A Visual Introduction to Curvature

Evan Curry Wilbur

May 7, 2024

Abstract

This paper seeks to cover the materials necessary prior to taking a course in Riemannian geometry as well as an exploratory look into the materials which would be covered in that class in a way that is both visually stimulating and enlightening. It is directed primarily to early undergraduate students of math who have familiarity with topics from linear algebra. We will assume the reader's understanding of inner product spaces, linear transformation, as well as a familiarity with the geometric understanding of all these ideas. However, a precocious high schooler could follow along after reading chapters 1,2,3, and 6 of Shelden Axler's Linear Algebra Done Right [1]. As well, practice in computing derivatives and integrals and the geometric meaning therein would be helpful from a first year course in calculus is necessary. However, a course in vector calculus is not required.

By no means should this be seen as a full and rigorous exploration into the vast topic of Riemannian and differential geometry; many tomes have be written doing such a task. Instead, it will leverage visual intuition of the space we occupy to make broad and powerful statements about surfaces more generally. Of course, visual intuition is limited by the dimension we occupy, so many generalizations will be stated, without proof, with references to where one could find proofs.

Introduction

The development of a general mathematical theory often begins with an attempt to model or understand something realized in the physical world. Calculus, for example, was first developed by Newton in order to model the motion of the planets, and now it has developed into a theory which can completely divorce itself from anything to do with physical reality. This is just as true for the theory of curvature and curved spaces.

The modern understanding of curvature is often traced back to Gauss and his work in developing the beginnings of differential geometry. He was tasked with producing a navigable map for ships to use at sea which preserved relative distances and angles. Gauss worked on the problem until he made a "remarkable discovery" which he called the Theorema Egregium, finding that no such map could possibly exist due to the curvature of the Earth and the lack of curvature of a flat map. This is seen as the start of a fruitful and beautiful subjects of differential and Riemmanian geometry.

We begin by asking an intuitively obvious, but mathematically difficult question: *what even is curvature?* We know from our daily experience what curvature is. The way a snake moves, the path a ball takes flying through the air, or the ripples of a pond a thrown pebble makes are all examples of this idea of curvature. Compare this to the jagged teeth of a saw blade, a zig-zag path, or even mathematical objects like triangles and squares. These seem decidedly less curvy than the previous examples. Intuitively, we understand the differences between these two. In

the former there seems to be a smooth transition of paths, while the latter seems to have more abrupt and sudden changes. How exactly can we say this in a more mathematically precise way?

We will start answering these questions by developing the theory of manifolds, abstractions of surfaces to higher dimensions. We will think of manifolds as being embedded in a canonically flat space we are familiar with, Euclidean space. We will impose certain structures on manifolds which we will see allow us to categorize what types of objects are curved, and what are flat or jagged. From there, additional work will go into understanding smoothness by constructing and studying the tangent space to a manifold. Finally, we get to the purpose of the paper and start to define and discuss curvature: first with one-dimensional manifolds, then to two, and finally to higher dimensions.

Our goal is not to provide a rigorous exploration into these topics. A list of references will be given at the end of this paper for that. Instead, this will provide a visual intuition for these topics to assist the reader in their further studies of this topic.

1 Making a Manifold

Any proper discussion of curvature must first answer the question *what is being curved?* After all, asking how curvy the integers are is meaningless, while asking the same for a sphere is seemingly not. The difference lies in the notion of space and geometry for the latter that the former lacks. A language capturing this intuition is necessary if any progress is to be made on the theory of curvature. Thankfully, mathematicians have developed such a language for discussing all manner of surfaces. A complete discussion of the theory and language would require an exploration into topology in a way more rigorous than necessary to develop an intuition for the subject at hand, so instead, many of the gaps in rigor should be filled in by the reader's intuition.

We will start the discussion by giving the definition for charts, which intuitively we will think of as a GPS system, or coordinates, allowing us to "walk along" a space in a way similar to how a spaceship would navigate through the solar system or how an ant would walk along a table. It will be clear by how we define charts that they are not suitable by themselves to encapsulate even all the surfaces we encounter in our everyday lives, much less those of higher dimensions.

To remedy this, we will "stitch together" different charts to create a manifold, the central object of discussion for the remainder of the paper. Once some examples of manifolds have been given, we will begin discussing ideas of smoothness and differentiability on a manifolds.

Charts

Before we introduce the definition of a chart, we start with an animation demonstrating their purpose. The animation below gives an intuition of how a chart can map out part of the surface $z = \sin(x + y)$.

Example 1.1.

At its most distilled, a chart is a map of some space around a point. Anyone who has picked up a map, used a GPS, or fantasized being a navigator in Star Trek is already familiar with the idea of charts. The chart is not the object of study in and of itself, much in the same way a map is not the area it seeks to represent. Rather, with a chart in hand we may point to a position on it, and we shall get a corresponding point on the surface we are studying.
From the animation, we can start to make some observations and develop an intuition for what charts do. First, the curved space appears flat in the chart. The chart exists in $\mathbb{R}^2$, and so is flat, yet despite this, it represents a rather curvaceous function! For those who have held a map or used GPS, this should seem like a natural requirement to have for charts. After all, if a map were required to conform to the natural topography of the region it's meant to represent, it would be rather cumbersome to use. Another thing one might notice is how the path traced out on the chart corresponds rather well to the other path on the surface. Again, this should feel like a natural property that charts should have. If a path followed on a map doesn't match a similar path traced on a surface, then its a bad map. The same holds true for charts. With these observations in mind, we give the formal definition for a chart.

Definition 1.1. A chart for a topological space M is a pair (U, φ). The set $U \subseteq \mathbb{R}^n$ is an open set of $\mathbb{R}^n$. The function $\varphi : U \to V \subseteq M$ is a homeomorphism. [5]

For those without an introduction into topology, this definition is likely meaningless. There are three main pieces of vocabulary to unpack: *topological spaces*, *homeomorphism*, and *open sets*.

We start with the definition of a topological space

Definition 1.2. A topological space is a set M along with a collection τ of open sets of M such that

1. Both M and the empty set are in τ. Succinctly, $M \in \tau, \varnothing \in \tau$.

2. For any index set $U_i \in \tau$, $\cup U_i \in \tau$. That is, an arbitrary union of open sets is open.

3. For any finite collection $U_k \in \tau$, $\cap_{k=0}^n U_k \in \tau$. That is, finite intersections of open sets are open. [5]

This definition, at first glance, might seem horribly abstract, and that's because it is. This definition is meant to encapsulate any conceivable notion of a "space", even those which are completely inconceivable. Not to mention that it seems that we need to understand what an open set is before we can understand what a topological space is. Let us ignore what the definition of an open set it. For now, we will consider it simply a property that certain subsets of M have. If we are comfortable with this, then the above definition is defining a very simple idea

1. A topological space consists of two sets: M and a collection of subsets of M called the *topology* on M

2. We call the elements of τ *open sets*

3. For any collection of open sets of τ, we know their union must also be in τ

4. For any finite collection of open sets of τ, their intersection must also be in τ.

5. Both the entire set and the empty set are open

That's it. It's not obvious how this definition relates to our metaphysical understanding of a "space". It's far too abstract. That's largely because the notion of "space" comes not from the definition of a topological space, but rather how you define the open sets. It's the open sets which determine how a set becomes a space - the definition of a topological space seeks only to limit the arbitrary labeling of sets as open in such a way so that our normal intuitions of "space" are held.

Thus, to understand what a topological space is, it is necessary to develop an intuition of open sets. At its most fundamental, open sets determine what it means for a function to be *continuous*.

Definition 1.3. Let M and N be topological spaces, a function $f : M \to N$ is continuous if for every open set $V \subseteq N$, the inverse image $f^{-1}(V) \subseteq M$ is an open set of M. [5]

Most students first learn of continuity in high school, where they are taught that a function is continuous if the graph can be drawn without removing a pencil from the paper. That is, there are no gaps or leaps of the graph of a function. In calculus, continuity of a function is determined by whether the limit of a function at a point equals the function evaluated at that point, for every point. Both these definitions are saying, in their own respective contexts, that continuous functions are functions which send nearby points to nearby points. Definition 1.3 is no different. It is saying precisely that, only in the language of topology. Open sets, therefore, determine what it means for points to be "nearby" each other.

Let's take a step away from the abstraction and focus on a space we are already familiar with, $\mathbb{R}^n$, Euclidean space. There are, in fact, many ways of defining open sets on $\mathbb{R}^n$ so that it forms a topology. There are even many different ways of defining open sets on $\mathbb{R}^n$ so that it produces *the same topology* on $\mathbb{R}^n$. We will show a common way of defining the standard topology on $\mathbb{R}^n$ right now. This topology comports to our usual notion of space, closeness, and distance that we are used to seeing in our everyday lives. It will also be the assumed topology for the rest of this paper whenever we discuss $\mathbb{R}^n$ and its open sets.

Recall that we think of open sets as *defining how* points are close together. One way of expressing this mathematically (but certainly not the only way!) is to say a subset $U \subseteq \mathbb{R}^n$ has points that are close together (i.e. U is

open) if, around every point $p \in U$ there is another set V of points which are close together (i.e. V is open), with $p \in V$ and such that $V \subset U$. Let's reify this idea.

Definition 1.4. An open ball of radius $\varepsilon \in \mathbb{R}$ centered at $p \in \mathbb{R}^n$ is the set $B_\varepsilon(p) = \{x : x \in \mathbb{R}^n, |x - p| < \varepsilon\}$. When p is omitted from the expression, it is assumed to be centered at the origin.

Definition 1.5. A set $U \subseteq \mathbb{R}^n$ is open if for every $p \in U$ there exists an $\varepsilon \in \mathbb{R}$ such that $B_\varepsilon(p) \subseteq U$.

Example 1.2. The dashed circle is an open ball $B_\varepsilon \subset \mathbb{R}^2$. Notice how no matter what point in the open ball is selected, it is always possible to surround it with another open ball. Thus, open balls are open sets.

Of course, not every set has this property

Example 1.3. The ball of radius 1 which contains its boundary, $\bar{B} = \{x : x \in \mathbb{R}^2, |x| \leq 1\}$, is not open.

Unlike the open ball, $\bar{B}$ contains it's boundary. We see from the animation that it is impossible to surround a point on the boundary with an open ball in such a way so that it lies entirely within $\bar{B}$. Thus, $\bar{B}$ cannot be open.

By considering the collection of all open sets U as defined in Definition 1.5, we have built the standard topology on $\mathbb{R}^n$. Check to see that this construction does indeed form a valid topology as defined in Definition 1.2. It is clear that around any point in $\mathbb{R}^n$ and $\varnothing$ there is an open ball surrounding every point contained in $\mathbb{R}^n$ and $\varnothing$ (the latter being vacuously true). Showing that this forms a topology, however, will be left as an exercise.

Finally, we get to the last of the vocabulary to unpack, *homeomorphism*. When we think of two objects being homeomorphic, then we may continuously transform one into the other without ripping, tearing, or gluing our shape, and then back again.

Due to the animation on the left, mathematicians often joke that the topologist cannot distinguish between a coffee cup and a donut since the two are homeomorphic. While said in jest, it does underpin a critical point in topology. When two objects are homeomorphic to each other *they are functionally the same object.*

Another important idea illustrated in the animations is that, while we cannot rip, tear, or otherwise damage the object during a homeomorphism, we are freely allowed to stretch, warp, and even self-intersect as much as we would like. With that, here is the definition

Definition 1.6. A function $f : M \to N$ between topological spaces M and N is a homeomorphism if
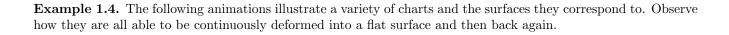
1. f is bijective.

2. f is continuous.

3. f^{-1} is continuous. [5]

These three conditions encapsulate exactly the conditions mentioned previously, just in the language of topology. The requirement that f and f^{-1} be continuous forbids us from ripping, tearing, and otherwise damaging our space, with f^{-1} also allowing us to "undo" the transformation. The requirement that f is bijective means that we cannot "squish" open sets of our surface down to a point. It is possible to continuously transform a circle down to a point, there is no way to do this without collapsing the open sets of the circle. In the animation below, the parentheses indicate one of the open sets of the circle. It's clear that as the circle transforms continuously to a point, the open set collapses to a point as well. This violates (1) of the definition of homeomorphism, thus it cannot be a homeomorphism, despite being a continuous transformation.

With the language out of the way, we now are in a position to understand charts in a way much more clear than the rigorous definition.

Definition 1.7. A chart is an open ball $U \subseteq \mathbb{R}^n$ representing some surface M in such a way that U can be continuously transformed into M and back again.

In essence, a chart is a flat map of a (potentially) curvy space.

Example 1.4. The following animations illustrate a variety of charts and the surfaces they correspond to. Observe how they are all able to be continuously deformed into a flat surface and then back again.

Manifolds

In the previous section, we saw that we were able to construct many surfaces using just a single chart. It might be tempting to think that a theory of charts is all we need to study all conceivable surfaces. This, however, is not the case. The issue lies in the requirement that our charts be homeomorphic to the space they represent. This is a strong requirement that prevents us from studying many surfaces which may not be homeomorphic to flat space with charts alone.

Example 1.5. The sphere is not homeomorphic to flat Euclidean space, thus it cannot be represented by a single chart.

Illustrated in the animation is the unfortunate fact that we cannot transform the sphere into a flat surface without overlapping two different regions on the sphere onto the same region of a chart, in violation of rule (1) of the definition of homeomorphism. If the sphere were missing one of the poles, then it would in fact be possible to cover it in one chart - an example visualized in Example 1.3. However, this requires that we puncture the sphere, a violation of rule (2). This is by no means meant to be a proof of the fact that we cannot cover the sphere in a single chart, rather it is only meant to illustrate what goes wrong when we try to. A curious reader would find that with only a brief exploration into topology, and in particular Brouwer's fixed point theorem, the proof that it is impossible to cover the sphere in a single chart is rather enlightening.

Since the theory of charts is not enough to study surfaces as typical and canonical as the sphere, we may wonder if is even worth talking about them at all. Is all hopelessly lost? Did the author really just spend the past 8 pages talking about something that has no bearing on the study of surfaces? Of course not. As is so common in mathematics, when our theory runs into a wall, we extend the theory so we may be able to walk up walls.

The key idea is not to limit ourselves by using only a *single* chart. Instead, we will use multiple charts to cover our surfaces. Each chart will represent a local patch of the surface. This local patch, by virtue of being covered by a chart, will look like an open set of Euclidean space. Because of this, we often say that our surfaces are *locally Euclidean*. A collection of charts which cover a surface, when taken in aggregate, will be powerful enough to describe a vast amount of spaces. This collection is called an *atlas*, and the space which they cover is called a *manifold*. With that, we present two definitions.

Definition 1.8. An atlas is a collection of charts of the same dimension.

Definition 1.9. An n-dimensional manifold M is a topological space such that for each point $p \in M$, there is a chart $U \subseteq \mathbb{R}^n$ and a homeomorphism $\varphi : U \to M$ such that $p \in \varphi(U)$.

With this definition now firmly established, we will cease to use the terms "surfaces" and instead use the more general term, manifold. The key idea here is that by considering collections of charts, we can study a vast collection of topological spaces.

Example 1.6. By covering the sphere in two charts, we are able to turn it into a manifold.

The animation in Example 1.4 shows that the charts of a manifold may overlap. On the region where charts overlap, we will have points which can be described by more than one chart. These regions of overlap will have a special consideration when we discuss smooth manifolds in the next section. For now, we will only consider how we can move between two charts which describe the same region of our manifold.

Definition 1.10. Given two charts $(U, \varphi), (V, \psi)$, of a manifold M, if $\varphi(U) \cap \psi(V) \neq \varnothing$, then the transition map is the homeomorphism $\varphi^{-1} \circ \psi : \mathbb{R}^n \to \mathbb{R}^n$. [3]

Example 1.7. Shown is the manifold M with two charts. The non-overlapping regions of V and U are colored red and blue respectively. Their shared region is colored green. We see how the transition map $\varphi^{-1} \circ \psi$ maps *between the charts*.

Smooth Manifolds

So far, our discussion has been centered around building up an intuition and language for manifolds. We have actually answered the question posed at the very beginning of this paper, *what is being curve*. The answer: manifolds. As is so common in mathematics, when one question is answered, others naturally arise. While we now know *what* is being curved, we do not yet know *how* it is being curved. Unfortunately, we are not in a position to answer this question just yet. The issue is that our definition for a manifold is not powerful enough to discuss topics as complex as curvature.

If you recall, the way we built up to the definition of manifolds was by stitching together regions of flat space. We then were able to discuss and study manifold by referencing their charts. However, the charts are just open sets of Euclidean space. Not only are they flat, they are essentially as flat as mathematically possible. So how can we possibly discuss curvature with only our collection of charts? The answer is that we can't. The definition is simply not capable of expressing these ideas.

Not all is lost, however. Much like we were able to extend our definition of charts to manifolds and expand the types of topological spaces we could talk about, we can build upon our current definition of manifolds in a way powerful enough to discuss curvature.

Before we do that, however, let us see what goes wrong when using our current definition of manifolds. To do so, let us first introduce the concept of a path. We understand intuitively that a path is just some trajectory through space that doesn't have gaps or holes in it. The following definition expresses mathematically this idea

Definition 1.11. Let M be a manifold. A path is a continuous function $[0, 1] \to M$.

This definition of path encompasses many different notions of what a path can be. We could have a path from two points, one that goes on forever off to infinity, or one between two points but never quiet reaching them. In general, we can think of a path as some curve through space.

Recall that our overall goal is to develop a theory so we may discuss curvature. Curvature, we understand, is a bending or change of direction of an object (or in our case, manifolds). However, it's not just any change of direction or bending. For example, a zig-zag path changes direction frequently, however, one wouldn't consider it "curving" through space. For something to be curved, it requires a smooth change of direction.

Example 1.8. Both paths follow a very similar trajectory through space. However, the path on the left consists of sharp and jagged turns while the one on the right takes smoother turns. The one on the left does not appear to curve while the one on the right does.

Let's imagine a person walking along the zig-zag curve on the left. At first, they will walk in a straight line until they reach the first corner. When they arrive at the corner, they must pause, change their direction, and then proceed walking. Compare this to that same person walking along the path on the right. Unlike before, they do not need to stop in order to change direction. They instead are able to walk along the path while slowly adjusting their direction as they walk.

Example 1.9. One key observation that we can make is how the direction the dot on the zig-zag path is moving changes abruptly when it reaches a corner. Compare this to the dot moving along the smooth curve. The direction it is moving does not change abruptly. If we wish to talk about curvature, we require that these tangent directions do not change abruptly.

Let's explore this idea a bit more as it is essential to understanding smooth manifolds and curvature more broadly. In a first year calculus course, we learn of the derivative and how it is used to calculate the slope of the tangent line to functions. The tangent line indicates the instantaneous direction of the function at any given point. So, the directions indicated by the arrows in Example 1.8 are nothing more than the derivative of the paths, indicated by the direction of travel. In fact, the derivatives of the equations for the paths were exactly what were used to compute the direction of the arrows in the animations!

Definition 1.12. Let M be a manifold. The derivative of a path $f : U \to M$ at time $t_0 \in U$ is a vector. The direction of the vector indicates the direction of travel along the path as a particle moves in a direction in U. The magnitude of the vector is called the speed of the particle at time t_0.

Definition 1.13. A smooth path is a path whose derivative is continuous. That is, the direction and speed of the derivative are continuous.

The problem with the zig-zag path can now be stated more mathematically. The corners where we make abrupt changes in direction are discontinuities of the derivative of the path. Thus, we shall require that all paths have continuous derivatives. In other words, the vector along a path has no sudden or abrupt changes in speed or direction.

We are now in a position to define what a smooth manifold is by utilizing the notion of a smooth path. We understand that a manifold can be realized as a collection of charts. If we have a smooth path on a chart, then for a manifold to be smooth it will be necessary that the corresponding path on the manifold be smooth as well.

Example 1.10. Despite the fact that the path on the chart is smooth, because the manifold has sharp corner, the corresponding path is not smooth. Thus, the manifold is not smooth.

Example 1.11. The sphere is a smooth manifold. Every smooth path drawn on a chart of the sphere is itself smooth.

With respect to our discussion of smooth paths, it might be tempting to define a smooth manifold as one where every smooth path on a chart produces a smooth path on the manifold. Intuitively, this idea is correct. However, it turns out that when defining these concepts more rigorously, such a definition requires us to be able to compute the derivatives of these paths on the manifold. This isn't a problem with the manifolds we've seen so far because we've been visualizing all our manifolds as being embedded in a surrounding space. Mathematicians, however, prefer to divorce the idea of a manifold and the space which it occupies. They wish to treat the manifold *as a space unto itself*. If we were to accept this definition for smooth manifolds, it would require that we reinvent calculus on every manifold to define what it even means to take a derivative on a manifold. To avoid undertaking such an insurmountable task just to define what it means for a surface to be smooth, mathematicians instead consider an alternative approach involving the differentiability of transition maps between charts of a manifold. Since we understand what it means for a function to be differentiable in $\mathbb{R}^n$, and transition maps are simply functions $\mathbb{R}^n \to \mathbb{R}^n$, we can define a smooth manifold as one where the transition maps are differentiable. Realize, however, that defining smoothness via the transition maps or by requiring that smooth paths on charts be smooth on the manifold gives exactly the same realization of smoothness.

Definition 1.14. An n-dimensional manifold M is smooth if, whenever two charts (U, φ) and (V, ψ) of M overlap, the transition map $\varphi^{-1} \circ \psi : \mathbb{R}^n \to \mathbb{R}^n$ is differentiable. [3]

2 The Measure of a Manifold

So far, it seems that every time we introduce a concept to help us understand curvature, the rug gets pulled and it turns out that what we've been studying wasn't *quite* enough to speak on the nature of curvature. First, the notions of charts were discussed and studied. However, it was pointed out that charts alone weren't enough to discuss many surfaces that we would otherwise wish to. To remedy this, we bundled and stitched charts together to create manifolds. We then saw how the definition of a manifold was actually too broad to have a coherent discussion on curvature. It allowed for manifolds with sharp curves or edges, which we pointed out were not the curvature we are interested in studying. So we refined the definition of the manifold to one more suitable to a discussion of curvature: smooth manifolds. Exasperation would not be an unreasonable emotion for a reader at this point. Thankfully, it turns out that smooth manifolds are sufficient for our purpose of understanding curvature.

This section will be dedicated to developing tools to let us better understand smooth manifolds. We will start with defining the tangent space to a manifold. This idea is central to a proper understanding of smooth manifolds as it allows us to extend ideas from single variable calculus to any smooth manifolds. This will let us utilize the tools and concepts that calculus provides us to better understand curvature.

We will explore more deeply the connection between the tangent space of a manifold and the underlying charts which define it through the concept of the pushforward, or differential, map. With this, we can compare directions on the charts for a manifold to the manifold itself by "pushing them forward".

Once we have defined the tangent space to a manifold and see how it compares to the tangent space of the underlying charts, we will realize that it is nothing more than a vector space attached to each point on a smooth manifold. Because of this, we can utilize concepts from linear algebra, like inner product spaces, to better understand the structure of manifolds.

The Tangent Space

In introductory calculus, we learn about derivatives, tangent lines, and their applications and interpretations. The derivative, we saw, gave us a function's instantaneous rate of change. Tangent lines were special lines which lie tangent to a graph at a given moment and serve as the best linear approximation for a function at that point. When we took the derivative *of the derivative* of a function, we were able to see how quickly the function diverges from the tangent line. This idea will be essential later when we finally and properly study curvature. Before we do that, however, we will need to extend these ideas from introductory calculus more generally to manifolds. The tangent space of a manifold is the first step in doing this.

Much like how the tangent line to a graph gave us a linear approximation of the graph at a given point, so too does the tangent space of a manifold give us a linear approximation to the manifold at a given point. In fact, we will see that the tangent space is simply a generalization of concept of a tangent line we've already seen.

We can think of the graph of a real-valued function as a one dimensional manifold. If we were to attach at each point of the graph the tangent line at that point, we would begin to have something that looks like the tangent space of the graph.

Example 2.1. Attaching a tangent line to every point of a graph resembles the tangent space to the graph

Similarly to the one dimensional example, we can extend this idea to that of two dimensions. However, instead of tangent lines we instead consider tangent planes.

Example 2.2.

Here we see the tangent plane to a sphere taken at various points. The plane always lies tangent to the sphere, and so serves as a linear approximation to the sphere at any given point. This idea generalizes to any smooth manifold. In fact, the primary motivation for requiring our manifolds to be smooth was so we can have a well defined tangent space at every point of the manifold. As the manifold increases in dimension, so too does the dimension of the tangent space at a point.

It might be tempting to define the tangent space as simply the manifold along with all the tangent spaces attached at each point. This isn't too far off from the actual definition. However, it turns out to be more effective to define it as a linear subspace at each point. In other words, rather than attaching planes and lines to the manifold at every point, we instead attach a vector space at every point.

It should be clear to anyone who has taken linear algebra why this would be more useful. Lines and planes are simple geometric objects. However, each of these geometric objects have implicitly a vector space which spans them, and using vector spaces allows us to tap into the powerful tools in linear algebra.

This begs the question: how exactly do we define a vector space at every point of the manifold. Intuitively, we know what needs to be done. However, using only the charts we have which define our manifold, it's not immediately obvious how we should go about constructing this. There are actually many ways to do this rigorously. We will show one here which uses the tangent space of paths to build a tangent space to a point, but we will avoid the rigor in lieu of building the intuition first. For the reader who wants a more detailed, thorough, and rigorous construction of the tangent space, John Lee's Introduction to Smooth Manifolds is a fantastic resource.

We've seen some hints how we are going to go about constructing the tangent space from examples 1.8 and 1.9 from the previous section. The key idea is to leverage what we know about the tangent spaces of paths on manifolds to construct the tangent space at a point.

We already were introduced to paths, derivatives of paths, and smooth paths in definitions 1.11, 1.12, and 1.13. These work as definitions, but they don't exactly tell us how to go about *computing* these derivatives. Here we will realize how.

Example 2.3. The graph of the function $f : [0,1] \to \mathbb{R}^3$

$$f = \begin{pmatrix} \sin(2t\pi) \\ \cos(2t\pi) \\ \cos(4t\pi) \end{pmatrix}$$

We see here an example of a parametrization of a path. You can see how we can plug in a variable from $[0,1]$ and we will output a vector which points to the position on the curve. The parametrization of the path reifies the abstract notion of a path into one where we can actually work with and compute. With the parametrization of a path in hand, we can compute the derivative simply by computing the derivative of each component of the path.

Example 2.4. The graph of the function $f : [0, 1] \rightarrow \mathbb{R}^3$ from Example 2.3 and its derivative

$$f = \begin{pmatrix} \sin(2t\pi) \\ \cos(2t\pi) \\ \cos(4t\pi) \end{pmatrix} \qquad\qquad f' = \begin{pmatrix} 2\pi\cos(2t\pi) \\ -2\pi\sin(2t\pi) \\ -4\pi\sin(4t\pi) \end{pmatrix}$$

The parametrization as well as its derivative allows us to express the idea of a tangent space using just the equations of the path. Let α be a path and let α' be the derivative of the path. At some point $t_0 \in [0, 1]$, $\alpha(t)$ describes a position along the path while $\alpha'(t)$ gives the tangent direction of the path. With this, we are able to construct a one dimensional vector space T at $\alpha(t_0)$ by

$$T = \alpha(t_0) + \lambda\alpha'(t_0) \qquad \lambda \in \mathbb{R}.$$

Example 2.5. The graph of the function f from Example 2.3 along with the tangent space at $t = \frac{1}{8}$ as we vary the scalar $\lambda \in \mathbb{R}$

$$f\left(\frac{1}{8}\right) = \begin{pmatrix} \sin\left(\frac{\pi}{4}\right) \\ \cos\left(\frac{\pi}{4}\right) \\ \cos\frac{\pi}{2} \end{pmatrix} \qquad\qquad f' = \begin{pmatrix} 2\pi\cos(2t\pi) \\ -2\pi\sin(t2\pi) \\ -4\pi\sin(2t\pi) \end{pmatrix}$$

You can see illustrated in the example how we can attach a vector space at a point on the path. Doing this for every point along the path gives us the tangent space of the path.

By taking the derivative of a smooth path, we've seen how we can construct a vector space for every point of the path. When taken in aggregate, the vector spaces become of the tangent space to the path, when we view the path as a manifold. We have the power now to construct the tangent space for every one dimensional smooth manifolds. This is all well and good. However, many interesting manifolds exist in higher dimensions and we've only seen how to construct a tangent space to one dimensional manifolds.

One of mathematics' greatest strengths is its ability to generalize concepts to those not imagined previously. We will tap into that strength right now and extend the concept of a tangent space to manifolds of higher dimensions. To do this, we will build up the tangent space to a point on the manifold by considering *multiple paths* through that point. When taken in aggregate, the paths will create a vector space, or tangent space, to the manifold at a point.

Suppose we have an n-dimensional manifold M. We wish to build the tangent space on M at a point $p \in M$. As a matter of notation, we will refer to this tangent space as T_pM. Since p is a point on the manifold, there exists a chart (U, φ) and a point $u \in U$ such that $\varphi(u) = p$. Consider now a smooth path $f : [0, 1] \to U$ such that $f(t_0) = u$ for some $t_0 \in (0, 1)$. We've seen already that the derivative of f gives us the tangent direction of the path at a point. So then $f'(t_0)$ is the direction of f through the point u. By using the map φ we can carry this path over to M. That is, the composition $\varphi \circ f \subseteq M$ is just a path in M. But we know exactly how to compute derivatives of paths. So by computing the derivative of $\varphi \circ f$ at t_0 gives us a tangent direction to the point p on the manifold! We are able to do this for any smooth path passing through the point $u \in U$. By considering all paths through the point u and considering the tangent vectors we find after the composition $\varphi \circ f$, we build T_pM.

Example 2.6. Taking multiple paths on a chart through a point allows us to construct the tangent plane to the torus

Thus, the tangent space to a point can be thought of as the set of all paths going through a point on the chart. Doing this for every point on the manifold gives us the total tangent space to the manifold, regardless of the dimension.

The Differential Map

We saw in the previous part that we were able to construct the tangent space to a point on a manifold by considering paths on a chart. We took the paths going through a point on a chart and considered the corresponding path on the manifold. We then *defined* the tangent space to a manifold at a point as the collection of these paths.

It should come as no surprise that the collection of these paths actually forms a vector space. The proof of this will not be given in this paper. The differential map, therefore, is a linear map from the tangent space of a point on a chart, to the corresponding tangent space on the manifold. With this map, we no longer have to view the tangent space of the manifold as a collection of paths, and instead as simple a traditional vector space.

Recall from the construction of the tangent space that we were able to map tangent vectors from a path on a chart to tangent vectors to a path on a manifold through the chart map. This map from the vector space at a point in a chart to the vector space at a point in the manifold is linear. Thus, we have a linear map from the tangent space of a chart to the tangent space of a manifold. We call this map the differential map.

Definition 2.1. Let M be a manifold and $p \in M$. Let (U, φ) be a chart with $u \in U$ and $\varphi(u) = p$. The linear map just described is called the differential map. It is represented as $d\varphi \mid_u : T_u U \to T_p M$. It maps vectors in the tangent space in U at u to the tangent space in M at p.

Riemannian Metric

On a manifold M, we've seen how we can construct a vector space at each point on M by considering the set of paths. The differential map allows us to view these sets of paths as a vector space. This vector space we interpret as the set of all directions which lie tangent to M. We are now squarely within the realm of linear algebra, and so may utilize concepts from it to expand our repertoire of tools.

In any beginner linear algebra course, the dot product and the theory of inner product spaces is taught. We saw during these courses how the inner product could give us a way of measuring distances and angles in a vector space. That is, given vectors v and w in a vector space, we define

$$|v| = \sqrt{\langle v, v \rangle} \qquad\qquad \cos(\theta) = \frac{\langle v, w \rangle}{|v||w|}.$$

That is, the length of a vector is the square root of it's inner product with itself. The cosine of the angle between to vectors is the inner product scaled by the reciprocal of the product of their respective magnitudes.

The tangent space to a manifold is a vector space. So we may put an inner product on it and have a way of determining lengths and distances on the manifold, at least locally. Going back to our analogy of charts as maps, putting an inner product space on the tangent space to a manifold is akin to putting a compass direction and a scale to compare lengths and angles on the charts to a manifold.

However, unlike a traditional map we are used to seeing in the real world, we are under no obligation to use the exact same scale and compass for every point in the chart. In fact, we may define a specific inner product for every point on M. So long as this inner product varies continuously and smoothly (in the sense that points close together give outputs of their inner product close together as well). When done this way, we call the inner product a Riemannian metric.

Definition 2.2. A Riemannian metric at a point $p \in M$ is a function $g_p : T_p M \times T_p M \to \mathbb{R}$. If it is defined smoothly and continuously for every point of M, then we call g the Riemannian metric on M. [4]

3 Curvature

Our theory is developed enough to finally begin exploring the main topic of this paper: curvature. In much the same way we built up a theory of manifolds starting from the basic charts, building all the way up to smooth manifolds, we develop our theory of curvature by studying how simple manifolds curve. We will then generalize these ideas to more complicated manifolds to broaden our theory.

Curvature of Paths

Imagine you are driving in a car along a straight highway at a constant speed. Our experience with the physical world tells us that you will not feel any force acting on you while driving. Eventually, you approach a curve. Without slowing down you take the turn. Suddenly, you feel a force towards the center of your turn. Your speed didn't change, but nonetheless a force was felt. After the turn, you continue driving straight ahead and resume not feeling any force on you. Thinking to yourself, you realize that, as long as you maintain your speed, any force felt by you or the passengers would indicate a turn, or change of direction. Interestingly, even someone completely blindfolded, or not looking outside the car at all, could determine when they are turning and when they are not, simply by feeling the forces acting on them while inside the car. This is the key idea. Provided that the speed of the car is maintained, a force is felt only when the car is turning. Hence, we can detect turns on our path by measuring the forces we feel along the path.

Those familiar with Newton's laws know that the force of an object is given by the object's mass times it's acceleration. The acceleration of an object is the second derivative of the object's position. Thus, simply by taking the second derivative along the path, we can compute the acceleration for any point along a path. With some additional caveats which will be mentioned shortly, whenever the second derivative along a path is nonzero, we will say that the path is curving at that point. The direction of curvature will be the direction of the second derivative.

It is critical that the speed traveled along the path remains constant. If you are traveling along a straight road with no turns, it is possible to still feel a force while you are driving even if there are no turns by hitting the breaks or gas. Thus, in order for this test to accurately detect curves along a path, we *require* that we are traveling at a constant speed along the path.

Many paths defined parametrically (like the examples in section 2) are not traveling at a constant speed. In other words, the magnitude of the tangent vector varies as we travel along the curve.

Example 3.1. Consider the path $f : [0,1] \to \mathbb{R}$

$$f = \begin{pmatrix} \sin(2t\pi) \\ \cos(2t\pi) \\ \cos(4t\pi) \end{pmatrix}$$

from the examples in section 2. The derivative of this path is

$$f' = \begin{pmatrix} 2\pi\cos(2t\pi) \\ -2\pi\sin(2t\pi) \\ -4\pi\sin(4t\pi) \end{pmatrix}.$$

We can compute the magnitude of the tangent vector at any point using the normal Euclidean definition

$$|f'| = \sqrt{\langle f', f' \rangle}$$
$$= \sqrt{4\pi^2 \cos^2(2t\pi) + 4\pi^2 \sin^2(2t\pi) + 16\pi^2 \sin^2(4t\pi)}$$
$$= \sqrt{4\pi^2 + 16\pi^2 \sin^2(4t\pi)}$$
$$= 2\pi\sqrt{1 + 4\sin^2(4t\pi)}$$

which is obviously not constant. Thus, it is not sufficient to show that the second derivative of this path is nonzero in order to demonstrate that the path curves.

Those readers who are savvy with their linear algebra might be tempted to simply divide f' by its magnitude. The reasoning being that this will always scale the vector f' so that its magnitude is always equal to 1. From there, it's a simple matter of computing the derivative of this scaled tangent vector, $\frac{f'}{|f'|}$. This is completely correct and valid reasoning provided that $|f'| \neq 0$. In order to show this rigorously, however, it is required that some existence and uniqueness theorems of differential equations be used, concepts outside the scope of this paper. Suffice to say that whenever we have a path, we can always "scale" the speed we travel along the path so that we are always traveling at a constant speed.

Definition 3.1. A path f has constant speed whenever $|f'| = c$ for some constant c. If $c = 1$ then f is said to have unit speed.

In light of the previous discussion, all paths whose derivatives do not vanish may be converted to an identical path that is of constant speed.

We are now in a position to give a definition of curvature.

Definition 3.2. A constant speed path f is said to *curve* at a point $t_0 \in (0, 1)$ if $f''(t_0) \neq 0$. The direction of f'' is called the *normal* direction of the curve.

Example 3.2. The purple arrow is the acceleration of the unit speed path. Observe how the arrow scales in size when it approaches the parts of the graph which curve.

Example 3.3. A normal vector to a path can be determined in higher dimensions as well

These two examples illustrate some facts about the curvature of paths. We can see how the tighter the curve, the greater the normal vector. So the magnitude of the normal vector gives us information on "how curvy" a curve is.

Gaussian Curvature

We now have a way to define curvature, at least for paths and one dimensional manifolds. We begin now with the generalization of this idea with the concept of Gaussian curvature for two dimensional manifolds, or surfaces.

First, we shall define the normal vector to a surface. We think of the normal vector to a surface as a unit vector which is perpendicular to the surface. Expressed equivalently, the normal vector to a surface is orthogonal to the tangent space of the surface. For two dimensional surfaces embedded in three dimensions, the normal vector can be computed directly by taking the cross product of the tangent vectors of two paths passing through a point and then scaling the result so it has unit length. Astute readers may notice that this does not provide a well-defined vector, as it is possible that the vector computed may vary up to sign. For our purposes, however, this will not be an issue. So long as we are able to find a normal vector to a surface at a point, the following discussion does not change depending on which we choose. However, there are many results in differential geometry which require that we pick an orientation.

We can imagine ourselves standing on a point on a surface. The tangent space gives us all the possible directions we may head from this point on the manifolds. For each direction, we consider a constant unit path on the surface through the point we are standing on whose derivative at the point we are on on the manifold is the direction in the tangent space. Each path has a well-defined normal curvature which we can compute just as we did in the previous subsection. We now pick a normal vector to the surface. For every path, we compute the inner product of the normal vector of the path with the normal vector of the surface. Results from topology guarantee that if we do this for every path, there will be at least two paths, f_1 and f_2 for which this inner product achieves a maximum and minimum respectively. The directions f_1' and f_2' are called the principle directions of curvature. The inner product with the surface normal is called the principle curvatures. If the surface normal is s, the normal vector to f_1 is n_1 and the normal vector to f_2 then $k_1 = \langle s, n_1 \rangle$ and $k_2 = \langle s, n_2 \rangle$ are called the principle curvatures. The Gaussian curvature at a point on the surface is then defined as $k_1 k_2$.

This definition may seem a bit mangled. So we will explain what is going on a bit more geometrically. For each direction we head on the surface, there is a unit path which has a normal vector indicating the curvature of the path. The surface normal s serves as a yardstick of sorts. It is there to keep track of the direction of curvature relative to the surface. When the normal direction of the path points away from the surface normal, the value of k will be negative as seen from taking the inner product. When the two coincide in direction, the inner product will be positive. If it happens that all the paths have normal curvature pointing either away from, or towards s, then either $k_1 < 0$ and $k_2 < 0$ or $k_1 > 0$ and $k_2 > 0$ so in either case, $k_1 k_2 > 0$ as the Gaussian curvature at the point is positive. If, however, the surface curves towards s for some paths and away from s in another, then $k_1 > 0$ and $k_2 < 0$ or $k_1 < 0$ and $k_2 > 0$. In either case, $k_1 k_2 < 0$ and the point is said to have negative Gaussian curvature. Finally, if one of the paths has no normal curvature, then either $k_1 = 0$ or $k_2 = 0$ and the point has zero Gaussian curvature.

Example 3.4. In purple is the surface normal vector. The blue and red vectors correspond to the normal vectors to the blue and red paths respectively. Observe how both the normals to the paths point in the same direction away from the surface normal to the surface. This indicates positive Gaussian curvature

Example 3.5. Here we see that the path normal of the blue path points in the same direction as the surface normal in purple. Additionally, the path normal for the path in red points in the opposite direction of the surface normal. This implies negative Gaussian curvature.

Example 3.6. We see here how the maximum Gaussian curvature is zero since the blue path has no normal path component.

While it still might not seem immediately obvious why we care about Gaussian curvature, it turns out that Guass, who first discovered it, found that Gaussian curvature is intrinsic to the surface. What this means is that a manifold with, say, positive Gaussian curvature, can never be deformed smoothly in such a way so that the result has negative or zero Gaussian curvature. This is exactly why you can fold a pizza in half so that it doesn't bend when you eat it. The pizza exists on a flat plane, and folding it does not change its Gaussian curvature. It will not be able to further bend for that would require the pizza to have negative Gaussian curvature, which Gauss showed is not possible.

Geodesics

We've seen up to now how to find curvature of one and two dimensional manifolds through normal curvature for paths and Gaussian curvature for surfaces. In order to generalize further to manifolds of higher dimension, a brief discussion must be had on geodesics as they are essential for defining sectional curvature, the analog to Gaussian curvature for higher dimensional manifolds. A proper discussion of geodesics requires a study into differential equations as well as a more thorough examination of the topics we have covered so far. Thus, here we introduce the concept intuitively and geometrically, so that the main ideas can be understood for later sections.

A geodesic of a manifold are paths on a manifold which curve only the amount they need to to stay on the manifold. Imagine a car which has the steering wheel locked so it may only travel straight. If we were to put this car on a flat plane, then the path it travels would be a straight line. If this car were to travel alone a sphere, it would travel along a great circle. Though the path the car traces out while traveling along the great circle on a sphere is curved, it is only curved as much as it needs to in order to stay on the sphere. This is how we should visualize geodesics of a surface.

Example 3.7. The geodesics of a flat plane are straight lines. This makes sense intuitively from our description of geodesics as paths where no force is felt.

Example 3.8. The geodesics of a sphere are the "great circles" or planar intersections of the sphere passing through the center.

Example 3.9. The three categories of geodesics to a cylinder.

It is an important fact that for every point on a smooth manifold, for every direction in the tangent space to that point, there exists a unique geodesic going that direction and passing through that point. Thus, no matter where we are on the manifold, we may pick a direction and be guaranteed that a geodesic passes through the point in that direction. In essence, going back to the analogy of the car, on every smooth manifold, no matter what direction we position our car with the steering wheel lock, it will always be able to drive, at least a short distance, in the path of a geodesic.

Sectional Curvature

Before we begin with our last topic of discussion regarding curvature, let's reflect how we got here which will in turn motivate our discussion of sectional curvature.

We began our discussion by investigating why the intuitive concept of curvature seems to make sense for surfaces and shapes, but not for other mathematical objects like integers. We introduced the notion of space, and in particular a *topological space*. This gave us a mathematical way of discussing space. More work still needed to be done. We refined these ideas more by defining and describing topological charts. Still, this wasn't enough, so we built up the theory of manifolds, then smooth manifolds. Each step we developed on the ideas we previously realized - refining and modifying on the previous concepts to develop new ones.

This carried over even to our discussion of curvature. We started by studying the curvature of paths. We extended this idea to curvature of surfaces when we considered the curvature of multiple paths, culminating into the definition of Gaussian curvature. This final section defines sectional curvature by continuing exactly what we've been doing since the start of this paper, using previous concepts and ideas to extend and generalize our definitions.

Imagine we are mathematicians developing this theory of curvature from scratch. So far, we've been able to give mathematical definitions for curvature in one and two dimensions by defining normal and Gaussian curvature respectively. How could we imagine extending these definitions even further to those of higher dimension? One potential way we could think about doing this is by "cutting out" two dimensional "sections" of our higher dimensional manifold and then computing the Gaussian curvature of each of these sections.

Of course, we would want to cut these sections out in a way that is not totally arbitrary. Simply taking random two-dimensional parts of a manifold might have absolutely no relevance to the properties of the original manifold, thus giving us no information. We must consider ways of slicing up our manifold so that the resulting sections encode information about the curvature of the manifold.

When we started out discussion into curvature, we considered how the acceleration of a path could be used as a test of curvature on a manifold. Geodesics, then, were the paths on the manifold that had no acceleration on the manifold. That is, a geodesic path on a manifold is one that curves exactly as it must in order to stay on the manifold. For example, the geodesic paths on a sphere are great circles, paths which are definitely curved. However, they are paths which curve only enough to stay on the sphere - hence they are geodesics. In this sense, the geodesic

paths encode information about the curvature of the manifold. Because of this property of the geodesic encoding curvature of the manifold, we shall use them to carve out our sections.

So geodesics will be the tool we use to carve out sections of our manifold. But which geodesics should we use exactly? To answer this, we turn to the tangent space of a manifold. Recall that the tangent space of a smooth manifold is a vector space defined at every point on the manifold which encodes all possible directions one may travel on it. Thus, the tangent space of a manifold has the same dimension as the underlying manifold. Just like in any vector space, we may consider the two dimensional subspaces of the tangent space. For every two-dimensional subspace, we get a set of directions by considering the set of all unit vectors on the plane. We recall from the section on geodesics that for every unit direction of the tangent space, there is a unique geodesic which goes in that direction. With this, we have the tools to construct a section.

For a smooth manifold M at a point p, we construct a section as follows. Let $L \subset T_pM$ be a two-dimensional linear subspace. This yields a set of unit vectors, each of which yields a unique geodesic. These geodesics, when taken in aggregate, create a surface. Since this surface is constructed from the geodesics, it serves as a two-dimensional representation of the curvature of that plane of directions. This section is a two-dimensional surface and so we may compute the Gaussian curvature of it. The Gaussian curvature of the geodesics generated from this plane is called the sectional curvature.

Accounting for all two-dimensional linear subspaces of the tangent space of a manifold in fact *totally defines* the curvature of a manifold at that point. In this respect, the sectional curvature of a manifold is the "correct" way to extend curvature to manifolds of higher dimensions.

References

[1] AXLER, S. *Linear Algebra Done Right.* Springer International Publishing, 2024.

[2] CARMO, M. D. *Differential Geometry of Curves and Surfaces.* CRC Press, Taylor & Francis Group, 2023.

[3] LEE, J. *Introduction to Smooth Manifolds.* Springer, 2013.

[4] LEE, J. *Introduction to Riemannian Manifolds.* Springer, 2018.

[5] MUNKRES, J. *Topology.* Prentice Hall, 2000.

[6] VIEIRA, L. Mug and torus morph, Jan. 2007.