4\pi\rho$ everywhere except at $\mathbf{x}$, the location of the point mass. Thus understanding spherically symmetric solutions of the relativistic versions of the vacuum Poisson equation is akin to understanding the geometry surrounding relativistic point masses, e.g relativistic stars, in otherwise empty space-time.

In any case, the vacuum version of the Poisson equation states that $\mathrm{Ricci}(\dot{\gamma}, \dot{\gamma})$ vanishes for all observers γ in free-fall. It can be shown (see exercise (9) below) that this is equivalent to

$$\mathrm{Ricci} = 0.$$

The latter is known as the *vacuum Einstein equation.*

The components of the Ricci tensor are highly non-linear expressions involving up to second order derivatives of the metric g, and so Einstein equation is one though non-linear system of PDE's in which the unknowns are the metric components. We will solve this system in spherical symmetry tomorrow. The overall bottom line is that in general relativity the unknown is the geometry of the space-time itself, and that Einstein equations are non-linear PDE's which place restrictions on the geometry in terms of the matter content.

A brief comment regarding non-vacuum Einstein equations. At this point it is probably very tempting to set $\mathrm{Ricci} = T$ where T is some symmetric covariant 2-tensor which incapsulates the matter distribution. The bad news is that one expects there to be conservation laws for the matter and thus one expects such tensors T to be divergence-free (in the sense that the covector field obtained by tracing ∇T vanishes). Since Ricci curvature is in general not divergence-free, we cannot simply go ahead and set $\mathrm{Ricci} = T$. The good news, however, is that $\mathrm{Ricci} - \frac{1}{2}\mathrm{Scal}\, g$ is divergence free by virtue of the second Bianchi identity. (Compare with problem (6) from the last exercise sheet.) In fact, one can show that the most general divergence-free geometric (made out of g and R) symmetric covariant 2-tensor takes the form of

$$\mathrm{Ricci} - \tfrac{1}{2}\mathrm{Scal}\, g + \Lambda g$$

for some constant Λ. So, the most general Einstein equation should somehow equate T with $\mathrm{Ricci} - \frac{1}{2}\mathrm{Scal}\, g + \Lambda g$. We do not need to get into this right now, but there are some analogy-based reasons for including a proportionality constant of 8π, and so the general Einstein equation is:

$$\mathrm{Ricci} - \tfrac{1}{2}\mathrm{Scal}\, g + \Lambda g = -8\pi T.$$

Whether or not the constant Λ should be included is still subject to debate. It appears that one might want to include Λ when studying phenomena on a yet larger scale (and not, say, laws of geometry near a system of bodies in otherwise empty space). The constant Λ, if included, is called *the cosmological constant.* For exercise (see (10)) you should show that vacuum Einstein equations with no cosmological constant are equivalent to the vacuum equation $\mathrm{Ricci} = 0$ announced above.

Exercises for Lecture 9.

(1) Suppose a situation where two inertial observers which are moving with relative constant velocity $\mathbf{v}$ with respect to one another coincide at an event, e.g. the origin of both of their coordinate systems. Let (t, x, y, z) and $(\tilde{t}, \tilde{x}, \tilde{y}, \tilde{z})$ denote the respective inertial coordinates. [Basic]

(a) Convince yourself that the $\tilde{t}$-axis (trajectory of the ˜-observer) can be viewed as a line of slope $|\mathbf{v}|$ with respect to t-axis.

(b) Next, investigate the relative position of the xyz-subspace and $\tilde{x}\tilde{y}\tilde{z}$-subspace. Note the significance of these subspaces: Being given by equations $t = 0$ and $\tilde{t} = 0$ they can be viewed as simultaneity planes for their respective observers. Convince yourself the future light cone emitted at an event $(-\tilde{t}_*, 0, 0, 0) = (-t_*, -x_*, -y_*, -z_*)$ and the past light cone received at $(\tilde{t}_*, 0, 0, 0) = (t_*, x_*, y_*, z_*)$ intersect at events which are temporally located "half-way in-between". In other words, convince yourself that the simultaneity plane $\tilde{t} = 0$ can be viewed as the set of intersections of such light cones.

(c) Show that the simultaneity plane from above is described by the following equation.

$$-t_*t + x_*x + y_*y + z_*z = 0.$$

(d) Discuss: is there such a thing as an absolute notion of simultaneity?

(2) Adopt the set-up of the previous problem. Based on the invariance of interval (see equation (5) in the text) argue that the event with ˜-coordinates $(\tilde{t}, 0, 0, 0)$ corresponds to (t, x, y, z) with

$$t = \frac{\tilde{t}}{\sqrt{1 - |\mathbf{v}|^2}}.$$

Explain why from the perspective of an observer "at rest" the clocks of moving observers run slow. (This effect is known as time dilation.) [Basic]

(3) Let $\langle ., . \rangle$ be a non-degenerate inner-product on a vector space V of dimension n. Let $\mathbf{u}$ be a fixed (non-zero) space-like or time-like vector. Investigate the orthogonal complement of $\mathbf{u}$ following the guidelines given below. [Basic]

(a) Show that $\mathbf{v} \mapsto \langle \mathbf{u}, \mathbf{v} \rangle$ is a linear map of rank 1 and nullity $n - 1$.

(b) Show that $\mathbf{u}^\perp = \{\mathbf{v} \in V \,|\, \langle \mathbf{v}, \mathbf{u} \rangle = 0\}$ is a subspace of dimension $n - 1$.

(c) Show that each vector $\mathbf{v} \in V$ can uniquely be decomposed as

$$\mathbf{v} = \mathbf{u}_1 + \mathbf{u}_2,$$

where $\mathbf{u}_1 \in \mathrm{Span}(\mathbf{u})$ and $\mathbf{u}_2 \in \mathbf{u}^\perp$.

(d) Show that the restriction of the inner-product $\langle ., . \rangle$ to $\mathbf{u}^\perp$ is still non-degenerate.

(e) How many of the above claims still hold if $\mathbf{u}$ is non-zero null?

(4) Let $\langle ., . \rangle$ be a non-degenerate inner-product on a vector space V of dimension n.

(a) Iterate the arguments of the previous problem to show the existence of a pseudo-orthonormal basis $\{e_1^-, \ldots, e_r^-, e_1^+, \ldots, e_s^+\}$ with $\langle e_i^-, e_i^- \rangle = -1$ and $\langle e_i^+, e_i^+ \rangle = 1$ for all i. [Intermediate]

(b) Let $\{f_1^-, \ldots, f_p^-, f_1^+, \ldots, f_q^+\}$, with $\langle f_i^-, f_i^- \rangle = -1$ and $\langle f_i^+, f_i^+ \rangle = 1$ for all i, be another pseudo-orthonormal basis for V. Show that $(r, s) = (p, q)$. [Intermediate]

(5) Consider the Minkowski innerproduct

$$\mathfrak{m}(\mathbf{u}_1, \mathbf{u}_2) = -t_1 t_2 + x_1 x_2 \quad \text{on} \quad \mathbb{R}^{1+1} = \{\mathbf{u} = (t, x)^T\}.$$

Find all linear transformations $L : \mathbb{R}^{1+1} \to \mathbb{R}^{1+1}$ which are *isometries* of $\mathfrak{m}$, i.e which satisfy $\mathfrak{m}(L\mathbf{u}, L\mathbf{v}) = \mathfrak{m}(\mathbf{u}, \mathbf{v})$. (These are called Lorentz transformations.) [Intermediate]

(6) Let V be a vector space equipped with a non-degenerate inner-product $\langle ., . \rangle$. We define the unit time-like (resp. space-like) sphere in V as the set of vectors $\mathbf{u}$ with $\langle \mathbf{u}, \mathbf{u} \rangle = -1$ (resp. $\langle \mathbf{u}, \mathbf{u} \rangle = 1$). Describe the unit spheres in $\mathbb{R}^{1+n} = \{\mathbf{u} = (x^0, x^1, \dots, x^n)^T\}$ equipped with the Minkowski inner-product

$$\mathfrak{m}(\mathbf{u}_1, \mathbf{u}_2) = -x_1^0 x_2^0 + x_1^1 x_2^1 + \cdots + x_1^n x_2^n.$$

Specifically, sketch these spheres in low dimensional cases when $n = 1, 2$. [Basic]

(7) Consider the top sheet M of the unit timelike sphere

$$-(x^0)^2 + (x^1)^2 + \cdots + (x^n)^2 = -1$$

of the Minkowski space $\mathbb{R}^{1+n}$; also consider the point $L = (-1, 0, \dots, 0)$. For a point P on M consider the point $(0, u^1, \dots, u^n)$ at which the line LP intersects the $x^0 = 0$-coordinate plane. (This is analogous to the stereographic projection for spheres in $\mathbb{R}^{n+1}$.) [Intermediate]

(a) Convince yourself that the $(u^1, \dots, u^n)$ coordinatize M. What region of $\mathbb{R}^n$ is occupied by the coordinatizing $(u^1, \dots, u^n)$?

(b) Find the rules which relate $(x^0, \dots, x^n)$ to $(u^1, \dots, u^n)$ and vice versa.

(c) Argue that all tangent vectors to M are space-like, and that the metric inherited on M from the Minkowski metric on $\mathbb{R}^{1+n}$ is Riemannian.

(d) Use an argument similar to the one we used on S^n at the beginning of the course to identify geodesics on M (with respect to the metric inherited from the Minkowski metric on $\mathbb{R}^{1+n}$).

(e) Compute an explicit formula for the Riemannian metric on M in $(u^1, \dots, u^n)$ coordinates. You should get a familiar result.

(f) What is the curvature of M?

(8) This problem investigates taking the trace in the presence of a general non-degenerate (and thus not necessarily positive-definite) inner-product. Throughout the problem $\{e_i\}$ denotes a pseudo-orthonormal basis, while $\epsilon_i = \langle e_i, e_i \rangle$. [Intermediate]

(a) Let L be a linear transformation on a finite dimensional non-degenerate inner-product space $(V, \langle ., . \rangle)$, and let $L = [L^i_j]$ be any of its matrix representations. Recall that $\mathrm{Tr}(L) = \sum_i L^i_i$. Show that for pseudo-orthonormal bases $\{e_i\}$ one also has:

$$\mathrm{Tr}(L) = \sum_i \epsilon_i \langle L e_i, e_i \rangle.$$

(b) Show that the value of $\sum_i \epsilon_i R(e_i, V, W, e_i)$ is independent of the choice of a pseudo-orthonormal basis $\{e_i\}$. This common value defines the Ricci tensor, $\mathrm{Ricci}(V, W)$.

(c) How would you go about defining the scalar curvature Scal of a pseudo-Riemannian manifold? State and prove the corresponding result.

(d) What are some difficulties with the integral definition of trace in the pseudo-Riemannian context?

(9) Show that $\mathrm{Ricci}(V, V) = 0$ for all unit timelike V if and only if $\mathrm{Ricci} = 0$. [INTERMEDIATE]

(10) Show that the vacuum Einstein equation with no cosmological constant in the form of $\mathrm{Ricci} - \frac{1}{2}\mathrm{Scal}\, g = 0$ is equivalent to $\mathrm{Ricci} = 0$. Hint: consider taking the trace. [INTERMEDIATE]

10. Geometry of Schwarzschild SpaceTime

Introduction. Today we find our first solution of the vacuum Einstein equation. More precisely, we find a particular spherically symmetric solution of the vacuum Einstein equation which describes the geometry surrounding a relativistic star in an otherwise empty space-time. We then investigate particle orbits in this relativistic setting and compare them to predictions of Newtonian celestial mechanics. The exposition in this lecture is inspired by *"Semi-Riemannian Geometry With Applications to Relativity"* by B. O'Neill and *"General Relativity"* by N. M. J. Woodhouse.

Solving the vacuum Einstein equation in spherical symmetry. We start by making some simplifying assumptions. We are looking for a solution which is *static* in the sense that the Lorentzian metric:

a) does not change with time, i.e. is invariant under $t \mapsto t + t_*$;
b) is time reversible, i.e. is invariant under $t \mapsto -t$.

In particular, we are looking for a metric which can be expressed in the form

$$-E\, dt^2 + h$$

where E is a positive function of spatial variables only and where h is some fixed metric on $\mathbb{R}^3$ (or at least a portion there of). In addition, we are going to make an assumption that the metric is *spherically symmetric* i.e. that symmetries of the standard sphere S^2 preserve our metric as well. This puts the unknown metric into the schematic form of

$$g = -E(r)dt^2 + F(r)dr^2 + G(r)g_{S^2},$$

where $E, F, G > 0$ and where g_{S^2} denotes the standard metric on the unit 2-sphere. The function $4\pi G(r)$ describes the surface area of spatial spheres $|\mathbf{x}| = r$. (Careful: these need not actually be spheres of radius r; it all depends on what F is). Yet another simplifying assumption is that these volumes increase with r. Under this assumption one can make the change of variables

$$r \leftrightarrow G(r)^{1/2},$$

which brings us to the Ansatz

$$g = -E(r)dt^2 + F(r)dr^2 + r^2 g_{S^2}.$$

At this point one has no choice but role up their sleeves and compute curvature components for the metric g. This computation shows that the only non-vanishing Ricci curvature components are

$$\begin{aligned}
\mathrm{Ricci}(\partial_t, \partial_t) &= -\frac{\partial_r^2 E}{2F} + \frac{(\partial_r E)(\partial_r F)}{4F^2} + \frac{(\partial_r E)^2}{4EF} - \frac{\partial_r E}{rF} \\
\mathrm{Ricci}(\partial_r, \partial_r) &= \frac{\partial_r^2 E}{2E} - \frac{(\partial_r E)^2}{4E^2} - \frac{(\partial_r E)(\partial_r F)}{4EF} - \frac{\partial_r F}{rF} \\
\mathrm{Ricci}(\partial_\phi, \partial_\phi) &= \frac{1}{\sin^2\phi}\mathrm{Ricci}(\partial_\theta, \partial_\theta) = -\frac{r\partial_r E}{2EF} + \frac{r\partial_r F}{2F^2} - \frac{1}{F} + 1.
\end{aligned}$$

A lot of cancelations occur when computing

$$F\,\mathrm{Ricci}(\partial_t, \partial_t) + E\,\mathrm{Ricci}(\partial_r, \partial_r) = -\tfrac{1}{rF}(E\partial_r F + F\partial_r E) = -\tfrac{1}{rF}\partial_r(EF).$$

At this point the vacuum Einstein equation $\mathrm{Ricci} = 0$ implies that $F = \frac{C}{E}$ for some positive constant C. Inserting this into $\mathrm{Ricci}(\partial_\phi, \partial_\phi) = 0$ we get:

$$\partial_r(rE) = C \quad \text{i.e.} \quad E = C(1 + \tfrac{D}{r}), \quad F = (1 + \tfrac{D}{r})^{-1}$$

for some constant D. For reasons we are going to see later in this lecture the constant D is expressed as $D = -2M$ for some other constant M. Finally, by making a change of variables

$$t \leftrightarrow t\sqrt{C}$$

the constant C can be absorbed into the dt^2-factor to produce the metric

$$g = -(1 - \tfrac{2M}{r})dt^2 + (1 - \tfrac{2M}{r})^{-1}dr^2 + r^2 g_{S^2}.$$

(To be honest, one also has to go back and make sure that for these choices of E and F the Ricci components $\mathrm{Ricci}(\partial_t, \partial_t)$ and $\mathrm{Ricci}(\partial_r, \partial_r)$ also vanish; this indeed is the case.)

The metric we just discovered is called the *Schwarzschild metric*, and the resulting space-time is called the *Schwarzschild space-time*. They are named in honor of the German astrophysicist Karl Schwarzschild who discovered the metric in this exact form in late 1915[9]. The remainder of the lecture is dedicated to understanding geodesics of Schwarzschild space-time; we will see that they provide an insight into relativistic celestial mechanics. To provide some context for these results we should first review basic Newtonian celestial mechanics.

Newtonian celestial mechanics. In the last lecture we discussed the equations of motion in the gravitational field generated by a point mass M located at the origin. After choosing units in which the gravitational constant becomes $G = 1$, these equations become:

$$\tfrac{d^2}{dt^2}\mathbf{y} = -M\tfrac{\mathbf{y}}{|\mathbf{y}|^3}.$$

Note that this equation is invariant under the reflection $\phi \mapsto \pi - \phi$ about the equatorial plane $\phi = \frac{\pi}{2}$. The basic uniqueness result for the initial value problems now implies that any solution of this equation corresponding to initial conditions

[9]Karl Schwarzschild made this discovery while serving on the Russian front during World War I. Unfortunately, he died shortly thereafter due to a painful disease he developed while at the front.

within the equatorial plane has to remain in the equatorial plane. Thus, without loss of generality we may assume that the orbit of the body under investigation lies within the equatorial plane:

$$\phi = \tfrac{\pi}{2}.$$

Expressed in terms of the polar coordinates r and θ the above equations of motion now become:

$$\begin{cases} \ddot{r} - r\dot{\theta}^2 = -\frac{M}{r^2} \\ r\ddot{\theta} + 2\dot{r}\dot{\theta} = 0. \end{cases}$$

Clearly, the last equation can be re-written as $\frac{d}{dt}(r^2\dot{\theta}) = 0$, meaning that we have a conserved quantity

$$J = r^2\dot{\theta}.$$

This quantity is called *the angular momentum.* Inserting this into the equation for $\ddot{r}$ produces

$$\ddot{r} + \frac{M}{r^2} - \frac{J^2}{r^3} = 0.$$

Equations of this form have a conserved quantity as well; in this case it is the *energy*

$$\mathscr{E} = \frac{1}{2}\dot{r}^2 - \frac{M}{r} + \frac{J^2}{2r^2}. \tag{7}$$

The algebraic form of this energy suggests the use of a new variable $u = \frac{M}{r}$. In terms of this new variable the energy is equal to

$$\mathscr{E} = \frac{J^2}{2M^2}\left(\frac{du}{d\theta}\right)^2 - u + \frac{J^2}{2M^2}u^2,$$

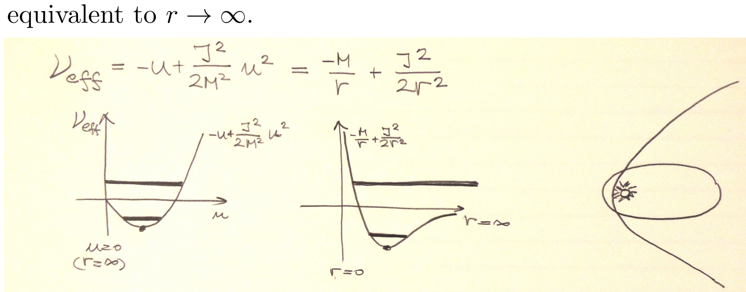
where we used the fact that $\dot{r} = -\frac{r^2}{M}\dot{u} = -\frac{J\dot{u}}{M\dot{\theta}} = -\frac{J}{M}\left(\frac{du}{d\theta}\right)$. On one hand, the conservation of $\mathscr{E}$ (i.e. $\partial_\theta \mathscr{E} = 0$) implies the differential equation

$$\frac{d^2u}{d\theta^2} + u = \left(\frac{M}{J}\right)^2$$

which can be solved exactly; this leads to Kepler's first law. On the other hand, a qualitative analysis of the solutions $u(\theta)$ for some fixed angular momentum J can be done directly from

$$\mathscr{E} \geq -u + \frac{J^2}{2M^2}u^2,$$

with the equality holding if and only if $\frac{du}{d\theta} = 0$. The following is the graph of the function $\mathcal{V}_{\text{eff}}(u) = -u + \frac{J^2}{2M^2}u^2$ in the case when $J \neq 0$. (The function $\mathcal{V}_{\text{eff}}$ is often called *the effective potential.*) Note that we are only focusing on $u > 0$, and that $u \to 0$ is equivalent to $r \to \infty$.



Each energy level above (or on) $\mathcal{V}_{\text{eff}}$ corresponds to an orbit. The lowest energy level can only be obtained when $\frac{du}{d\theta}$ identically vanishes, and thus signifies an orbit where u (and, consequently, r) remains constant. This is a circular orbit. The orbits corresponding to slightly higher energy levels are bounded orbits (both u and r are bounded, and bounded away from zero) while the high energy orbits describe an escape to infinity ($u \to 0$ i.e $r \to \infty$). The explicit solution for $r(\theta)$ reveals that the bounded orbits are ellipses, that the orbit corresponding to $\mathscr{E} = 0$ is a parabola, and the orbits corresponding to $\mathscr{E} > 0$ are hyperbolas. The elliptical orbits are periodic in the sense that r is a 2π-periodic function of θ. In particular, the so-called *perihelion* – the point on the orbit closest to the point mass M – is always the same.

Orbits in Schwarzschild Space-time. The find orbits of small bodies (whose contribution to the gravitational field is negligible) in Schwarzschild space-time we need to find geodesics parametrized by proper time τ. For the same reasons as in the Newtonian case it suffices to study the geodesics in the equatorial plane $\phi = \frac{\pi}{2}$. In particular, it suffices to compute the geodesics of the 3-dimensional Lorentzian metric

$$-(1 - \tfrac{2M}{r})dt^2 + (1 - \tfrac{2M}{r})^{-1}dr^2 + r^2 d\theta^2.$$

The idea of explicitly writing down the geodesic equations may appear intimidating at first, but the blessing is that the Christoffel symbols Γ^t_{ij} and Γ^θ_{ij} are easy to compute due to the fact that the metric is diagonal and that its components are independent of t and θ. One gets

$$\begin{cases} \ddot{t} + \frac{\partial_r(1-\frac{2M}{r})}{1-\frac{2M}{r}}\dot{t}\,\dot{r} = 0 \\ \ddot{r} + (\text{mess}) = 0 \\ \ddot{\theta} + \frac{2}{r}\dot{\theta}\,\dot{r} = 0, \end{cases}$$

where $\dot{}$ denotes the derivative with respect to proper time τ. The first and the third equation can be expressed as

$$\frac{d}{d\tau}\left((1 - \tfrac{2M}{r})\dot{t}\right) = 0 \quad \text{and} \quad \frac{d}{d\tau}\left(r^2\dot{\theta}\right) = 0,$$

leading to the conservation laws for $(1 - \frac{2M}{r})\dot{t}$ and $r^2\dot{\theta}$. The latter clearly corresponds to the angular momentum J; as we shall soon see the former corresponds to energy. From now on we set

$$\mathscr{E} = (1 - \tfrac{2M}{r})\dot{t}, \quad J = r^2\dot{\theta}.$$

At this point we need a differential equation for r. Although we could go back to the geodesic equation starting with $\ddot{r}$, it is easier to extract an equivalent equation from the fact that geodesics are always parametrized so that their velocity is constant. Specifically, we may assume that

$$-(1 - \tfrac{2M}{r})\dot{t}^2 + (1 - \tfrac{2M}{r})^{-1}\dot{r}^2 + r^2\dot{\theta}^2 = -1$$

so that

$$-\frac{\mathscr{E}^2}{1 - \frac{2M}{r}} + (1 - \tfrac{2M}{r})^{-1}\dot{r}^2 + \frac{J^2}{r^2} = -1.$$

After an algebraic simplification we are brought to the conserved quantity

$$\tilde{\mathscr{E}} = \frac{1}{2}\left(\mathscr{E}^2 - 1\right) = \frac{1}{2}\dot{r}^2 - \frac{M}{r} + \frac{J^2}{2r^2} - \frac{MJ^2}{r^3}.$$

The reader should compare this formula with (7). It is this similarity that prompted the choice of the constant D in the derivation of the Schwarzschild metric to be $D = -2M$. It is this connection that allows us to think of the Schwarzschild metric with parameter M as the relativistic model of an otherwise isolated massive object (e.g a star) of mass M.

Analysis of orbits in Schwarzschild space-time can be done by introducing the auxiliary variable $u = \frac{M}{r}$, just as in the Newtonian theory. We obtain

$$\tilde{\mathscr{E}} = \frac{J^2}{2M^2}\left(\frac{du}{d\theta}\right)^2 - u + \frac{J^2}{2M^2}u^2 - \frac{J^2}{M^2}u^3.$$

The conservation law $\partial_\theta \mathscr{E} = 0$ implies the differential equation

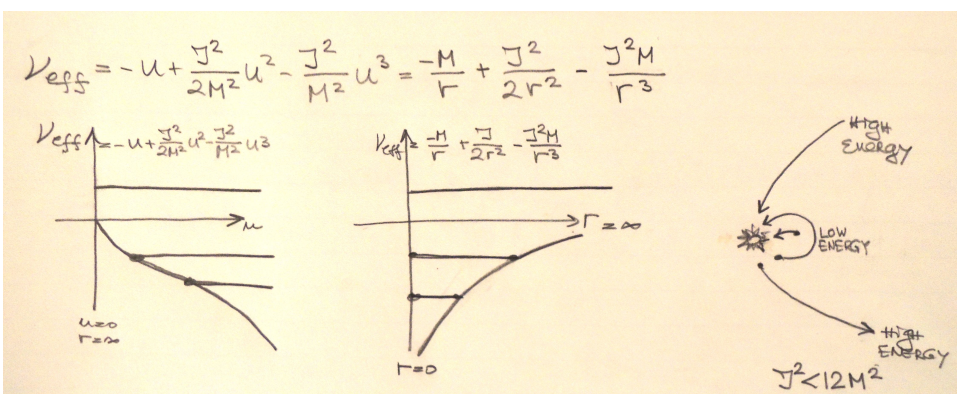
$$\frac{d^2u}{d\theta^2} + u - 3u^2 = \left(\frac{M}{J}\right)^2.$$

The analysis of the solutions of this differential equation can be done qualitatively, using $\tilde{\mathscr{E}} \geq \mathcal{V}_{\text{eff}}$ where

$$\mathcal{V}_{\text{eff}}(u) = -u + \frac{J^2}{2M^2}u^2 - \frac{J^2}{M^2}u^3.$$

We consider the case of non-zero fixed angular momentum $J \neq 0$, and determine "the fate of a particle" based on its energy-like conserved quantity $\tilde{\mathscr{E}}$.

Unlike the Newtonian scenario the general shape of the graph of $\mathcal{V}_{\text{eff}}$ varies depending on the coefficients i.e on the ratio $\frac{J}{M}$ of the angular momentum of the particle to the mass of the "star". When $\frac{J^2}{M^2} < 12$ the function $\mathcal{V}_{\text{eff}}$ has no critical points, when $\frac{J^2}{M^2} = 12$ it has exactly one critical (saddle) point, and when $\frac{J^2}{M^2} > 12$ there are two critical points (a local minimum and a local maximum). Critical points, you will recall, correspond to constant (equilibrium) solutions $u(\theta)$. Expressed in terms of $r(\theta)$ this is to say that critical points correspond to circular orbits. We now distinguish several of these cases.

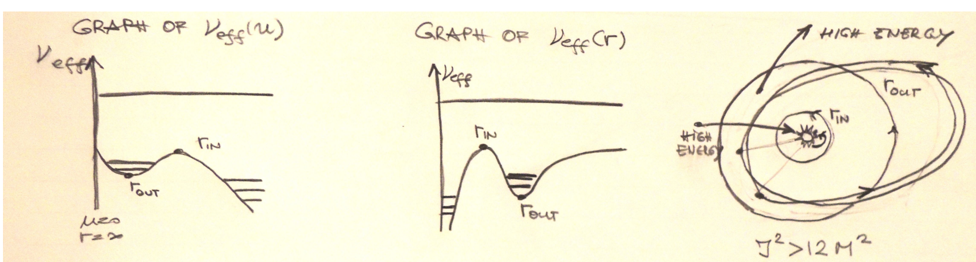


THE CASE OF $J^2 < 12M^2$. In this case there are no circular orbits. All orbits either escape to infinity (e.g. outgoing orbits with high energy) or crash[10]

[10]Due to conservation of angular momentum, small r means high $\dot{\theta}$ which in turn means that the crash is probably more accurately described by the word "in-spiral". Having said this, the geometric nature of Schwarzschild space-time for small r (i.e. $r < 2M$) deserves a more thorough investigation, and without it all comments about crashes and in-spirals are on a shaky ground.

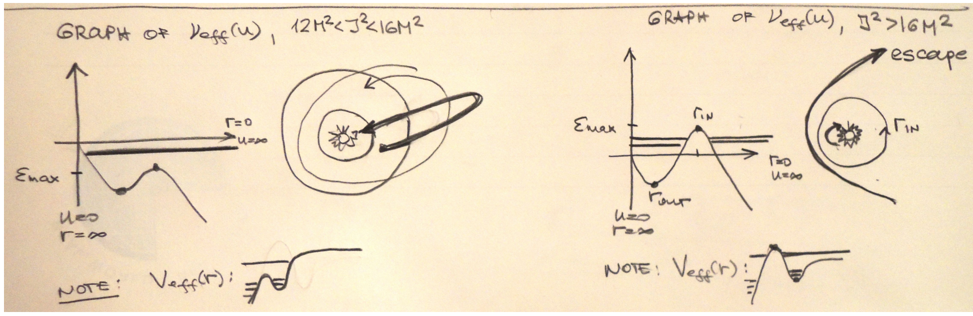
into the "star" (e.g. orbits in low energy or ingoing orbits with high energy). This is different from the Newtonian case where any non-zero amount of angular momentum prevents the particle from crashing into the "star". The bottom line is that in this relativistic case the angular momentum is simply too small.

THE CASE OF $J^2 > 12M^2$. In this case there are two circular orbits, one corresponding to the local minimum and one to the local maximum of $\mathcal{V}_{\text{eff}}$. The circular orbit corresponding to the local minimum is outer ($r = r_{\text{out}}$), and the circular orbit corresponding to the local maximum is inner ($r = r_{\text{in}}$). The outer circular orbit qualitatively resembles the Newtonian circular orbit in that nearby orbits are bounded; we investigate these nearby orbits below. The inner circular orbit only exists in relativistic framework. It provides an inner bound for all bounded orbits. More precisely, the fate of the particle of relatively low energy ($\tilde{\mathscr{E}} < 0$ and $\tilde{\mathscr{E}} < \tilde{\mathscr{E}}_{\text{max}}$, where $\tilde{\mathscr{E}}_{\text{max}}$ corresponds to the inner circular orbit) depends on whether the particle is on the inside or on the outside of the inner circular orbit. Particles inside $r = r_{\text{in}}$ are on a crash orbit, while the particles outside are on a bounded orbit. Note that due to the preservation of angular momentum, $\dot{\theta}$ is bounded away from zero which ultimately means that the particle is in one way of the other spiraling (within bounds). In contrast, particles in really high energy are on escape/crash orbits – all depending on whether they are ingoing or outgoing. The situation for intermediate levels of energy depends on whether the energy level $\tilde{\mathscr{E}}_{\text{max}}$ is bigger or smaller than zero. A computation shows that $\tilde{\mathscr{E}}_{\text{max}} < 0$ when $12M^2 < J^2 < 16M^2$, and that $\tilde{\mathscr{E}}_{\text{max}} > 0$ when $J^2 > 16M^2$.



Consider first the case of $12M^2 < J^2 < 16M^2$ and the intermediate energy level $\tilde{\mathscr{E}}_{\text{max}} < \tilde{\mathscr{E}} < 0$. In this case the particle is on a crash orbit, even if initially pointed outward. Basically, for how much angular momentum it has the particle has too much energy to stay "near" the stable orbit and yet not enough energy to escape.

Now consider the case of $J^2 > 16M^2$ and the corresponding intermediate energy level $0 < \tilde{\mathscr{E}} < \tilde{\mathscr{E}}_{\text{max}}$. If at some point the particle is located outside the unstable circular inner orbit $r = r_{\text{in}}$, then the particle in on an escape orbit. (This is very much like the hyperboloidal orbits in Newtonian gravity). On the other hand, if the particle finds itself inside $r = r_{\text{in}}$ then it is on a crash orbit. This is true even if the particle is initially pointed outward. In both cases the particle has too much angular momentum for how much energy it has and is forced to turn around (if pointed "the wrong way").



Perihelion advance. As promised earlier, we now investigate small perturbations of the stable outer circular orbits. Note that the these orbits correspond to the smaller of the two solutions of $u - 3u^2 = (\frac{M}{J})^2$. The said solution will be labeled by u_*; its actual value depends on J and M but it always less than $\frac{1}{6}$.

The orbits under the investigation are governed by

$$\frac{d^2u}{d\theta^2} + u - 3u^2 = \left(\frac{M}{J}\right)^2.$$

with $u \approx u_*$. This is not the kind of an equation whose solution we can hope to find in closed form. So, we retreat to linearization at u_*. Let $u = u_* + v$. If $v \approx 0$ then $3u^2 \approx 3u_* + 6u_*v$ and

$$\frac{d^2v}{d\theta^2} + (1 - 6u_*)v \approx 0.$$

The approximate equation is a simple harmonic oscillator equation; its solutions are periodic with period

$$\frac{2\pi}{\sqrt{1 - 6u_*}} \approx 2\pi + 6u_*\pi.$$

The orbits we are investigating are not circular. In fact, they are not even closed! The locations where the minimum (or the maximum) values of r are achieved (or, more intuitively, where the particle in a bound orbit is closest / furthest from the "star") themselves rotate. This phenomenon is called the *perihelion advance.* (Perihelion refers to the point on the orbit which is closest to the "star".)

This is something that one could check/observe within our Solar System. To be honest, perihelion advance exists in the Newtonian description of the Solar System as well: it is caused by gravitational interaction of a planet (say, Mercury) with other planets (say, Jupiter). Such "Newtonian" perihelion advance can and has been numerically estimated. In the case of Mercury Newtonian mechanics predicts the perihelion advance of about 532 seconds of an arc[11] per century. Observations of the French astronomer Urbain Le Verrier, reported in 1859, show that the perihelion advance of Mercury is actually higher: somewhere about 575 seconds per century. The only Newtonian explanation of this anomaly would involve an existence of another planet, possibly closer to the Sun. This actually was the explanation suggested by Le Verrier. He named this hypothetical planet (Star Trek fans rejoice!) Vulcan and made a prediction of when and where it would be visible from the Earth. However, no planet was observed in accordance with Le Verrier's

[11]This is a unit of angular measurement. There are 3 600 seconds in one degree of an arc.

predictions. In 1915 Einstein presented "Explanation of the Perihelion Motion of Mercury from the General Theory of Relativity"[12] which addressed the relativistic perihelion advance discussed in this lecture. Our analysis, when applied to Mercury, suggests a perihelion advance of about 40 seconds of an arc per century. A more careful analysis based on an approximation near an elliptical rather than a perfectly circular orbit puts the perihelion advance at about 43 seconds of an arc per century! This not only solves the anomaly of Mercury's perihelion, but also serves as an evidence of the validity of the theory of general relativity.

Exercises for Lecture 10.

(1) Find Ricci curvature components of $g = -E(r)dt^2 + F(r)dr^2 + r^2 g_{S^2}$. [Intermediate]

(2) Prove Kepler's First Law: Orbit of every planet is an ellipse with Sun at one of the focal points. [Basic]

(3) Go through all of the algebraic details and the special cases of our treatment of particle orbits in Schwarzschild space-time. Make sure you address the details such as: why are $\frac{J^2}{M^2} = 12$ and $\frac{J^2}{M^2} = 16$ critical, what happens when $\frac{J^2}{M^2} = 12$ or $\frac{J^2}{M^2} = 16$, what happens when $J = 0$, etc. [Basic]

(4) Find photon orbits in Schwarzschild space-time. This can be done in the exact same manner as what was done in the lecture except that the line element needs to be null along a photon orbit. [Intermediate]

(5) Investigate the gravitational field of a relativistic star in a spherically symmetric vacuum with cosmological constant Λ. Follow the guidelines given below. [Intermediate]

    (a) Show that $g = -(1-\frac{2M}{r}-\frac{\Lambda}{3}r^2)dt^2+(1-\frac{2M}{r}-\frac{\Lambda}{3}r^2)^{-1}dr^2+r^2 g_{S^2}$ solves the vacuum Einstein equations with the cosmological constant Λ. (For $\Lambda > 0$ this space-time is called Schwarzschild-deSitter space-time and for $\Lambda < 0$ this space-time is called Schwarzschild-anti deSitter space-time.)

    (b) Discuss particle orbits in Schwarzschild-anti de Sitter spacetimes, following the same steps as those taken in today's lecture.

[12] Erklärung der Perihelbewegung des Merkur aus der allgemeinen Relativitätstheorie

CHAPTER 3

Insight into Geometric Analysis

11. The PDE toolbox

Introduction. The most basic PDEs are the linear ones, i.e. the ones of the form $Lu = f$ where L is

- linear, in the sense that $L(c_1 u + c_2 v) = c_1 Lu + c_2 Lv$ for constants c_1, c_2;
- differential operator, in the sense that it is made out covariant derivatives of u by means of algebraic manipulations.

It turns out that certain types of linear differential operators, called the *elliptic operators*, exhibit the behavior akin to those of linear transformations between finite dimensional vector spaces. This is particularly true of *self-adjoint linear elliptic operators* such as $L = \Delta - c$ (cf. Theorem 8). For instance, there is a self-adjoint linear elliptic PDE analogue of what some basic linear algebra texts call the *Fundamental Subspaces Theorem*: $\mathrm{Im}(L) = \mathrm{Ker}(L)^{\perp}$ for symmetric $n \times n$ matrices L acting on $\mathbb{R}^n$ with the Euclidean inner-product. Another way to phrase this theorem is to say that the linear equation $Lu = f$ in Euclidean $\mathbb{R}^n$ has a solution if and only if f is orthogonal to $\mathrm{Ker}(L)$. Our goal for today is to gain a basic understanding of the inner-product spaces appropriate for solving linear elliptic equations on *compact manifolds without boundary*. For simplicity we assume the metric g is smooth.

Disclaimer No. 3. A rigorous treatment of the subject relies on the knowledge of measure theory and functional analysis. The purpose of this lecture is to provide a very elementary big-picture overview of the linear algebra framework used in (elliptic) PDE's. Most of the discussion will consist of heuristic arguments which will hopefully provide some context and motivation for a serious technical study of functional analysis and linear PDE's down the road.

The function space $L^2(M)$. We start by generalizing the Euclidean inner-product space $\mathbb{R}^n$ to an inner-product space that linear differential operators can operate on. To that end we take the following perspective on vectors in $\mathbb{R}^n$: they are *functions* $\mathbf{u} : \{1, 2, ..., n\} \to \mathbb{R}$. (Under this convention $\mathbf{u} = (u^1, ..., u^n)^T$ corresponds to the function $\mathbf{u}(i) = u^i$.) From this point of view the Euclidean inner-product takes the form of

$$\langle \mathbf{u}, \mathbf{v} \rangle = \sum_i \mathbf{u}(i)\mathbf{v}(i),$$

and the vector magnitude / norm becomes $\|\mathbf{u}\|^2 = \sum_i \mathbf{u}(i)^2$.

More generally, functions $u : M \to \mathbb{R}$ can also be thought of as vectors. In such a setting vector addition and scalar multiplication are given by point-wise operations $(u + v)(P) = u(P) + v(P)$ and $(c \cdot u)(P) = c \cdot u(P)$ for $P \in M$ and constants c. One natural extension of the Euclidean inner-product described above is then given by

$$\langle u, v\rangle = \int_M uv \,\mathrm{dvol}.$$

The functions u and v need not be smooth or even continuous for this inner-product to be defined. Due to Cauchy-Schwarz inequality these functions only need to be measurable (in some sense of the word) and square integrable:

$$\int_M u^2 \,\mathrm{dvol} < \infty.$$

It should be emphasized that, as the example of functions $f(x) = x^\alpha$ for $\alpha > -\frac{1}{2}$ on $(0,1)$ suggests, square integrability permits unboundedness:

$$\int_0^1 f(x)\,dx = \int_0^1 x^{2\alpha}\,dx = \frac{1}{2\alpha+1} < \infty, \qquad \lim_{x\to 0^+} f(x) = \infty.$$

The collection of square integrable functions on a (compact) Riemannian manifold (M, g) is denoted by $L^2(M; g)$ or just L^2 when the manifold is clear from context. When confusion is possible we use $\langle ., .\rangle_{L^2}$ to distinguish the L^2 from other inner-products.

It is insightful to think of the L^2-norm $\|u\|_{L^2}$ as a measure of the average size of u. We explore this on two examples.

Examples.

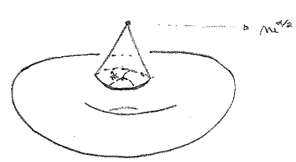
- Let $m \gg 1$ and let $B_{1/m}$ denote the geodesic ball of radius $1/m$ centered at some point P of an n-dimensional Riemannian manifold (M, g). Recall that $\mathrm{vol}(B_{1/m}) = O((\frac{1}{m})^n)$ as $m \to \infty$, in the sense that

$$\lim_{m\to\infty} \frac{\mathrm{vol}(B_{1/m})}{(\frac{1}{m})^n} = \mathrm{vol}(S^{n-1}).$$

  Consider the function $u_m : M \to \mathbb{R}$ which is non-zero only within $B_{1/m}$ where it is equal to

$$u_m = m^{n/2} - m^{(n/2)+1} r,$$

  with r denoting the geodesic distance to P. The graphs of u_m are cone-like:



  We have $\int_M u_m^2 \,\mathrm{dvol} \le m^n \mathrm{vol}(B_{1/m}) \to \mathrm{vol}(S^{n-1})$ as $m \to \infty$. In particular, the L^2-norms $\|u_m\|_{L^2}$ are all bounded in m. The point here is that even though point-wise u_m reach large values ($\max u_m = m^{n/2}$), the functions u_m are not all that big in average. The large size of the value of u_m is compensated for by the small size of the geodesic ball $B_{1/m}$.

- Similar to the above consider the function $v_m : M \to \mathbb{R}$ which is non-zero only within $B_{1/m}$ where it is equal to

$$v_m = m^{\alpha/2} - m^{(\alpha/2)+1} r \ \text{ for } \ \alpha \approx n,\ \alpha < n.$$

As above we compute that $\int_M v_m^2 \,\mathrm{dvol} \le m^\alpha \mathrm{vol}(B_{1/m}) \to 0$. In average, the functions v_m are more and more like 0.

This brings us to the issue of *convergence in* L^2. The L^2-norm gives rise to the notion of L^2-distance

$$\|u - v\|_{L^2} = \int_M (u-v)^2 \,\mathrm{dvol}.$$

As in any other analysis setting we say that $u_m \to u$ *in* L^2 if for all $\varepsilon > 0$ there is m_* such that $\|u_k - u\|_{L^2} < \varepsilon$ for all $k > m_*$. Thus, in the second example we had $v_m \to 0$ in L^2.

Here is one very important distinction between convergence in $\mathbb{R}^n$ and convergence in L^2. In $\mathbb{R}^n$ we have $\mathbf{u}_m \to \mathbf{u}$ if and only if $\mathbf{u}_m(i) \to \mathbf{u}(i)$ for all $1 \le i \le n$. An analogous statement is false in L^2: the L^2-convergence does not imply point-wise convergence. This is, in fact, evident from the second example above where we had $v_m(P) \to \infty$ while $v_m \to 0$ in L^2. Nevertheless, there are still some similarities between $\mathbb{R}^n$ and L^2. An important feature that the two spaces have in common is that they are both *complete*. In other words, Cauchy sequences[1] (characterized by the property that for all $\varepsilon > 0$ there is some m_* such that for all $m, l > m_*$ we have $\|u_m - u_l\|_{L^2} < \varepsilon$) are necessarily convergent. Inner-product spaces which are complete are called *Hilbert spaces*. Overall, L^2 is an infinite dimensional Hilbert space.

The inclusion of *all* square integrable functions in the definition of L^2 is crucial for its completeness; the completeness result would no longer hold if one focused just on "nice" (e.g. smooth) square integrable functions. What is true, however, is that smooth functions are *dense in* L^2 i.e. that for all $u \in L^2$ there is a sequence of smooth u_m such that $u_m \to u$ in L^2.

Finally, it should be noted that L^2-notions of inner-product, norm and distance can be established for tensors of all types not just for scalar-valued functions. In the setting of vector fields the L^2 inner-product can be computed as

$$\langle V, W\rangle_{L^2} := \int_M g(V, W) \,\mathrm{dvol}$$

and the L^2-norm is then given by $\|V\|_{L^2}^2 = \int_M |V|_g^2 \,\mathrm{dvol}$. We say that $V \in L^2(M)$ if $\int_M |V|_g^2 \,\mathrm{dvol} < \infty$. Once the point-wise inner-product and norm are appropriately interpreted, this exact approach can be carried out in the cases of other tensor types. For instance, in the case of covectors ω, η the point-wise inner-product can be defined in coordinates by $\langle \omega, \eta\rangle_g = \sum_{ij} g^{ij}\omega_i \eta_j$. One then verifies that this expression is independent of the choice of coordinates and that it, therefore, defines a point-wise inner-product on covectors. (Compare with exercise 2.) Integration of point-wise inner-product gives us the L^2-inner product, and we write $\omega \in L^2(M)$ if $\int_M |\omega|_g^2 \,\mathrm{dvol} < \infty$. The general pattern can be induced from the following point-wise inner-product formula for $(2,1)$-tensors:

$$\langle T, S\rangle_g = \sum_{ijklpq} g^{ij} g^{kl} g_{pq} T_{ik}^p S_{jl}^q.$$

[1] Sequences whose terms, loosely speaking, are getting closer and closer to one another

Sobolev spaces. An obvious complaint about the examples brought up above is that they feature functions whose "rates of change" within $B_{1/m}$ are becoming increasingly large. Very roughly speaking, in the second example we may think of $-m^{(\alpha/2)+1}$ as measuring this "rate of change within $B_{1/m}$". The fact that $\int_{B_{1/m}} m^{\alpha+2}\,\mathrm{dvol} = O(m^{\alpha-n+2})$ then suggests that the "rate of change" of v_m is not small – not even in average L^2 sense of the word. Of course, the excessive use of quotes is due to the fact that v_m are not differentiable. Nevertheless, the above provides an insight into the nature of convergence in L^2 and a motivation to study the average size of the derivatives of a function, along with the average size of a function.

For integers $k \geq 0$ and smooth functions $u, v : M \to \mathbb{R}$ we define

$$\langle u, v\rangle_{H^k} := \int_M uv + \langle \nabla u, \nabla v\rangle_g + \cdots + \langle \nabla^k u, \nabla^k v\rangle_g \,\mathrm{dvol}_g$$

and the corresponding H^k-norm $\|u\|^2_{H^k} = \langle u, u\rangle_{H^k}$. Thus, H^k-norms measure the average size of a function and its first k covariant derivatives. Note that H^0 and L^2 mean the same.

The new H^k-norm defines a new concept of distance: $\|u - v\|_{H^k}$. Convergence $u_m \to u$ with respect to this distance function not only means that u_m in average becomes more and more like u but also that the derivatives u_m in average become more and more like the corresponding derivatives of u. The discussion above suggests that $u_m \to u$ in H^k for high enough k might imply point-wise convergence.

There is a subtle point here. The H^k-norm on the set of smooth functions does not give rise to a complete space: there are Cauchy sequences u_m of smooth functions in H^k which do not converge (in H^k-sense) to a smooth function. In particular, it is possible to have a sequence of smooth functions u_m which converge to some non-differentiable $u \in L^2$, such that ∇u_m converge to some covector field $\omega \in L^2$. The following example illustrates such a situation. For reasons of notational ease the example is chosen to be over the interval $[-1, 1]$; similar examples exists over any (compact) manifold.

Example. Consider the sequence $u_m(x) = \sqrt{x^2 + \frac{1}{m^2}}$ of functions of $x \in [-1, 1]$, and consider the function $u(x) = |x|$ over the same interval. It follows easily from $|x| \leq u_m(x) \leq |x| + \frac{1}{m}$, that $u_m \to u$ in the L^2-sense. We would like to study the behavior of the derivatives $u_m'(x) = \frac{x}{\sqrt{x^2+\frac{1}{m^2}}}$. In average, these derivatives converge to the function

$$v(x) = \begin{cases} 1 & \text{if } \ x \geq 0; \\ -1 & \text{if } \ x < 0. \end{cases}$$

Indeed, we have:

$$\int_{-1}^{1} (u_m' - v)^2\,dx = \int_0^1 \frac{2\,dx}{((mx)^2+1)\left(mx+\sqrt{(mx)^2+1}\right)^2} < \frac{2}{m}\int_{-\infty}^{\infty} \frac{ds}{(s^2+1)\,(s+\sqrt{s^2+1})^2} \to 0.$$

This should make some intuitive sense as well: the derivatives of $|x|$ on $(0, 1)$ and $(-1, 0)$ are 1 and -1, respectively. A natural conjecture at this point would be that the second derivatives $u_m''(x) = \frac{1}{m^2}(x^2 + \frac{1}{m^2})^{-3/2}$ in average converge to zero.

However, this is false:

$$\int_{-1}^{1} (u_m'')^2 \,\mathrm{dvol} = m^2 \int_{-1}^{1} \frac{dx}{((mx)^2+1)^3} = m \int_{-m}^{m} \frac{ds}{(s^2+1)^3} \to +\infty,$$

as the integral $\int_{-\infty}^{\infty} \frac{ds}{(s^2+1)^3}$ converges (to a positive value). In fact, being unbounded in L^2-sense the sequence u_m'' cannot be convergent in L^2-sense. The overall conclusion here is that the sequence u_m is Cauchy in H^1-sense but not in H^2-sense.

Suppose $\{u_m\}$ is a sequence of functions which are Cauchy in H^k-sense. We then have that $\{u_m\}$, $\{\nabla u_m\}$, ..., $\{\nabla^k u_m\}$ are all Cauchy in L^2. Consequently, we have that $u_m \to u$, $\nabla u_m \to \omega$, ..., $\nabla^k u_m \to T$ in L^2 for some u, ω, ..., T. In such a situation we say that ω is the *weak covariant derivative* of u, ..., and that T is the *weak k-th covariant derivative u*. So, for example, the piecewise-defined function v from the previous example is the weak derivative of $u(x) = |x|$. On the other hand, the function $u(x) = |x|$ does not have two weak covariant derivatives in L^2. (It does have two weak derivatives in the sense of distributions, but this concept goes beyond the scope of our lectures.) A relatively easy integration by parts argument shows that (the existence of) the weak covariant derivatives of u do not depend on the choice of the approximating smooth sequence u_m. (See exercise 3.) Thus, the concepts of weak covariant derivatives are well-defined.

The *Sobolev*[2] *space* $H^k(M)$ consists of all L^2 functions which have k weak covariant derivatives. It is a complete inner-product space, i.e. it is a Hilbert space. The subspace of smooth functions is dense in $H^k(M)$. Clearly, Sobolev spaces are infinite dimensional.

Sobolev Embedding. It was hinted above that convergence in H^k-sense for large k implies point-wise convergence. The latter is perhaps more directly addressed by the following normed vector spaces. We let $C^k(M)$ denote the vector space of all functions which have k continuous covariant derivatives. We introduce the C^k-norm[3] by:

$$\|u\|_{C^k} := \max_M \{|u| + |\nabla u|_g + ... + |\nabla^k u|_g\}.$$

This norm gives rise to the concept of C^k-distance between functions, $\|u - v\|_{C^k}$, and thus the concept of C^k-convergence. A classic exercise of elementary analysis consists of showing that all Cauchy sequences in $C^k(M)$ converge, i.e. that C^k is complete. As a complete normed vector space, $C^k(M)$ is a *Banach space.*

Clearly, we have $C^k(M) \subseteq H^k(M)$. In fact, if $u \in C^k(M)$ then the point-wise inequality $|\nabla^i u|_g \leq \|u\|_{C^k}$ implies that

$$\|u\|_{H^k} \leq C\|u\|_{C^k}$$

for some constant C depending only on k and the volume of M. It is now easy to see that convergence in $C^k(M)$ implies convergence in $H^k(M)$. Seeing how Sobolev spaces fit inside C^k-spaces is much more difficult. The proof of the following Theorem goes much beyond the level of these lectures.

---

[2]Named in honor of the Soviet mathematician Sergei Lvovich Sobolev who, through introduction of these spaces in 1930's, made substantial progress in PDE theory.

[3]Our use of max is only appropriate as we are assuming that M is a compact manifold without boundary; otherwise M one should use the supremum $\sup_M$ instead.

THEOREM 25 (SOBOLEV EMBEDDING THEOREM). *Assume* $k > l + \frac{n}{2}$. *We have* $H^k(M) \subseteq C^l(M)$ *and* $\|u\|_{C^l} \leq C\|u\|_{H^k}$ *for a constant* C *independent of* u.

There are several very deep aspects of this result. One is that a function with sufficiently many weak covariant derivatives is necessarily continuous. More precisely, if a function has $k > l + \frac{n}{2}$ weak covariant derivatives then it has l continuous (strong) derivatives. Another aspect is that the average size of a function and sufficiently many of its derivatives controls the size of the size of a function (and several of its derivatives) at each point. Finally, if $k > l + \frac{n}{2}$ then H^k-convergence (i.e. in-average convergence of a sequence of functions and k of its weak derivatives) is sufficient to conclude the C^l-convergence (i.e. uniform point-wise convergence of l first covariant derivatives of u.)

Rellich Lemma. There is a number of theorems of analysis which amount to saying that controlling the behavior of the derivatives of a sequence of functions $\{u_m\}$ can tell us something about its convergence. One such example is the Arzelá-Ascoli Theorem, proven through combined efforts of Cesare Arzelá and Giulio Ascoli in 1880's, which states that on compact sets every bounded family of equicontinuous[4] functions has a uniformly convergent subsequence. In view of the classic Mean Value Theorems this is to say that every sequence $\{u_m\}$ of functions with bounded $C^1(M)$-norm:

$$\|u_m\|_{C^1} \leq C \text{ for some constant } C \text{ independent of } m,$$

on a compact manifold M without boundary, has a subsequence $\{u_{m_i}\}$ which converges in $C^0(M)$. In general, boundedness in $C^k(M)$ implies subsequential convergence in $C^{k-1}(M)$.

A similar result exists for Sobolev spaces; it is commonly attributed to Franz Rellich and is known under several different names such as Rellich Lemma or Rellich-Kondrakov Compactness Theorem. As stated below it applies to compact manifolds without boundary only.

THEOREM 26 (RELLICH LEMMA). *Every* H^k*-bounded sequence has a convergent subsequence in* $H^{k-1}(M)$.

Elliptic Operators. Interpreting the derivatives in terms of weak derivatives if necessary, we see that linear differential operators give rise to $\mathbb{R}$-linear maps between Sobolev spaces. More precisely, if a linear differential operator L includes covariant derivatives of order up to and including l then we say that the operator is of order l and we have that

$$L : H^{k+l}(M) \to H^k(M), \quad k \geq 0.$$

As announced in the introduction to this lecture, there is a great deal of similarity between certain linear differential operators acting on Sobolev spaces and linear transformations acting on $\mathbb{R}^n$. The operators being alluded to here are *elliptic differential operators.* In the interests of brevity, more than anything else, we will avoid the general treatment of notion of ellipticity and focus on examples.

Second order linear operators L can in coordinates be expressed as

$$Lu = \sum_{ij} a^{ij}\partial^2_{ij}u + \sum_i b^i\partial_i u + cu$$

[4]Equicontinuity requires that, for a given ε, the same δ can be used at all points of the domain and for each function of the family.

for some functions a, b and c. Such an operator is elliptic if and only if the matrix $[a^{ij}]$ of each of its coordinate representations is always positive (or always negative) definite. For example, the Laplacian takes the form of

$$\Delta u = \frac{1}{\sqrt{|g|}} \sum_{ij} \partial_i (g^{ij} \sqrt{|g|} \partial_j u) = \sum_{ij} g^{ij} \partial^2_{ij} u + \text{ lower order terms.}$$

Therefore the Laplace operator, or more generally any operator L of the form $L = \Delta - c$ for some function c, is elliptic. Although there exist elliptic operators of orders other than 2, we will always be working with operators of order 2.

Elliptic Regularity. One extremely remarkable feature of elliptic operators is their ability to detect and control the size of individual covariant derivatives of a function. This is the content of miscellaneous *Regularity Theorems*. One such theorem is given below; as stated it only holds on compact manifolds without boundary.

THEOREM 27 (ELLIPTIC REGULARITY). *Let L be a linear elliptic operator. If for some $u \in H^2(M)$ we have $Lu \in H^k(M)$ then $u \in H^{k+2}(M)$ and*

$$\|u\|_{H^{k+2}} \leq C \left(\|Lu\|_{H^k} + \|u\|_{L^2} \right).$$

for some constant C independent of u.

To see the power of this result consider $u \in H^2(M)$ such that $Lu = 0$. Elliptic regularity implies $u \in H^k(M)$ for all k. By Sobolev embedding we have that $u \in C^l(M)$ for all l, i.e. that u is smooth. In particular, the kernel of the operator $L : H^{k+2}(M) \to H^k(M)$, denoted $\mathrm{Ker}(L)$, is independent of k and its elements are smooth. Furthermore, in the cases when $\mathrm{Ker}(L) = \{0\}$ the elliptic estimate can be improved. The corresponding proof is a very nice application of the above; see exercise 7 following this lecture.

THEOREM 28. *Let L be a linear elliptic operator with* $\mathrm{Ker}(L) = \{0\}$. *Then*

$$\|u\|_{H^{k+2}} \leq C \|Lu\|_{H^k}$$

for some constant C independent of $u \in H^{k+2}(M)$.

Fredholm Theory. At the beginning of this lecture we mentioned that there are versions of the Fundamental Subspaces Theorem for linear elliptic operators acting on Sobolev spaces. This is the subject of something called Fredholm[5] Theory. We state the Theorem for the case of self-adjoint elliptic operators on compact manifolds without boundary, i.e. operators L which satisfy

$$\langle Lu, v \rangle_{L^2} = \int_M (Lu) v \,\mathrm{dvol} = \int_M u(Lv) \,\mathrm{dvol} = \langle u, Lv \rangle_{L^2}.$$

We saw in Theorem 8 that the Laplace operator, as well as any other operator of the form $L = \Delta - c$ for functions c, is self-adjoint. Self-adjointness of an operator generalizes the symmetry property of matrices.

[5]Named in honor of the Swedish mathematician Erik Ivar Fredholm whose related work in integral equations anticipated that of David Hilbert and served as a motivation for the introduction of Hilbert spaces.

THEOREM 29. *Let L be a self-adjoint second order linear elliptic operator on a compact manifold without boundary. The equation $Lu = f$ with $f \in L^2(M)$ has a solution if and only if $f \in (\mathrm{Ker}(L))^\perp$, in which case $u \in H^2(M)$. In particular, if $\mathrm{Ker}(L) = \{0\}$ then for each $f \in L^2(M)$ there exists a unique solution $u \in H^2(M)$ of $Lu = f$.*

We end by noting that the condition $\mathrm{Ker}(L) = \{0\}$ can, in many situations, be verified through means of a maximum principle. For instance, Weak Maximum Principle (see Theorem 10) shows that $\mathrm{Ker}(\Delta - c) = \{0\}$ for positive functions c. In view of Elliptic Regularity we see that equations of the form

$$\Delta u - cu = f$$

have unique solutions $u \in H^{k+2}(M)$ for each and every $f \in H^k(M)$.

Exercises for Lecture 11.

(1) Let r be a smooth positive function on n-dimensional compact (M, g) such that near a point $P \in M$ the function r is identical to the distance away from P and such that away from P the function r is constant. For what α is r^α an element of $L^2(M)$? [INTERMEDIATE]

(2) (a) Let ω and η be two covector fields. Show that the value of $\sum_{ij} g^{ij}\omega_i\eta_j$ is independent of the choice of coordinates. Based on this, show that the formula $\langle \omega, \eta\rangle_g := \sum_{ij} g^{ij}\omega_i\eta_j$ defines a positive-definite point-wise inner-product of covectors. [BASIC]

(b) Repeat the above for the formula $\langle T, S\rangle_g := \sum_{ijklpq} g^{ij}g^{kl}g_{pq}T^p_{ik}S^q_{jl}$ on $(2,1)$-tensors. [BASIC]

(c) Generalize the above to $(l, 0)$ and $(l, 1)$-tensors. That is, state the formula for the point-wise inner-product of two $(l, 0)$ or $(l, 1)$-tensors. [BASIC]

(d) Define the L^2 and the H^k inner-products and norms of arbitrary (smooth) $(l, 0)$ or $(l, 1)$ tensors. [BASIC]

(e) Compute the H^k-norm of the metric tensor g on a compact manifold (M, g). [BASIC]

(3) Let $\{u_m\}$ and $\{v_m\}$ be sequences of smooth functions on a compact manifold M without boundary. Assume the following convergences in $L^2(M)$:

$$u_m, v_m \to u, \quad \nabla u_m \to \omega, \quad \nabla v_m \to \eta.$$

Apply Integration by Parts to $\int_M \nabla_X(u_m - v_m)\,\mathrm{dvol}$ to obtain that

$$\int_M \omega(X) - \eta(X)\,\mathrm{dvol} = 0$$

for all (smooth) vector fields X; conclude that $\omega = \eta$. [INTERMEDIATE]

(4) Show that convergence $u_m \to u$ in C^k-sense implies the convergence in the H^k-sense. [BASIC]

(5) (a) Show that if $u_m \to u$ in $C^k(M)$ and $v_m \to v$ in $H^k(M)$ then we have $u_m v_m \to uv$ in $H^k(M)$. [INTERMEDIATE]

(b) Assume that $\dim(M) = n$ and that $k > \frac{n}{2}$. If both $u_m \to u$ and $v_m \to v$ in $H^k(M)$ then $u_m v_m \to uv$ in $H^k(M)$. [INTERMEDIATE]

(6) Let L be an linear elliptic second order operator on compact (M, g) without boundary.

(a) Use Elliptic Regularity and Rellich Lemma to show that every bounded sequence of elements of $\mathrm{Ker}(L) \subseteq L^2(M)$ has a convergent subsequence in $L^2(M)$. [INTERMEDIATE]

(b) Show that $\mathrm{Ker}(L)$ is finite dimensional. Hint: assume the opposite and consider an orthonormal basis for $\mathrm{Ker}(L) \subseteq L^2(M)$. [INTERMEDIATE]

(c) If u_* is a solution of $Lu = f$ and if $u_1, \ldots, u_l$ form a basis for $\mathrm{Ker}(L)$ then the general solution of $Lu = f$ can be expressed in the form of $u = u_* + c_1 u_1 + ... + c_l u_l$. [BASIC]

(7) Let L be an linear elliptic second order operator on compact (M, g) without boundary.

(a) Suppose that a sequence of smooth functions $\{u_m\}$ on compact (M, g) is such that $\|u_m\|_{H^2} = 1$ and $\|Lu_m\|_{L^2} \to 0$. Use Rellich Lemma and Elliptic Regularity to show that there exists a convergent subsequence $u_{m_i} \to u$ in $H^2(M)$ with $Lu = 0$. [INTERMEDIATE]

(b) Use the above to prove Theorem 28. [INTERMEDIATE]

(8) Let $\mathcal{V}$ be an infinite dimensional vector space, and let $L : \mathcal{V} \to \mathcal{V}$ be a linear operator.

(a) If L is one-to-one, does it have to be onto?

(b) If L is not one-to-one, does it have to have a finite dimensional kernel?

If your answer is no, please provide a counter example. [BASIC]

(9) Let $\langle ., . \rangle$ denote the Euclidean inner-product in $\mathbb{R}^n$ and let A be an $n \times n$ matrix. Show that for all vectors $\mathbf{v}, \mathbf{w} \in \mathbb{R}^n$ we have

$$\langle A\mathbf{v}, \mathbf{w} \rangle = \langle \mathbf{v}, A^T \mathbf{w} \rangle.$$

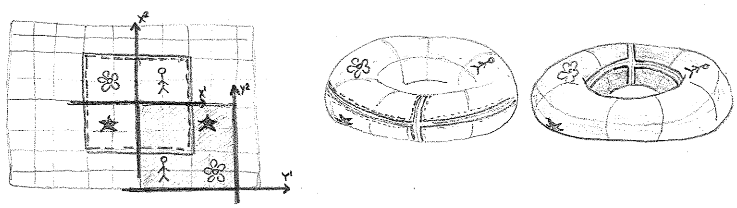
In particular, conclude that the linear transformations defined by symmetric matrices are self-adjoint. [BASIC]

(10) Use the Strong Maximum Principle to show that the equation $\Delta u = f$ has a solution if and only if $\int_M f \, \mathrm{dvol} = 0$. [INTERMEDIATE]

12. Example of a problem from geometric analysis

Introduction. Many a problem in Riemannian geometry can be reduced to a problem in differential equations. We have already seen how the study of geodesics and Jacobi vector fields leads to an analysis of certain ODE's. In this last lecture we solve a geometric problem through a PDE. As is the case with many geometric PDE's, the equation we will be working with is non-linear and demands some creativity with the PDE toolbox: the Maximum Principle, Sobolev spaces, Sobolev Embedding, Rellich Lemma, Elliptic Regularity and Fredholm Theory. The analysis we undertake in this lecture is probably not the "best" way of proving the claimed result, but it is an extremely representative (and yet very accessible) example from the field of *geometric analysis*.

A problem from geometry. We have worked out in detail the curvature of the standard 2-dimensional sphere S^2 in $\mathbb{R}^3$: it has constant positive[6] curvature. The 2-dimensional torus sitting inside of $\mathbb{R}^3$ does not inherit the metric of constant curvature (e.g. compare the behavior of the small geodesic balls on the inside and on the outside portion of a donut). However, it can (!) be endowed with a Riemannian metric of constant curvature. One can coordinatize the torus using several Euclidean unit squares in such a way that the transition functions are translations.

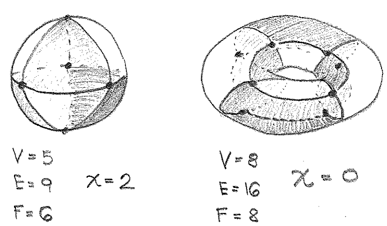


Thus $ds^2 = d(x^1)^2 + d(x^2)^2$ gives rise to a well-defined global Riemannian metric on the torus; its curvature is zero. What about tori with several holes? Can they be given metrics of constant curvature?

Topological considerations. A feature shared by all the tori with multiple holes is that they have a negative Euler characteristic

$$\chi = V - E + F$$

where V denotes the number of vertices, E denotes the number of edges and F denotes the number of faces of a polygonal decomposition.



In fact, one can prove that tori with multiple holes have the Euler characteristic of $2 - 2G$ where G denotes the number of holes. The Gauss-Bonnet Theorem $\int_M \mathscr{K}\,\mathrm{dvol} = 2\pi\chi(M)$ shows that although there is a great deal of freedom in choosing a Riemannian metric on a manifold M all such metrics must have the same total sectional curvature. Total sectional curvature of any metric on S^2 is always 4π; total curvature of any metric on a torus in always zero; total curvature of any metric of a torus with multiple holes is always negative. Should a torus with multiple holes carry a metric of constant curvature, this constant curvature has to be negative!

One should note the irony of the situation. By the end of the lecture we will prove that tori with multiple holes do indeed cary metrics of constant negative

[6]This refers to both sectional and scalar curvature; note that in dimension 2 we have the relationship Scal $= 2\mathscr{K}$.

curvature. In fact, the proof applies to any 2-dimensional manifold of negative Euler characteristic. In some sense of the word, such manifolds are more common than those whose Euler characteristic is non-negative. The result being referred to here is the classification of 2-dimensional compact manifolds without boundary, according to which there are only four manifolds (up to elastic deformations) of non-negative Euler characteristic: 2-dimensional sphere, real projective plane, torus and the Klein bottle. If constant curvature geometry is the "optimal" or "natural" geometry for a 2-dimensional compact manifold, then the "optimal" geometry is more commonly that of constant negative curvature, that is, (locally) hyperbolic. Given how controversial the discovery of hyperbolic geometry was, the fact that it is so common is a quite a quirk of history!

Reduction to a problem in geometric analysis. One way to find a metric of constant curvature on a manifold is to start with any metric g and make a *pointwise conformal deformation* $h := e^u g$ such that the resulting metric h has constant curvature. The function u and the scalar curvatures of g and h relate as follows

$$\mathrm{Scal}_h = e^{-u}\left(\mathrm{Scal}_g - \Delta_g u\right), \quad \text{i.e.} \quad \Delta_g u = \mathrm{Scal}_g - \mathrm{Scal}_h e^u,$$

where Δ_g denotes the Laplace operator defined with respect to the metric g. Viewing the latter as a PDE on (M, g) for the unknown u, we see that any solution of the above with $\mathrm{Scal}_h = \mathrm{const}$ gives rise to a metric h with constant scalar curvature.

Recall that in the cases of interest – when M is compact, two dimensional, without boundary and of negative Euler characteristic – the obstruction coming from the Gauss-Bonnet Theorem forces us to consider $\mathrm{Scal}_h = \mathrm{const} < 0$. For this reason, we choose to work with $\mathrm{Scal}_h = -2$. Likewise, we know that the starting metric g automatically has negative total curvature $\int_M \mathrm{Scal}_g \,\mathrm{dvol}_g < 0$. Overall, we have reduced a geometric problem of finding a metric of (negative) constant curvature to finding solutions of a non-linear elliptic PDE of the form

$$\Delta u = s + 2e^u, \quad \text{where} \quad \int_M s \,\mathrm{dvol} < 0.$$

Rather than just stating and applying a theorem from the theory of PDE's guaranteeing the existence of solution(s) of our equation, we shall examine one particular strategy of solving non-linear PDE's. The reason for this is perhaps best articulated by Jerry L. Kazdan in his "Applications of Partial Differential Equations to Problems in Geometry"[7].

> In many ways, partial differential equations is a subject whose essence is more a body of techniques rather than a body of theorems. One of the easiest ways to learn these techniques is to see how they can be applied to simple examples.

In the spirit of this, we examine how the basic elements of the PDE toolbox fit together into one particularly simple method of solving non-linear elliptic PDE's.

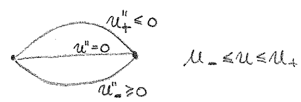
The Method of Sub- and Super-solutions. To begin, observe that the Weak Maximum Principle also serves as a Comparison Principle. Indeed, under the assumptions of the Weak Maximum Principle we have that $Lu \leq Lv$ (i.e.

[7] I highly recommend this set of lecture notes. It is available at J. Kazdan's website `http://www.math.upenn.edu/~kazdan/`

$L(u-v) \leq 0$) implies $u \geq v$. In particular, if u_-, u and u_+ are functions such that $Lu_- \geq f$, $Lu = f$ and $Lu_+ \leq f$ then also

$$u_- \leq u \leq u_+.$$

An admittedly oversimplified heuristic visual interpretation of these inequalities is that graphs of concave up functions ($u_-'' \geq 0$) and located under the graphs of linear functions ($u'' = 0$), which is turn are located under the graphs of concave down functions ($u_+'' \leq 0$), assuming the same boundary conditions.



Conversely, if for some reason one has two functions $u_- \leq u_+$ such that

$$Lu_- \geq f \quad \text{and} \quad Lu_+ \leq f,$$

then one should expect to find the solution of $Lu = f$ between them: $u_- \leq u \leq u_+$. The functions u_- ad u_+ in such a context are called *the sub- and the super- solutions*, respectively. The key observation now is that a similar line of reasoning applies to non-linear elliptic equations of the form $Lu = G(u)$ with G non-increasing, and more generally to equations of the form $\Delta u = F(x, u)$.

THEOREM 30. *Let M be a compact Riemannian manifold without boundary. Consider the non-linear equation $\Delta u = F(x, u)$ where $F(x, u)$ is a smooth function of $x \in M$ and the scalar variable u. Assume that for some smooth functions u_- and u_+ we have:*

$$u_-(x) \leq u_+(x), \quad F(x, u_-) \leq \Delta u_-, \quad \Delta u_+ \leq F(x, u_+)$$

at each $x \in M$. Then there is a smooth function u_* such that $u_-(x) \leq u(x) \leq u_+(x)$ and $\Delta u_* = F(x, u_*)$.

We present the proof in low dimensions where the Sobolev embedding implies $H^2(M) \subseteq C^0(M)$. The proof in higher dimensions is essentially the same, except that in place of Sobolev spaces $H^k(M)$ one has to use L^p-Sobolev spaces $H^{2,p}(M)$ for $p \gg 1$.

Re-write the problem so to have a monotone non-linearity and so that the PDE tools such as the Maximum Principles apply. Let $a \leq \min_M u_-$ and $\max_M u_+ \leq b$. Furthermore, let $c > 0$ be a constant such that $\partial_u F\big|_{M \times [a,b]} \leq c$. Our equation is equivalent to

$$Lu = G(x, u),$$

where $Lu = \Delta u - cu$ and where $G(x, u) = F(x, u) - cu$ for each $x \in M$ is a non-increasing function of u. Note that we still have the inequalities

$$G(x, u_-) \leq Lu_- \quad \text{and} \quad Lu_+ \leq G(x, u_+).$$

Also observe that L is a linear self-adjoint elliptic operator and that $\mathrm{Ker}(L) = \{0\}$ due to the Weak Maximum Principle, Theorem 10. It then follows from Elliptic Regularity (Theorem 28) and Fredholm theory (Theorem 29) that the equations of the form $Lu = f$ for $f \in H^k(M)$ are uniquely solvable in $H^{k+2}(M)$ and that

$$\|u\|_{H^{k+2}} \leq C \|Lu\|_{H^k}$$

for some constant C independent of u. In fact, by Elliptic Regularity we also know that if f is smooth so is u. Finally, we would like to point out that $Lu \leq Lv$, i.e. $L(u-v) \leq 0$, implies $u \geq v$ by the Weak Maximum Principle, Theorem 10.

Use Fredholm Theory to create a sequence of smooth approximate solutions. The iterative scheme

$$u_0 := u_+, \quad Lu_m = G(x, u_{m-1})$$

defines a sequence of smooth functions $\{u_m\}$. Our goal is to show that $\{u_m\}$ converges to a smooth solution of $Lu = G(x,u)$.

Use Maximum Principles to show that the sequence of approximates is monotone and bounded. Specifically, one can use induction to show:

$$a \leq u_- \leq \cdots \leq u_m \leq u_{m-1} \leq \cdots \leq u_+ \leq b.$$

To that end assume $u_- \leq u_m \leq u_+$. From the monotonicity of G we obtain $G(x,u_+) \leq G(x,u_m) \leq G(x,u_-)$ and consequently $Lu_+ \leq Lu_{m+1} \leq Lu_-$. By the Weak Maximum Principle it follows that $u_- \leq u_{m+1} \leq u_+$. Similarly, $u_m \leq u_{m-1}$ implies $G(u_{m-1}) \leq G(u_m)$ which in turn gives us $Lu_m \leq Lu_{m+1}$ and $u_{m+1} \leq u_m$.

The limit u_*. The sequence of numbers $\{u_m(P)\}$ for $P \in M$ is a monotone and bounded sequence. As such, it converges to some value $u_*(P)$. We now need to show that as a function u_* is smooth and satisfies $Lu_* = G(x,u_*)$.

Use Ellitpic Regularity, Rellich Lemma and monotonicity to show that $\{u_m\}$ is Cauchy in $L^2(M)$. It follows from the point-wise boundedness of $\{u_m\}$ that the sequence $\{G(x,u_m)\}$ is bounded in $L^2(M)$. Thus, the sequence $\{Lu_m\}$ is bounded in $L^2(M)$. Using the elliptic estimate for invertible operators, Theorem 28, we see that $\{u_m\}$ is bounded in $H^2(M)$. Rellich Lemma (Theorem 27) now guarantees the existence of a subsequence $\{u_{m_i}\}$ which is convergent in $H^1(M)$ and, in particular, Cauchy in $L^2(M)$. Using monotonicity of $\{u_m\}$ we are able to show that, in fact, the sequence $\{u_m\}$ is Cauchy in $L^2(M)$: If $\|u_{m_i} - u_{m_j}\|_{L^2} < \varepsilon$ for all $i,j \geq i_*$ then for all $m, l > m_{i*}$ there is some $m_i \gg 1$ with $m_i > m, l > m_{i*}$ and

$$\|u_m - u_l\|_{L^2} \leq \|u_{m_i} - u_{m_i*}\|_{L^2} < \varepsilon.$$

Use the Mean Value Theorem, the Elliptic Regularity and the Sobolev Embedding to show that $\{u_m\}$ is Cauchy in both $C^0(M)$ and $H^2(M)$. By the Mean Value Theorem we have the point-wise estimate $|G(x,u_m) - G(x,u_l)| \leq C'|u_m - u_l|$, where $C' = \max_{M\times[a,b]} |\partial_u G|$. After integration we get that $\{G(x,u_m)\}$ is Cauchy in $L^2(M)$. This means that $\{Lu_m\}$ is Cauchy in $L^2(M)$. Another application of the elliptic estimate for invertible operators, Theorem 28, gives that $\{u_m\}$ is Cauchy in $H^2(M)$. In addition, the Sobolev embedding now implies that $\{u_m\}$ is also Cauchy in $C^0(M)$. <u>*WARNING:*</u> the latter relies on the small dimension of M (e.g. $n = 2$ or $n = 3$.) The proof in higher dimensions takes the same basic steps, but to circumvent this issue the proof is executed in slightly more general L^p Sobolev spaces.

The bootstrap argument: $\{u_m\}$ is Cauchy in each $C^{k-1}(M)$ and $H^k(M)$. An easy, albeit notationally difficult, argument based on the Mean Value Theorem shows that if $\{u_m\}$ Cauchy is in both $C^{k-1}(M)$ and $H^k(M)$ then $\{G(x,u_m)\}$ must be Cauchy in $H^k(M)$. For example, to show that $\{G(x,u_m)\}$ is Cauchy in $H^1(M)$ it remains to show that $\{\nabla G(x,u_m(x))\}$ is Cauchy in $L^2(M)$. Since in principle $\nabla G(x,u(x))$ is the sum of the covariant derivative of $G(x,u)$ with respect to the x

variable and the terms of the form $(\partial_u G)\nabla u$, the Mean Value Theorem implies an estimate of the form

$$\begin{aligned}&\|\nabla G(x, u_m(x)) - \nabla G(x, u_l(x))\|_{L^2}\\ &\le \text{Const}\|u_m - u_l\|_{L^2} + \text{Const}\|\nabla u_m - \nabla u_l\|_{L^2} + \text{Const}\|u_m - u_l\|_{C^0}\|\nabla u_l\|_{L^2},\end{aligned}$$

where all the constants involved are bounds on miscellaneous derivatives of G. Since $\|u_m - u_l\|$ can be made arbitrarily small in both $C^0(M)$ and $H^1(M)$ sense, we see that $\{G(x, u_m)\}$ is Cauchy in $H^1(M)$. Similar arguments apply for general k.

We now execute a bootstrap[8] argument for raising the regularity / differentiability level. Since $\{G(x, u_m)\}$ is Cauchy in $H^1(M)$, Elliptic Regularity gives us that $\{u_m\}$ is Cauchy in $H^3(M)$. The Sobolev Embedding (see the *WARNING* above) further proves that $\{u_m\}$ is Cauchy in $C^1(M)$. Using the fact that $\{u_m\}$ is Cauchy in both $C^1(M)$ and $H^2(M)$, we show that $\{G(x, u_m)\}$ is Cauchy in $H^2(M)$. Repeating the argument indefinitely / inductively shows that $\{u_m\}$ is Cauchy in each $C^{k-1}(M)$ and each $H^k(M)$.

The smooth solution u_*. As $\{u_m\}$ is Cauchy in each $C^k(M)$, it converges in each $C^k(M)$. Therefore, the point-wise limit u_* discussed above represents a smooth function. Furthermore, since all the derivatives of $\{u_m\}$ converge to the corresponding derivatives of u_* we have $Lu_m \to Lu_*$ and $G(x, u_m) \to G(x, u_*)$, point-wise. Taking the point-wise limit of $Lu_m = G(x, u_{m-1})$ shows that $Lu_* = G(x, u_*)$.

The solution of our problem. We now go back to

$$\Delta u = s + 2e^u, \quad \text{with} \quad \int_M s \, \text{dvol} < 0.$$

The idea is to find sub and super solutions u_- and u_+ for this equation. The easiest possible sub and supersolutions are constants. Does our equation permit constant sub and supersolutions? If s changes sign the answer is no: at points where $s > 0$ we cannot arrange $\Delta u \ge s + 2e^u$ for constant u. However, we can make a change of variables which preserves the general form of our PDE but whose "s-term" is always negative. The modified equation will then permit constant sub and super solutions.

To make the change of variables, consider a linear PDE of the form $\Delta z = f$. Fredholm theory, Theorem 29, shows that this equation is solvable if and only if $f \in \text{Ker}(\Delta)^\perp$ in the L^2-sense. By the Strong Maximum Principle, Theorem 11, we have that $\text{Ker}(\Delta) = \text{Span}\{1\}$ consists of constants. Therefore, equations of the form $\Delta z = f$ are solvable (non-uniquely) if and only if $\int_M f \, \text{dvol} = 0$. Since the latter is fulfilled for $f = s - \frac{1}{\text{Vol}(M)} \int_M s \, \text{dvol}$, there exist smooth functions z such that

$$\Delta z = s - \frac{1}{\text{Vol}(M)} \int_M s \, \text{dvol}.$$

Consider one such function z and introduce a new unknown $v := u - z$. It satisfies the PDE

$$\Delta v = \bar{s} + pe^v,$$

where $\bar{s} := \frac{1}{\text{Vol}(M)} \int_M s \, \text{dvol} < 0$ and $p = 2e^z > 0$.

---

[8]This is basically an inductive argument, but in PDE community it is traditional to call it a *bootstrap argument*.

Since p reaches a strictly positive minimum, we can find a sufficiently large positive constant v_+ such that $\bar{s}+pe^{v_+} > 0$. This constant serves as a supersolution of our equation:

$$\Delta v_+ \leq \bar{s} + pe^{v_+}.$$

Likewise, if v_- is sufficiently large negative constant so that $\bar{s} + pe^{v_-} < 0$, the constant function v_- serves a subsolution of our equation:

$$\Delta v_- \geq \bar{s} + pe^{v_-}.$$

Since $v_- \leq v_+$ the hypotheses of Theorem 30 are met; we conclude that both $\Delta v = \bar{s} + pe^v$ and $\Delta u = s + 2e^u$ can be solved. This completes the proof that all compact, 2-dimensional manifolds M without boundary and of negative Euler characteristic can be given a Riemannian metric with constant negative curvature.

Exercises for Lecture 12.

(1) Provide an intuitive/visual explanation of what conformal changes of metric $g \to \phi^2 g$ do. You may want to include some familiar examples such as $ds^2 = y^{-2}\left(dx^2 + dy^2\right)$, $ds^2 = \frac{4}{(1\pm(x^2+y^2))^2}\left(dx^2 + dy^2\right)$, etc. [Basic]

(2) Confirm the formula $\mathrm{Scal}_h = e^{-u}\left(\mathrm{Scal}_g - \Delta_g u\right)$ for the conformal change of metrics $h = e^u g$ on a 2-dimensional manifold M. [Intermediate]

(3) Let $(\bar{M}, \bar{g})$ be a compact Riemannian manifold *with* boundary ∂M and interior M. Let ρ be a function which vanishes on ∂M, is positive in the interior M, and has unit gradient along ∂M: $|\mathrm{grad}_{\bar{g}}\rho|_{\bar{g}} = 1$. Consider the open Riemannian manifold (M, g) where

$$g := \rho^{-2}\bar{g}.$$

Show that sectional curvatures of g approach -1 near ∂M, and sketch a rough sketch of (M, g). [Intermediate]

(4) Apply the method of sub and super solutions to an ODE boundary value problem; use the following guidelines. [Intermediate]

(a) Let $f(x)$ be a smooth function of $x \in [a, b]$. Show that the equation $u''(x) = f(x)$ has a unique smooth solution within the set of functions satisfying the boundary conditions $u(a) = u(b) = 0$.

(b) Show that if $u(x)$ and $v(x)$ are two smooth functions satisfying the boundary conditions $u(a) = u(b) = v(a) = v(b) = 0$ and $u'' \leq v''$ then $u \geq v$.

(c) Suppose that $F(u)$ is a non-increasing smooth function and that for some smooth functions $u_- \leq u_+$ satisfying the boundary conditions $u_-(a) = u_-(b) = u_+(a) = u_+(b) = 0$ one has

$$u_-'' \geq F(u_-) \quad \text{and} \quad u_+'' \leq F(u_+).$$

Show that the smooth sequence defined by

$$u_0 = u_-, \quad u_n'' = F(u_{n-1}), \quad u_n(a) = u_n(b) = 0$$

is monotone and bounded: $u_- \leq u_0 \leq u_1 \leq ... \leq u_+$.

(d) Show that the sequence of derivatives $\{u_n'\}$ is bounded in $C^0[a, b]$.

(e) Apply the Arzelá-Ascoli Theorem to get a convergent subsequence $\{u_{n_k}\}$ in $C^0[a, b]$.

(f) Use monotonicity to show that $\{u_n\}$ converges in $C^0[a,b]$.

(g) Use the Mean Value Theorems to show that the sequence $\{u_n\}$ converges to a function u in $C^2[a,b]$.

(h) Show that u is a smooth solution of the boundary value problem

$$u'' = F(u), \quad u(a) = u(b) = 0.$$

(5) Assume that $F(u)$ is an increasing smooth function of the scalar variable u. Show that the equation $\Delta u = F(u)$ on a compact manifold M without boundary has a solution if and only if there is a smooth function v such that $\int_M F(v)\,\mathrm{dvol} = 0$. [Intermediate]