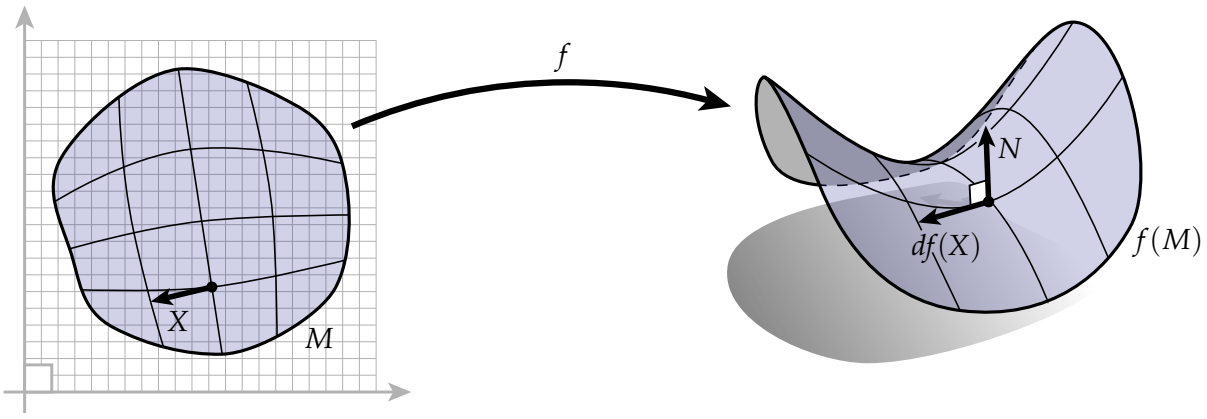
CHAPTER 2

A Quick and Dirty Introduction to Differential Geometry

2.1. The Geometry of Surfaces

There are many ways to think about the geometry of a surface (using *charts*, for instance) but here's a picture that is well-suited to the way we work with surfaces in the discrete setting. Consider a little patch of material floating in space, as depicted below. Its geometry can be described via a map $f : M \to \mathbb{R}^3$ from a region M in the Euclidean plane $\mathbb{R}^2$ to a subset $f(M)$ of $\mathbb{R}^3$:



The *differential* of such a map, denoted by df, tells us how to map a vector X in the plane to the corresponding vector $df(X)$ on the surface. Loosely speaking, imagine that M is a rubber sheet and X is a little black line segment drawn on M. As we stretch and deform M into $f(M)$, the segment X also gets stretched and deformed into a different segment, which we call $df(X)$. Later on we can talk about how to explicitly express $df(X)$ in coordinates and so on, but it's important to realize that fundamentally there's *nothing deeper to know about the differential than the picture you see here*—the differential simply tells you how to stretch out or *"push forward"* vectors as you go from one space to another. For example, the length of a tangent vector X pushed forward by f can be expressed as

$$\sqrt{df(X) \cdot df(X)},$$

where $\cdot$ is the standard inner product (a.k.a. *dot product* or *scalar product*) on $\mathbb{R}^3$. Note that this length is typically *different* than the length of the vector we started with! To keep things clear, we'll use angle brackets to denote the inner product in the plane, e.g., the length of the original vector would be $\sqrt{\langle X, X \rangle}$. More generally, we can measure the inner product between any *two* tangent vectors $df(X)$ and $df(Y)$:

$$g(X, Y) = df(X) \cdot df(Y).$$

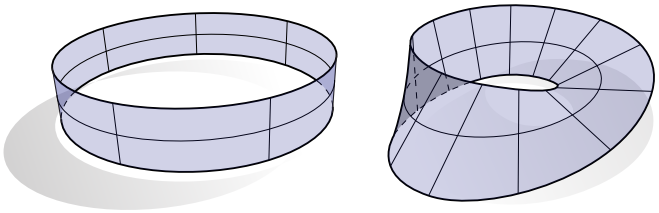
The map g is called the *metric* of the surface, or to be more pedantic, the *metric induced by* f. Note that throughout we will use $df(X)$ interchangeably to denote both the pushforward of a single

vector or an entire *vector field*, *i.e.*, a vector at every point of M. In most of the expressions we'll consider this distinction won't make a big difference, but it's worth being aware of. Throughout we'll use TM to denote the *tangent bundle* of M, *i.e.*, the set of all tangent vectors.

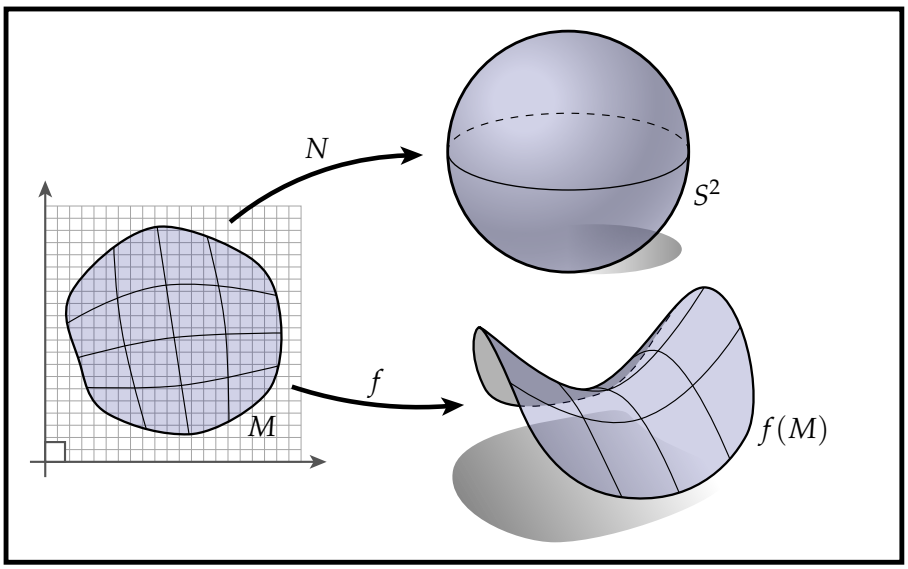
So far we've been talking about *tangent* vectors, i.e., vectors that lay flat along the surface. We're also interested in vectors that are orthogonal to the surface. In particular, we say that a vector $u \in \mathbb{R}^3$ is *normal* to the surface at a point p if

$$df(X) \cdot u = 0$$

for all tangent vectors X at p. For convenience, we often single out a particular normal vector N called the *unit normal*, which has length one. Of course, at any given point there are two distinct unit normal vectors: $+N$ and $-N$. Which one should we use? If we can pick a consistent direction for N then we say that M is *orientable*. For instance, the circular band on the left is orientable, but the *Möbius band* on the right is not:



For orientable surfaces, we can actually think of N as a continuous map $N : M \to S^2$ (called the *Gauss map*) which associates each point with its unit normal, viewed as a point on the unit sphere S^2. In fact, if we think of S^2 as a subset of $\mathbb{R}^3$ (consisting of all the points unit distance from the origin), then we can do all the same things with N that we did with our map f. In particular, the differential dN (called the *Weingarten map*) tells us about the change in the normal direction as we move from one point to the other. For instance, we can look at the change in normal along a particular tangent direction X by evaluating $dN(X)$—this interpretation will become useful when we talk about the curvature of surfaces. Overall we end up with the following picture, which captures the most fundamental ideas about the geometry of surfaces:

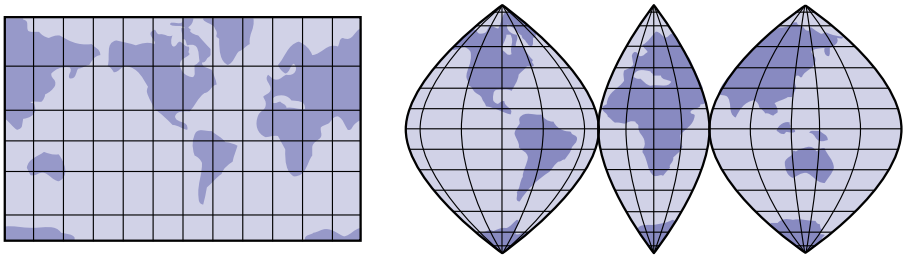


2.1.1. Conformal Coordinates. When working with curves, one often introduces the idea of an *isometric* (a.k.a. *arc-length* or *unit speed*) parameterization. The idea there is to make certain expressions simpler by assuming that no "stretching" occurs as we go from the domain into $\mathbb{R}^3$. One way to state this requirement is

$$|df(X)| = |X|,$$

i.e., we ask that the norm of any vector X is preserved.

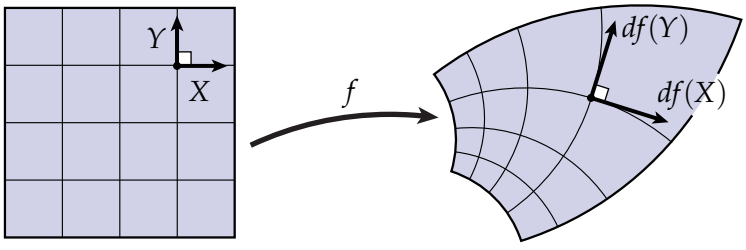
For surfaces, an isometric parameterization does not always exist (not even locally!). Most of the time you simply have to stretch things out. For instance, you may know that it's impossible to flatten the surface of the Earth onto the plane without distortion—that's why we end up with all sorts of different funky projections of the globe.



However, there is a setup that (like arc-length parameterization for curves) makes life a lot easier when dealing with certain expressions, namely *conformal coordinates*. Put quite simply, a map f is conformal if it preserves the *angle* between any two vectors. More specifically, a conformal map $f : \mathbb{R}^2 \supset M \to \mathbb{R}^3$ satisfies

$$df(X) \cdot df(Y) = a\langle X, Y\rangle$$

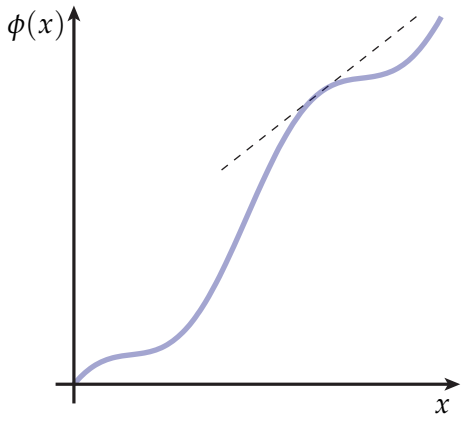
for all tangent vectors X, Y, where a is a **positive** function and $\langle \cdot, \cdot \rangle$ is the usual inner product on $\mathbb{R}^2$. In practice, the function a is often replaced with e^u for some real-valued function u—this way, one never has to worry about whether the scaling is positive. Notice that vectors can still get *stretched out*, but the surface never gets *sheared*—for instance, orthogonal vectors always stay orthogonal:



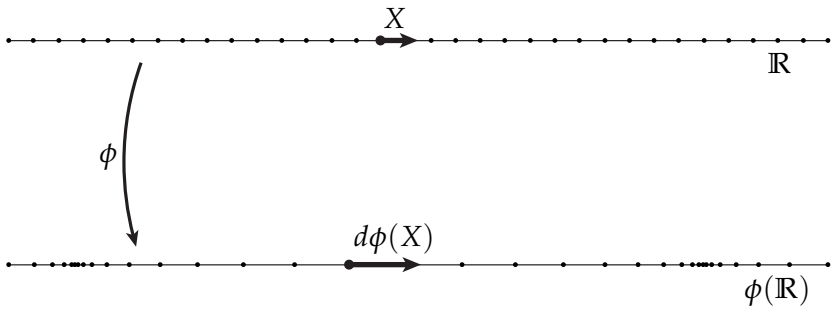
A key fact about conformal maps is that they always exist, as guaranteed by the uniformization theorem. In a nutshell, the uniformization theorem says that any disk can be conformally mapped to the plane. So if we consider any point p on our surface $f(M)$, we know that we can always find a conformal parameterization in some small, disk-like neighborhood around p. As with unit-speed curves, it is often enough to simply know that a conformal parameterization *exists*—we do not have to construct the map explicitly. And, as with arc-length parameterization, we have to keep track of the least possible amount of information about how the domain gets stretched out: just a single number at each point (as opposed to, say, an entire Jacobian matrix).

2.2. Derivatives and Tangent Vectors

2.2.1. Derivatives on the Real Line. So far we've been thinking about the differential in a very geometric way: it tells us how to stretch out or *push forward* tangent vectors as we go from one place to another. In fact, we can apply this geometric viewpoint to pretty much any situation involving derivatives. For instance, think about a good old fashioned real-valued function $\phi(x)$ on the real line. We typically visualize ϕ by plotting its value as a height over the x-axis:



In this case, the derivative ϕ' can be interpreted as the slope of the height function, as suggested by the dashed line in the picture above. Alternatively, we can imagine that ϕ stretches out the real line itself, indicated by the change in node spacing in this picture:



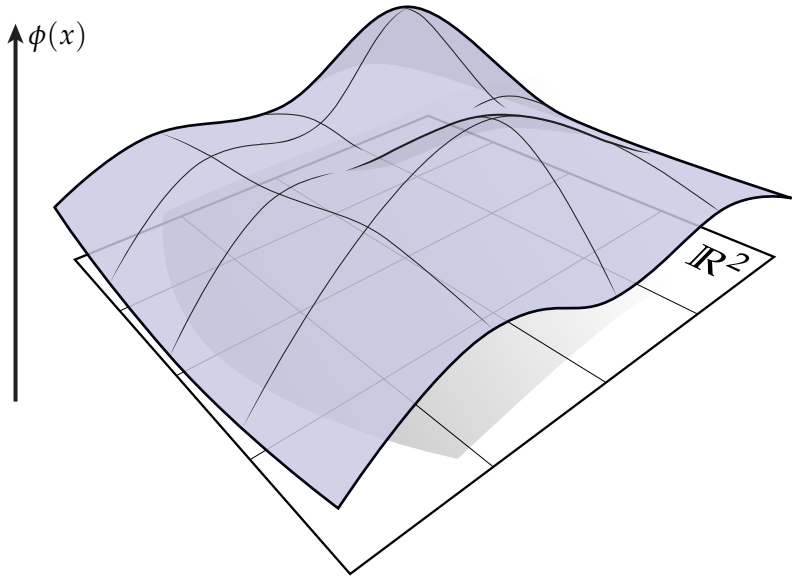
Where the derivative is large, nodes are spaced far apart; where the derivative is small, nodes are spaced close together. This picture inspires us to write the derivative of ϕ in terms of the push-forward $d\phi(X)$ of a unit tangent vector X pointing along the positive x-axis:

$$\phi' = d\phi(X).$$

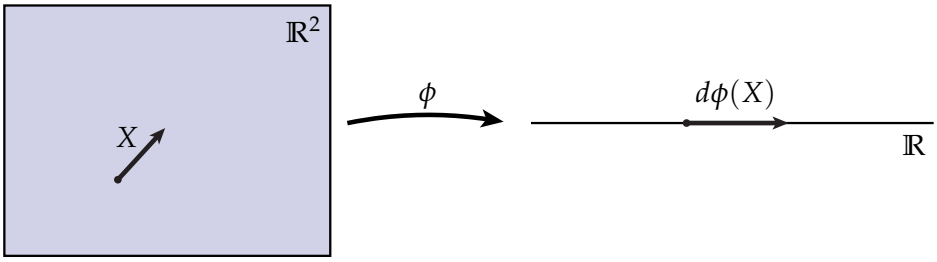
In other words, the derivative of ϕ is just the "stretch factor" as we go from one copy of $\mathbb{R}$ to the other. But wait a minute—does this equality even make sense? The thing on the left is a scalar, but the thing on the right is a vector! Of course, any tangent vector on the real line can be represented as just a single value, quantifying its extent in the positive or negative direction. So this expression does make sense—as long as we understand that we're identifying tangent vectors on $\mathbb{R}$ with real numbers. Often this kind of "type checking" can help verify that formulas and expressions are correct, similar to the way you might check for matching units in a physical equation.

Here's another question: how is this interpretation of the derivative any different from our usual interpretation in terms of height functions? Aren't we also stretching out the real line in that case? Well, yes and no—certainly the real line still gets stretched out into some other curve. But this curve is now a subset of the plane $\mathbb{R}^2$—in particular, it's the curve $\gamma = (x, \phi(x))$. So for one thing, "type checking" fails in this case: ϕ' is a scalar, but $d\gamma(X)$ is a 2-vector. But most importantly, the *amount* of stretching experienced by the curve doesn't correspond to our usual notion of the derivative of ϕ—for instance, if we look at the magnitude of $|d\gamma(X)|$ we get $\sqrt{1+(\phi')^2}$. (Why is this statement true geometrically? How *could* you write ϕ' in terms of $d\gamma(X)$? Can you come up with an expression that recovers the proper sign?)

2.2.2. Directional Derivatives. So far so good: we can think of the derivative of a real-valued function on $\mathbb{R}$ as the pushforward of a (positively-oriented) unit tangent vector X. But what does $d\phi(X)$ mean if ϕ is defined over some other domain, like the plane $\mathbb{R}^2$? This question may "stretch" your mind a little, but if you can understand this example then you're well on your way to understanding derivatives in terms of tangent vectors. Let's take a look at the geometry of the problem—again, there are two ways we could plot ϕ. The usual approach is to draw a height function over the plane:



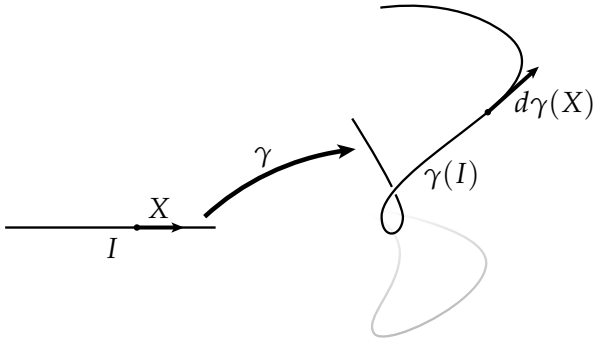
The derivative has something to do with the slope of this hill, but in which direction? To answer this question, we can introduce the idea of a *directional derivative*—i.e., we pick a vector X and see how quickly we travel uphill (or downhill) in that direction. And again we can consider an alternative picture:



Since ϕ is a map from $\mathbb{R}^2$ to $\mathbb{R}$, we can imagine that it takes a flat sheet of rubber and stretches it out into a long, skinny, one-dimensional object along the real line. Therefore if we draw an arrow X on the original sheet, then the "stretched-out" arrow $d\phi(X)$ gives us the rate of change in ϕ along the direction X, i.e., the directional derivative. What about type checking? As before, everything matches up: $d\phi(X)$ is a tangent vector on $\mathbb{R}$, so it can be represented by a single real number. (What if we had continued to work with the height function above? How could we recover the directional derivative in this case?)

By the way, don't worry if this discussion seems horribly informal! We'll see a more explicit, algebraic treatment of these ideas when we start talking about exterior calculus. The important thing for now is to build some geometric intuition about derivatives. In particular: a map from any space to any other space can be viewed as some kind of bending and twisting and stretching (or possibly tearing!); derivatives can be understood in terms of what happens to little arrows along the way.

2.3. The Geometry of Curves



The picture we looked at for surfaces is actually a nice way of thinking about shapes of any dimension. For instance, we can think of a one-dimensional curve as a map $\gamma : I \to \mathbb{R}^3$ from an interval $I = [0, T] \subset \mathbb{R}$ of the real line to $\mathbb{R}^3$. Again the differential $d\gamma$ tells us how tangent vectors get stretched out by γ, and again the induced length of a tangent vector X is given by

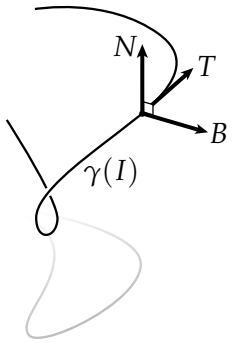
$$|d\gamma(X)| = \sqrt{d\gamma(X) \cdot d\gamma(X)}.$$

Working with curves is often easier if γ preserves *length*, i.e., if for every tangent vector X we have

$$|d\gamma(X)| = |X|.$$

There are various names for such a parameterization (*"unit speed"*, *"arc-length"*, *"isometric"*) but the idea is simply that the curve doesn't get stretched out when we go from $\mathbb{R}$ to $\mathbb{R}^3$—think of γ as a completely relaxed rubber band. This unit-speed view is also often the right one for the discrete setting where we have no notion of a base domain I—from the very beginning, the curve is given to us as a subset of $\mathbb{R}^3$ and all we can do is assume that it sits there in a relaxed state.

2.3.1. The Curvature of a Curve.



Suppose we have a unit-speed curve γ and a positively-oriented unit vector X on the interval I. Then

$$T = d\gamma(X)$$

is a unit vector in $\mathbb{R}^3$ tangent to the curve. Carrying this idea one step further, we can look at the change in tangent direction as we move along γ. Since T may change at any rate (or not at all!) we

split up the change into two pieces: a unit vector N called the *principal normal* that expresses the direction of change, and a scalar $\kappa \in \mathbb{R}$ called the *curvature* that expresses the magnitude of change:

$$dT(X) = -\kappa N.$$

One thing to realize is that T and N are always orthogonal. Why? Because if the change in T were *parallel* to T, then it would cease to have unit length! (This argument is a good one to keep in mind any time you work with unit vector fields.) By convention, we choose N to be the normal pointing to the "left" of the curve, *i.e.*, if at any point we consider a plane spanned by the tangent and the normal, N is a quarter turn in the counter-clockwise direction from N.

How does this frame change as we move along the curve? The answer is given by the *Frenet-Serret formula*:

$$\underbrace{\begin{bmatrix} T' \\ N' \\ B' \end{bmatrix}}_{Q' \in \mathbb{R}^{3\times 3}} = \underbrace{\begin{bmatrix} 0 & -\kappa & 0 \\ \kappa & 0 & -\tau \\ 0 & \tau & 0 \end{bmatrix}}_{A \in \mathbb{R}^{3\times 3}} \underbrace{\begin{bmatrix} T \\ N \\ B \end{bmatrix}}_{Q \in \mathbb{R}^{3\times 3}}.$$

Here T, N, and B are interpreted as row vectors, and a *prime* indicates the change in a quantity as we move along the curve at unit speed. For instance, $T' = dT(X)$, where X is a positively-oriented unit vector on I. The quantity τ is called the *torsion*, and describes the way the normal and binormal twist around the curve.

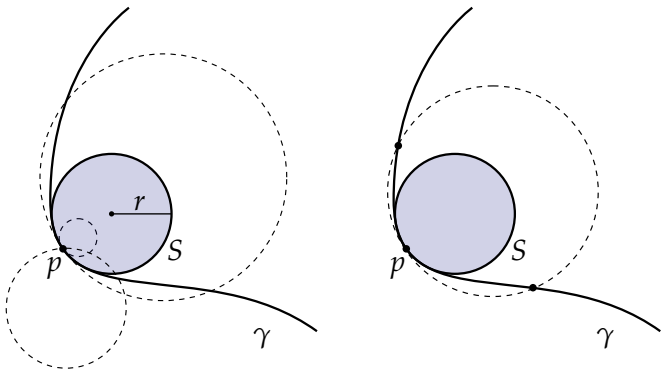
A concise proof of this formula was given by Cartan. First, since the vectors T, N, and B are mutually orthogonal, one can easily verify that $QQ^T = I$, i.e., Q is an orthogonal matrix. Differentiating this relationship in time, the identity vanishes and we're left with $Q'Q^T = -(Q'Q^T)^T$, i.e., the matrix $Q'Q^T$ is *skew-symmetric*. But since $A = Q'Q^T$, A must also be skew-symmetric. Skew symmetry implies that the diagonal of A is zero (why?) and moreover, we already know what the top row (and hence the left column) looks like from our definition of κ and N. The remaining value $A_{23} = -A_{32}$ is not constrained in any way, so we simply give it a name: $\tau \in \mathbb{R}$.

What do you think about this proof? On the one hand it's easy to verify; on the other hand, it provides little geometric understanding. For instance, why does N change in the direction of both T and B, but B changes only in the direction of N? Can you come up with more geometric arguments?

2.3.2. Visualizing Curvature. What's the curvature of a circle S? Well, if S has radius r then it takes time $2\pi r$ to go all the way around the circle at unit speed. During this time the tangent turns around by an angle 2π. Of course, since T has unit length the instantaneous change in T is described exclusively by the instantaneous change in angle. So we end up with

$$\kappa = |\kappa N| = |dT(X)| = 2\pi/2\pi r = 1/r.$$

In other words, the curvature of a circle is simply the reciprocal of its radius. This fact should make some intuitive sense: if we watch a circle grow bigger and bigger, it eventually looks just like a straight line with zero curvature: $\lim_{r\to\infty} 1/r = 0$. Similarly, if we watch a circle get smaller and smaller it eventually looks like a single point with infinite curvature: $\lim_{r\to 0} 1/r = \infty$.



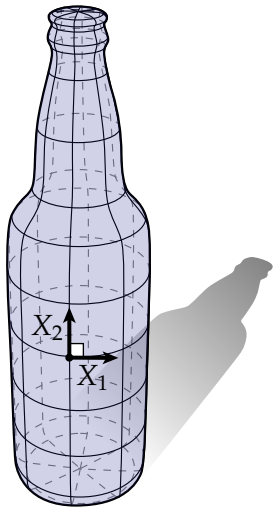
Now consider a smooth curve γ in the plane. At any point $p \in \gamma$ there is a circle S called the *osculating circle* that best approximates γ, meaning that it has the same tangent direction T and curvature vector κN. In other words, the circle and the curve agree "up to second order." (The phrase "agree up to nth order" is just shorthand for saying that the first n derivatives are equal.) How do we know such a circle exists? Easy: we can always construct a circle with the appropriate curvature by setting $r = 1/\kappa$; moreover *every* circle has some tangent pointing in the direction T. Alternatively, we can consider a circle passing through p and two other points: one approaching from the left, another approaching from the right. Since these three points are shared by both γ and S, the first and second derivatives will agree in the limit (consider that these points can be used to obtain consistent finite difference approximations of T and κN).

The radius and center of the osculating circle are often referred to as the *radius of curvature* and *center of curvature*, respectively. We can tell this same story for any curve in $\mathbb{R}^3$ by considering the *osculating plane* $T \times N$, since this plane contains both the tangent and the curvature vector.

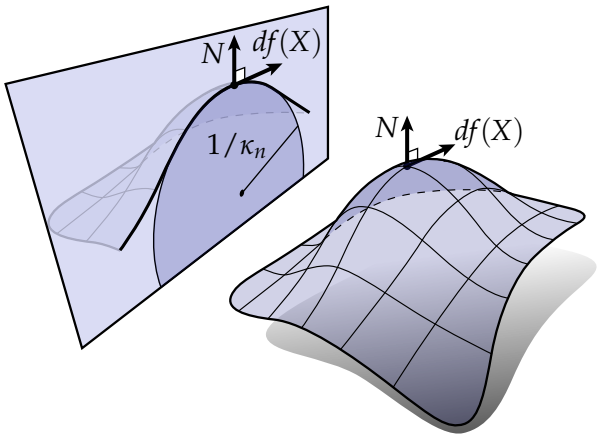
For curves it makes little difference whether we express curvature in terms of a change in the tangent vector or a change in the (principal) normal, since the two vectors are the same up to a quarter-rotation in the osculating plane. For surfaces, however, it will often make more sense to think of curvature as the change in the normal vector, since we typically don't have a distinguished tangent vector to work with.

2.4. Curvature of Surfaces

Let's take a more in-depth look at the curvature of surfaces. The word "curvature" really corresponds to our everyday understanding of what it means for something to be curved: eggshells, donuts, and cavatappi pasta have a lot of curvature; floors, ceilings, and cardboard boxes do not. But what about something like a beer bottle? Along one direction the bottle quickly curves around in a circle; along another direction it's completely flat and travels along a straight line:



This way of looking at curvature—in terms of curves contained in the surface—is often how we treat curvature in general. In particular, let $df(X)$ be a unit tangent direction at some distinguished point on the surface, and consider a plane containing both $df(X)$ and the corresponding normal N. This plane intersects the surface in a curve, and the curvature κ_n of this curve is called the *normal curvature* in the direction X:

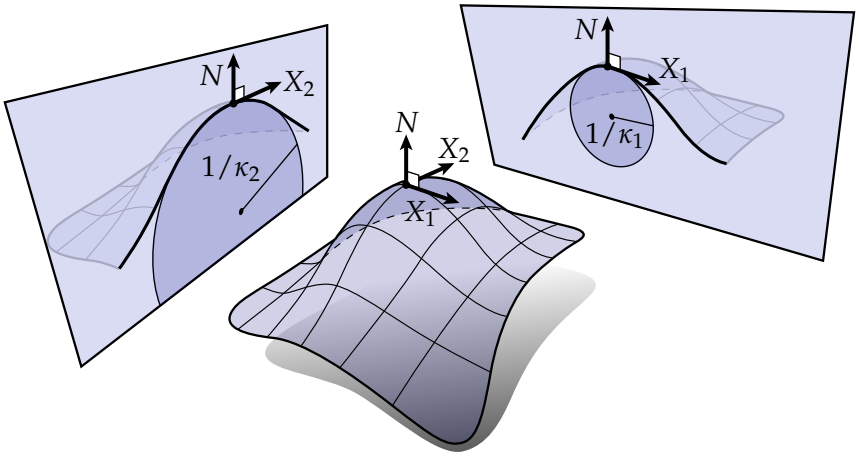


Remember the Frenet-Serret formulas? They tell us that the change in the normal along a *curve* is given by $dN = \kappa T - \tau B$. We can therefore get the normal curvature along X by extracting the tangential part of dN:

$$\kappa_n(X) = \frac{df(X) \cdot dN(X)}{|df(X)|^2}.$$

The factor $|df(X)|^2$ in the denominator simply normalizes any "stretching out" that occurs as we go from the domain M into $\mathbb{R}^3$—a derivation of this formula can be found in Appendix A. Note that normal curvature is *signed*, meaning the surface can bend toward the normal or away from it.

2.4.1. Principal, Mean, and Gaussian Curvature.



At any given point we can ask: along which directions does the surface bend the most? The unit vectors X_1 and X_2 along which we find the maximum and minimum normal curvatures κ_1 and κ_2 are called the *principal directions*; the curvatures κ_i are called the *principal curvatures*. For instance, the beer bottle above might have principal curvatures $\kappa_1 = 1$, $\kappa_2 = 0$ at the marked point.

We can also talk about principal curvature in terms of the *shape operator*, which is the unique map $S : TM \to TM$ satisfying

$$df(SX) = dN(X)$$

for all tangent vectors X. The shape operator S and the Weingarten map dN essentially represent the same idea: they both tell us how the normal changes as we travel along a direction X. The only difference is that S specifies this change in terms of a tangent vector on M, whereas dN gives us the change as a tangent vector in $\mathbb{R}^3$. It's worth noting that many authors do not make this distinction, and simply assume an isometric identification of tangent vectors on M and the corresponding tangent vectors in $\mathbb{R}^3$. However, we choose to be more careful so that we can explicitly account for the dependence of various quantities on the immersion f—this dependence becomes particularly important if you actually want to compute something! (By the way, why can we always express the change in N in terms of a *tangent* vector? It's because N is the *unit* normal, hence it cannot grow or shrink in the normal direction.)

One important fact about the principal directions and principal curvatures is that they correspond to eigenvectors and eigenvalues (respectively) of the shape operator:

$$SX_i = \kappa_i X_i.$$

Moreover, the principal directions are orthogonal with respect to the induced metric: $g(X_1, X_2) = df(X_1) \cdot df(X_2) = 0$—see Appendix B for a proof of these two facts. The principal curvatures therefore tell us everything there is to know about normal curvature at a point, since we can express any tangent vector Y as a linear combination of the principal directions X_1 and X_2. In particular, if

Y is a unit vector offset from X_1 by an angle θ, then the associated normal curvature is

$$\kappa_n(Y) = \kappa_1 \cos^2\theta + \kappa_2 \sin^2\theta,$$

as you should be able to easily verify using the relationships above. Often, however, working directly with principal curvatures is fairly inconvenient—especially in the discrete setting.

On the other hand, two closely related quantities—called the *mean curvature* and the *Gaussian curvature* will show up over and over again (and have some particularly nice interpretations in the discrete world). The mean curvature H is the arithmetic mean of principal curvatures:

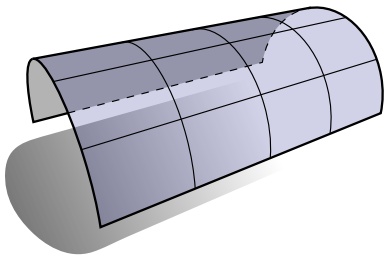
$$H = \frac{\kappa_1 + \kappa_2}{2},$$

and the Gaussian curvature is the (square of the) geometric mean:

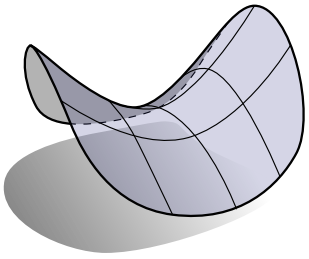
$$K = \kappa_1 \kappa_2.$$

What do the values of H and K imply about the shape of the surface? Perhaps the most elementary interpretation is that Gaussian curvature is like a logical "and" (is there curvature along *both* directions?) whereas mean curvature is more like a logical "or" (is there curvature along *at least one* direction?) Of course, you have to be a little careful here since you can also get zero mean curvature when $\kappa_1 = -\kappa_2$.

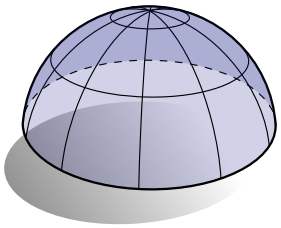
It also helps to see pictures of surfaces with zero mean and Gaussian curvature. Zero-curvature surfaces are so well-studied in mathematics that they have special names. Surfaces with zero Gaussian curvature are called *developable surfaces* because they can be "developed" or flattened out into the plane without any stretching or tearing. For instance, any piece of a cylinder is developable since one of the principal curvatures is zero:



Surfaces with zero mean curvature are called *minimal surfaces* because (as we'll see later) they minimize surface area (with respect to certain constraints). Minimal surfaces tend to be saddle-like since principal curvatures have equal magnitude but opposite sign:



The saddle is also a good example of a surface with negative *Gaussian* curvature. What does a surface with positive Gaussian curvature look like? The hemisphere is one example:



Note that in this case $\kappa_1 = \kappa_2$ and so principal directions are not uniquely defined—maximum (and minimum) curvature is achieved along *any* direction X. Any such point on a surface is called an *umbilic point*.

There are plenty of cute theorems and relationships involving curvature, but those are the basic facts: the curvature of a surface is completely characterized by the *principal curvatures*, which are the maximum and minimum *normal curvatures*. The Gaussian and mean curvature are simply averages of the two principal curvatures, but (as we'll see) are often easier to get your hands on in practice.

2.4.2. The Fundamental Forms. For historical reasons, there are two objects we should probably mention: *first fundamental form* I and the *second fundamental form* $I\!I$. I'm actually not sure what's so fundamental about these forms, since they're nothing more than a mashup of the metric g and the shape operator S, which themselves are simple functions of two *truly* fundamental objects: the immersion f and the Gauss map N. In fact, the first fundamental form is literally just the induced metric, *i.e.*,

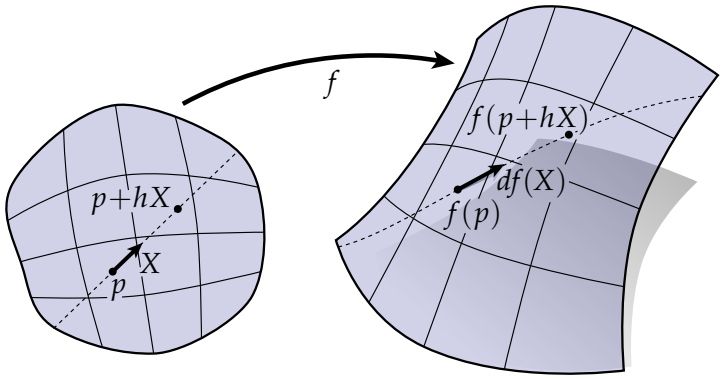
$$I(X, Y) := g(X, Y).$$

The second fundamental form looks quite similar to our existing expression for normal curvature:

$$I\!I(X, Y) := -g(SX, Y) = -dN(X) \cdot df(Y).$$

The most important thing to realize is that I and $I\!I$ do not introduce any new *geometric* ideas—just another way of writing down things we've already seen.

2.5. Geometry in Coordinates



So far we've given fairly abstract descriptions of the geometric objects we've been working with. For instance, we said that the differential df of an immersion $f : M \to \mathbb{R}^3$ tells us how to stretch out tangent vectors as we go from the domain $M \subset \mathbb{R}^2$ into the image $f(M) \subset \mathbb{R}^3$. Alluding to the picture above, we can be a bit more precise and define $df(X)$ in terms of limits:

$$df_p(X) = \lim_{h \to 0} \frac{f(p + hX) - f(p)}{h}.$$

Still, this formula remains a bit abstract—we may want something more concrete to work with in practice. When we start working with discrete surfaces we'll see that $df(X)$ often has an incredibly concrete meaning—for instance, it might correspond to an edge in our mesh. But in the smooth setting a more typical representation of df is the *Jacobian matrix*

$$\mathsf{J} = \begin{bmatrix} \partial f^1/\partial x^1 & \partial f^1/\partial x^2 \\ \partial f^2/\partial x^1 & \partial f^2/\partial x^2 \\ \partial f^3/\partial x^1 & \partial f^3/\partial x^2 \end{bmatrix}.$$

Here we pick coordinates on $\mathbb{R}^2$ and $\mathbb{R}^3$, and imagine that

$$f(x^1, x^2) = (f_1(x^1, x^2), f_2(x^1, x^2), f_3(x^1, x^2))$$

for some triple of scalar functions $f_1, f_2, f_3 : M \to \mathbb{R}$. So if you wanted to evaluate $df(X)$, you could simply apply J to some vector $X = [X^1 \;\; X^2]^T$.

2.5.1. Coordinate Representations Considered Harmful. You can already see one drawback of the approach taken above: expressions get a lot longer and more complicated to write out. But there are other good reasons to avoid explicit matrix representations. The most profound reason is that matrices can be used to represent many different types of objects, and these objects can behave in very different ways. For instance, can you guess what the following matrix represents?

$$\begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$$

Give up? It's quite clear, actually: it's the adjacency matrix for the complete graph on two vertices. No, wait a minute—it must be the Pauli matrix σ_x, representing spin angular momentum along the x-axis. Or is it the matrix representation for an element of the dihedral group D_4? You get the idea: when working with matrices, it's easy to forget where they come from—which makes it very easy

to forget which rules they should obey! (Don't you already have enough things to keep track of?) The real philosophical point here is that *matrices are not objects: they are merely* ***representations*** *of objects!* Or to paraphrase Plato: matrices are merely shadows on the wall of the cave, which give us nothing more than a murky impression of the real objects we wish to illuminate.

A more concrete example that often shows up in geometry is the distinction between linear operators and bilinear forms. As a reminder, a *linear operator* is a map from one vector space to another, e.g.,

$$f : \mathbb{R}^2 \to \mathbb{R}^2; u \mapsto f(u),$$

whereas a *bilinear form* is a map from a pair of vectors to a scalar, e.g.,

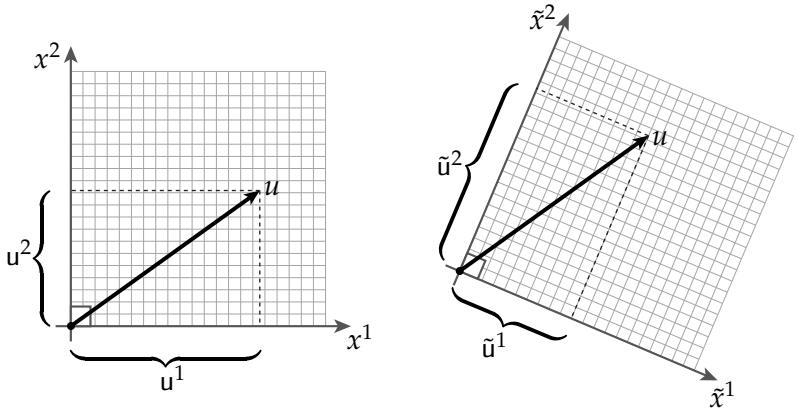
$$g : \mathbb{R}^2 \times \mathbb{R}^2 \to \mathbb{R}; (u, v) \mapsto g(u, v).$$

Sticking with these two examples let's imagine that we're working in a coordinate system (x^1, x^2), where f and g are represented by matrices $\mathsf{A}, \mathsf{B} \in \mathbb{R}^{2\times 2}$ and their arguments are represented by vectors $\mathsf{u}, \mathsf{v} \in \mathbb{R}^2$. In other words, we have

$$f(u) = \mathsf{Au}$$

and

$$g(u, v) = \mathsf{u}^\mathsf{T}\mathsf{Bv}.$$



Now suppose we need to work in a different coordinate system $(\tilde{x}^1, \tilde{x}^2)$, related to the first one by a change of basis $\mathsf{P} \in \mathbb{R}^{2\times 2}$. For instance, the vectors u and v get transformed via

$$\tilde{\mathsf{u}} = \mathsf{Pu},$$

$$\tilde{\mathsf{v}} = \mathsf{Pv}.$$

How do we represent the maps f and g in this new coordinate system? We can't simply evaluate $\mathsf{A}\tilde{\mathsf{u}}$, for instance, since A and $\tilde{\mathsf{u}}$ are expressed in different bases. What we need to do is evaluate

$$f(u) = \mathsf{PAu} = \mathsf{PAP}^{-1}\tilde{\mathsf{u}}$$

and similarly

$$g(u, v) = \mathsf{u}^T\mathsf{Bv} = (\mathsf{P}^{-1}\tilde{\mathsf{u}})^T\mathsf{B}(\mathsf{P}^{-1}\tilde{\mathsf{v}}) = \tilde{\mathsf{u}}^T(\mathsf{P}^{-\mathsf{T}}\mathsf{BP}^{-1})\tilde{\mathsf{v}}.$$

In other words, linear operators transform like

$$\mathsf{A} \mapsto \mathsf{P}\mathsf{A}\mathsf{P}^{-1},$$

whereas bilinear forms transform like

$$\mathsf{B} \mapsto \mathsf{P}^{-\mathsf{T}}\mathsf{B}\mathsf{P}^{-1}.$$

So what we discover is that *not all matrices transform the same way!* But if we're constantly scrawling out little grids of numbers, it's very easy to lose track of which transformations should be applied to which objects.

2.5.2. Standard Matrices in the Geometry of Surfaces. Admonitions about coordinates aside, it's useful to be aware of standard matrix representations for geometric objects because they provide an essential link to classical results. We've already seen a matrix representation for one object: the differential df can be encoded as the Jacobian matrix J containing first-order derivatives of the immersion f. What about the other objects we've encountered in our study of surfaces? Well, the induced metric g should be pretty easy to figure out since it's just a function of the differential—remember that

$$g(u, v) = df(u) \cdot df(v).$$

Equivalently, if we use a matrix $\mathsf{I} \in \mathbb{R}^{2\times 2}$ to represent g, then we have

$$\mathsf{u}^{\mathsf{T}}\mathsf{I}\mathsf{v} = (\mathsf{J}\mathsf{u})^{\mathsf{T}}(\mathsf{J}\mathsf{v})$$

which means that

$$\mathsf{I} = \mathsf{J}^{\mathsf{T}}\mathsf{J}.$$

We use the letter "I" to denote the matrix of the induced metric, which was historically referred to as the *first fundamental form* —fewer authors use this terminology today. In older books on differential geometry you may also see people talking about "E", "F", and "G", which refer to particular entries of I:

$$\mathsf{I} = \begin{bmatrix} \mathsf{E} & \mathsf{F} \\ \mathsf{F} & \mathsf{G} \end{bmatrix}.$$

(Is it clear why "F" appears twice?) One might conjecture that these fifth, sixth, and seventh letters of the alphabet have fallen out of fashion precisely because they are so coordinate-dependent and hence carry little geometric meaning on their own. Nonetheless, it is useful to be able to recognize these critters, because they do show up out there in the wild.

Earlier on, we also looked at the *shape operator,* defined as the unique map $S : TM \to TM$ satisfying

$$dN(X) = df(SX),$$

and the *second fundamental form,* defined as

$$II(u, v) = g(Su, v).$$

(Remember that S turned out to be self-adjoint with respect to g, and likewise II turned out to be symmetric with respect to its arguments u and v.) If we let $\mathsf{S}, \mathbb{I\!I} \in \mathbb{R}^{2\times 2}$ be the matrix representations of S and II, respectively, then we have

$$\mathsf{u}^{\mathsf{T}}\mathbb{I\!I}\mathsf{v} = \mathsf{u}^{\mathsf{T}}\mathsf{I}\mathsf{S}\mathsf{v}$$

for all vectors $\mathsf{u}, \mathsf{v} \in \mathbb{R}^2$, or equivalently,

$$\mathbb{I\!I} = \mathsf{I}\mathsf{S}.$$

Components of $\mathrm{I\!I}$ are classically associated with *lowercase* letters from the Roman alphabet, namely

$$\mathrm{I\!I} = \begin{bmatrix} \mathsf{e} & \mathsf{f} \\ \mathsf{f} & g \end{bmatrix},$$

which in coordinates (x, y) are given explicitly by

$$\begin{aligned} \mathsf{e} &= N \cdot f_{xx}, \\ \mathsf{f} &= N \cdot f_{xy}, \\ \mathsf{g} &= N \cdot f_{yy}, \end{aligned}$$

where N is the unit surface normal and f_{xy} denotes the second partial derivative along directions x and y.

At this point we might want to stop and ask: how does a matrix like IS transform with respect to a change of basis? The first term, I, is a bilinear form, but the second term S is a linear map! As emphasized above, we can't determine the answer by just staring at the matrices themselves—we need to remember what they represent. In this case, we know that IS corresponds to the second fundamental form, so it should transform like any other bilinear form: $\mathsf{IS} \mapsto \mathsf{P}^{-\mathsf{T}} I S \mathsf{P}^{-1}$.

Finally, we can verify that classical geometric expressions using matrices correspond to the expressions we derived earlier using the differential. For instance, the classical expression for normal curvature is

$$\kappa_n(u) = \frac{I\!I(u,u)}{I(u,u)},$$

which we can rewrite as

$$\frac{\mathsf{u}^{\mathsf{T}} \mathrm{I\!I} \mathsf{u}}{\mathsf{u}^{\mathsf{T}} \mathsf{I} \mathsf{u}} = \frac{\mathsf{u}^{\mathsf{T}} \mathsf{IS} \mathsf{u}}{\mathsf{u}^{\mathsf{T}} \mathsf{I} \mathsf{u}} = \frac{(\mathsf{J}\mathsf{u})^{\mathsf{T}} (\mathsf{J}\mathsf{S}\mathsf{u})}{(\mathsf{J}\mathsf{u})^{\mathsf{T}} (\mathsf{J}\mathsf{u})} = \frac{df(u) \cdot dN(u)}{|df(u)|^2}.$$

Up to a choice of sign, this expression is the same one we obtained earlier by considering a curve embedded in the surface.