Differential Geometry

Part I: Curves and Surfaces in Two and Three Dimensions

Prison Mathematics Project

Introduction

Hello and welcome to Part I of the module on Differential Geometry! What follows is a module intended to support the reader in learning this fascinating topic. The Prison Mathematics Project (PMP) realises that you may be practising mathematics in an environment that is highly restrictive, so this text can both be used independently and does not require a calculator.

What is Differential Geometry?

You may or may not be used to elementary geometry (for familiarity, you can also find notes made through the PMP on this topic) but it doesn't matter so much. We will be using calculus to analyse curves (one-dimensional) and surfaces (two-dimensional). If any of these words seem a bit foreign, don't fret! We will provide a self-contained introduction to everything required for the understanding of this topic. These notes cover basic preliminaries, curves and surfaces embedded in two- and three-dimensional space, and their properties.

Learning in this Module

The best way to learn mathematics is to do mathematics. Indeed, education isn't something that happens more than it is something we should all participate in. You will find various exercise questions and worked examples in these notes so that you may try to solve problems and deepen your understanding of this topic. Although the aim is for everything to only require the content of this module, you are encouraged to use any other sources you have at your disposal.

Acknowledgements

These notes are based on a lecture course by J.M. Speight at the University of Leeds.

Contents

1 Preliminaries

Historians know that Archimedes managed to compute volumes of certain surfaces of revolution back in the 3rd Century BC! It wasn't until Gottfried Leibniz and Isaac Newton independently developed calculus in the 17th Century AD that a more systematic approach to computing volumes of revolutions was possible, so what a feat this was for Archimedes. Things really kicked off when Euler developed the notion of curvature, and Gauss and Riemann both studied various remarkable properties of surfaces.

Goal: These are two of the main aims of the module.

1. Mathematically describe curves and surfaces in a precise way.
2. Translate a mathematical system into a visual picture.

1.1 Sets

We recall some basic notions. A *set* X is a collection of objects (called *elements*); they could be numbers, or functions, or presidents, or a combination of many different types of thing! We usually label sets by capital letters and elements by lowercase letters. If we want to say that "x is an element of the set X", then we just write $x \in X$. If we want to emphasise that "x is **not** an element of the set X", then we write $x \notin X$. We use curly brackets { and } when we write out a set in full, e.g. $\{2, 4, 0, -8, \pi, 2 + i\}$. A colon : in a set is shorthand for "where we assume".

Of course, we can pick out some elements from a set and look at that as a set itself; this is the idea of a *subset*. For example, if we have the set $\{1, 2, 3, 4, 5\}$, then there are a number of subsets we can form: $\{1\}$ is a subset, $\{1, 3\}$ is another subset, $\{4, 5, 3\} = \{3, 4, 5\}$ is another (these sets are equal because we don't care about order or repetition of the elements). We can use the symbols $\subset$ and $\subseteq$ to denote a subset. For example, $X \subseteq Y$ says that "the set X is a subset of set Y"; this means that everything appearing within X also appears within Y. We use the symbol $\not\subset$ or $\not\subseteq$ to say that something is **not** a subset. The *empty set* is the subset $\emptyset$ containing nothing!

Definition 1.1 The set of real numbers is $\mathbb{R}$. How these are constructed is complicated, but essentially these contain all the whole numbers, decimals, fractions and irrational numbers (like π and e and $\sqrt{2}$). It can be useful to have symbols for some common subsets of $\mathbb{R}$:

$$\begin{aligned}
\mathbb{N} &= \text{the set of natural numbers } \{0, 1, 2, 3, 4, ...\},\\
\mathbb{Z}^+ &= \text{the set of positive integers } \{1, 2, 3, 4, ...\},\\
\mathbb{Z} &= \text{the set of integers } \{..., -2, -1, 0, 1, 2...\},\\
\mathbb{Q} &= \text{the set of rational numbers } \{\tfrac{a}{b} : a, b \in \mathbb{Z} \text{ with } b \neq 0\}.
\end{aligned}$$

Sometimes, we may have a set and want to talk about certain elements that **don't** appear in a subset. For instance, the irrational numbers mentioned in Definition 1.1 are just the numbers that aren't rational (duh). But instead of making a new symbol for it, we write $\mathbb{R} \setminus \mathbb{Q}$ to mean "everything in $\mathbb{R}$ that is **not** in $\mathbb{Q}$". This backslash is the *set-minus operation* or *set complement*.

> **Note:** Some authors sensibly use a minus sign for this operation: $\mathbb{R} \setminus \mathbb{Q}$ is written $\mathbb{R} - \mathbb{Q}$.

Throughout this module, we use the idea of an interval; these are special subsets of $\mathbb{R}$. We will write down an official definition soon but this is a good place to develop some geometric intuition. We first describe an interval without obscuring the definition with lots of symbols and letters.

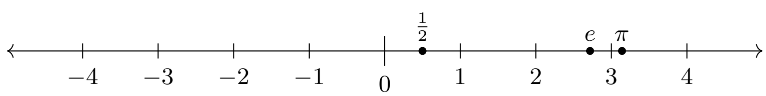


Figure 1: The real number line $\mathbb{R}$.

We have drawn a geometric picture in Figure 1 of the set $\mathbb{R}$ and it is nothing other than a line! We can now discuss the ideal of an interval before we write down a precise definition: an *interval* is a subset of $\mathbb{R}$ whose geometric picture is either a single line segment (a "finite line") or a single ray (a line which goes infinitely in **one** direction).

Definition 1.2 An interval is a subset of $\mathbb{R}$ of one of the following forms, where $a, b \in \mathbb{R}$:

- $[a, b] = \{x \in \mathbb{R} : a \leq x \leq b\}$.
- $[a, b) = \{x \in \mathbb{R} : a \leq x < b\}$.
- $(a, b] = \{x \in \mathbb{R} : a < x \leq b\}$.
- $(a, b) = \{x \in \mathbb{R} : a < x < b\}$.
- $[a, \infty) = \{x \in \mathbb{R} : x \geq a\}$.
- $(a, \infty) = \{x \in \mathbb{R} : x > a\}$.
- $(-\infty, b] = \{x \in \mathbb{R} : x \leq b\}$.
- $(-\infty, b) = \{x \in \mathbb{R} : x < b\}$.

We use square brackets [or] when we want to **include** the endpoint in our interval and we use standard parentheses (or) when we want to **exclude** the endpoint. We can again sketch a interval geometrically, where we use a filled circle $\bullet$ for including an endpoint and a non-filled circle $\circ$ for excluding an endpoint. The *infinity* symbol ∞ tells us the interval goes on forever in a certain direction (to the left if $-\infty$ and to the right if ∞). We always use regular parentheses when we have infinity because it is never included **in** the interval: we "never reach it".

Example 1.3 We sketch the interval $(-1.5, 3]$ on the real number line from Figure 1; see below.

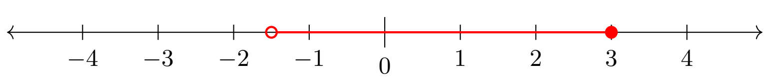


Figure 2: The interval $(-1.5, 3]$.

Definition 1.4 An interval is open if either both endpoints are excluded **or** one endpoint is excluded and the other is infinity. An interval is closed if either both endpoints are included **or** one endpoint is included and the other is infinity. Concretely, here we go:

- $[a, b]$ is closed.
- $[a, b)$ is **neither**.
- $(a, b]$ is **neither**.
- (a, b) is open.
- $[a, \infty)$ is closed.
- (a, ∞) is open.
- $(-\infty, b]$ is closed.
- $(-\infty, b)$ is open.

We can actually define what it means for more general sets (not just intervals) to be *open* and *closed*; this is done later in the module for certain types of subset. But if there is one message to have in mind, it is this: **open and closed are not opposites**! We can have sets (intervals) that are open, closed, neither or both. There are not many sets that are both, but we can think of the the real numbers as the interval $\mathbb{R} = (-\infty, \infty)$ and this is both.

Note: Some people use the silly word clopen for sets that are both... but it is quite catchy!

If we have a collection of sets, sometimes it is helpful to look at the set of *tuples*, that is we pick one element from each set and this gives us an element of a new set. Written out here it sounds horrible but it isn't that bad! We will start with the definition where we only have two sets and then we can naturally extend it.

Definition 1.5 Let X and Y be sets. The Cartesian product of these sets is the new set

$$X \times Y \coloneqq \{(x, y) : x \in X \text{ and } y \in Y\}.$$

Notation 1.6 If we let $Y = X$, that is the Cartesian product with itself, we write $X^2 \coloneqq X \times X$. In general, the *n-fold* Cartesian product $X^n = X \times X \times \cdots \times X$ has X appearing n times.

Note: The symbol (x, y) does **not** mean an interval from x to y; we also use regular parentheses to mean collections of elements, so (x, y) is a pair of elements. Similarly, we could have (x, y, z) which is a triple of elements, or (w, x, y, z) which is a quadruple, etc. In general, an n-tuple is a grouping of n elements which we write as $(x_1, x_2, ..., x_n)$.

Notation 1.7 We use the symbol $\coloneqq$ in Definition 1.5. This is an equals $=$ except what it means is "equal to by definition". In other words, they are equal because we say that they are!

Example 1.8 Consider $X = \{1, 3, 4\}$ and $Y = \{3, 7\}$. Then, their Cartesian product is the set

$$X \times Y = \{(1, 3), (1, 7), (3, 3), (3, 7), (4, 3), (4, 7)\}.$$

The last thing we need to discuss about sets at this state is their size. For the moment, we will focus only on **finite** sets (but we have already encountered a lot of sets that are infinitely large, e.g. all the ones listed in Definition 1.1).

Definition 1.9 The cardinality of a set X is the number of elements it has, denoted $|X|$.

At this level, we can write $|Y| = \infty$ whenever Y is a non-finite set. But there is a bit of a complication with this, since this suggests that $|\mathbb{R}| = \infty$ and $|\mathbb{Z}| = \infty$, so surely $|\mathbb{R}| = |\mathbb{Z}|$ right? Hold onto your hats but this is **false**: the infinity that is $\mathbb{R}$ is "larger" than the infinity that is $\mathbb{Z}$!

Lemma 1.10 *Let X and Y be finite sets. Then, $|X \times Y| = |X| \cdot |Y|$.*

Proof: Because X and Y are finite, we can assume that $|X| = m$ and $|Y| = n$, where $m, n \in \mathbb{N}$. Now, the elements $(x, y) \in X \times Y$ arise by choosing an element $x \in X$ (of which there are m choices) and an element $y \in Y$ (of which there are n choices). Because each choice is independent of the other, there are a total of mn choices, so there are a total of mn elements in $X \times Y$. □

Note: A lemma is a small result; they are mostly used as a stepping stone to bigger results.

1.2 Functions

Now that we have a good understanding of sets, we can talk about *functions*: these are rules that tell us how one element from a set is related to an element of a (potentially) different set. Their formal definition looks a little bit complicated and it is **hardly** used in practice, so we give a more practical definition.

Definition 1.11 Let X and Y be sets. A function $f : X \to Y$ is a rule which assigns to every element $x \in X$ a unique element $y \in Y$. To make the association clear, we would write $f(x) := y$. There are two more terms to define that we should be mindful of:

(i) The set X is called the domain of f.

(ii) The set Y is called the co-domain of f.

It is often very important to specify the domain and co-domain of a function. For instance, just writing $f(x) = x^3$ is not enough information; if we write $f : \mathbb{R} \to \mathbb{R}$ given by $f(x) = x^3$, this is significantly better. Although we have written the co-domain, is there a smaller set (a subset) which also acts as the co-domain? As an example, I could define $g : \mathbb{Z} \to \mathbb{R}$ by $g(n) = n + 1$, but if we test out this function, we see that we only get integers out. This is the idea of the *range* of a function: this is the set of stuff we actually get out.

Definition 1.12 The image (or range) of a function $f : X \to Y$ is the subset of its co-domain

$$\mathrm{im}(f) := \{f(x) : x \in X\} \subseteq Y.$$

Example 1.13 Consider the function $g : \mathbb{Z} \to \mathbb{R}$ given by $g(n) = n + 1$ that we had above. Its co-domain is $\mathbb{R}$, but we already discussed that $\mathrm{im}(g) = \mathbb{Z}$.

If we have a subset of a function's domain, sometimes we want to see what the function does to that specific subset. We call this the *image of a subset*. Similarly, if we have a subset of a function's co-domain, we may want to see from where the function sends things to end up in this subset. This is called the *pre-image of a subset*. We will write out the definitions now and have a go at an example.

Definition 1.14 Let $f : X \to Y$ be a function and consider subsets $U \subseteq X$ and $V \subseteq Y$.

(i) The image of U under f is the set $f(U) := \{f(x) : x \in U\}$.

(ii) The pre-image of V under f is the set $f^{-1}(V) := \{x \in X : f(x) \in V\}$.

Example 1.15 Consider the function $h : \mathbb{R} \to \mathbb{R}$ given by $h(x) = x^5$. We see that the pre-image of the subset $W := \{..., -4, -2, 0, 2, 4, ...\} \subseteq \mathbb{R}$ of even integers under the function f is the set

$$f^{-1}(W) = \{..., -\sqrt[5]{4}, -\sqrt[5]{2}, 0, \sqrt[5]{2}, \sqrt[5]{4}, ...\}.$$

Note: Strictly speaking, we don't know that the fifth root of a real number exists; I promise that it does but you will have to suspend disbelief for now. Such things are often justified in a real analysis course (where you study real numbers and real functions).

Some functions are a little bit "bad" in that they may send two different numbers to the same output. e.g. $f : \mathbb{R} \to \mathbb{R}$ given by $f(x) = x^2$ is such that $f(2) = 4$ and $f(-2) = 4$. There's nothing really wrong with this (and f is still definitely a valid function) but it just means that it is not *injective*; these are a nice class of function. In part, it is related to us being able to undo a function (if I am in the co-domain of an injective function, I can always find my way back home). Related is the idea of a *surjective* function, but this just means that its image is equal to its co-domain (there is no extra information hanging about for no reason).

Definition 1.16 Let $f : X \to Y$ be a function.

(i) It is injective if for every $x_1, x_2 \in X$, whenever $f(x_1) = f(x_2)$, it follows that $x_1 = x_2$.

(ii) It is surjective if for every $y \in Y$, there exists $x \in X$ such that $f(x) = y$.

(iii) It is bijective if it is both injective and surjective.

Example 1.17 We consider three functions and determine if they are injective/surjective or not.

(i) Let the function $f : \mathbb{N} \to \mathbb{N}$ be given by $f(n) = n^3$. This is injective but **not** surjective. Indeed, it is injective because $f(n_1) = f(n_2)$ is equivalent to $n_1^3 = n_2^3$, which is true if and only if $n_1 = n_2$. It is not surjective because for $2 \in \mathbb{N}$, say, there is **no** $n \in \mathbb{N}$ with $f(n) = 2$.

(ii) Let the function $g : \mathbb{R} \to [0, \infty)$ be given by $g(x) = (x - 1)^2$. This is **not** injective but is surjective. Indeed, it is not injective because $0 \neq 2$ yet $g(0) = 1 = g(2)$ are equal. It is surjective because, for $y \in [0, \infty)$, we can see that $g(\sqrt{y} + 1) = y$, and $\sqrt{y} + 1 \in \mathbb{R}$.

(iii) Let the function $h : \mathbb{R} \to \mathbb{R}$ be given by $h(x) = x + 1$. This is bijective. Indeed, it is injective because $h(x_1) = h(x_2)$ is equivalent to $x_1 + 1 = x_2 + 1$, which clearly implies that $x_1 = x_2$. It is surjective because any $y \in \mathbb{R}$ is obtained by $h(y - 1)$, and $y - 1 \in \mathbb{R}$.

Note: A bijection basically tells us that two sets are the *same*. This is very helpful in a lot of areas of mathematics when we are working with different-looking objects; if we can find a bijection between them, they actually describe the same thing!

Notation 1.18 Let $f : X \to Y$ be a function. We know now that $f(x)$ is one way to denote the image of $x \in X$ under the map f. However, the notation $\mapsto$ is also commonly used, which says the thing at the straight-end of is "sent to" the thing at the arrow end:

$$\begin{array}{cccccc} f & : & X & \longrightarrow & Y & \\ & & x & \longmapsto & f(x) & . \end{array}$$

The final thing we want to talk about is when a function is invertible, whatever that means. But before we do that, we will discuss doing a function of another function; this is called *function composition*! Once we get our heads around this, the end of this talk will make a lot more sense.

Definition 1.19 The composition of $f : X \to Y$ and $g : Y \to Z$ is a function $g \circ f : X \to Z$.

Informally, this means that we first apply f to an element $x \in X$ and then we end up with an element $f(x) \in Y$, but we then apply to this the function g and then we end up with the element $g(f(x)) \in Z$. The notation $g \circ f$ means "first f, then g"; it seems a bit backwards but this is the normal thing to do.

Note: It is important that the co-domain (or at least the range) of f is equal to the domain of g; this is why they are both labelled by the same set Y in Definition 1.19. Because of this, knowing that $g \circ f$ exists does **not** guarantee that $f \circ g$ exists.

Lemma 1.20 *If $f : X \to Y$ and $g : Y \to Z$ are injective, then $g \circ f : X \to Y$ is injective.*

Proof: We need to show that $(g \circ f)(x_1) = (g \circ f)(x_2)$ implies that $x_1 = x_2$ for any $x_1, x_2 \in X$. Well, we know from the informal discussion after Definition 1.19 that $(g \circ f)(x_1)$ just means $g(f(x_1))$ and, similarly, $(g \circ f)(x_2)$ is the same as $g(f(x_2))$. Therefore, the equality we start with is just $g(f(x_1)) = g(f(x_2))$. Now for the magic: because g is injective, we know the things on the inside are equal, that is $f(x_1) = f(x_2)$. But even better, because f is injective, we know the things on the inside of this are equal, so $x_1 = x_2$. This is exactly what we wanted to show. □

Lemma 1.21 *If $f : X \to Y$ and $g : Y \to Z$ are surjective, then $g \circ f : X \to Y$ is surjective.*

Proof: We need to show that for every $z \in Z$, there exists $x \in X$ such that $(g \circ f)(x) = z$. Again by using the informal discussion above, we can re-write the left-hand side as $g(f(x))$. But because f is surjective, there exists $y \in Y$ such that $f(x) = y$, so the left-hand side simplifies to $g(y)$. But because g is surjective, there exists $z \in Z$ such that $g(y) = z$. □

Corollary 1.22 *If $f : X \to Y$ and $g : Y \to Z$ are bijective, then $g \circ f : X \to Y$ is bijective.*

Proof: This is immediate by the definition of bijective and Lemmas 1.20 and 1.21. □

Note: A corollary is a result which is a (near-)immediate consequence of an earlier result.

We are now ready to explain the idea of an inverse of a function; the first thing to point out is that we can only take inverses of bijective functions (we can tweak this slightly: if we have an injective function that is **not** surjective, we can always shrink its co-domain down so that it turns bijective, and then we can take its inverse).

Without naming names, the idea is that if we have a bijection $X \to Y$, then there exists another bijection $Y \to X$ in the opposite way; these two guys should exactly undo each other (i.e. their compositions in either order result in the *identity*).

Definition 1.23 Let X be a set. The identity function is $\mathrm{id} : X \to X$ given by $\mathrm{id}(x) = x$.

In words, it is one of the most boring functions in the world: whatever you put in it just spits straight back out! Using Notation 1.18, we can write identity functions as $x \mapsto x$.

Notation 1.24 Sometimes (very soon) it is necessary to talk about identity functions for different sets, so we often denote them like $\mathrm{id}_X : X \to X$ and $\mathrm{id}_Y : Y \to Y$ to avoid confusion.

Theorem 1.25 *Let $f : X \to Y$ be a bijection. Then, there exists a bijection $g : Y \to X$.*

Proof: Omitted. □

Note: A theorem is reserved for the most powerful or important results in mathematics.

We won't bother with the proof here; it actually uses the formal definition of a function (which I said is hardly used... I lied!) but it isn't the most enlightening use of our time. The point of Theorem 1.25 is that every bijection can be "reversed".

Definition 1.26 The bijection $g : Y \to X$ we have in Theorem 1.25 is the inverse of f.

The identity function plays a crucial role in what it means to be an inverse. Indeed, we make the "reversing" mentioned above precise: a bijection $f : X \to Y$ has an inverse $g : Y \to X$ if

$$g \circ f = \mathrm{id}_X \qquad \text{and} \qquad f \circ g = \mathrm{id}_Y \,.$$

Proposition 1.27 *The inverse of a bijection $f : X \to Y$ is unique.*

Proof: This is the typical strategy for uniqueness proofs: assume that f has **two** inverses, namely $g_1 : Y \to X$ and $g_2 : Y \to X$. We aim to show that these guys are actually equal (i.e. there is in fact only **one** inverse). Using the first of the composition formulae above, we know that

$$g_1 \circ f = \mathrm{id}_X \qquad \text{and} \qquad g_2 \circ f = \mathrm{id}_X \,.$$

In particular, $g_1 \circ f = g_2 \circ f$. If we now compose this by g_1 on the right, then we see that

$$g_1 \circ f \circ g_1 = g_2 \circ f \circ g_1 \qquad \text{which is the same as} \qquad g_1 \circ \mathrm{id}_Y = g_2 \circ \mathrm{id}_Y \,,$$

because g_1 is an inverse of f which means that $f \circ g_1$ is the identity function. But the identity function doesn't do anything, so composing by it doesn't matter. Therefore, the final equality above becomes $g_1 = g_2$; we have shown exactly what we wanted. □

Note: A proposition is a good mathematical result but is often not as major as a theorem.

Notation 1.28 To really hammer home that $g : Y \to X$ belongs to f, we now write $f^{-1} := g$. But do take care not to mix up the inverse of a function with the pre-image of a subset from Definition 1.14. It's unfortunate this is the notation commonly used but it is what it is!

1.3 Limits

We now have a good understanding of functions, so we decide to look at those whose co-domain is $\mathbb{R}$. It turns out that a number of these functions have some amazing properties; this is where calculus comes into its own. Before we can talk about the main ideas like differentiation and integration, we must discuss what is meant by a *limit*.

Definition 1.29 Let $D \subseteq \mathbb{R}$ be the domain of a function $f : D \to \mathbb{R}$. Suppose that $f(x)$ is defined for all $x \in D$ that is "arbitrarily close" to some number $a \in \mathbb{R}$; we do not require that $f(a)$ is defined. If f is "arbitrarily close" to some number $L \in \mathbb{R}$ whilst x is "sufficiently close" but **not** equal to a, then we call L the limit of f as x approaches a. We denote this by $\lim_{x\to a} f(x) = L$.

The problem here is that phrases like "arbitrarily close" and "sufficiently close" aren't given a precise mathematical meaning. The way to fix this is to use real analysis and talk about sequences; this is a far cry from what we want to talk about so we won't go any further here.

Example 1.30 Consider the function $f : [0, 2] \to \mathbb{R}$ defined by

$$f(x) = \begin{cases} 3 & \text{if } x = 1 \\ -5 & \text{if } x \neq 1 \end{cases}.$$

We claim $\lim_{x\to 1} f(x) = -5$. Indeed, according to Definition 1.29, we consider $x \in [0, 2]$ close to 1 but **not** equal to it, that is $x \neq 1$. But f then outputs -5 for all $x \neq 1$, so this must be the limit.

Note: The point of Example 1.30 is to show there is no expectation that $\lim_{x\to a} f(x) = f(a)$ in general. In fact, we soon define a *continuous* function, for which this is always true.

Theorem 1.31 (Algebra of Limits) *Suppose f and g are functions where $\lim_{x\to a} f(x) = L$ and $\lim_{x\to a} g(x) = K$ exist. Then, the following properties of limits are true:*

(i) $\lim_{x\to a} \big(f(x) + g(x)\big) = L + K$.

(ii) $\lim_{x\to a} \big(f(x)g(x)\big) = LK$.

(iii) $\lim_{x\to a} \big(f(x)/g(x)\big) = L/K$ **if** $g(x) \neq 0$ *for all* x *and* $K \neq 0$.

(iv) $\lim_{x\to a} kf(x) = kL$, *for any* $k \in \mathbb{R}$.

(v) $\lim_{x\to a} f(x)^n = L^n$ *for any* $n \in \mathbb{N}$.

Proof: Omitted. □

Example 1.32 Let's use the Algebra of Limits to compute the following:

$$\lim_{x\to 3} f(x), \qquad f(x) = \frac{x^2-9}{x-3}$$

First, the numerator is the difference of two squares, so it can be factorised as $(x-3)(x+3)$. Thankfully, we see that $f(x) = x+3$; if we didn't realise this, on may begin to panic that the denominator was approaching zero. Therefore, using the Algebra of Limits(i), with $g(x) \equiv 3$ being the constant function, we see that $\lim\limits_{x\to 3} f(x) = 3+3 = 6$.

Note: The function f in Example 1.32 has a special property: if we let $x = 3$, then we are dividing by zero and this is undefined. However, the cancelled form $f(x) = x + 3$ **is** defined at $x = 3$. This is a *removable singularity*; although it looks bad, we can remove it!

Back in high school or (early) college, you may have encountered what it means for a function to be *continuous* in the following sense: "continuous functions are ones that can be drawn without taking your pen off of the page". This isn't the worst idea in the world... but it is far from accurate!

Definition 1.33 Let $D \subseteq \mathbb{R}$. We say $f : D \to \mathbb{R}$ is continuous at $a \in D$ if $\lim\limits_{x\to a} f(x) = f(a)$. If not, f is discontinuous at $a \in D$. We say f is continuous if it is continuous at **all** $a \in D$.

Remark 1.34 If any of your non-mathematical friends ask you to describe a continuous function to them in a way they can understand **and** is more correct than talking about pens and pages, you must be prepared: "continuous functions are ones where if you slightly vary the input, then the output only slightly varies too".

Lemma 1.35 *The* reciprocal function $f : \mathbb{R}\setminus\{0\} \to \mathbb{R}$ *given by* $f(x) = \frac{1}{x}$ *is continuous.*

Proof: Let $a \in \mathbb{R}\setminus\{0\}$, meaning $a \neq 0$ is any non-zero number. As $\frac{1}{x}$ is defined at $x = a$, we get

$$\lim_{x\to a} f(x) = \frac{1}{a} = f(a).$$

□

Note: This function is one we cannot draw without taking the pen off the page but we have shown it **is** continuous. Of course, it is **not** continuous on the domain $\mathbb{R}$, but that's because it isn't even defined there! Even choosing an output for $x = 0$ will give a discontinuity.

1.4 Calculus

We are finally ready to discuss *differentiable* and *integrable* functions. This will again not be from the most rigorous standing because we were a bit obscure with what limits are in Definition 1.29. However, it is enough for our purposes; I strongly urge you to read up on real analysis if you feel dissatisfied by my avoidance of proper rigour in this subsection and the last.

Definition 1.36 Let $D \subseteq \mathbb{R}$. We say $f : D \to \mathbb{R}$ is differentiable at $a \in D$ if this limit exists:

$$f'(a) := \lim_{x \to a} \frac{f(x) - f(a)}{x - a}.$$

We say f is differentiable if the derivative exists at **all** $a \in D$.

Note: There is an equivalent definition of the derivative, often more practical to use:

$$f'(a) = \lim_{h \to 0} \frac{f(a+h) - f(a)}{h}.$$

Example 1.37 Recall the boring identity function we introduced just after Definition 1.26. We look at the one for real numbers, namely $\mathrm{id} : \mathbb{R} \to \mathbb{R}$ given by $\mathrm{id}(x) = x$. We will show that this function is differentiable (at **every** $a \in \mathbb{R}$). First thing's first, let $a \in \mathbb{R}$ be arbitrary. Therefore,

$$\begin{aligned} \lim_{h \to 0} \frac{\mathrm{id}(a+h) - \mathrm{id}(a)}{h} &= \lim_{h \to 0} \frac{a + h - a}{h} \\ &= \lim_{h \to 0} \frac{h}{h} \\ &= \lim_{h \to 0} 1 \\ &= 1. \end{aligned}$$

Not only have we shown that id is differentiable, but that its derivative is $\mathrm{id}'(a) = 1$ for all $a \in \mathbb{R}$.

Proposition 1.38 *If $f : D \to \mathbb{R}$ is differentiable at $a \in D$, then is is continuous at $a \in D$.*

Sketch of Proof: Although it is somewhat perverse, re-write $f(x)$ in this form for all $x \neq a$:

$$f(x) = \frac{f(x) - f(a)}{x - a}(x - a) + f(a).$$

Taking the limit as $x \to a$ and using the Algebra of Limits, we can immediately deduce that $\lim_{x \to a} f(x) = f(a)$; this is the very definition of continuity at $a \in D$. □

Before we list a number of useful derivatives to always have at our finger tips, we will write some facts about differentiation which are immensely useful.

Proposition 1.39 *Let f and g be differentiable functions. Then, the following are true:*

(i) $(f+g)'(x) = f'(x) + g'(x)$. (**Sum Rule**)

(ii) $(fg)'(x) = f'(x)g(x) + f(x)g'(x)$. (**Product Rule**)

(iii) $(f/g)'(x) = \left(f'(x)g(x) - f(x)g'(x)\right)/g(x)^2$ **if** $g(x) \neq 0$. (**Quotient Rule**)

(iv) $(f \circ g)'(x) = f'(g(x))g'(x)$. (**Chain Rule**)

Proof: The first three are direct from the Algebra of Limits, so no real work is needed. The Chain Rule is a little more complicated and thus the proof is omitted. □

Definition 1.40 A polynomial in x is a function $p : \mathbb{R} \to \mathbb{R}$ with $p(x) = a_0 + a_1 x + \cdots + a_n x^n$, where $a_0, ..., a_n \in \mathbb{R}$ are the coefficients. If $a_n \neq 0$, we call n the degree of p, denoted $\deg(p)$.

These are a very common type of function and appear throughout mathematics. The constant function which sends everything to a single number, the identity function, linear functions, quadratics, cubics etc. are all examples of polynomials! The Sum Rule of Differentiation from Proposition 1.39 means we can differentiate a polynomial (assuming we know that the function $x \mapsto x^k$ is differentiable; it actually is).

You have maybe encountered trigonometric functions sin, cos, tan and the exponential and logarithmic functions exp, log before. We won't go into the details on how they are defined (they are actually given by infinite sums) but we will recall their derivatives.

Reminder: The following derivatives are well-known and may come in handy.

1. If $f(x) = x^n$ with $n \in \mathbb{Z}$, we have $f'(x) = nx^{n-1}$.
2. If $f(x) = \sin(x)$, we have $f'(x) = \cos(x)$.
3. If $f(x) = \cos(x)$, we have $f'(x) = -\sin(x)$.
4. If $f(x) = \tan(x)$, we have $f'(x) = \sec^2(x)$.
5. If $f(x) = e^x$, we have $f'(x) = e^x$.
6. If $f(x) = \log(x)$, we have $f'(x) = \frac{1}{x}$, where log is the natural logarithm.

The last thing we may want is the an answer to this question: if f is differentiable, then treating its derivative f' as a function in its own right, can we differentiate f'? In other words, can we differentiate f twice, three times, four times, or even more?

Definition 1.41 Let f be a function where taking higher derivatives makes sense.

(i) The second derivative of f is the function $f''(x) = (f')'(x)$.

(ii) The third derivative of f is the function $f'''(x) = (f'')'(x)$.

(iii) The k^{th} derivative of f is the function $f^{(k)}(x) = (f^{(k-1)})'(x)$.

Before now, we only considered *first-order derivatives*, that is differentiating a function once. There was an important phrase used in Definition 1.41 however, namely "where taking higher derivatives makes sense". It might happen that f is a perfectly well-behaved differentiable function, but f' is some nasty thing! However, there are some cases where the higher-order derivatives behave better.

Definition 1.42 Let $D \subseteq \mathbb{R}$ and $f : D \to \mathbb{R}$ be differentiable.

(i) We call f continuously differentiable if f' is continuous.

(ii) We call f k-times continuously differentiable if $f', f'', ..., f^{(k)}$ are all continuous.

(iii) We call f smooth if every derivative $f', f'', f''', ...$ exists.

Notice there is no mention of continuity in the definition of smooth; this is because Proposition 1.38 tells us that any differentiable function is automatically continuous. Because all derivatives exist, they must therefore all be continuous.

Notation 1.43 We use the following shorthand notation for the different types of function:

$$\begin{aligned} C^0(D) &= \text{the class of continuous functions on } D, \\ C^1(D) &= \text{the class of continuously differentiable functions on } D, \\ C^k(D) &= \text{the class of } k\text{-times continuously differentiable functions on } D, \\ C^\infty(D) &= \text{the class of smooth functions on } D. \end{aligned}$$

Note: If we don't need to reference the function's domain, we can just write C^k or C^∞.

We conclude the story of derivatives and instead look at the other side of this particular coin. It stands to reason that if we know the derivative of id is 1 (from Example 1.37), then can we start with the constant function 1 and work backwards to get id? This is the idea of an *integral*.

Definition 1.44 Let $f : [a, b] \to \mathbb{R}$ be a continuous function defined on a closed interval. The integral of f over $[a, b]$ is the limit of the area of the rectangles that split the interval $[a, b]$ and fit just under/over the graph of the function f. Denote this area (number) by

$$\int_a^b f(x)\,\mathrm{d}x.$$

We call f the integrand and $\mathrm{d}x$ the differential of x. In the case that this area exists, we say that f is integrable on $[a, b]$. We call a and b the (lower and upper) limits of integration.

This is not even close to being rigorous; although it is true the integral can be interpreted as telling you about the area under a graph and what not, it isn't the best practice to define the integral by this property. But let's continue down this awkward line of thought since the true details are not important to us here.

Note: Let f be integrable on the interval $[a, b]$. Then, we define the following relation:

$$\int_b^a f(x)\,\mathrm{d}x := -\int_a^b f(x)\,\mathrm{d}x.$$

This tells us we can swap the limits of integration but that means we get a minus sign.

Proposition 1.45 *Let f and g be integrable on the interval $[a, b]$. The following are true:*

(i) $\displaystyle\int_a^b \alpha f(x) + \beta g(x)\,\mathrm{d}x = \alpha \int_a^b f(x)\,\mathrm{d}x + \beta \int_a^b g(x)\,\mathrm{d}x$ *for any* $\alpha, \beta \in \mathbb{R}$.

(ii) $\displaystyle\int_a^b f(x)\,\mathrm{d}x = \int_a^k f(x)\,\mathrm{d}x + \int_k^b f(x)\,\mathrm{d}x$ *for any* $k \in [a, b]$.

Proof: Omitted. □

In fact, the set of all functions that are integrable on the interval $[a, b]$ is denoted $L([a, b])$ and is a special type of structure called a *vector space* (over $\mathbb{R}$). Even more than that, we can define a function which takes as an **input** a function f and spits out $\int_a^b f(x)\,\mathrm{d}x$; this is a *linear map* of the form $L([a, b]) \to \mathbb{R}$. Don't worry if this is too much because we absolutely don't need it (but if any of these words are familiar to you already, then we are still in semi-recognisable territory).

Definition 1.46 Let $f : [a, b] \to \mathbb{R}$ be continuous. An anti-derivative of f (or primitive) is a function $F : [a, b] \to \mathbb{R}$ such that $F'(x) = f(x)$ for all $x \in (a, b)$, that is a function whose derivative everywhere **except** the endpoints of the interval is the same as f.

Lemma 1.47 *Let $f : [a, b] \to \mathbb{R}$ be continuous and F and G be two anti-derivatives of f. Then, their difference $G - F$ is a constant function.*

Proof: This is another typical strategy: we will show that its derivative is zero (because this implies the original thing we differentiated is a constant; spoilers for Problem **1.21** below). Well,

$$(G - F)'(x) = G'(x) - F'(x) = f(x) - f(x) = 0,$$

by the definition of an anti-derivative. Hence, we know that $G - F$ is constant. □

Note: For any continuous function of the form $f : [a, b] \to \mathbb{R}$, there exists a primitive F of f. The proof isn't that hard to sketch but it really isn't required for us to think about it.

Theorem 1.48 (Fundamental Theorem of Calculus) *Let $f : [a, b] \to \mathbb{R}$ be continuous and F be* **any** *anti-derivative of f. Then, the integral of f in terms of the anti-derivative is*

$$\int_a^b f(x)\,\mathrm{d}x = F(b) - F(a).$$

Proof: Omitted. □

Example 1.49 Consider the function $f : [-2, 3] \to \mathbb{R}$ given by $f(x) = x^3$. An anti-derivative is something which differentiates to yield x^3. From what we know about polynomials, a good 'guess' is the function $x \mapsto x^4$. However, differentiating this gives us $4x^3$, so it is *almost* f but with the wrong coefficient; dividing by four will give the result. Hence, $F(x) = \frac{1}{4}x^4$ is an anti-derivative. By the Fundamental Theorem of Calculus, we conclude that that

$$\int_{-2}^3 f(x)\,\mathrm{d}x = F(3) - F(-2) = \frac{81}{4} - \frac{16}{4} = \frac{65}{4}.$$

Notation 1.50 When computing a definite integral, we use the notation $[F]_a^b := F(b) - F(a)$.

Thus far, the integrals we have discussed have been so-called *definite integrals*; this is where we integrate on an interval and the result is a number. However, we now develop the notion of an *indefinite integral* and formulate a way to find an anti-derivative of a given function. Be warned: there is no "one size fits all" method!

Reminder: The derivative of the constant function is zero (again, Problem **1.21** spoilers).

Definition 1.51 Let $f : [a, b] \to \mathbb{R}$ be continuous. The indefinite integral of f is the family of anti-derivatives F of f plus a constant, namely $F(x) + C$ for $C \in \mathbb{R}$. This is denoted

$$\int f(x)\,\mathrm{d}x.$$

The point is that $\int f(x)\,\mathrm{d}x$ is alternate notation for an anti-derivative of f, but we can add any constant we want to it; when we differentiate, the anti-derivative differentiates to f and the constant to zero, so the end result is indeed f.

Note: In practice, we often only include $+\,C$ at the end of indefinite integral calculations.

Example 1.52 Consider the function $f : [0, 2\pi] \to \mathbb{R}$ given by $f(x) = \cos(x)$. We can see from a previous reminder that an anti-derivative is $F(x) = \sin(x)$. However, if we consider the function $G(x) = \sin(x) + 2023$, this is also an anti-derivative. Therefore, the family of all anti-derivatives is $\int \cos(x)\,\mathrm{d}x = \sin(x) + C$.

Method – Integration by Inspection: Let f be a continuous function with anti-derivative F and g any function with continuous derivative (i.e. class C^1). The Chain Rule implies

$$F'(g(x)) = F'(g(x))g'(x) = f(g(x))g'(x),$$

because F is an anti-derivative of f. The Fundamental Theorem of Calculus implies

$$F(g(x)) = \int F'(g(x))\,\mathrm{d}x = \int f(g(x))g'(x)\,\mathrm{d}x.$$

Example 1.53 Suppose we wish to compute the (indefinite) integral of $h(x) = \sin(2x)$; we can write this as the composition $f(g(x))$ where $f(x) = \sin(x)$ and $g(x) = 2x$. Using Integration by Inspection, an anti-derivative of f is $F(x) = -\cos(x)$. The last line of the method says

$$F(g(x)) = \int f(g(x))g'(x)\,\mathrm{d}x = \int 2f(g(x))\,\mathrm{d}x,$$

using $g'(x) = 2$. Because we know that $F(g(x)) = -\cos(2x)$, we can obtain the integral we want by dividing both sides by two (which is allowed by Proposition 1.45). Consequently, we get

$$\int h(x)\,\mathrm{d}x = -\frac{1}{2}\cos(2x) + C.$$

Note: The moral of the story in Example 1.53 is to notice the function we are integrating 'looks like' sin which has known anti-derivative $-\cos$. If we differentiate $x \mapsto -\cos(2x)$, we get $2\sin(2x)$ which is **almost** what we want but not quite; we must divide by two.

The above method is a specific example of a more general *integration by substitution* approach. We will write this method out in full; it is very useful and the idea is to re-label part of the integrand so that the whole thing becomes simpler upon differentiating the re-labelled part.

Method – Integration by Substitution: Let $f(x)$ be a continuous function with anti-derivative F. The idea is to express the variable x as a function of a new variable u, that is $x = x(u)$, such that the function $x(u)$ is either strictly increasing or strictly decreasing. If $x \in [a, b]$ is the original domain, then this corresponds to the new domain $u \in [\alpha, \beta]$. Now, the inverse of $x(u)$ is $u(x)$, that is u as a function of x. The Chain Rule implies

$$F'(x(u)) = F'(x(u))x'(u) = f(x(u))x'(u),$$

by definition of F being a primitive of f. The Fundamental Theorem of Calculus implies

$$\int_\alpha^\beta F'(x(u))\,\mathrm{d}u = \int_\alpha^\beta f(x(u))x'(u)\,\mathrm{d}u = F(x(\beta)) - F(x(\alpha)) = F(b) - F(a) = \int_a^b f(x)\,\mathrm{d}x.$$

If we look again at Proposition 1.45, there is **no** mention on how to approach the integral of a product of two functions. In a general setting, we have the following non-equality:

$$\int f(x)g(x)\,\mathrm{d}x \neq \left(\int f(x)\,\mathrm{d}x\right)\left(\int g(x)\,\mathrm{d}x\right).$$

So is all hope lost? Of course not! There is in fact a method to deal with such situations.

Method – Integration by Parts: Let f and g be functions with continuous derivatives (i.e. class C^1). Recall the Product Rule of Differentiation says $(fg)'(x) = f'(x)g(x) + f(x)g'(x)$. The Fundamental Theorem of Calculus implies

$$f(x)g(x) = \int f'(x)g(x) + f(x)g'(x)\,\mathrm{d}x.$$

If we split the integral into two parts, then the above rearranges to the useful formula

$$\int f(x)g'(x)\,\mathrm{d}x = f(x)g(x) - \int f'(x)g(x)\,\mathrm{d}x.$$

Reminder: Recall we have the following trigonometric values in radians (**not** degrees).

	0	$\frac{\pi}{6}$	$\frac{\pi}{4}$	$\frac{\pi}{3}$	$\frac{\pi}{2}$	$\frac{2\pi}{3}$	π	$\frac{3\pi}{2}$	2π
$\sin(x)$	0	$\frac{1}{2}$	$\frac{\sqrt{2}}{2}$	$\frac{\sqrt{3}}{2}$	1	$\frac{\sqrt{3}}{2}$	0	-1	0
$\cos(x)$	1	$\frac{\sqrt{3}}{2}$	$\frac{\sqrt{2}}{2}$	$\frac{1}{2}$	0	$-\frac{1}{2}$	-1	0	1
$\tan(x)$	0	$\frac{\sqrt{3}}{3}$	1	$\sqrt{3}$	**N/A**	$-\sqrt{3}$	0	**N/A**	0

Table 1: Some useful trigonometric values to know.

Example 1.54 Suppose we want to compute $\int_0^\pi x\cos(x)\,\mathrm{d}x$. For the moment, we shall ignore the limits of integration and focus on finding an anti-derivative of $x\cos(x)$ (so that we can then apply the Fundamental Theorem of Calculus to get an exact number at the end). Indeed, we just want $\int x\cos(x)\,\mathrm{d}x$ at the moment. Using the labels from the above method, we see that

$$\begin{aligned} f(x) &= x & &\Rightarrow & f'(x) &= 1, \\ g'(x) &= \cos(x) & &\Rightarrow & g(x) &= \sin(x). \end{aligned}$$

The hardest part thus far is finding an anti-derivative to $\cos(x)$, but we know that $\sin(x)$ fits the bill. Substituting all this into the integration by parts formula, we get

$$\begin{aligned} \int x\cos(x)\,\mathrm{d}x &= x\sin(x) - \int 1\sin(x)\,\mathrm{d}x \\ &= x\sin(x) + \cos(x) + C. \end{aligned}$$

This is the **indefinite** integral, but our question has limits. Thus, we choose **any** anti-derivative (for ease, $F(x) = x\sin(x) + \cos(x)$ where $C = 0$) and use the Fundamental Theorem of Calculus:

$$\int_0^\pi x\cos(x)\,\mathrm{d}x = F(\pi) - F(0) = -1 - 1 = -2.$$

There is one more rule to mention, but we first introduce a function which comes up everywhere.

Definition 1.55 The absolute value is the function $|\cdot| : \mathbb{R} \to \mathbb{R}$ defined by the formula

$$|x| = \begin{cases} x & \text{if } x \geq 0 \\ -x & \text{if } x < 0 \end{cases}.$$

In words, the absolute value will always spit out a non-negative number; it will either output what you put in (if $x \geq 0$) or it will multiply it by -1 and cancel the minus sign and thus output the positive version of what you put in (if $x < 0$). For example, $|2| = 2$ and $|-6| = 6$.

Note: A quicker way to write the absolute value function is $|x| = \sqrt{x^2}$. If this is a bit confusing, don't worry. But the moral here is that there is a subtle difference between finding a square root (a single non-negative number) and solving an equation like $x^2 = k$.

Lemma 1.56 *Let $f : [a,b] \to \mathbb{R}$ be continuous with $f(x) \neq 0$ for any $x \in [a,b]$. Then,*

$$\int \frac{f'(x)}{f(x)} \, \mathrm{d}x = \log\big(|f(x)|\big) + C.$$

Sketch of Proof: Consider the function $g(x) = \log\big(f(x)\big)$. If we use the Chain Rule, we obtain

$$g'(x) = \log'(f(x))f'(x) = \frac{1}{f(x)}f'(x).$$

Taking the integral of this equation (and using the fact that $\int \log(x)\,\mathrm{d}x = \frac{1}{x}$ which you can prove using integration by parts of all things; more spoilers this time for Problem **1.27**), we see that

$$\int \frac{f'(x)}{f(x)} \, \mathrm{d}x = \int g'(x)\,\mathrm{d}x = \log\big(f(x)\big) + C.$$

The absolute value ensures the inputs into the logarithm are positive as its domain is $(0, \infty)$. □

Reminder: The following indefinite integrals are well-known and may come in handy.

1. If $f(x) = x^n$ with $n \in \mathbb{Z} \setminus \{-1\}$, we have $\displaystyle\int f(x)\,\mathrm{d}x = \frac{1}{n+1}x^{n+1} + C$.
2. If $f(x) = x^{-1}$, we have $\displaystyle\int f(x)\,\mathrm{d}x = \log\big(|x|\big) + C$.
3. If $f(x) = \sin(x)$, we have $\displaystyle\int f(x)\,\mathrm{d}x = -\cos(x) + C$.
4. If $f(x) = \cos(x)$, we have $\displaystyle\int f(x)\,\mathrm{d}x = \sin(x) + C$.
5. If $f(x) = \tan(x)$, we have $\displaystyle\int f(x)\,\mathrm{d}x = -\log\big(|\cos(x)|\big) + C$.
6. If $f(x) = e^x$, we have $\displaystyle\int f(x)\,\mathrm{d}x = e^x + C$.

We have now reached the conclusion of the start of the journey. In the words of Obi-Wan Kenobi, you have taken your first step into a larger world. However, this is just the tip of the iceberg!

1.5 Matrices

We will move away to something more algebraic for the moment. Although we won't need this stuff until we look at surfaces in Section 3, it is a good idea to introduce it gently here. Fortunately, there isn't much hard mathematics at play (although some of the ideas run deep and can be far-reaching); this is really a place to introduce some notation and new words.

Definition 1.57 An $m \times n$ matrix A is a grid with m rows and n columns of the form

$$A = \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{pmatrix}.$$

We denote by a_{ij} the number in the i^{th} row and j^{th} column. For shorthand, $A = (a_{ij})$.

Note: The set of matrices whose entries $a_{ij} \in \mathbb{R}$ are all real numbers is denoted $\mathbb{M}_{m\times n}(\mathbb{R})$.

It turns out that we can add and multiply matrices together, but only in certain circumstances:

- **Matrix Addition**: this is well-defined if we have two matrices of the **same size**, say $A = (a_{ij})$ and $B = (b_{ij})$ that are both $m \times n$. Assuming this, the sum of the matrices is given by

$$A + B = (a_{ij} + b_{ij}),$$

  meaning we simply add the corresponding entries of each matrix.

- **Matrix Multiplication**: this is well-defined if the number of **columns** of the **first** matrix is equal to the number of **rows** of the **second** matrix, say $A = (a_{ik})$ and $B = (b_{kj})$ that are $m \times p$ and $p \times n$. Assuming this, the product of the matrices is the $m \times n$ matrix given by

$$AB = (c_{ij}), \qquad \text{where } c_{ij} = \sum_{k=1}^{n} a_{ik} b_{kj},$$

  meaning we add products between entries of the i^{th} row of A and the j^{th} column of B.

- **Scalar Multiplication**: the scalar multiple of a matrix $A = (a_{ij})$ by some $k \in \mathbb{R}$ is given by

$$kA = (ka_{ij}),$$

  meaning we simply multiply each entry of A by the number k.

Reminder: An operation $*$ on a set X is commutative if $x * y = y * x$ for all $x, y \in X$.

We know that multiplication of real numbers is commutative (it doesn't matter what order we do it in), but can the same be said about matrices? Absolutely **not**! Firstly, if the matrix product AB is well-defined, there is no guarantee that BA is even possible using the above definition. Secondly, even **if** AB and BA are both well-defined, they can be unequal.

Example 1.58 Let's first consider the matrices

$$A = \begin{pmatrix} 1 & 0 & 1 \\ 2 & -1 & 4 \end{pmatrix} \qquad \text{and} \qquad B = \begin{pmatrix} 2 & 9 & 5 \\ 0 & 0 & 1 \\ -1 & 8 & 8 \end{pmatrix}.$$

These are 2×3 and 3×3, respectively. In particular, we see that AB is well-defined, with

$$AB = \begin{pmatrix} 1 & 17 & 13 \\ 0 & 50 & -23 \end{pmatrix}.$$

As for BA, note that the number of columns of B does **not** equal the number of rows of A $(3 \neq 2)$, so this isn't even well-defined. On the other hand, consider the following matrices:

$$C = \begin{pmatrix} 1 & 2 \\ 2 & 3 \end{pmatrix} \qquad \text{and} \qquad D = \begin{pmatrix} 3 & 4 \\ 0 & 1 \end{pmatrix}.$$

These are each 2×2 matrices, meaning both of CD and DC are well-defined. However, we have

$$CD = \begin{pmatrix} 3 & 6 \\ 6 & 11 \end{pmatrix} \neq \begin{pmatrix} 11 & 18 \\ 2 & 3 \end{pmatrix} = DC.$$

Definition 1.59 A square matrix $A = (a_{ij})$ is an $n \times n$ matrix, that is where $m = n$. The main diagonal of a square matrix is the diagonal consisting of a_{ii} for all $i = 1, ..., n$.

Two of the main objects we wish to introduce are defined for square matrices, hence the need for Definition 1.59. Before we make these definitions that we want to make, it would be remiss of us not to mention, albeit briefly, one of the most important square matrices: the so-called *identity matrix*. This has zeros everywhere **except** for a main diagonal that is filled by ones.

Note: The $n \times n$ identity matrix I_n has $a_{ii} = 1$ for all $i = 1, ..., n$ and $a_{ij} = 0$ for all $i \neq j$:

$$I_n = \begin{pmatrix} 1 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 1 \end{pmatrix}.$$

We think of I_n as the *multiplicative unit* of $\mathbb{M}_{n\times n}(\mathbb{R})$, meaning multiplication by the identity does nothing. The proof of this is not too challenging if we use the definition of matrix multiplication.

Lemma 1.60 *Multiplication by the identity matrix I_n does nothing to the starting matrix.*

Proof: Let $A = (a_{ik}) \in \mathbb{M}_{n\times n}(\mathbb{R})$ be an arbitrary square matrix and denote the entries of the identity matrix by b_{kj}. By definition, the matrix product $AI_n = (c_{ij})$ where

$$c_{ij} = \sum_{k=1}^{n} a_{ik}b_{kj} = a_{ij}b_{jj} = a_{ij},$$

since $b_{kj} = 0$ for all $k \neq j$ and $b_{jj} = 1$ for all $j = 1, ..., n$. Similarly, $I_nA = (d_{ij})$ where

$$d_{ij} = \sum_{k=1}^{n} b_{ik}a_{kj} = b_{ii}a_{ij} = a_{ij}$$

for identical reasoning as above: $b_{ik} = 0$ for all $i \neq k$ and $b_{ii} = 1$ for all $i = 1, ..., n$. □

Remark 1.61 We can be a bit more general than Lemma 1.60: if $A \in \mathbb{M}_{m\times n}(\mathbb{R})$ is a **non-square** matrix, the matrix products AI_n and I_mA are well-defined (note the different sizes of identity matrix). That said, we actually have straight-up equality of these two products, namely

$$AI_n = I_mA = A.$$

Definition 1.62 Let $A = (a_{ij}) \in \mathbb{M}_{n\times n}(\mathbb{R})$. The trace is the sum of its diagonal entries:

$$\mathrm{tr}(A) = a_{11} + a_{22} + \cdots + a_{nn}.$$

Example 1.63 Consider the square matrices B, C and D from Example 1.58. We can see that

$$\mathrm{tr}(B) = 2 + 0 + 8 = 10, \qquad \mathrm{tr}(C) = 1 + 3 = 4, \qquad \mathrm{tr}(D) = 3 + 1 = 4.$$

Note: It is known that $\operatorname{tr}(A+B) = \operatorname{tr}(A) + \operatorname{tr}(B)$ and $\operatorname{tr}(kA) = k\operatorname{tr}(A)$ for all $k \in \mathbb{R}$.

The final definition we make is specific to 2×2 matrices (be aware that a generalisation exists for all $n \times n$ matrices, but it is not in the scope of this module to properly discuss it here). We will give a brief geometric interpretation of this special case, however, which can naturally extend to higher dimensions in the more general setting.

Definition 1.64 The determinant of a 2×2 matrix $A = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$ is $\det(A) = ad - bc$.

Remark 1.65 Geometrically, a 2×2 matrix A 'acts' on a (column) vector by multiplication (if we view the vector as a 2×1 matrix, we see that this matrix multiplication is in fact well-defined). In this way, A acts on everything in the two-dimensional plane $\mathbb{R}^2$ by multiplication. In particular, this will map the unit square to a parallelogram, which has area precisely $\big|\det(A)\big|$. In this sense, the determinant is simply the scale factor by which areas are transformed by the matrix A.

Example 1.66 Consider the 2×2 matrices C and D from Example 1.58. We can see that

$$\det(C) = (1)(3) - (2)(2) = 3 - 4 = -1 \qquad \text{and} \qquad (3)(1) - (0)(4) = 3.$$

Note: It is known that $\det(AB) = \det(A)\det(B)$ and $\det(kA) = k^n \det(A)$ for all $k \in \mathbb{R}$, where n is the size of the matrix A. As said earlier, it will always be $n = 2$ in this module.

We conclude with an interesting relationship between the determinant and the trace of a matrix.

Proposition 1.67 *Let $A \in \mathbb{M}_{n\times n}(\mathbb{R})$ and $f : \mathbb{R} \to \mathbb{R}$ where $f(t) = \det(I_n + tA)$. Then,*

$$f'(0) = \operatorname{tr}(A).$$

Proof (of the 2×2 Case): Let A be as in Definition 1.64. Then, from the definition, we have

$$f(t) = \det\begin{pmatrix} 1+ta & tb \\ tc & 1+td \end{pmatrix} = (1+ta)(1+td) - (tb)(tc) = t^2(ad-bc) + t(a+d) + 1.$$

Taking the derivative, we get $f'(t) = 2t(ad - bc) + a + d$. Hence, $f'(0) = a + d = \operatorname{tr}(A)$. □

1.6 Problem Set 1

We now provide a number of problems to complement the notes in Section 1. Solutions can be found at the end of the notes, but do try to work things out as best as possible without looking the answers up. Of course, if you are very stuck, take a peak at the solutions and try to unravel what you were finding difficult.

Problems for Section 1.1

1.1. State true or false to this statement: $32 \in \{2, \{32\}, \{5\}, 5\}$.

1.2. State true or false to this statement: $64 \in \{\text{even integers}\}$.

1.3. State true or false to this statement: $\{4\} \in \{\pi, 100, \{4\}, 7\}$.

1.4. State true or false to this statement: $\emptyset = \{\text{prime numbers that are even}\}$.

1.5. (a) Write down all distinct subsets of the empty set $\emptyset$.

(b) How many distinct subsets are there in total?

1.6. (a) Write down all distinct subsets of the set $\{1\}$.

(b) How many distinct subsets are there in total?

1.7. (a) Write down all distinct subsets of the set $\{1, 2\}$.

(b) How many distinct subsets are there in total?

1.8. (a) Write down all distinct subsets of the set $\{1, 2, 3\}$.

(b) How many distinct subsets are there in total?

1.9. Using your answers above, suggest how many distinct subsets there are of the set $\{1, 2, ..., n\}$.

1.10. Let $X = \{2, 1\}$ and $Y = \{3, 6\}$ and $Z = \{8, 7\}$. Write out $X \times Y \times Z$ in full.

1.11. The *power set* of a set X is the set of all subsets, denoted $\mathcal{P}(X)$. Write out the power set of $X = \{a, \{b, c\}\}$ and determine its cardinality $\left|\mathcal{P}(X)\right|$. Does this agree with Problem **1.7**?

[**Hint:** The set X has two elements, namely a and $\{b, c\}$. Even though the latter is itself a set, it is still an element with respect to X. So, you can give it an easier label, say $d := \{b, c\}$, so that the set X is just $\{a, d\}$.]

Problems for Section 1.2

1.12. Write down the range of the function $f : \mathbb{N} \to \mathbb{N}$ given by $f(n) = n^3$.

1.13. Write down the range of the function $f : \mathbb{R} \to \mathbb{R}$ given by $f(n) = n^3$.

1.14. Consider the function $p : \mathbb{Z}^+ \to \mathbb{N}$ where $p(n)$ is the number of distinct prime factors of n and recall $0 \in \mathbb{N}$ by Definition 1.1. Explain why we require 0 to be in the co-domain of p.

1.15. Recall that $\mathbb{R}^2 = \mathbb{R}\times\mathbb{R}$ and consider the function $\varphi : \mathbb{R}^2 \to \mathbb{R}^2$ given by $\varphi(x, y) = (x-y, xy)$.

(a) Determine if φ is injective.

(b) Determine if φ is surjective.

1.16. Determine which of the following functions are injective and which aren't.

(a) $\begin{array}{cccc} f & : & \mathbb{R} & \longrightarrow & \mathbb{R} \\ & & x & \longmapsto & x^2 \end{array}$.

(b) $\begin{array}{cccc} g & : & \mathbb{Q} & \longrightarrow & \mathbb{Q} \\ & & x & \longmapsto & x^2 \end{array}$.

(c) $\begin{array}{cccc} h & : & \mathbb{Z} & \longrightarrow & \mathbb{Z} \\ & & x & \longmapsto & x^2 \end{array}$

(d) $\begin{array}{cccc} k & : & \mathbb{N} & \longrightarrow & \mathbb{N} \\ & & x & \longmapsto & x^2 \end{array}$

(e) $\begin{array}{cccc} s & : & \mathbb{N} & \longrightarrow & \mathbb{Z} \\ & & x & \longmapsto & x^2 \end{array}$

(f) $\begin{array}{cccc} t & : & \mathbb{Z} & \longrightarrow & \mathbb{N} \\ & & x & \longmapsto & x^2 \end{array}$

[**Hint:** The injectivity of a function not only depends on the rule $x \mapsto \cdots$ but its domain also.]

Problems for Section 1.3

1.17. Use the Algebra of Limits to compute $\lim_{x\to 2} \dfrac{x^2+3x-4}{x^3-7}$.

1.18. Use the Algebra of Limits to compute $\lim_{x\to 6}(x+2)^4$.

1.19. Use the Algebra of Limits to compute $\lim_{x\to\pi}(x-\pi)\dfrac{\sin(x)+\tan\big(x^2+3\big)\cos(6)}{x^{-8}+e^{x\tan(x)}-2}$.

Problems for Section 1.4

1.20. Prove that the function $f : \mathbb{R} \to \mathbb{R}$ given by $f(x) = x^2$ is differentiable.

1.21. (a) Prove that the function $g : \mathbb{R} \to \mathbb{R}$ given by $g(x) = 4$ is differentiable.

(b) Can you adapt your proof if we replace 4 by any real number $k \in \mathbb{R}$?

1.22. Using Proposition 1.39, find the first-order derivatives of the following functions.

(a) $a : \mathbb{R} \to \mathbb{R}$ given by $a(x) = x^2 + 5x + 6$.

(b) $b : \mathbb{R} \to \mathbb{R}$ given by $b(x) = x\sin(x)$.

(c) $c : \mathbb{R} \to \mathbb{R}$ given by $c(x) = \cos\big(x^2\big)$.

(d) $d : \mathbb{R} \setminus \{0\} \to \mathbb{R}$ given by $d(x) = \dfrac{e^x}{2x^5}$.

1.23. Find the derivative of $f : \mathbb{R} \to \mathbb{R}$ given by $f(x) = x^2\cos\big(x^3\big)$ by using the Chain Rule and Product Rule of Differentiation.

1.24. Give one example of each of the following.

(a) A smooth function.

(b) A class C^0 function which is **not** of class C^1 (or higher).

(c) A class C^1 function which is **not** of class C^2 (or higher).

1.25. Compute the integral $\displaystyle\int_{-1}^{3} x^2 \, \mathrm{d}x$ by using the Fundamental Theorem of Calculus.

1.26. Compute the following definite integrals using inspection.

(a) $\displaystyle\int_0^2 e^{kx} \, \mathrm{d}x$ where $k \in \mathbb{R}$.

(b) $\displaystyle\int_{-\pi}^{\frac{\pi}{2}} \cos(kx) \, \mathrm{d}x$ where $k \in \mathbb{R}$.

1.27. Compute the following indefinite integrals using integration by parts.

(a) $\displaystyle\int x^2 \sin(x) \, \mathrm{d}x$.

[**Hint:** You will have to do integration by parts twice here; once attacking the problem head-on and then a second time for the integral that you get in the by-parts formula you have just written out!]

(b) $\displaystyle\int \log(x) \, \mathrm{d}x$.

[**Hint:** Use the standard trick of viewing the integrand as the product $\log(x) = \log(x) \cdot 1$.]

1.28. Compute the integral $\displaystyle\int_0^4 \frac{3x^2 + 4}{x^3 + 4x + 2} \, \mathrm{d}x$.

Problems for Section 1.5

1.29. Consider the following collection of matrices:

$$A = \begin{pmatrix} 1 & 2 \\ 0 & 0 \\ 4 & 3 \end{pmatrix}, \qquad B = \begin{pmatrix} 7 & 0 \\ 6 & 3 \\ -1 & 1 \end{pmatrix}, \qquad C = \begin{pmatrix} 0 & -9 & 10 \\ 2 & 5 & 2 \end{pmatrix},$$
$$D = \begin{pmatrix} 2 & 4 \\ -1 & -6 \end{pmatrix}, \qquad E = \begin{pmatrix} -4 & 2 & 5 \\ 4 & 5 & 11 \end{pmatrix}.$$

Evaluate the following if they exist. If they do not exist, explain why not:

(a) $A + B$.

(b) $A - B$.

(c) $E + C$.

(d) $D - E$.

(e) AB.

(f) DC.

(g) AE.

(h) EA.

1.30. Consider the following collection of 2×2 matrices:

$$A = \begin{pmatrix} 3 & 1 \\ 4 & 3 \end{pmatrix}, \qquad B = \begin{pmatrix} -2 & 2 \\ -7 & -13 \end{pmatrix}, \qquad C = \begin{pmatrix} 1 & 0 \\ 5 & 7 \end{pmatrix}.$$

Evaluate the following:

(a) $\det(A)$.

(b) $\det(B)$.

(c) $\det(3C)$.

(d) $\operatorname{tr}(A)$.

(e) $\operatorname{tr}(B)$.

(f) $\operatorname{tr}(A + C)$.

1.31. Let A be an arbitrary 2×2 matrix.

(a) If B is the matrix obtained by swapping the **rows** of A, prove $\det(B) = -\det(A)$.

(b) If C is the matrix obtained by swapping the **columns** of A, prove $\det(C) = -\det(A)$.

[**Hint:** Write A out as in Definition 1.64, from which you can write B and C and the determinants thereafter.]

1.32. For any $A, B \in \mathbb{M}_{2\times 2}(\mathbb{R})$, does it follow that $\det(AB) = \det(BA)$? Justify your answer.

1.33. Let $f : \mathbb{R} \to \mathbb{R}$ be defined as $f(t) = \operatorname{tr}(A + tI_n)$ for any $A \in \mathbb{M}_{n\times n}(\mathbb{R})$. Prove that $f'(t) = n$.

2 Parametrised Curves

We are now ready to begin discussing what is promised by the title of this module. The first place we begin is one-dimensional (literally, not figuratively; there is a lot to say): we will discuss curves that lie on a plane or in space. This should be easy for us to visualise by the end.

2.1 General Curves

We start by discussing a more general situation, where the curves we consider live in **any** number of dimensions. This just amounts to thinking about tuples of real numbers (or *real vectors*).

Reminder: A function is smooth if all its derivatives exist at all points in its domain.

Definition 2.1 Let $I \subseteq \mathbb{R}$ be an open interval. A parametrised curve is a smooth map $\gamma : I \to \mathbb{R}^n$ and we label the input variables as t (think of these as 'times'). Now let $t \in I$.

(i) We call t a regular point if $\gamma'(t) \neq \mathbf{0}$.

(ii) We call t a singular point if $\gamma'(t) = \mathbf{0}$.

If **every** $t \in I$ is a regular point, then we call γ a regularly parametrised curve.

Notation 2.2 Be aware that $\gamma(t)$ and $\gamma'(t)$ are vectors in (elements of) $\mathbb{R}^n$, so we can alternatively write (i) $\gamma'(t) \neq (0, ..., 0)$ or (ii) $\gamma'(t) = (0, ..., 0)$ in Definition 2.1 above. For shorthand, we use bold letters for vectors. This is standard in algebra but not often in differential geometry.

In other words, a parametrised curve is a regularly parametrised curve (is regular) if and only if there does **not** exist $t \in I$ such that $\gamma'(t) = \mathbf{0}$. Using this, it is now quite easy to tell if a given parametrised curve is regular. We will give this a shot now and leave an exercise for you.

Example 2.3 Consider the curve $\gamma : \mathbb{R} \to \mathbb{R}^2$ given by $\gamma(t) = (t, t^2)$. Its derivative $\gamma'(t) = (1, 2t)$ is found by differentiating each function in the tuple (which is just a pair in this case). But notice that $\gamma'(t) = (0, 0)$ if and only if $0 = 1$; this is impossible, so we know there does **not** exist $t \in \mathbb{R}$ making $\gamma'(t) = \mathbf{0}$ and thus it is a regular curve!

Exercise 1 Determine if the curve $\gamma : \mathbb{R} \to \mathbb{R}^2$ given by $\gamma(t) = (t^3, t^6)$ is regular.

Why the naming in Definition 2.1? Well, a parametrised curve might be a smooth map (function) but it may **not** represent a "smooth" curve in general; it may be that drawing out the curve has some sort of nasty point (e.g. a cusp).

Note: A regularly parametrised curve does represent a "smooth" curve (no nasty points).

Example 2.4 Consider the smooth map $\gamma : \mathbb{R} \to \mathbb{R}^2$ given by $\gamma(t) = (t^2, t^3)$; this is a parametrised curve. However, we see $\gamma'(t) = (2t, 3t^2)$ and therefore $t = 0$ is a singular point since $\gamma'(0) = (0, 0)$. If we sketch this curve in $\mathbb{R}^2$ as we do below, we see that there is a cusp at the point $(0, 0)$.

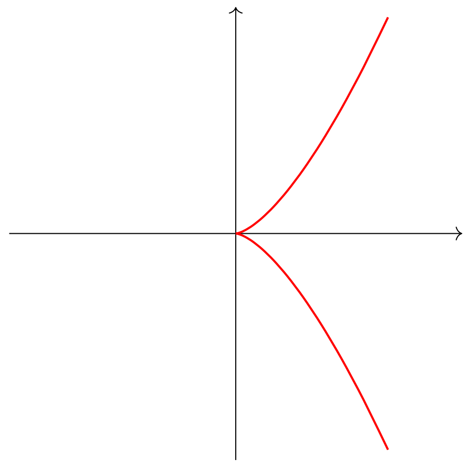


Figure 3: The parametrised curve $\gamma(t) = (t^2, t^3)$.

Reminder: The Euclidean norm of $\mathbf{v} = (v_1, v_2, ..., v_n) \in \mathbb{R}^n$ is the non-negative number

$$\|\mathbf{v}\| := \sqrt{v_1^2 + v_2^2 + \cdots + v_n^2}.$$

Throughout, think of this analogy: a point is a particle which has position $\gamma(t)$ at a particular time $t \in I$. Therefore, if we vary t, we think of our particle moving around. In keeping with this theme, we define the following ideas that are influenced directly from classical mechanics.

Definition 2.5 Let $I \subseteq \mathbb{R}$ be an interval and $\gamma : I \to \mathbb{R}^n$ be a parametrised curve.

(i) The velocity of the curve is $\gamma' : I \to \mathbb{R}^n$.

(ii) The speed of the curve is $\|\gamma'\| : I \to [0, \infty)$.

(iii) The acceleration of the curve is $\gamma'' : I \to \mathbb{R}^n$.

Using this language, we can again re-phrase what it means for a curve to be regular. Indeed, a parametrised curve is a regularly parametrised curve if and only if its velocity (or speed) **never** vanishes.

Example 2.6 Let $\mathbf{x}, \mathbf{v} \in \mathbb{R}^n$ be fixed vectors. Then, the map $\gamma : \mathbb{R} \to \mathbb{R}^n$ given by $\gamma(t) = \mathbf{x} + t\mathbf{v}$ is a parametrised curve. In fact, it is a line which passes through $\mathbf{x}$ and travels in the direction of $\mathbf{v}$. Notice that $\gamma'(t) = \mathbf{v}$ is constant; it is regular if and only if $\mathbf{v} \neq \mathbf{0}$.

Recall that a *tangent line* to a curve at a point is a straight line which intersects the curve only at that point. It turns out that every regularly parametrised curve has a well-defined tangent line for each $t \in I$.

Definition 2.7 Let $\gamma : I \to \mathbb{R}^n$ be a regularly parametrised curve. The tangent line at $t_0 \in I$ is the parametrised curve $\widehat{\gamma}_{t_0} : \mathbb{R} \to \mathbb{R}^n$ given by $\widehat{\gamma}_{t_0}(t) = \gamma(t_0) + t\gamma'(t_0)$.

Exercise 2 Prove that the tangent line $\widehat{\gamma}_{t_0}$ to a regular curve γ at any t_0 is itself regular.

Example 2.8 Consider the parametrised curve $\gamma : \mathbb{R} \to \mathbb{R}^2$ given by $\gamma(t) = (t^3 - t, t^2 - 1)$. Note that $\gamma'(t) = (3t^2 - 1, 2t) = \mathbf{0}$ if and only if $3t^2 = 1$ and $2t = 0$; these are incompatible so there is **no** $t \in \mathbb{R}$ such that its velocity vanishes and thus this curve is regular. Its tangent lines are

$$\widehat{\gamma}_{t_0}(t) = (t_0^3 - t_0, t_0^2 - 1) + t(3t_0^2 - 1, 2t_0).$$

This is of the form of the curve in Example 2.6; the direction of the line is $(3t_0^2 - 1, 2t_0)$. We can determine precisely where the tangent lines are vertical as follows: verticality means that the first coordinate (which is the horizontal component) is zero, which means

$$3t_0^2 - 1 = 0 \qquad \Rightarrow \qquad t_0 = \pm\frac{\sqrt{3}}{3}.$$

Consequently, there are precisely two vertical tangent lines: $\widehat{\gamma}_{\frac{\sqrt{3}}{3}}(t)$ and $\widehat{\gamma}_{-\frac{\sqrt{3}}{3}}(t)$.

Definition 2.9 Let $\gamma : I \to \mathbb{R}^n$ be a parametrised curve. A self-intersection point is one where $\gamma(t_1) = \gamma(t_2)$ but $t_1 \neq t_2$, that is the same point occurs at two different times.

Example 2.10 (Revisited) We look at the same curve $\gamma : \mathbb{R} \to \mathbb{R}^2$ given by $\gamma(t) = (t^3 - t, t^2 - 1)$ from Example 2.8 above to find the self-intersection points. Indeed, $\gamma(t_1) = \gamma(t_2)$ is equivalent to $(t_1^3 - t_1, t_1^2 - 1) = (t_2^3 - t_2, t_2^2 - 1)$. We have the following by comparing second coordinates:

$$t_1^2 - 1 = t_2^2 - 1 \qquad \Rightarrow \qquad t_1^2 = t_2^2 \qquad \Rightarrow \qquad t_2 = -t_1.$$

Substituting these into the first coordinates, we see that

$$t_1^3 - t_1 = t_2^3 - t_2 \qquad \Rightarrow \qquad t_1^3 - t_1 = -(t_1^3 - t_1).$$

It is a general fact that the only solution (in the real numbers) to an equation of the form $x = -x$ is $x = 0$. Therefore, we conclude that $t_1^3 - t_1 = 0$; factorising the left-hand side results in $t_1(t_1^2 - 1) = 0$ and thus we get three possibilities:

(i) If $t_1 = 0$, then $t_2 = -t_1 = 0$ so this is **not** a self-intersection point (we must have $t_1 \neq t_2$).

(ii) If $t_1 = 1$, then $t_2 = -t_1 = -1$ and this **is** a self-intersection point, namely at $(0, 0)$.

(iii) If $t_1 = -1$, then $t_2 = -t_1 = 1$ and this **is** also a self-intersection point, the same as in (ii).

Exercise 3 Find any self-intersection points, if they in fact exist, of the parametrised curve $\gamma : \mathbb{R} \to \mathbb{R}^3$ given by $\gamma(t) = (t^2, 4t^3, t - 1)$.

We will now discuss the length along a segment of a curve, starting with motivation which is akin to what is found in a calculus course. Indeed, let $\gamma : I \to \mathbb{R}^n$ be a regularly parametrised curve and suppose we want to compute the length of γ starting from $t = t_0$ and ending at $t = t_n$. To do this, we split the interval $[t_0, t_n]$ into n pieces $[t_{k-1}, t_k]$ where $k = 1, 2, ..., n$ of equal length.

Note: The length of each of these intervals is $\delta t := \dfrac{t_n - t_0}{n}$, the 'mean' length of $[t_0, t_n]$.

We see that when we split our interval into a lot of chunks (when n is very large), then the quantity δt is very small (because we are dividing by a large number). Moreover, the length of the straight line segment between $\gamma(t_{k-1})$ and $\gamma(t_k)$ is given by

$$\delta s_k := \big\|\gamma(t_k) - \gamma(t_{k-1})\big\| = \big\|\gamma(t_{k-1} + \delta t) - \gamma(t_{k-1})\big\| \approx \big\|\gamma'(t_{k-1})\big\|\delta t,$$

where $\approx$ means approximately and this estimate comes from the definition of the derivative we had in Section 1.4, specifically the note after Definition 1.36. If we add up all of these straight lines, we get an estimate for the arc length via *piecewise straight line*, namely

$$s_n := \sum_{k=1}^{n} \delta s_k \approx \sum_{k=1}^{n} \big\|\gamma'(t_{k-1})\big\|\delta t.$$

Reminder: The integral of a function can be thought of as an infinite limit of a finite sum.

If we take the limit as $n \to \infty$, then $\delta t \to 0$ as we mentioned already and the piecewise straight line is exactly the curve γ in the limit. Per the reminder then, $\lim_{n\to\infty} s_n$ will be an integral.

Definition 2.11 The arc length from t_0 to t_1 along a regularly parametrised curve γ is

$$s := \int_{t_0}^{t_1} \big\|\gamma'(t)\big\| \, \mathrm{d}t.$$

Example 2.12 (Revisited) Consider again the regular curve $\gamma : \mathbb{R} \to \mathbb{R}^2$ where $\gamma(t) = (t^3 - t, t^2 - 1)$ and suppose we wish to compute the arc length from $t = 0$ to $t = 1$. Then, we have the following:

$$\begin{aligned} \gamma'(t) &= (3t^2 - 1, 2t), \\ \|\gamma'(t)\| &= \sqrt{(3t^2 - 1)^2 + (2t)^2} = \sqrt{9t^4 - 2t^2 + 1}, \\ s &= \int_0^1 \sqrt{9t^4 - 2t^2 + 1}\,\mathrm{d}t = 1.3577959.... \end{aligned}$$

The point here is that it is technically possible to compute arc lengths. However, the integral above has no closed expression and we had to use a computer to get a solution. Don't worry, any questions we include here will be possible using standard methods!

> **Exercise 4** Compute the arc length of the regularly parametrised curve $\gamma : (0, \infty) \to \mathbb{R}^2$ given by $\gamma(t) = (t, \frac{2}{3}t^{\frac{3}{2}})$ from $t = 3$ to $t = 15$.

Now that we understand the arc length, we can actually define a function which works as follows: choose and fix a *basepoint* $t_0 \in I$. Then, we will input times t into our function and it will output the arc length along γ from the basepoint t_0 to our input point t.

> **Definition 2.13** Let $\gamma : I \to \mathbb{R}^n$ be a regularly parametrised curve. The arc length function based at $t_0 \in I$ is the real function $\sigma_{t_0} : I \to \mathbb{R}$ defined by
>
> $$\sigma_{t_0}(t) = \int_{t_0}^{t} \|\gamma'(u)\|\,\mathrm{d}u.$$

Remark 2.14 In the definition of σ_{t_0}, the integral uses the variable u; this is a so-called *dummy variable*. What does this mean? Well, as standard we use the variable t like we did in Definition 2.11 but we **cannot** do so here because t is now a limit of integration. Therefore, we just picked u. We could call it any letter; it doesn't matter because the integral is definite (a real number) so no variable comes out the other side.

Example 2.15 Consider the regular curve $\gamma : (0, \infty) \to \mathbb{R}^2$ given by $\gamma(t) = (t, \frac{2}{3}t^{\frac{3}{2}})$ from Exercise 4 and choose $t_0 = 1$ as the basepoint. We should first find the integrand explicitly:

$$\gamma'(u) = (1, u^{\frac{1}{2}}) \qquad \Rightarrow \qquad \|\gamma'(u)\| = \sqrt{1 + u}.$$

Therefore, the arc length function is

$$\sigma_1(t) = \int_1^t \sqrt{1 + u}\,\mathrm{d}u = \left[\frac{2}{3}(1 + u)^{\frac{3}{2}}\right]_1^t = \frac{2}{3}\left((1 + t)^{\frac{3}{2}} - 2^{\frac{3}{2}}\right).$$

Exercise 5 Let $\sigma_{t_0}(t)$ be the arc length function associated to some regular curve γ.

(i) Write down the value $\sigma_{t_0}(t_0)$ of the arc length function at the basepoint.

(ii) Write down the derivative $\sigma'_{t_0}(t)$ of the arc length function.

We want to write down some useful properties of the arc length function and will indeed get to this soon. However, there is first a useful result from calculus that we must mention here.

Theorem 2.16 (Mean Value Theorem) *Let $f : [a, b] \to \mathbb{R}$ be continuous and differentiable on the open interval (a, b). Then, there exists $c \in (a, b)$ such that*

$$f'(c) = \frac{f(b) - f(a)}{b - a}.$$

Proof: Omitted. □

Remark 2.17 Consider the graph $y = f(x)$ of a differentiable function $f : \mathbb{R} \to \mathbb{R}$. For each pair of numbers (a, b) where $a < b$, we can construct the chord (line segment) between the points $(a, f(a))$ and $(b, f(b))$. The Mean Value Theorem asserts that at some point $(c, f(c))$ on the graph between the previous two points, the tangent line to the graph is parallel to the chord. This is made clear by Figure 4 below. Note the statement guarantees the existence of a point $c \in (a, b)$ with the required value of $f'(c)$, but it does **not** say it is unique.

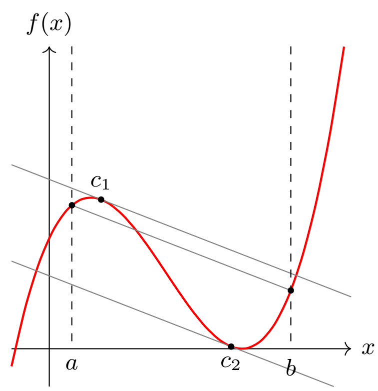


Figure 4: The geometric interpretation of the Mean Value Theorem.

Corollary 2.18 *Let $\sigma_{t_0} : I \to \mathbb{R}$ be the arc length function associated to some regular curve γ. Then, σ_{t_0} is* strictly increasing, *i.e. $t_1 < t_2$ implies $\sigma_{t_0}(t_1) < \sigma_{t_0}(t_2)$, and thus injective.*

Proof: Assume that σ_{t_0} is **not** strictly increasing. Then, there exist $t_1, t_2 \in I$ where $t_1 < t_2$ but that $\sigma_{t_0}(t_1) \geq \sigma_{t_0}(t_2)$. By the Mean Value Theorem, there exists a point $c \in (t_1, t_2)$ such that

$$\sigma'_{t_0}(c) = \frac{\sigma_{t_0}(t_2) - \sigma_{t_0}(t_1)}{t_2 - t_1} \leq 0.$$

However, we can see from Exercise 5 that $\sigma'_{t_0}(t) > 0$ for all $t \in I$. Because our assumption that σ_{t_0} is **not** strictly increasing leads to a contradiction with an earlier result, we must conclude that our assumption is false: σ_{t_0} is therefore strictly increasing. It is a known result from real analysis that all strictly increasing functions are injective. □

Note: This type of proof is called *proof by contradiction*; the idea is to assert the opposite of what you aim to prove and eventually get to some sort of inconsistency or nonsense.

It turns out that σ_{t_0} isn't surjective, but remember what we said in Section 1.2 about injective functions that are **not** surjective: we can always shrink the co-domain so that it turns bijective and thus we can talk about inverses. This is exactly what we do here (but we won't prove it rigorously as it requires a high level theorem which is beyond our discussion for now).

Proposition 2.19 *Let $\sigma_{t_0} : I \to \mathbb{R}$ be the arc length function associated to some regular curve γ and $J := \operatorname{im}(\sigma_{t_0}) \subseteq \mathbb{R}$ be its range. Then, there is an inverse function $\tau_{t_0} : J \to I$.*

Proof: Omitted. □

Exercise 6 Let τ_{t_0} be the inverse of σ_{t_0} as above. Write down the value $\tau_{t_0}(0)$.

Reminder: We have the following crucial trigonometric identity: $\sin^2(x) + \cos^2(x) = 1$.

Example 2.20 Consider the curve $\gamma : \mathbb{R} \to \mathbb{R}^2$ given by $\gamma(t) = (3\cos t, 3\sin t)$. We will find both the arc length function and the associated inverse at the basepoint $t_0 = 0$. To begin to apply Definition 2.13, we first see that

$$\gamma'(u) = (-3\sin t, 3\cos t) \qquad \Rightarrow \qquad \left\|\gamma'(u)\right\| = \sqrt{9} = 3.$$

Therefore, the arc length function is

$$\sigma_0(t) = \int_0^t 3 \,\mathrm{d}u = [3u]_0^t = 3t.$$

Finally, the arc length function here is the map which multiplies the input by three. Hence, the opposite of this is dividing the input by three, so we conclude that $\tau_{t_0}(s) = \frac{1}{3}s$.

Method – Finding Inverse Functions: Suppose the function $f : D \to E \subseteq \mathbb{R}$ is invertible.

1. Write down $f(s) = t$.
2. Rearrange the equation from Step **1** to make s the subject.
3. The result from Step **2** is $s = g(t)$; the function on the right is in fact $f^{-1}(t)$.

It turns out that the inverse is also injective; we should know this already because we can only invert bijective functions (and the inverse of an invertible function is itself invertible), so the inverse is also injective and surjective. Nevertheless, we can show this using the Chain Rule.

Lemma 2.21 *Let $\sigma_{t_0} : I \to \mathbb{R}$ be the arc length function associated to some regular curve γ. Then, the inverse $\tau_{t_0} : \mathrm{im}(\sigma_{t_0}) \to I$ is also strictly increasing and hence injective.*

Proof: By definition of the inverse and what we said in Section 1.2, we know $\sigma_{t_0} \circ \tau_{t_0} = \mathrm{id}_{\mathrm{im}(\sigma_{t_0})}$. In other words, we have $\sigma_{t_0}(\tau_{t_0}(s)) = s$ for every $s \in \mathrm{im}(\sigma_{t_0})$. Now apply the Chain Rule:

$$\sigma'_{t_0}(\tau_{t_0}(s))\tau'_{t_0}(s) = 1 \qquad \Rightarrow \qquad \tau'_{t_0}(s) = \frac{1}{\sigma'_{t_0}(\tau_{t_0}(s))} = \frac{1}{\left\|\gamma'(\tau_{t_0}(s))\right\|} > 0.$$

Combined with the Mean Value Theorem, this implies that τ_{t_0} is strictly increasing (we can run a similar contradiction argument as we did for Corollary 2.18), and thus τ_{t_0} is also injective. $\square$

Exercise 7 Consider the regular curve $\gamma : (0, \infty) \to \mathbb{R}^2$ given by $\gamma(t) = (t, \frac{2}{3}t^{\frac{3}{2}})$ from Exercise 4 and fix the basepoint $t_0 = 1$. Find expressions for both $\tau_1(s)$ and $\tau'_1(s)$.

[**Hint:** You are free to use the expression $\sigma_1(t)$ we obtained in Example 2.15. You should write down the range $J = \mathrm{im}(\sigma_1)$ in your answer, which can be found by considering the limits $t \to 0$ and $t \to \infty$ of $\sigma_1(t)$.]

Thus far, we have stuck to one time variable when we have looked at any particular curve. However, it is sometimes beneficial to "re-define" the time variable so that we are still looking at exactly the same curve (geometrically) except, from the point of view of a particle, it is moving around more slowly or more fast.

Definition 2.22 Let $\gamma : I \to \mathbb{R}^n$ be a parametrised curve. A reparametrisation is another parametrised curve $\widetilde{\gamma} : J \to \mathbb{R}^n$ with $\widetilde{\gamma} = \gamma \circ h$, where the parameter transformation $h : J \to I$ is a smooth surjective map and whose derivative $h'(u) > 0$ for all $u \in J$.

Let's just take a moment to process this: the curves γ and $\widetilde{\gamma}$ have the same image sets (they look the same geometrically). The difference is that we have changed one time interval I to another J. The map h is the rule which tells us how we are changing the time interval.

- We require h to be smooth so that $\widetilde{\gamma}$ is smooth and hence a parametrised curve.
- We require h to be surjective so that $\widetilde{\gamma}(J) = \gamma(I)$, that is their images are the same.
- We require h to have positive derivative so that it is injective, and thus bijective.

Example 2.23 Consider the curve $\gamma : \mathbb{R} \to \mathbb{R}^2$ given by $\gamma(t) = (t, e^t)$. We will consider two maps h, one which satisfies the conditions of Definition 2.22 and one which doesn't; the latter case will show us why the three conditions listed above are so important. Notice first that the derivative $\gamma'(t) = (1, e^t) \neq \mathbf{0}$ for any $t \in \mathbb{R}$ so it is a regular curve.

(i) Let $h : (0, \infty) \to \mathbb{R}$ be given by $h(u) = \log(u)$. Therefore, the reparametrisation is

$$\widetilde{\gamma}(u) = \gamma(h(u)) = (\log u, u).$$

Now, $\widetilde{\gamma}'(u) = (\frac{1}{u}, 1) \neq \mathbf{0}$ so the reparametrisation **is** also a regular curve.

(ii) Let $h : \mathbb{R} \to \mathbb{R}$ be given by $h(u) = \sin(u)$. Therefore, the reparametrisation is

$$\widetilde{\gamma}(u) = \gamma(h(u)) = (\sin u, e^{\sin u}).$$

But $\widetilde{\gamma}'(u) = (\cos u, e^{\sin u} \cos u) = \mathbf{0}$ when $t = \frac{\pi}{2}$, so the reparametrisation is **not** regular.

Note: The formal definition of reparametrisation excludes nasty cases like the one above. We can prove concretely that regularity of curves doesn't change under reparametrisation.

Lemma 2.24 *Any reparametrisation of a regular curve is itself regular.*

Proof: Let γ be a regularly parametrised curve with reparametrisation $\widetilde{\gamma}(u) = \gamma(h(u))$. By the Chain Rule, we see that $\widetilde{\gamma}'(u) = \gamma'(h(u))h'(u)$, which is zero if and only if either $\gamma'(h(u)) = 0$ (which it isn't because γ is regular) or $h'(u) = 0$ (which it isn't because h has positive derivative). Hence, there is **no** time at which the velocity $\widetilde{\gamma}'$ is zero, so it is regular. □

Lemma 2.25 *Any reparametrisation preserves arc length.*

Proof: Let γ be a parametrised curve with reparametrisation $\widetilde{\gamma}(u) = \gamma(h(u))$. Suppose that s is the arc length along γ from t_0 to t_1. Suppose that $u_0 := h(t_0)$ and $u_1 := h(t_1)$; we must show that s is also the arc length along $\widetilde{\gamma}$ from u_0 to u_1. To this end, we see that

$$s = \int_{t_0}^{t_1} \left\|\gamma'(t)\right\| \mathrm{d}t$$

$$= \int_{h^{-1}(t_0)}^{h^{-1}(t_1)} \big\|\gamma'(h(u))\big\| h'(u)\,\mathrm{d}u$$
$$= \int_{u_0}^{u_1} \big\|\gamma'(h(u))h'(u)\big\|\,\mathrm{d}u$$
$$= \int_{u_0}^{u_1} \big\|\widetilde{\gamma}'(u)\big\|\,\mathrm{d}u.$$

Note that the second equality comes by substituting $t = h(u)$ and using the fact that we can (implicitly) differentiate this equation with respect to u. The third equality is simply because $h'(u) > 0$ is a positive real number and so can therefore be placed inside the Euclidean norm. The final equality is an application of the Chain Rule. □

Now that we have a rigorous understanding of reparametrisation, what if we can reparametrise to ensure that the speed of the curve $\big\|\gamma'(t)\big\| = 1$? It turns out that this is possible, but we will first write out the definition and prove this fact soon.

Definition 2.26 A unit speed curve is a smooth $\gamma : I \to \mathbb{R}^n$ with $\big\|\gamma'(s)\big\| = 1$ for all $s \in I$.

Notation 2.27 It doesn't really matter because we can label things as we like, but it is common to use s to denote the time of a unit speed curve as opposed to t as for curves in general.

Note: It is no coincidence that s is also used for the arc length in Definition 2.11; this is because the distance along a unit speed curve γ is given rather easily as

$$\int_{s_0}^{s_1} \big\|\gamma'(s)\big\|\,\mathrm{d}s = \int_{s_0}^{s_1} 1\,\mathrm{d}s = s_1 - s_0.$$

Although unit speed curves seem quite special, there are actually completely universal in the sense that we will now prove.

Theorem 2.28 *Any regular curve $\gamma : I \to \mathbb{R}^n$ has a* unit speed reparametrisation $\widetilde{\gamma} : J \to \mathbb{R}^n$ *which is unique up to time translation. In other words, if $\overline{\gamma} : K \to \mathbb{R}^n$ is another unit speed reparametrisation, then there exists a constant $c \in \mathbb{R}$ such that*

$$\widetilde{\gamma}(s) = \overline{\gamma}(s - c).$$

Proof: Results such as this have two parts to prove: (i) existence and (ii) uniqueness.

(i) Fix a basepoint $t_0 \in I$ and let $\sigma_{t_0} : I \to J$ be the signed arc length function of γ and recall there is a smooth, well-defined inverse $\tau_{t_0} : J \to I$ which has positive derivative (Lemma 2.21).

Then, we can set $h = \tau_{t_0}$ in Definition 2.22; we now show that $\widetilde{\gamma}(s) = \gamma(\tau_{t_0}(s))$ has unit speed:

$$\widetilde{\gamma}'(s) = \gamma'(\tau_{t_0}(s))\tau_{t_0}'(s) = \gamma'(\tau_{t_0}(s))\frac{1}{\left\|\gamma'(\tau_{t_0}(s))\right\|} \qquad \Rightarrow \qquad \left\|\widetilde{\gamma}(s)\right\| = 1.$$

(ii) Suppose $\overline{\gamma} : K \to \mathbb{R}^n$ is another unit speed reparametrisation of γ and consider the arc length along γ from $\gamma(t_0) = \widetilde{\gamma}(s_0) = \overline{\gamma}(\mathfrak{s}_0)$ to an arbitrary $\gamma(t) = \widetilde{\gamma}(s) = \overline{\gamma}(\mathfrak{s})$. Using Lemma 2.25 and the previous note which tells us the arc length along a unit speed curve, the arc length is

$$s - s_0 = \mathfrak{s} - \mathfrak{s}_0 \qquad \Rightarrow \qquad \mathfrak{s} = s - (s_0 - \mathfrak{s}_0).$$

Consequently, $\widetilde{\gamma}(s) = \overline{\gamma}(\mathfrak{s}) = \overline{\gamma}(s - (s_0 - \mathfrak{s}_0))$, so we can take $c = s_0 - \mathfrak{s}_0$. □

We can now write out a solid method for finding the unit speed reparametrisation of any regular curve. Note that sometimes it can't be done explicitly (we will see an example of this soon) but, in principle, this will always work; any questions in these notes will also be designed to be done fully by-hand so don't panic!

Method – Finding a Unit Speed Reparametrisation: Let $\gamma : I \to \mathbb{R}^n$ be a regular curve.

1. Find the arc length function $\sigma_0(t)$ based at $t = 0$.
2. Invert the arc length function to find an expression for $\tau_0(s)$.
3. The unit speed reparametrisation is $\gamma(\tau_0(s))$.

Example 2.29 Let $\gamma : \mathbb{R} \to \mathbb{R}^n$ be given by $\gamma(t) = (r\cos t, r\sin t)$ where $r > 0$ is any positive real number. We will demonstrate the above method to find a unit speed reparametrisation. First,

$$\gamma'(t) = (-r\sin t, r\cos t) \qquad \Rightarrow \qquad \left\|\gamma'(t)\right\| = \sqrt{r^2\sin^2(t) + r^2\cos^2(t)} = r,$$

so we know that γ is regular because $r > 0$. Consequently, we can find the arc length function and its inverse like we have done earlier, see Example 2.15 for instance. To this end,

$$\sigma_0(t) = \int_0^t r\,\mathrm{d}u = rt \qquad \Rightarrow \qquad \tau_0(s) = \frac{s}{r}.$$

It remains to substitute this into the expression for γ to get the unit speed reparametrisation:

$$\widetilde{\gamma}(s) = \gamma(\tau_0(s)) = \left(r\cos\frac{s}{r}, r\sin\frac{s}{r}\right).$$

Note: We can fix any basepoint as in the proof of Theorem 2.28 but zero seems easiest.

Example 2.30 We now consider the curve $\gamma : \mathbb{R} \to \mathbb{R}^n$ given by $\gamma(t) = (a \cos t, \sin t)$ for $a > 1$. The point of this example is to see that we can still follow the method above but that there isn't always hope to write things down explicitly! Indeed, we have

$$\gamma'(t) = (-a \sin t, \cos t) \qquad \Rightarrow \qquad \|\gamma'(t)\| = \sqrt{a^2 \sin^2(t) + \cos^2(t)}.$$

Already, this doesn't simplify as nice as in Example 2.29. Now for the arc length function:

$$\sigma_0(t) = \int_0^t \sqrt{a^2 \sin^2(u) + \cos^2(u)}\, \mathrm{d}u,$$

which is a so-called *elliptic integral of the second kind*. This can be expressed as an infinite sum, but then its inverse isn't possible to write down generally with standard mathematical functions. Thus, the unit speed reparametrisation $\widetilde{\gamma}$ exists but we can't actually see it!

Exercise 8 Find a unit speed reparametrisation of the regular curve $\gamma : \mathbb{R} \to \mathbb{R}^4$ where

$$\gamma(t) = (t, \cos(2t), \sin(2t), 7).$$

We will now spend some time talking about *curvature* of curves. Of course, the 'curve' $\gamma(t) = (t, t)$ is a curve but it is geometrically a straight line. On the other hand, we may have something like $\gamma(t) = (t, t + \sin t)$ which is very wiggly and thus we expect its curvature to change (which it does so infinitely often). It is now our job to motivate the next part of the discussion.

Note: This is the basic idea behind regions of high, low and no curvature of some γ:

- High curvature is where tangent lines change rapidly.
- Low curvature is where tangent lines change slowly.
- No curvature is where tangent lines don't change at all.

Recall from Definition 2.7 that the tangent to a regular curve γ at the point $t_0 \in I$ is given by

$$\widehat{\gamma}_{t_0}(t) = \gamma(t_0) + t\gamma'(t_0).$$

Note that the important part is $\gamma'(t_0)$ which determines the direction of the tangent line; this is what we want to analyse. At first, we may expect that we want to see the rate of change of this vector, that is its derivative $(\gamma')'(t_0) = \gamma''(t_0)$. But this is **not** quite sufficient as we see below.

Example 2.31 Consider the curve $\gamma : \mathbb{R} \to \mathbb{R}^2$ given by $\gamma(t) = (\log t, 2\log t + 1)$ which is a straight line; it isn't curved at all. However, its velocity and acceleration vectors are

$$\gamma'(t) = (t^{-1}, 2t^{-1}) \qquad \text{and} \qquad \gamma''(t) = (-t^{-2}, -2t^{-2}) \neq \mathbf{0}.$$

Remark 2.32 The problem is that γ'' tells us not only about the rate of change of direction of γ' but also how its **length** changes too. But we do not care about length (and more people should follow this teaching). Therefore, we can compute curvatures this way for unit speed curves since $\|\gamma'(s)\| = 1$ for all s. The goal now is to extend this to **any** arbitrary regular curve.

Definition 2.33 Let $\gamma : I \to \mathbb{R}^n$ be a unit speed curve. Then, the curvature vector of γ is the vector-valued function $k : I \to \mathbb{R}^n$ given by $k(s) = \gamma''(s)$. The norm of the curvature vector $\|k(s)\|$ is called the curvature of γ.

Technically speaking, it isn't hard to extend Definition 2.33. Indeed, if γ is a regular curve that does **not** have unit speed, we can refer to Theorem 2.28 to see that there exists a unit speed reparametrisation $\widetilde{\gamma} = \gamma \circ h$. Thus, the curvature vector of γ and $t = h(s)$ is defined to be the curvature vector of $\widetilde{\gamma}$ at s, which is precisely $\widetilde{\gamma}''(s)$.

Note: This definition is well-defined in the sense that it doesn't matter what unit speed reparametrisation we choose. This is because any two unit speed reparametrisations $\widetilde{\gamma}$ and $\overline{\gamma}$ differ by a constant; this has no effect on the second (or first) derivative of $\widetilde{\gamma}$.

Example 2.34 Let $\gamma : \mathbb{R} \to \mathbb{R}^2$ be the curve $\gamma(t) = (2\cos t, 2\sin t)$. Then, we see that

$$\gamma'(t) = (-2\sin t, 2\cos t) \qquad \Rightarrow \qquad \|\gamma'(t)\| = \sqrt{4\sin^2(t) + 4\cos^2(t)} = 2.$$

Much like Example 2.29, we can write down the arc length function and its inverse:

$$\sigma_0(t) = \int_0^t 2\,\mathrm{d}u = 2t \qquad \Rightarrow \qquad \tau_0(s) = \frac{s}{2}.$$

As such, the unit speed reparametrisation is $\widetilde{\gamma}(s) = \gamma(\tau_0(s)) = (2\cos\frac{s}{2}, 2\sin\frac{s}{2})$. Because this is a unit speed curve, we can apply Definition 2.33 to see that the curvature vector is

$$\widetilde{\gamma}'(s) = \left(-\sin\frac{s}{2}, \cos\frac{s}{2}\right) \qquad \Rightarrow \qquad \widetilde{\gamma}''(s) = \left(-\frac{1}{2}\cos\frac{s}{2}, -\frac{1}{2}\sin\frac{s}{2}\right).$$

We are technically done but, really, we should change back to the original time variable:

$$k(t) = \widetilde{\gamma}''(\sigma_0(t)) = \left(-\frac{1}{2}\cos t, -\frac{1}{2}\sin t\right).$$

Exercise 9 Find the curvature vector of $\gamma : \mathbb{R} \to \mathbb{R}^3$ with $\gamma(s) = (\frac{1}{\sqrt{2}}\cos s, \frac{1}{\sqrt{2}}\sin s, \frac{1}{\sqrt{2}}s)$.

Although this process works a treat, it is quite clumsy in practice and allows for many mistakes to arise. In particular, it is usually impossible to write down the arc length function or the unit speed reparametrisation explicitly (recall Example 2.30). But there is another way.

Reminder: We have the following well-known ideas for vectors/vector-valued functions.

1. Let $\mathbf{u}, \mathbf{v} \in \mathbb{R}^n$. Their dot product is the number $\mathbf{u} \cdot \mathbf{v} := u_1 v_1 + u_2 v_2 + \cdots + u_n v_n$.
2. Let $\mathbf{u} \in \mathbb{R}^n$. We can write its norm using the dot product as $\|\mathbf{u}\| = \sqrt{\mathbf{u} \cdot \mathbf{u}} = (\mathbf{u} \cdot \mathbf{u})^{\frac{1}{2}}$.
3. Let $u, v : I \to \mathbb{R}^n$ be vector-valued functions. The Product Rule tells us that
$$\frac{\mathrm{d}}{\mathrm{d}t}[u(t) \cdot v(t)] = u'(t) \cdot v(t) + u(t) \cdot v'(t).$$
4. Let $u : I \to \mathbb{R}^n$ be a vector-valued function. The Chain Rule tells us that
$$\frac{\mathrm{d}}{\mathrm{d}t}\|\mathbf{u}\| = \frac{\mathrm{d}}{\mathrm{d}t}(u(t) \cdot u(t))^{\frac{1}{2}} = \frac{u'(t) \cdot u(t) + u(t) \cdot u'(t)}{2(u(t) \cdot u(t))^{\frac{1}{2}}} = \frac{u(t) \cdot u'(t)}{\|u(t)\|}.$$

We are now in a position to prove a more general curvature vector formula.

Proposition 2.35 *The curvature vector for any arbitrary regular curve $\gamma : I \to \mathbb{R}^n$ is the vector-valued function $k : I \to \mathbb{R}^n$ given by*

$$k(t) = \frac{1}{\|\gamma'(t)\|^2}\left(\gamma''(t) - \frac{\gamma'(t) \cdot \gamma''(t)}{\|\gamma'(t)\|^2}\gamma'(t)\right).$$

Proof: Let γ be a regular curve and $\widetilde{\gamma}(s) = \gamma(h(s))$ be a unit speed reparametrisation of γ. Writing $t = h(s)$, the Chain Rule (and implicit differentiation) yields the following expression:

$$\widetilde{\gamma}'(s) = \gamma'(h(s))h'(s) = \frac{\mathrm{d}\gamma}{\mathrm{d}t}\frac{\mathrm{d}t}{\mathrm{d}s} = \frac{\gamma'(t)}{\mathrm{d}s/\mathrm{d}t}.$$

Because $\widetilde{\gamma}$ is a unit speed curve, we must have that $\mathrm{d}s/\mathrm{d}t = \|\gamma'(t)\|$ to ensure the norm of the above equation is one. In particular then, this means that

$$\widetilde{\gamma}'(s) = \frac{\gamma'(t)}{\|\gamma'(t)\|}.$$

Differentiating this with respect to s once again gives us the following second derivative:

$$\widetilde{\gamma}''(s) = \frac{\mathrm{d}}{\mathrm{d}s}\left[\frac{\gamma'(t)}{\|\gamma'(t)\|}\right]$$

$$= \frac{\mathrm{d}}{\mathrm{d}t}\left[\frac{\gamma'(t)}{\|\gamma'(t)\|}\right]\frac{\mathrm{d}t}{\mathrm{d}s}$$

$$= \left(\frac{\gamma''(t)\|\gamma'(t)\| - \gamma'(t)\frac{\mathrm{d}}{\mathrm{d}t}\left[\|\gamma'(t)\|\right]}{\|\gamma'(t)\|^2}\right)\frac{1}{\|\gamma'(t)\|}$$

$$= \left(\frac{\gamma''(t)}{\|\gamma'(t)\|} - \frac{\gamma'(t)}{\|\gamma'(t)\|^2}\frac{\mathrm{d}}{\mathrm{d}t}\left[\|\gamma'(t)\|\right]\right)\frac{1}{\|\gamma'(t)\|}$$

$$= \left(\frac{\gamma''(t)}{\|\gamma'(t)\|} - \frac{\gamma'(t)}{\|\gamma'(t)\|^2}\frac{\gamma'(t)\cdot\gamma''(t)}{\|\gamma'(t)\|}\right)\frac{1}{\|\gamma'(t)\|}$$

$$= \left(\gamma''(t) - \frac{\gamma'(t)}{\|\gamma'(t)\|}\gamma'(t)\cdot\gamma''(t)\right)\frac{1}{\|\gamma'(t)\|^2},$$

which is exactly as is written in the statement albeit we have things in a different but equivalent order. Note that the third equality uses the Quotient Rule and the fifth equality uses the final fact in the above reminder. □

Note: A good habit to adopt is to look precisely how to get from one line of a proof to the next; it may be that it is very obvious or straightforward right from the get-go. Sometimes, it requires you to look back to piece together the story. That said, I will always try to write things to help guide you as best as possible when doing this.

Example 2.36 Consider $\gamma : \mathbb{R} \to \mathbb{R}^2$ where $\gamma(t) = (t, t^2)$; this is a simple parabola that we have probably seen before (if every you have seen $y = x^2$, this is that very same curve). We will calculate the curvature vector à la Proposition 2.35. Indeed, the first things to compute are

$$\gamma'(t) = (1, 2t) \qquad \Rightarrow \qquad \|\gamma'(t)\|^2 = 1 + 4t^2 \qquad \text{and} \qquad \gamma''(t) = (0, 2).$$

The dot product $\gamma'(t)\cdot\gamma''(t) = 4t$. Hence, we are now able to write the curvature vector:

$$k(t) = \frac{1}{1+4t^2}\left((0,2) - \frac{4t}{1+4t^2}(1,2t)\right)$$

$$= \frac{1}{(1+4t^2)^2}(0, 2+8t^2) - \frac{4t}{(1+4t^2)^2}(1,2t)$$

$$= \frac{1}{(1+4t^2)^2}(-4t, 2+8t^2-8t^2)$$

$$= \frac{1}{(1+4t^2)^2}(-4t, 2).$$

Exercise 10 Find the curvature vector of $\gamma : \mathbb{R} \to \mathbb{R}^2$ with $\gamma(t) = (2\cos t, \sin t)$.

Of course, the method for finding the curvature vector can still be pretty complicated; there are a lot of delicate things to be aware of (and I myself have made many mistakes, and still do, when doing intricate mathematics)! However, there is a very easy test that can tell us immediately if our computations are incorrect.

Lemma 2.37 *The curvature vector k of a regular curve γ is* orthogonal *to its velocity, which means that their dot product $k(t) \cdot \gamma'(t) = 0$ for all $t \in I$.*

Proof: This is immediate from the expression we have in Proposition 2.35. Indeed,

$$\begin{aligned} k(t) \cdot \gamma'(t) &= \frac{1}{\|\gamma'(t)\|^2} \left(\gamma''(t) \cdot \gamma'(t) - \frac{\gamma'(t) \cdot \gamma''(t)}{\|\gamma'(t)\|^2} \gamma'(t) \cdot \gamma'(t) \right) \\ &= \frac{1}{\|\gamma'(t)\|^2} \left(\gamma''(t) \cdot \gamma'(t) - \frac{\gamma'(t) \cdot \gamma''(t)}{\|\gamma'(t)\|^2} \|\gamma'(t)\|^2 \right) \\ &= \frac{1}{\|\gamma'(t)\|^2} \left(\gamma''(t) \cdot \gamma'(t) - \gamma'(t) \cdot \gamma''(t) \right) \\ &= 0. \end{aligned}$$

□

Example 2.38 (Revisited) Even though we have proved the result concretely, it's always nice to come back to reality with an explicit example which shows that k and γ' are orthogonal. In this way, we will revisit Example 2.34, where we found that the curvature vector of $\gamma : \mathbb{R} \to \mathbb{R}^2$ given by $\gamma(t) = (2\cos t, 2\sin t)$ is $k(t) = -\frac{1}{2}(\cos t, \sin t)$. Recall also that $\gamma'(t) = (-2\sin t, 2\cos t)$, so we conclude immediately that

$$k(t) \cdot \gamma'(t) = \left(-\frac{1}{2}\cos t\right)(-2\sin t) + \left(-\frac{1}{2}\sin t\right)(2\cos t) = \sin(t)\cos(t) - \sin(t)\cos(t) = 0.$$

We conclude the discussion by defining a few more objects under the motivation of making the formula in Proposition 2.35 just a little bit more bearable than the beast it currently is.

Definition 2.39 Let $\gamma : I \to \mathbb{R}^n$ be a regular curve. Its unit tangent vector $u : I \to \mathbb{R}^n$ is

$$u(t) = \frac{\gamma'(t)}{\|\gamma'(t)\|}.$$

Note: The norm of the unit tangent vector $\|u(t)\| = 1$ for all t. In fact, this is why it is called the *unit* tangent vector; it has unit length. Moreover, its norm is one "by construction", which is mathematical slang for "we want a certain object to have a certain property, so we define it in a way to ensure this happens".

Because we declare that γ is a regularly parametrised curve in Definition 2.39, it ensures that u is well-defined (i.e. we are not ever dividing by zero because $\|\gamma'(t)\| > 0$ for all t).

Definition 2.40 Let $u : I \to \mathbb{R}^n$ be the unit tangent vector of some regular curve γ and $v : I \to \mathbb{R}^n$ be any vector-valued function. The normal projection $v_\perp : I \to \mathbb{R}^n$ is given by

$$v_\perp(t) = v(t) - \big(v(t) \cdot u(t)\big)\, u(t).$$

The geometric picture is this: $v_\perp$ is the part of the vector v left over after we have subtracted from it the component which travels in the direction of u. Another way to think of it is to draw a vector from a point on the line with direction u to the end of the vector v such that it forms a right-angle with the line with direction u. We now draw a geometric diagram representing this.

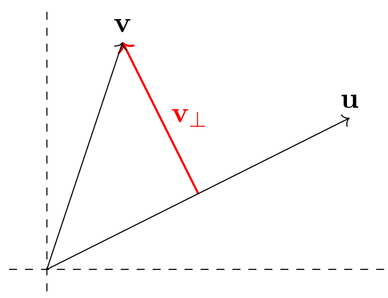


Figure 5: The normal projection of $\mathbf{v}$ onto $\mathbf{u}$.

Example 2.41 Consider $\gamma : \mathbb{R} \to \mathbb{R}^3$ given by $\gamma(t) = (\frac{1}{2}t^2, \sin t, \cos t)$. We will construct the unit tangent vector and then determine $\gamma''_\perp$. The reason that the normal projection of acceleration is interesting will be seen soon. Well, $\gamma'(t) = (t, \cos t, -\sin t)$ and so $\|\gamma'(t)\| = \sqrt{1+t^2}$. Thus,

$$u(t) = \frac{1}{\sqrt{1+t^2}}(t, \cos t, -\sin t).$$

To apply Definition 2.40 for the acceleration vector γ'', we first need an expression for both the acceleration itself as well as its dot product with the unit tangent vector:

$$\gamma''(t) = (1, -\sin t, -\cos t) \qquad \text{and} \qquad \gamma''(t) \cdot u(t) = \frac{t}{\sqrt{1+t^2}}.$$

Finally, it remains to use the definition of the normal projection, that is

$$\begin{aligned}
\gamma''_\perp(t) &= \gamma''(t) - \big(\gamma''(t)\cdot u(t)\big)\,u(t) \\
&= (1, -\sin t, -\cos t) - \frac{t}{\sqrt{1+t^2}}\frac{1}{\sqrt{1+t^2}}(t, \cos t, -\sin t) \\
&= (1, -\sin t, -\cos t) - \frac{t}{1+t^2}(t, \cos t, -\sin t) \\
&= \left(1 - \frac{t^2}{1+t^2}, -\sin t - \frac{t}{1+t^2}\cos t, -\cos t + \frac{t}{1+t^2}\sin t\right).
\end{aligned}$$

Note: Looking at the first equality above and remembering the definition of $u(t)$, we have

$$\gamma''_\perp(t) = \gamma''(t) - \left(\gamma''(t)\cdot\frac{\gamma'(t)}{\left\|\gamma'(t)\right\|}\right)\frac{\gamma'(t)}{\left\|\gamma'(t)\right\|} = \gamma''(t) - \frac{\gamma'(t)\cdot\gamma''(t)}{\left\|\gamma'(t)\right\|^2}\gamma'(t),$$

which is precisely the expression appearing in the brackets for $k(t)$ in Proposition 2.35.

We have now the most succinct way of writing the curvature vector. Be aware it is definitely **not** as practically useful; the expression in Proposition 2.35 tells us exactly what we need to calculate and where to substitute it in order to write $k(t)$. Nevertheless, here we go:

$$k(t) = \frac{\gamma''_\perp(t)}{\left\|\gamma'(t)\right\|^2}.$$

Exercise 11 (Harder) Let $f : \mathbb{R} \to \mathbb{R}$ be smooth and consider $\gamma : \mathbb{R} \to \mathbb{R}^3$ given by

$$\gamma(t) = \left(t^2, f(t), \int_0^t e^{f(u)}\,\mathrm{d}u\right).$$

(i) Show that γ is a regularly parametrised curve.

(ii) Given that $f(1) = \log(2)$, $f'(1) = 0$ and $f''(1) = 1$, find the curvature vector $k(1)$.

[**Hint:** Write γ' and γ'' in terms of f and its derivatives and then substitute $t = 1$.]

2.2 Planer Curves

We restrict our attention to curves that lie in the *plane*, meaning those whose co-domain is the two-dimensional space $\mathbb{R}^2$, not the general n-dimensional space discussed in Section 2.1. It turns out the theory of curvature can be developed deeper if we restrict ourselves to two dimensions.

Note: The curvature vector $k(t)$ under this restriction is a pair of real functions, namely $k(t) = \big(k_1(t), k_2(t)\big)$. Because we know that this is orthogonal to the unit tangent vector by Lemma 2.37, we automatically know the direction of k. The extra information we need is its length and its "sense", by which we mean if it points to the left or right of $u(t)$.

We begin by constructing a new vector which is orthogonal to the unit tangent vector by design; we will then soon define the *signed curvature* of a regularly parametrised curve by using it.

Definition 2.42 Let $\gamma : I \to \mathbb{R}^2$ be a regularly parametrised planar curve with unit tangent vector $u(t) = \big(u_1(t), u_2(t)\big)$. Its unit normal vector $n : I \to \mathbb{R}^n$ is $n(t) := \big(-u_2(t), u_1(t)\big)$.

Example 2.43 Consider the curve $\gamma : \mathbb{R} \to \mathbb{R}^2$ given by $\gamma(t) = (\sin t, t)$. To construct its unit normal vector, we must first find its unit tangent vector. With respect to this goal then, we see that $\gamma'(t) = (\cos t, 1)$ and so $\big\|\gamma'(t)\big\| = \sqrt{1 + \cos^2(t)}$. Therefore, the unit tangent vector is

$$u(t) = \frac{1}{\sqrt{1 + \cos^2(t)}}(\cos t, 1).$$

Consequently, the unit normal vector is obtained by swapping the entries in the vector above and multiplying the new first entry by minus one, which yields

$$n(t) = \frac{1}{\sqrt{1 + \cos^2(t)}}(-1, \cos t).$$

Definition 2.44 Let $\gamma : I \to \mathbb{R}^2$ be a regular curve. The signed curvature $\kappa : I \to \mathbb{R}$ is

$$\kappa(t) = k(t) \cdot n(t).$$

We know both $k(t)$ and $n(t)$ are orthogonal to $u(t)$. Because we are in two dimensions, this means the curvature vector and the unit normal vector are parallel (i.e. they are scalar multiples of each other, meaning one is nothing more than a number multiplied by the other). This leads to the equivalent definition of the signed curvature $\kappa(t)$ as the constant which makes this true:

$$k(t) = \kappa(t)n(t).$$

Note: Taking norms, it follows that $\big\|k(t)\big\| = \big|\kappa(t)\big|$, but κ contains more information than $\|k\|$ as its sign (positive or negative) is what tells us the "sense" of the curvature vector.

We will now demonstrate a slightly more convenient expression for the signed curvature.

Lemma 2.45 *Let $\gamma : I \to \mathbb{R}^2$ be a regular curve. Then, it signed curvature κ is given by*

$$\kappa(t) = \frac{\gamma''(t) \cdot n(t)}{\left\|\gamma'(t)\right\|^2}.$$

Proof: Recall from the end of Section 2.1 that the curvature vector is given by

$$k(t) = \frac{\gamma''_\perp(t)}{\left\|\gamma'(t)\right\|^2} \qquad \Rightarrow \qquad \kappa(t) = k(t) \cdot n(t) = \frac{\gamma''_\perp(t) \cdot n(t)}{\left\|\gamma'(t)\right\|^2}.$$

This is almost what we want, but the expression in the statement is about acceleration, **not** the normal projection of acceleration. That said, we quickly see from Definition 2.40 that

$$\gamma''_\perp(t) \cdot n(t) = \Big(\gamma''(t) - \big(\gamma''(t) \cdot u(t)\big)\, u(t)\Big) \cdot n(t) = \gamma''(t) \cdot n(t),$$

because n and u are orthogonal so expanding out the dot product means the only surviving thing (i.e. "things that aren't zero") is the dot product of acceleration and unit normal vector. ☐

Example 2.46 (Revisited) Continuing from Example 2.43 with $\gamma : \mathbb{R} \to \mathbb{R}^2$ where $\gamma(t) = (\sin t, t)$, we can compute its signed curvature. Indeed, we will use Lemma 2.45 to do this. We first need

$$\gamma''(t) = (-\sin t, 0) \qquad \Rightarrow \qquad \gamma''(t) \cdot n(t) = \frac{\sin(t)}{\sqrt{1 + \cos^2(t)}}.$$

Dividing this by the square of the speed yields the signed curvature $\kappa(t) = \sin(t)(1 + \cos^2 t)^{-\frac{3}{2}}$.

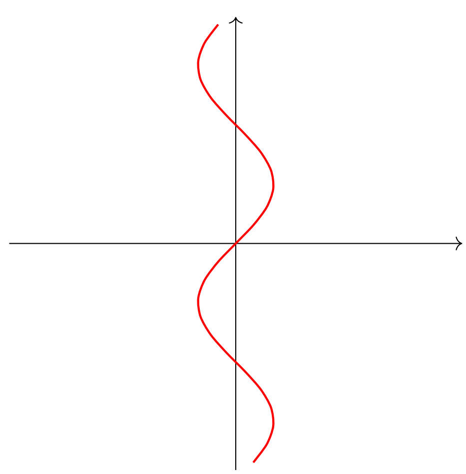


Figure 6: The parametrised curve $\gamma(t) = (\sin t, t)$.

Note: We have the following geometric intuition for the signed curvature:

(i) If $\kappa > 0$, then the vector k points in the same direction as n and γ turns **left**.

(ii) If $\kappa < 0$, then the vector k points in the opposite direction to n and γ turns **right**.

(iii) If $\kappa = 0$, then the vector k points orthogonally to n and γ is **straight**.

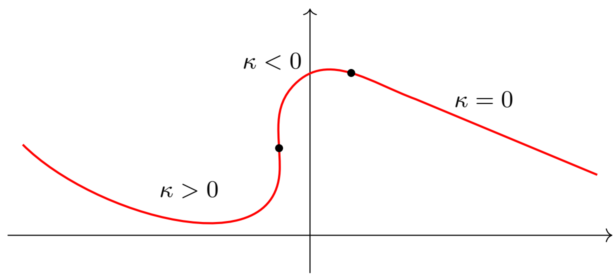


Figure 7: Regions of positive, negative and zero signed curvature.

Crucially, the observation made above depends on the *orientation* of the curve, that is the direction in which it is traversed. If we were to reparametrise so that time is 'reversed', then we would have to swap the geometric intuition.

Exercise 12 Compute the signed curvature for $\gamma : \mathbb{R} \to \mathbb{R}^2$ given by $\gamma(t) = (t, t^2)$.

As is seen in Figure 7 where the orientation of the curve is left-to-right, there are sometimes points where the signed curvature is exactly zero. Such points are given a special name.

Definition 2.47 Let $\gamma : I \to \mathbb{R}^2$ be a regular curve with signed curvature κ. If there exists $\tau \in I$ with $\kappa(\tau) = 0$ and κ changes sign at $t = \tau$, we call $\gamma(\tau)$ an inflexion point of γ.

Example 2.48 (Revisited) We once again a consider $\gamma : \mathbb{R} \to \mathbb{R}^2$ where $\gamma(t) = (\sin t, t)$. Looking at the expression we obtained for its signed curvature, we see that

$$\kappa(t) = \frac{\sin(t)}{(1 + \cos^2 t)^{\frac{3}{2}}} = 0 \qquad \Leftrightarrow \qquad t = n\pi \text{ for } n \in \mathbb{Z}.$$

Furthermore, the sign of $\kappa(t)$ changes at each of the times $t = n\pi$, so we have determined that γ has infinite-many inflexion points $\gamma(n\pi) = (0, n\pi)$.

Exercise 13 Find all inflexion points of $\gamma : \mathbb{R} \to \mathbb{R}^2$ given by $\gamma(t) = (t, t + \sin t)$.

Note: It is crucial that Definition 2.47 includes that κ changes sign at $t = \tau$; if we forget about this condition and only solve $\kappa(\tau) = 0$, we aren't guaranteed an inflexion point.

Example 2.49 Consider the curve $\gamma : \mathbb{R} \to \mathbb{R}^2$ given by $\gamma(t) = (t, t^4)$. We claim that $\kappa(0) = 0$ but that there is **no** inflexion point, in particular at $\gamma(0) = (0, 0)$. We will provide a sketch of this curve so that we can see its curvature $\kappa(t) \geq 0$ because it is always turning left, but we will leave it to Exercise 14 for you to find the signed curvature and show that $\kappa(0) = 0$.

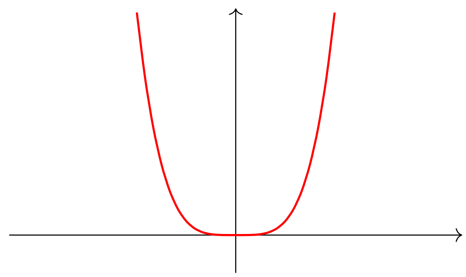


Figure 8: The parametrised curve $\gamma(t) = (t, t^4)$.

Exercise 14 Compute the signed curvature for $\gamma : \mathbb{R} \to \mathbb{R}^2$ given by $\gamma(t) = (t, t^4)$.

Henceforth, we assume that all our planar curves are unit speed curves; we are free to do this because Theorem 2.28 means any non-unit speed curve can be made into one by re-defining time.

Now, we have shown that given a curve γ, we can construct its signed curvature κ. But can we go the other way, that is given a function κ, can we construct a curve γ who has that function as its signed curvature? Assuming that γ has unit speed, this is possible! Well, provided two other pieces of information are known.

Theorem 2.50 (Fundamental Theorem of Local Theory of Plane Curves) *Given any smooth function $\kappa : I \to \mathbb{R}$ with $0 \in I$ and $\theta \in \mathbb{R}$ and $(a, b) \in \mathbb{R}^2$, there exists a unique unit speed curve $\gamma : I \to \mathbb{R}$ with signed curvature κ such that $\gamma(0) = (a, b)$ and $\gamma'(0) = (\cos\theta, \sin\theta)$.*

Proof: This proof is a bit long so we will talk about the strategy. Indeed, we start by assuming that a unit speed curve γ exists with signed curvature κ and this allows us to find a system of differential equations that κ must satisfy. Therefore, given a function κ, we can solve the system to find the curve γ (working in the reverse direction).

Assuming γ is a unit speed curve, we know that $\|\gamma'(s)\| = 1$ which means that its velocity is determined only by its **direction** In other words, there exists a smooth function $\varphi : I \to \mathbb{R}$ such that $\gamma'(s) = \big(\cos\varphi(s), \sin\varphi(s)\big)$. But the unit tangent vector $u(s) = \gamma'(s)$ for unit speed curves, so the unit normal vector is $n(t) = \big(-\sin\varphi(s), \cos\varphi(s)\big)$. Recall from Definition 2.33 that the curvature vector for a unit speed curve is $k(t) = \gamma''(t) = \theta'(s)\,\big(-\sin\varphi(s), \cos\varphi(s)\big)$. Thus,

$$\kappa(s) = k(s)\cdot n(s) = \varphi'(s).$$

We therefore have the following *coupled system of non-linear ordinary differential equations*, and is now our goal to prove the existence of a unique global solution of this system of differential equations under the conditions that $\gamma(0) = (a, b)$ and $\gamma'(0) = (\cos\theta, \sin\theta)$:

$$\begin{cases} \dfrac{\mathrm{d}\varphi}{\mathrm{d}s} = \kappa(s) \\ \dfrac{\mathrm{d}\gamma_1}{\mathrm{d}s} = \cos(\varphi) \\ \dfrac{\mathrm{d}\gamma_2}{\mathrm{d}s} = \sin(\varphi) \end{cases} . \tag{$*$}$$

To solve the system $(*)$, we realise it is *separable*; we can solve the first *initial value problems* (these are differential equations with a condition at $s = 0$) and substitute the solution into the second and third equations (the latter two are independent). Indeed, the first is

$$\frac{\mathrm{d}\varphi}{\mathrm{d}s} = \kappa(s), \qquad \varphi(0) = \theta.$$

If we integrate both sides of the differential equation and use the initial value, we get

$$\varphi(s) = \theta + \int_0^s \kappa(u)\,\mathrm{d}u.$$

This solution is actually unique (Exercise 15). It remains to solve the second and third initial value problem, but this is now easy because we can substitute the above solution and integrate the latter two problems. The second initial value problem is

$$\frac{\mathrm{d}\gamma_1}{\mathrm{d}s} = \cos(\varphi) \text{ with } \gamma_1(0) = a \qquad \Rightarrow \qquad \gamma_1(s) = a + \int_0^s \cos\varphi(u)\,\mathrm{d}u$$

and third initial value problem is

$$\frac{\mathrm{d}\gamma_2}{\mathrm{d}s} = \sin(\varphi) \text{ with } \gamma_2(0) = b \qquad \Rightarrow \qquad \gamma_2(s) = b + \int_0^s \sin\varphi(u)\,\mathrm{d}u.$$

So the unique γ exists; it has signed curvature κ and satisfies the $\gamma(0)$ and $\gamma'(0)$ conditions. □

Note: If this proof is giving you a bit difficulty when following it, you may just skip it!

Exercise 15 (Harder) Prove the above solution to $\frac{\mathrm{d}\varphi}{\mathrm{d}s} = \kappa(s)$ with $\varphi(0) = \theta$ is unique.

[**Hint:** Assume there is another solution $f \neq \varphi$ satisfying $f(0) = \theta$ and apply the Mean Value Theorem to the function $g(s) := \varphi(s) - f(s)$ with the aim of obtaining a contradiction.]

Not only does Theorem 2.50 tell us the existence of a unique γ but it gives a formula for it:

$$\gamma(s) = \left(a + \int_0^s \cos\varphi(u)\,\mathrm{d}u, b + \int_0^s \sin\varphi(u)\,\mathrm{d}u\right), \qquad \text{with } \varphi(u) = \theta + \int_0^u \kappa(w)\,\mathrm{d}w. \qquad (\dagger)$$

Example 2.51 Let $\kappa(s) = K \in \mathbb{R} \setminus \{0\}$ be a constant function. We will find a curve of constant signed curvature K under the conditions that $\gamma(0) = (0, 0)$ and $\gamma'(0) = (1, 0)$. Looking at the velocity, we see that $\varphi(0) = \theta = 0$. Using the above formula $(\dagger)$, we see that

$$\varphi(u) = 0 + \int_0^u \kappa(w)\,\mathrm{d}w = \int_0^u K\,\mathrm{d}w = Ku.$$

Therefore, we can find formulae for the coordinates defining points on the curve:

$$\begin{aligned}\gamma_1(s) &= 0 + \int_0^s \cos(Ku)\,\mathrm{d}u = \frac{1}{K}\sin(Ks),\\ \gamma_2(s) &= 0 + \int_0^s \sin(Ku)\,\mathrm{d}u = -\frac{1}{K}\cos(Ks) + \frac{1}{K}.\end{aligned}$$

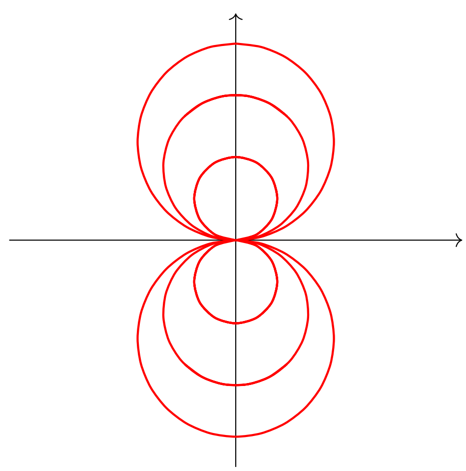


Figure 9: The parametrised curves with different signed curvatures $\kappa(s) = K$.

Note: This process is hard to do in general; we cannot even do it explicitly for $\kappa(s) = s$.

Exercise 16 Find a unit speed curve $\gamma : \mathbb{R} \to \mathbb{R}^2$ such that $\gamma(0) = (0,0)$ and $\gamma'(0) = (1,0)$ which has signed curvature $\kappa(s) = 0$ for every $s \in \mathbb{R}$ by using the formula (†).

Now, we spend some time developing our intuition of signed curvature to the point that we will be able to look at a given function $\kappa(s)$ and match it to a picture of a curve which should have that as its signed curvature. We have already seen that inflexion points relate to curvature; we now look at symmetries of κ and limits of κ as $s \to \pm\infty$.

Reminder: Let $f : D \to \mathbb{R}$ be any arbitrary function.

1. We say that f is even if $f(-x) = f(x)$ for all $x \in D$. A typical example is cos.
2. We say that f is odd if $f(-x) = -f(x)$ for all $x \in D$. A typical example is sin.

Phrased differently, the graph of an even function is symmetric across the y-axis and the graph of an odd function is symmetric under rotation by π (measured in *radians*, but 180° if you prefer) about the origin $(0,0)$. In fact, these apply directly to the signed curvature and γ.

Proposition 2.52 *Let $\gamma : I \to \mathbb{R}$ be a unit speed curve with signed curvature κ, where also $\gamma(0) = (0,0)$ and $\gamma'(s) = (1,0)$. Then, it has the following properties depending on κ.*

(i) *If κ is an even function, then γ is symmetric under reflection in the y-axis.*

(ii) *If κ is an odd function, then γ is symmetric under rotation by π about $(0,0)$.*

Proof: From the formula (†) and using that $\theta = 0$ under our assumptions, we see that

$$\varphi(-u) = \int_0^{-u} \kappa(w)\,\mathrm{d}w = -\int_0^{u} \kappa(-v)\,\mathrm{d}v,$$

where we define the new dummy variable $v := -w$ to get the minus sign down from the upper limit of integration. Therefore, we can make two conclusions from the above equalities:

- If κ is an even function, then $\varphi(-u) = -\varphi(u)$ which means φ is an odd function.
- If κ is an odd function, then $\varphi(u) = \varphi(u)$ which means φ is an even function.

Similarly, we see the following from the formula (†), remembering we have $(a,b) = (0,0)$ here:

$$\gamma_1(-s) = \int_0^{-s} \cos\varphi(u)\,\mathrm{d}u = -\int_0^{s} \cos\varphi(-v)\,\mathrm{d}v, \qquad \text{where } v := -u,$$
$$\gamma_2(-s) = \int_0^{-s} \sin\varphi(u)\,\mathrm{d}u = -\int_0^{s} \sin\varphi(-v)\,\mathrm{d}v, \qquad \text{where } v := -u.$$

We have two conclusions from each of these, which will give us the overall result in question:

- If κ is even, then φ is odd. Since cos is even and sin is odd, we see γ_1 is odd and γ_2 is even.
- If κ is odd, then φ is odd. Since cos is even and sin is odd, we see γ_1 is odd and γ_2 is odd.

Summarising these findings, we obtain the following:

$$\kappa \text{ even} \Rightarrow \gamma(-s) = \big(\gamma_1(-s), \gamma_2(-s)\big) = \big(-\gamma_1(s), \gamma_2(s)\big)\,,$$
$$\kappa \text{ odd} \Rightarrow \gamma(-s) = \big(\gamma_1(-s), \gamma_2(-s)\big) = \big(-\gamma_1(s), -\gamma_2(s)\big)\,.$$

Interpreting these geometrically, these are just the (i) reflection and (ii) rotation we expect. □

Exercise 17 Consider $\kappa(s) = s^2 - 1$. Determine the symmetries of the corresponding unit speed curve γ via Proposition 2.52 and find any inflexion points per Definition 2.47.

As is typical in mathematics, we often ask ourselves "how can we get something new from something old?" when we have gotten used to a notion. In this context, can we take a parametrised curve γ and generate from it some new parametrised curves? Yes, there are a number of different curves that we can define! Before we get there, we study a useful preliminary idea.

Definition 2.53 Let $\gamma : I \to \mathbb{R}^2$ be a regular curve. The centre of curvature at $t_0 \in I$ is

$$c(t_0) \coloneqq \gamma(t_0) + \frac{1}{\kappa(t_0)} n(t_0).$$

Note: The centre of curvature in Definition 2.53 only makes sense wherever $\kappa(t_0) \neq 0$.

Example 2.54 Consider the regular curve $\gamma : \mathbb{R} \to \mathbb{R}^2$ given by $\gamma(t) = (t, t^2)$. Recall we found in Exercise 12 (spoiler alert again) that the unit normal vector and signed curvature are

$$n(t) = \frac{1}{\sqrt{1+4t^2}}(-2t, 1) \qquad \text{and} \qquad \kappa(t) = \frac{2}{(1+4t^2)^{\frac{3}{2}}}.$$

The centre of curvature at $t_0 = 0$ is therefore found to be $c(0) = (0,0) + \frac{1}{2}(0,1) = (0, \frac{1}{2})$.

Exercise 18 Find the centre of curvature $c(1)$ of $\gamma : \mathbb{R} \to \mathbb{R}^2$ where $\gamma(t) = (t^2 - t, t + t^3)$.

We can give a neat geometric interpretation of the centre of curvature in terms of the behaviour of so-called *normal lines* to γ. But we first need a useful lemma to help us prove this result.

Reminder: A basis for $\mathbb{R}^2$ is a pair of vectors $\{\mathbf{u}, \mathbf{v}\}$ where every $\mathbf{x} = a\mathbf{u} + b\mathbf{v}$ for $a, b \in \mathbb{R}$.

Lemma 2.55 *For all $t \in I$, we have $n'(t) = -\kappa(t)\gamma'(t)$ and $u'(t) = \kappa(t)\|\gamma'(t)\|n(t)$.*

Proof: Because $u(t)$ and $n(t)$ are an *orthonormal* pair of vectors (meaning they are orthogonal and have unit length), we know that they form a basis for $\mathbb{R}^2$. In other words, we can write **any** vector in $\mathbb{R}^2$ in terms of u and n, i.e. $\mathbf{v} = \lambda(t)u(t) + \mu(t)n(t)$, where $\mathbf{v} \in \mathbb{R}^2$ is arbitrary and λ, μ are scalars (which depend on t smoothly). Applying this idea to the vector $n'(t)$, we can write

$$n'(t) = \lambda(t)u(t) + \mu(t)n(t).$$

If we take the dot product of this equation with $u(t)$, we see that

$$n'(t) \cdot u(t) = \lambda(t)\underbrace{u(t) \cdot u(t)}_{1} + \mu(t)\underbrace{n(t) \cdot u(t)}_{0} = \lambda(t),$$

the numbers coming from the orthonormality of u and n. Now, if we consider the orthogonality equation and differentiate it using the Product Rule (which applies to dot products, as we saw in the reminder-before-last), we get

$$n(t) \cdot u(t) = 0 \qquad \Rightarrow \qquad n'(t) \cdot u(t) + n(t) \cdot u'(t) = 0 \qquad \Rightarrow \qquad n'(t) \cdot u(t) = -u'(t) \cdot n(t).$$

Therefore, we can compare this to the previous equation which told us this dot product is $\lambda(t)$:

$$\begin{aligned}
\lambda(t) &= -u'(t) \cdot n(t) \\
&= -\frac{\mathrm{d}}{\mathrm{d}t}\left[\frac{\gamma'(t)}{\|\gamma'(t)\|}\right] \cdot n(t) \\
&= -\left(\frac{\gamma''(t)}{\|\gamma'(t)\|} - \gamma'(t)\frac{\mathrm{d}}{\mathrm{d}t}\left[\frac{1}{\|\gamma'(t)\|}\right]\right) \cdot n(t) \\
&= -\frac{\gamma''(t)}{\|\gamma'(t)\|} \cdot n(t) \\
&= -\kappa(t)\|\gamma'(t)\|,
\end{aligned}$$

where the fourth equality comes from the fact that n and u (so therefore γ') are orthogonal; the dot product with the complicated-looking derivative is just zero. It remains to do something similar to find the other scalar $\mu(t)$. Well, we see from orthonormality once more that

$$n'(t) \cdot n(t) = \lambda(t)\underbrace{u(t) \cdot n(t)}_{0} + \mu(t)\underbrace{n(t) \cdot n(t)}_{1} = \mu(t).$$

Again though, we can apply the Product Rule to the following dot product:

$$n(t) \cdot n(t) = 1 \qquad \Rightarrow \qquad n'(t) \cdot n(t) + n(t) \cdot n'(t) = 0 \qquad \Rightarrow \qquad n'(t) \cdot n(t) = 0.$$

Therefore, we know immediately that $\mu(t) = 0$. Substituting both λ and μ into the first expression above for n' gives us $n'(t) = -\kappa(t)\|\gamma'(t)\|u(t) + 0n(t)$, which is precisely the result we wanted to show. As for the second result, we can do something similar and write

$$u'(t) = \alpha(t)u(t) + \beta(t)n(t).$$

The details can be filled in very similarly to the argument for the first result (see Exercise 19) but the punchline is that we end up with the following scalars:

$$\alpha(t) = u'(t) \cdot u(t) = 0 \qquad \text{and} \qquad \beta(t) = u'(t) \cdot n(t) = -n'(t) \cdot u(t) = \kappa(t)\|\gamma'(t)\|,$$

where we use the first result to get the expression in the very last equality. □

Exercise 19 Follow the above proof and derive the expressions we got for $\alpha(t)$ and $\beta(t)$.

We can now state the theorem we hinted at earlier and use Lemma 2.55 to prove it.

Theorem 2.56 *Let $\gamma : I \to \mathbb{R}^2$ be a regular curve, fix $t_0 \in I$ and consider the* normal line *through $\gamma(t_0)$, that is the line through $\gamma(t_0)$ which is orthogonal to the tangent line at that point. If $\kappa(t_0) \neq 0$ and the distance $|t - t_0|$ is "sufficiently small", then the normal line intersects the normal line $\gamma(t)$ at t at the point $\alpha(t) \in \mathbb{R}^2$. Then,*

$$\lim_{t \to t_0} \alpha(t) = c(t_0).$$

Proof: We start by assuming the normal lines (which are themselves parametrised curves) have unit speed parametrisations. Because they are straight lines, we know they are described by a point they pass through and a direction they travel in. Say $N_t : \mathbb{R} \to \mathbb{R}^2$ is the normal line to γ at time t. Then, we can express the ones in question as

$$N_{t_0}(\mathfrak{s}) = \gamma(t_0) + \mathfrak{s}n(t_0) \qquad \text{and} \qquad N_t(s) = \gamma(s) + sn(t).$$

Their intersection point α is precisely when these two lines are equal. Note that because this depends on time t, the unit speed parameters $\mathfrak{s} = \mathfrak{s}(t)$ and $s = s(t)$ will also be functions of t.

So, we can write the following with unknown functions $\mathfrak{s}(t)$ and $s(t)$ which we will solve for:

$$\alpha(t) = N_{t_0}(\mathfrak{s}) = N_t(s) \qquad \Leftrightarrow \qquad \gamma(t_0) + \mathfrak{s}(t)n(t_0) = \gamma(t) + s(t)n(t).$$

Differentiating the above equation with respect to t (t_0 is fixed so is **not** itself a variable) gives

$$\mathfrak{s}'(t)n(t_0) = \gamma'(t) + s'(t)n(t) + s(t)n'(t).$$

If we then take the limit as $t \to t_0$, we obtain the equation

$$\mathfrak{s}'(t_0)n(t_0) = \gamma'(t_0) + s'(t_0)n(t_0) + s(t_0)n'(t_0).$$

Now, take the dot product of this with $u(t_0)$ and use orthonormality to conclude that

$$\begin{aligned} 0 &= \gamma'(t_0) \cdot u(t_0) + 0 + s(t_0)n'(t_0) \cdot u(t_0) \\ &= \gamma'(t_0) \cdot \frac{\gamma'(t_0)}{\left\|\gamma'(t_0)\right\|^2} - s(t_0)\kappa(t)\gamma'(t_0) \cdot \frac{\gamma'(t_0)}{\left\|\gamma'(t_0)\right\|^2} \\ &= \left\|\gamma'(t_0)\right\| - s(t_0)\kappa(t)\left\|\gamma'(t_0)\right\|. \end{aligned}$$

Notice that we used the definition of $u(t)$ and the expression for $n'(t)$ from Lemma 2.55 in the second equality above; we also used $\mathbf{v} \cdot \mathbf{v} = \|\mathbf{v}\|^2$ for a general vector. By regularity, $\left\|\gamma'(t_0)\right\| > 0$, so we can divide by it. Consequently, rearranging the above produces

$$s(\tau) = \frac{1}{\kappa(t_0)}.$$

We finally obtain precisely the limit that we want, namely

$$\lim_{t \to t_0} \alpha(t) = \gamma(t) + \frac{1}{\kappa(t_0)} n(\tau) = c(t_0). \qquad \square$$

Note: Let's again take a breather to assess what we have shown in Theorem 2.56 (and be assured that the proof isn't vital so understanding it can be left for another day). If we have two points on a regular curve γ that are "sufficiently close", then their normal lines will intersect. But as we push one point closer to the other, the intersection of their normal lines will slowly creep towards the centre of curvature.

The point now is to define a curve which passes through **every** centre of curvature of a curve γ; the definition is only well-defined provided its signed curvature is everywhere non-vanishing (it doesn't equal zero at any point).

Definition 2.57 Let $\gamma : I \to \mathbb{R}^2$ be a regular curve. Its evolute $E_\gamma : I \to \mathbb{R}^2$ is the curve

$$E_\gamma(t) = \gamma(t) + \frac{1}{\kappa(t)} n(t).$$

Remark 2.58 As mentioned, Definition 2.57 is well-defined everywhere with $\kappa(t) \neq 0$. In fact, this means γ must have **no** inflexion points. If γ **does** have inflexion points (and signed curvature is therefore zero), then the geometric picture of E_γ is that it will shoot off to infinity!

Exercise 20 (Harder) Is the evolute E_γ itself always regular? Prove it or explain why not.

[**Hint:** Although this is the standard thing to do, I'll write out what must be done: look at the derivative E'_γ and analyse when this is $(0,0)$; the key is to look at the expression involving κ and κ'.]

Example 2.59 Consider the curve $\gamma : \mathbb{R} \to \mathbb{R}^2$ given by $\gamma(t) = (a\cos t, \sin t)$ for $a > 1$. We will construct the evolute; we need the signed curvature and the unit normal vector. Indeed,

$$\gamma'(t) = (-a\sin t, \cos t) \qquad \Rightarrow \qquad n(t) = \frac{1}{\left\|\gamma'(t)\right\|}(-\cos t, -a\sin t),$$

where we won't bother actually writing out $\left\|\gamma'(t)\right\|$ in full right now because it isn't the nicest expression (but we sort-of do so at the end anyway). The only remaining stepping stone is to find the acceleration; we can then use the signed curvature formula from Lemma 2.45 as usual:

$$\gamma''(t) = (-a\cos t, -\sin t) \qquad \Rightarrow \qquad \kappa(t) = \frac{a}{\left\|\gamma'(t)\right\|^3}.$$

Substituting everything into Definition 2.57, we obtain

$$E_\gamma(t) = (a\cos t, \sin t) + \frac{\left\|\gamma'(t)\right\|^3}{a} \frac{1}{\left\|\gamma'(t)\right\|}(-\cos t, -a\sin t).$$

This is a bit messy, but if we now write out $\left\|\gamma'(t)\right\|^2 = a^2\sin^2(t) + \cos^2(t)$ and use a trigonometric identity (I don't expect you to know these at all, but I do want to show you a nice-looking answer for a change), the expression above becomes the still-complicated but much prettier

$$E_\gamma(t) = \frac{a^2-1}{a}(\cos^3 t, -a\sin^3 t).$$

Note: As $a \to 1$, the curve γ degenerates to a circle and the evolute E_γ to the point $(0,0)$.

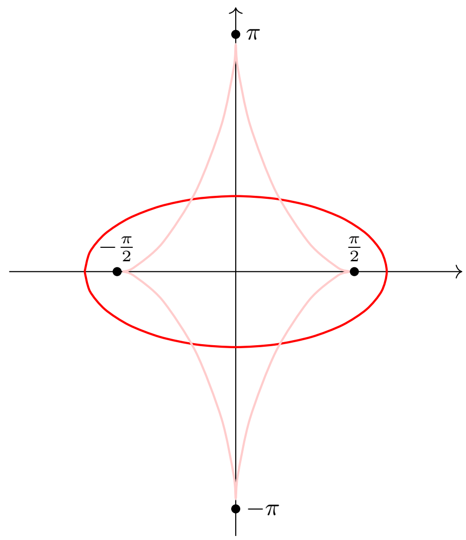


Figure 10: The parametrised curve $\gamma(t) = (2\cos t, \sin t)$ and its evolute E_γ.

Exercise 21 Construct the evolute of the regular curve $\gamma : \mathbb{R} \to \mathbb{R}^2$ given by $\gamma(t) = (t, t^2)$.

We have a number of geometric quantities such as the unit normal and tangent vectors, not forgetting the arc length function we introduced all the way back in Definition 2.13. The goal of the next result is to relate such geometric quantities of the evolute back to the curve itself.

Theorem 2.60 *Let $\gamma : I \to \mathbb{R}^2$ be a regular curve with evolute E_γ and suppose that the unit tangent vector, unit normal vector and arc length function based at $t_0 \in I$ of the* **evolute** *are denoted u^E, n^E and $\sigma^E_{t_0}$ respectively. If $\kappa'(t) < 0$ for all $t \in I$, we have the following:*

$$u^E(t) = n(t) \qquad \textit{and} \qquad n^E(t) = -u(t) \qquad \textit{and} \qquad \sigma^E_{t_0}(t) = \frac{1}{\kappa(t)} - \frac{1}{\kappa(t_0)}.$$

Proof: From Exercise 20, we have an expression for the derivative of the evolute:

$$E'_\gamma(t) = -\frac{\kappa'(t)}{\kappa(t)^2} n(t).$$

Because $\kappa'(t) < 0$ is always negative, we conclude that

$$\left\| E'_\gamma(t) \right\| = \frac{-\kappa'(t)}{\kappa(t)^2} > 0 \qquad \Rightarrow \qquad u^E(t) := \frac{E'_\gamma(t)}{\left\| E'_\gamma(t) \right\|} = n(t).$$

Next, $n^E(t) := (-u^E_2(t), u^E_1(t))$, but we can refer to the expression we have above to see that

$u_1^E(t) = n_1(t)$ and $u_2^E(t) = n_2(t)$. Hence, it is straightforward to get the result in question:

$$n^E(t) = \big(-n_2(t), n_1(t)\big) = -\big(u_1(t), u_2(t)\big) = -u(t).$$

Finally, the arc length function is given by the integral of speed (which we wrote above). Thus,

$$\begin{aligned} \sigma_{t_0}^E(t) &= \int_{t_0}^{t} -\frac{\kappa'(u)}{\kappa(u)^2}\,\mathrm{d}u \\ &= \int_{t_0}^{t} \frac{\mathrm{d}}{\mathrm{d}u}\left[\frac{1}{\kappa(u)}\right]\mathrm{d}u \\ &= \frac{1}{\kappa(t)} - \frac{1}{\kappa(t_0)}, \end{aligned}$$

by definition of an anti-derivative and an integral; this is the final thing we had to prove. □

Exercise 22 (Continued) Write the unit tangent vector, unit normal vector and arc length function of the evolute E_γ for the curve $\gamma : \mathbb{R} \to \mathbb{R}^2$ where $\gamma(t) = (t, t^2)$ from Exercise 21.

There is a sort-of inverse to taking the evolute of a curve. Before we define it, let's have the following picture in mind: imagine we have a piece of string stuck along a curve γ between two points $\gamma(t_0)$ and $\gamma(t_1)$, but we manage to tug the end at $\gamma(t_0)$ so we can start pulling it. If we keep the string perfectly taut, the released section of string as we pick it off of the curve will be tangent to γ at a point in the interval between the endpoints, namely $t \in (t_0, t_1)$.

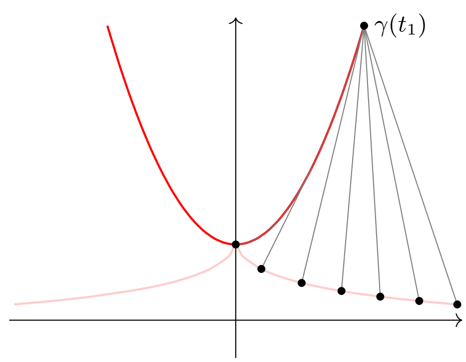


Figure 11: The geometric intuition behind the so-called *involute*.

In Figure 11, we do this "string approach" for the right-hand part of the curve (it works the same for the left but I find it easier to visualise just for a chunk of it). In doing this, the length of the released bit of string is the arc length along γ from t_0 to this value t, i.e. this is given by $\sigma_{t_0}(t)$. Since this is always a straight line touching the curve at $\gamma(t)$ and pointing away in the direction $u(t)$ at length $\sigma_{t_0}(t)$, we have a way to describe the resulting curve explicitly.

Definition 2.61 Let $\gamma : I \to \mathbb{R}$ be a regular curve and $t_0 \in I$. Its involute $I_\gamma : I \to \mathbb{R}^2$ is

$$I_\gamma(t) = \gamma(t) - \sigma_{t_0}(t)u(t).$$

Remark 2.62 In the case that γ is a unit speed curve, we know that $\sigma_{t_0}(t) = t - t_0$. Therefore, the equation for the involute given above reduces to $I_\gamma(t) = \gamma(t) - (t - t_0)\gamma'(t)$.

Note: Although γ is a unit speed curve above, we do **not** use the time parameter s because it is generally **not** true that its involute I_γ will also be a unit speed curve.

Example 2.63 We construct the involute at $t = 0$ of $\gamma : \mathbb{R} \to \mathbb{R}^2$ given by $\gamma(t) = (t, \cosh t)$; if you haven't come across such a function before, this is like a *hyperbolic analogue* of the function cos. We won't go into it here but there are a few noteworthy rules that it satisfies for every t:

$$\cosh'(t) = \sinh(t), \qquad \sinh'(t) = \cosh'(t), \qquad \cosh^2(t) - \sinh^2(t) = 1.$$

We see that $\gamma'(t) = (1, \sinh t)$ and so the speed is $\|\gamma'(t)\| = \sqrt{1 + \sinh^2(t)} = \cosh(t)$. Hence,

$$\sigma_0(t) = \int_0^t \cosh(u)\,\mathrm{d}u = \sinh(u).$$

Substituting everything into Definition 2.61, we get the expression

$$I_\gamma(t) = (t, \cosh t) - \frac{\sinh(t)}{\cosh(t)}(1, \sinh t).$$

We could write this more succinctly as $I_\gamma(t) = (t - \tanh t, \operatorname{sech} t)$. If you are comfortable with this, fantastic! If not, try not to dwell on this; if any of these functions come up again, we will explain what is needed beforehand.

Exercise 23 Construct the involute of $\gamma : \mathbb{R} \to \mathbb{R}^2$ given by $\gamma(t) = (\cos t, \sin t)$ at $t_0 = 0$.

We once again include some spoilers for an earlier exercise: you should have found in Exercise 20 that the evolute E_γ is **not always** itself a regularly parametrised curve. That said, it might be. However, we now quickly prove that the involute I_γ is **never** a regular curve.

Proposition 2.64 *The involute I_γ of a regular curve γ is* **never** *a regular curve itself.*

Proof: If we differentiate the defining equation of an involute, we see that

$$\begin{aligned} I'_\gamma(t) &= \gamma'(t) - \sigma'_{t_0}(t)u(t) - \sigma_{t_0}(t)u'(t) \\ &= \gamma'(t) - \|\gamma'(t)\|u(t) - \sigma_{t_0}(t)\kappa(t)\|\gamma'(t)\|n(t) \\ &= \gamma'(t) - \gamma'(t) - \sigma_{t_0}(t)\kappa(t)\|\gamma'(t)\|n(t) \\ &= -\sigma_{t_0}(t)\kappa(t)\|\gamma'(t)\|n(t). \end{aligned}$$

Note we used Lemma 2.55 to get the expression for u' in the second equality. Now, we see that $I_\gamma(t) = \mathbf{0}$ if and only if $\sigma_{t_0}(t) = 0$ or $\kappa(t) = 0$; note that γ is assumed to be regular which means both $\|\gamma'(t)\| > 0$ for all t and $n(t) \neq 0$. In particular then, $I_\gamma(t_0) = 0$ because $\sigma_{t_0}(t_0) = 0$. □

Note: To use our earlier string analogy, an involute also has zero derivative where the contact point between the string and the curve is an inflection point (since $\kappa(t) = 0$ here).

Much like how the Fundamental Theorem of Calculus tells us a **one-way** (sort-of) relationship between derivatives and integrals, namely that the derivative of an integral is the original function, we wish to relate the evolute and involute of a curve. First, we will find formulae for the unit normal vector and signed curvature of an involute in terms of geometric quantities of the original curve.

Lemma 2.65 *Let $\gamma : I \to \mathbb{R}^2$ be a unit speed curve with involute I_γ based at $t_0 \in I$ and suppose that the unit normal vector and signed curvature of the* **involute** *are denoted n^I and κ^I respectively. Then, we have the following:*

$$n^I(t) = \frac{(t-t_0)\kappa(t)}{|t-t_0|\,|\kappa(t)|}u(t) \qquad \text{and} \qquad \kappa^I(t) = \frac{\kappa(t)}{|\kappa(t)|}\frac{1}{|t-t_0|}.$$

Proof: Since γ has unit speed, the formula for I'_γ above can be written much simpler:

$$I'_\gamma(t) = -(t-t_0)\kappa(t)n(t) = -(t-t_0)k(t),$$

where k is the curvature vector; this follows from the alternate definition of the signed curvature given immediately below Definition 2.44. Therefore, the involute has unit tangent vector

$$u^I(t) := \frac{I'_\gamma(t)}{\|I'_\gamma(t)\|} = \frac{(t-t_0)\kappa(t)}{|t-t_0|\,|\kappa(t)|}n(t).$$

Note that this uses the first equality for I'_γ, not the one involving k (we need that soon though).

Continuing on, we can immediately find the expression for the unit normal vector we want:

$$n^I(t) := \big(-u_2(t), u_1(t)\big) = \frac{(t-t_0)\kappa(t)}{|t-t_0|\big|\kappa(t)\big|}u(t).$$

As for the signed curvature, we require the acceleration; we get this by differentiating the second part of the equation for I'_γ (the one including k):

$$I''_\gamma(t) = -k(t) - (t-t_0)k'(t).$$

Using the expression for I'_γ at the start of the proof, this tells us that

$$\Big\|I'_\gamma(t)\Big\|^2 = \big\|-(t-t_0)k(t)\big\|^2 = |t-t_0|^2\big\|k(t)\big\|^2 = |t-t_0|^2\big|\kappa(t)\big|^2,$$

recalling that the norm of the curvature vector is the absolute value of the signed curvature (the note after Definition 2.44). Applying the usual formula for the signed curvature (Lemma 2.45),

$$\begin{aligned}
\kappa^I(t) &= \frac{I''_\gamma(t)\cdot n^I(t)}{\Big\|I'_\gamma(t)\Big\|^2} \\
&= \frac{-k(t)-(t-t_0)k'(t)}{|t-t_0|^2\big|\kappa(t)\big|^2}\cdot\frac{(t-t_0)\kappa(t)}{|t-t_0|\big|\kappa(t)\big|}u(t) \\
&= -\frac{\kappa(t)}{|t-t_0|\big|\kappa(t)\big|^3}k'(t)\cdot u(t),
\end{aligned}$$

since $k(t)\cdot u(t) = 0$; this is essentially seen in Lemma 2.37. However, we can get an expression for $k'(t)\cdot u(t)$ by differentiating this orthogonality equation using the Product Rule:

$$k(t)\cdot u(t) = 0 \qquad \Rightarrow \qquad k'(t)\cdot u(t) + k(t)\cdot u'(t) = 0 \qquad \Rightarrow \qquad k'(t)\cdot u(t) = -u'(t)\cdot k(t).$$

As γ is a unit speed curve, $u = \gamma'$ and $u' = \gamma'' = k$, so the final equation above simplifies to $k'(t)\cdot u(t) = -\big\|k(t)\big\|^2 = -\kappa(t)^2$. Finally, substituting this into the expression for κ^I gives us

$$\kappa^I(t) = \frac{\kappa(t)^3}{|t-t_0|\big|\kappa(t)\big|^3} = \frac{\kappa(t)}{\big|\kappa(t)\big|}\frac{1}{|t-t_0|}.$$ □

Note: A few times in the above proof, we used the property that whenever $f(t) > 0$, then

$$\frac{f(t)}{\big|f(t)\big|} = 1.$$

We can now state and prove (rather easily) the relationship between evolute and involute.

Theorem 2.66 *Let $\gamma : I \to \mathbb{R}^2$ be a regular curve with involute I_γ. Then, the evolute E_{I_γ} of the involute I_γ is just the original regular curve γ.*

Proof: Let γ be a unit speed curve (see Theorem 2.28). Then, its involute at $t_0 \in I$ has evolute

$$\begin{aligned} E_{I_\gamma}(t) &= I(t) + \frac{1}{\kappa^I(t)} n^I(t) \\ &= \gamma(t) - (t-t_0)u(t) + \left(\frac{|\kappa(t)|}{\kappa(t)}|t-t_0|\right)\left(\frac{(t-t_0)\kappa(t)}{|t-t_0||\kappa(t)|}u(t)\right) \\ &= \gamma(t) - (t-t_0)u(t) + (t-t_0)u(t) \\ &= \gamma(t). \end{aligned}$$

□

This shows that the evolute "undoes" the involute, in the same way that differentiation "undoes" integration per the Fundamental Theorem of Calculus. But what happens when we reverse things, that is taking the involute of the evolute? In fact, the analogy to calculus serves us well because it is similar to that.

Reminder: When we differentiate a function f and integrate the resulting function f', we *almost* get back the original function, but there may be a non-zero constant $f(x) + c$ that appears. One can think of this as a "parallel" function, as it is a shift of the original .

We now make this *converse* way of doing things precise mathematically by introducing the notion of a *parallel curve*. We can then prove that the involute of an evolute is, at worst, parallel to the original curve in this sense.

Definition 2.67 Let $\gamma : I \to \mathbb{R}^2$ be a regular curve. For $\lambda \in \mathbb{R}$, a parallel curve to λ is a curve $\gamma_\lambda : I \to \mathbb{R}^2$ defined as $\gamma_\lambda(t) := \gamma(t) + \lambda n(t)$.

It turns out that parallel curves can be regular under a certain condition.

Lemma 2.68 *The parallel curve γ_λ to a regular curve γ is itself a regularly parametrised curve if and only if $\kappa(t) \neq \frac{1}{\lambda}$ for any $t \in I$.*

Proof: Differentiating the defining equation in Definition 2.67, $\gamma_\lambda'(t) = \gamma'(t) + \lambda n'(t)$. So Lemma 2.55 implies $\gamma_\lambda'(t) = \gamma'(t) - \lambda\kappa(t)\gamma'(t)$. Hence, $\gamma_\lambda'(t) = \mathbf{0}$ if and only if $\lambda\kappa(t) = 1$, i.e. $\kappa(t) = \frac{1}{\lambda}$. □

Example 2.69 Consider the curve $\gamma : \mathbb{R} \to \mathbb{R}^2$ given by $\gamma(t) = (t, t^2)$. We shall construct the general parallel curve to this, that is we will keep λ floating about instead of choosing a particular number for it. All we must do is construct the unit normal vector:

$$\gamma'(t) = (1, 2t) \qquad \Rightarrow \qquad n(t) = \frac{1}{\sqrt{1+4t^2}}(-2t, 1).$$

Therefore, the general parallel curve to this regular curve is given by

$$\gamma_\lambda(t) = (t, t^2) + \frac{\lambda}{\sqrt{1+4t^2}}(-2t, 1) = \left(t - \frac{2\lambda t}{\sqrt{1+4t^2}}, t^2 + \frac{\lambda}{\sqrt{1+4t^2}}\right).$$

Exercise 24 Suppose $\gamma : \mathbb{R} \to \mathbb{R}^2$ is given by $\gamma(t) = (t, t^2)$ as in Example 2.69. Continuing from the example, deduce which values of λ ensure that the parallel curve γ_λ is regular.

[**Hint:** You have already found the signed curvature κ in Exercise 12; just use this with Lemma 2.68.]

We can now prove the reverse relationship between the involute and evolute of a curve.

Theorem 2.70 *Let $\gamma : I \to \mathbb{R}^2$ be a regular curve with evolute E_γ. Then, the involute I_{E_γ} of the evolute E_γ is a parallel curve to the original regular curve γ.*

Proof: Let $\kappa'(t) < 0$ for all t (a similar argument works for $\kappa'(t) > 0$). For fixed $t_0 \in I$, recall

$$u^E(t) = n(t) \qquad \text{and} \qquad \sigma^E_{t_0}(t) = \frac{1}{\kappa(t)} - \frac{1}{\kappa(t_0)}$$

from Theorem 2.60. Then, the evolute has involute based at t_0 given by

$$\begin{aligned} I_{E_\gamma}(t) &= E_\gamma(t) - \sigma^E_{t_0}(t)u^E(t) \\ &= \gamma(t) + \frac{1}{\kappa(t)}n(t) - \left(\frac{1}{\kappa(t)} - \frac{1}{\kappa(t_0)}\right)n(t) \\ &= \gamma(t) - \frac{1}{\kappa(t_0)}n(t), \end{aligned}$$

but this is nothing more than Definition 2.67 where we set $\lambda = \frac{1}{\kappa(t_0)}$. □

2.3 Space Curves

We have done an awful lot of work for curves in two-dimensional space. That said, it's about time we kick it up a notch: we now look at *space* curves, i.e. ones that live in three-dimensional space. Okay, it is another dimension to worry about but at least we can easily visualise it (e.g. the world we live in is three-dimensional). The benefit of looking now at three dimensions is this: given an ordered pair of vectors, we can uniquely determine a third vector which is orthogonal to both; the method we use here is only possible in dimensions three and seven (yes, seven)!

Reminder: Let $\mathbf{u}, \mathbf{v} \in \mathbb{R}^3$. Their vector product is the somewhat horrible-looking vector $\mathbf{u} \times \mathbf{v} := (u_2v_3 - u_3v_2, u_3v_1 - u_1v_3, u_1v_2 - u_2v_1)$. We list a number of important properties of the vector product which will be useful in the discussion that is to come:

1. $\mathbf{u} \times \mathbf{v} = -\mathbf{v} \times \mathbf{u}$.
2. $(\lambda\mathbf{u} + \mu\mathbf{v}) \times \mathbf{w} = \lambda(\mathbf{u} \times \mathbf{w}) + \mu(\mathbf{v} \times \mathbf{w})$ for all $\lambda, \mu \in \mathbb{R}$.
3. If $\mathbf{u}$ and $\mathbf{v}$ are parallel, meaning $\mathbf{u} = \lambda\mathbf{v}$ for some $\lambda \in \mathbb{R}$, then $\mathbf{u} \times \mathbf{v} = \mathbf{0}$.
4. $\|\mathbf{u} \times \mathbf{v}\|^2 = \|\mathbf{u}\|^2\|\mathbf{v}\|^2 - (\mathbf{u} \cdot \mathbf{v})^2$.
5. $\mathbf{u} \cdot (\mathbf{v} \times \mathbf{w}) = \mathbf{v} \cdot (\mathbf{w} \times \mathbf{u}) = \mathbf{w} \cdot (\mathbf{u} \times \mathbf{v})$.
6. $\mathbf{u} \times \mathbf{v}$ is orthogonal to both $\mathbf{u}$ and $\mathbf{v}$.

We will now define analogues to some of the objects we have seen in Section 2.2 and this will be three vector quantities with any two orthogonal that we discussed before the above reminder.

Definition 2.71 Let $\gamma : I \to \mathbb{R}^3$ be a regularly parametrised space curve whose curvature **never** vanishes, that is $\|k(t)\| \neq 0$ for all $t \in I$. The Frenet frame of γ is the ordered triplet $[u(t), n(t), b(t)]$, where the unit tangent vector $u(t)$, the principal unit normal vector $n(t)$ and the binormal vector $b(t)$ are defined as follows:

$$u(t) := \frac{\gamma'(t)}{\|\gamma'(t)\|}, \qquad n(t) := \frac{k(t)}{\|k(t)\|}, \qquad b(t) := u(t) \times n(t).$$

It is clear u is defined the same here as in Definition 2.39. Notice however that n is analogous to Definition 2.42 but is slightly different; we have to use the notion of curvature as the norm of the curvature vector, and we call it the *principal* unit normal vector because there are infinitely-many unit vectors that are orthogonal to u (this is because we are now in $\mathbb{R}^3$, not just $\mathbb{R}^2$).

Note that the unit tangent vector does **not** have a principal direction, because there are only two choices: pointing in the direction of the curve's movement and pointing opposite to it. We don't ever really care about the backwards one. The binormal vector is then fixed because we have decided on u and n, so there is again **no** principal direction for this.

Lemma 2.72 *The Frenet frame is orthonormal, i.e. any two vectors of u, n, b are orthogonal and they all have unit length.*

Proof: It is clear that $\|u(t)\| = 1$ and $\|n(t)\| = 1$ by definition. Furthermore, we know $k(t)$ is orthogonal to $u(t)$ by Lemma 2.37, and we know that $n(t)$ and $u(t)$ are orthogonal. We therefore conclude that $n(t)$ is parallel to $k(t)$. There are only two things left to prove: (i) $\|b(t)\| = 1$ and (ii) $b(t)$ is orthogonal to each of $u(t)$ and $n(t)$. But these both follow from the above reminder. Indeed, (ii) is automatic by **6.** and (i) follows immediately from **4.**, because the formula says

$$\|b(t)\|^2 = \|u(t) \times n(t)\|^2 = \underbrace{\|u(t)\|^2}_{1} \underbrace{\|n(t)\|^2}_{1} - (\underbrace{u(t) \cdot n(t)}_{0})^2. \qquad \square$$

Note: Hence, the Frenet frame of any such regular curve is an orthonormal basis for $\mathbb{R}^3$.

Example 2.73 Consider the curve $\gamma : \mathbb{R} \to \mathbb{R}^3$ given by $\gamma(t) = (\frac{1}{2}t^2, \frac{1}{3}t^3, t)$. We shall construct the Frenet frame at $t = 0$; this is only possible because $k(0) \neq \mathbf{0}$. Indeed, we will actually use the succinct expression for $k(t)$ given at the end of Section 2.1. First,

$$\gamma'(t) = (t, t^2, 1) \qquad \text{and} \qquad \gamma''(t) = (1, 2t, 0).$$

Thus, at $t = 0$, we see that $\gamma'(0) = (0, 0, 1)$ and $\gamma''(0) = (1, 0, 0)$ which have unit length already. Using Definition 2.40 for the normal projection, $\gamma''_\perp(0) = \gamma''(0) = (1, 0, 0)$ in this case. As such,

$$k(0) = k(t) = \frac{\gamma''_\perp(0)}{\|\gamma'(0)\|^2} = (1, 0, 0).$$

Because this is non-zero, we know that the Frenet frame is well-defined. Thus, we see that

$$u(0) = (0, 0, 1), \qquad n(0) = (1, 0, 0), \qquad b(0) = (0, 0, 1) \times (1, 0, 0) = (0, 1, 0).$$

Exercise 25 Consider the curve $\gamma : \mathbb{R} \to \mathbb{R}^3$ given by $\gamma(s) = \frac{1}{5}(4s, \cos 3s, \sin 3s)$.

(i) Show that γ is a unit speed curve of non-vanishing curvature.

(ii) Construct the Frenet frame $[u(s), n(s), b(s)]$ of this curve.

In this part of the module, we want to describe the geometry of space curves by analysing the Frenet frame. This process is simplified if we assume our curves have unit speed (Theorem 2.28).

Note: The Frenet frame for a unit speed curve $\gamma : I \to \mathbb{R}^3$ consists of the three vectors

$$u(s) = \gamma'(s), \qquad n(s) = \frac{\gamma''(s)}{\|\gamma''(s)\|}, \qquad b(s) = u(s) \times n(s).$$

Notation 2.74 Within this discussion, it is conventional to denote the *scalar* curvature $\|k(s)\|$ by the symbol $\kappa(s)$. This should **not** be confused with the *signed* curvature; we have seen previously that $\|k(s)\|$ is the absolute value of the signed curvature. We call this an "abuse of notation"!

Note: As $k(s) = \gamma''(s) = u'(s)$ for a unit speed curve, we re-write the definition of $n(s)$ as

$$u'(s) = \kappa(s)n(s).$$

Because the Frenet frame is a basis for (and in particular spans) $\mathbb{R}^3$, then it should be possible to find formulae similar to the above for $n'(s)$ and $b'(s)$. In a similar vein as to Lemma 2.55, we use the fact that we can write any three-dimensional vector $\mathbf{v} = \lambda(s)u(s) + \mu(s)n(s) + \nu(s)b(s)$ where λ, μ, ν are scalars (functions here because they depend on s); this is the definition of spanning.

Lemma 2.75 *For all $s \in I$, we have $b'(s) = \mu(s)n(s)$ for some smooth $\mu : I \to \mathbb{R}$.*

Proof: Because $b'(s) \in \mathbb{R}^3$, it means there exist smooth functions λ, μ, ν such that

$$b'(s) = \lambda(s)u(s) + \mu(s)n(s) + \nu(s)b(s).$$

We take the dot product of this with each of b and u in turn and use orthonormality established from Lemma 2.72. First thing's first, we will take the dot product with b, which means

$$b'(s) \cdot b(s) = \lambda(s)\underbrace{u(s) \cdot b(s)}_{0} + \mu(s)\underbrace{n(s) \cdot b(s)}_{0} + \nu(s)\underbrace{b(s) \cdot b(s)}_{1} = \nu(s).$$

On the other hand, we know b has unit length; the Product Rule therefore gives us this:

$$b(s) \cdot b(s) = 1 \qquad \Rightarrow \qquad b'(s) \cdot b(s) + b(s) \cdot b'(s) = 0 \qquad \Rightarrow \qquad b'(s) \cdot b(s) = 0.$$

We conclude that $\nu(s) = 0$. Similarly, we take the dot product of b' with u and see that

$$b'(s) \cdot u(s) = \lambda(s)\underbrace{u(s) \cdot u(s)}_{1} + \mu(s)\underbrace{n(s) \cdot u(s)}_{0} + \nu(s)\underbrace{b(s) \cdot u(s)}_{0} = \lambda(s).$$

Again, we know b and u are orthogonal; the Product Rule again helps us out:

$$b(s) \cdot u(s) = 0 \qquad \Rightarrow \qquad b'(s) \cdot u(s) + b(s) \cdot u'(s) = 0 \qquad \Rightarrow \qquad b'(s) \cdot u(s) = -u'(s) \cdot b(s).$$

But we know that $u'(s) = \kappa(s)n(s)$ from the note above. If we substitute this into the above equation, we see that we actually have a dot product between n and b, but they are orthogonal so this is again zero. Hence, $\lambda(s) = 0$, so all we are left with at the start is $b'(s) = \mu(s)n(s)$. $\square$

Definition 2.76 Let $\gamma : I \to \mathbb{R}^3$ be a unit speed curve of non-vanishing curvature. The torsion of the curve is the function $\tau : I \to \mathbb{R}$ defined by the equation $b'(s) = -\tau(s)n(s)$.

Example 2.77 Consider the curve $\gamma : I \to \mathbb{R}^3$ given by $\gamma(s) = (\cos\frac{s}{\sqrt{2}}, \sin\frac{s}{\sqrt{2}}, \frac{s}{\sqrt{2}})$. We will compute the torsion of this curve. But first, we must verify that it is a unit speed curve and has everywhere-non-zero curvature. Indeed, the fact it is unit speed follows quickly:

$$\gamma'(s) = \frac{1}{\sqrt{2}}\left(-\sin\frac{s}{\sqrt{2}}, \cos\frac{s}{\sqrt{2}}, 1\right) \qquad \Rightarrow \qquad \|\gamma'(s)\| = 1.$$

But since it is unit speed, we know that $k(s) = \gamma''(s)$; the curvature vector is

$$k(s) = \frac{1}{2}\left(-\cos\frac{s}{\sqrt{2}}, -\sin\frac{s}{\sqrt{2}}, 0\right) \neq \mathbf{0}$$

because sin and cos are never simultaneously zero. Therefore, the curvature is non-vanishing and the torsion is well-defined. To find τ, we must first find the binormal vector, which means we will compute the whole Frenet frame. This is done using the simplified expressions noted earlier that apply when our curve has unit speed:

$$u(s) = \frac{1}{\sqrt{2}}\left(-\sin\frac{s}{\sqrt{2}}, \cos\frac{s}{\sqrt{2}}, 1\right),$$
$$n(s) = \left(-\cos\frac{s}{\sqrt{2}}, -\sin\frac{s}{\sqrt{2}}, 0\right),$$
$$b(s) = \frac{1}{\sqrt{2}}\left(\sin\frac{s}{\sqrt{2}}, -\cos\frac{s}{\sqrt{2}}, 1\right).$$

It remains to differentiate b and compare it to n; this will tell us what τ is. Indeed then,

$$b'(s) = \frac{1}{2}\left(\cos\frac{s}{\sqrt{2}}, \sin\frac{s}{\sqrt{2}}, 0\right) = -\frac{1}{2}n(s) \qquad \Rightarrow \qquad \tau(s) = \frac{1}{2}.$$

Exercise 26 (Continued) Let $\gamma : \mathbb{R} \to \mathbb{R}^3$ with $\gamma(s) = \frac{1}{5}(4s, \cos 3s, \sin 3s)$ from Exercise 25.

(i) Determine its torsion τ.

(ii) Verify that you get the same answer if you instead compute $-b'(s) \cdot n(s)$.

(iii) Does $\tau(s) = -b'(s) \cdot n(s)$ for **any** unit speed curve of non-vanishing curvature?

Let's discuss the geometric meaning of torsion. Recall that the curvature measure the rate of change of the direction of the tangent line to the curve; torsion has a similar interpretation but in terms of *planes* instead of lines. **If this discussion is a bit much, just skip to Lemma 2.80**.

Reminder: A plane in $\mathbb{R}^3$ is any subset spanned by two orthogonal vectors. In other words, we have two vectors $\{\mathbf{u}, \mathbf{v}\}$ whose dot product is zero such that every vector in said plane can be uniquely represented as $\lambda\mathbf{u} + \mu\mathbf{v}$ for some $\lambda, \mu \in \mathbb{R}$.

At some point $\gamma(s_0)$ on the curve, we define the *osculating plane* of the curve to be the plane through $\gamma(s_0)$ spanned by the vectors $\{u(s_0), n(s_0)\}$. Because γ has unit speed, this is equivalent to the plane being spanned by the vectors $\{\gamma'(s_0), \gamma''(s_0)\}$. In general, the osculating plane is **not** tangent to the curve; it actually slices through it.

Remark 2.78 The name of the plane is not misspelt: "osculating" is derived from Latin and means "to kiss", whereas "oscillating" means "to move back and forth". The reason we call the plane here osculating is because it just 'kisses' the curve (at the point mentioned in the reminder above about tangent planes).

Note: The plane's orientation is uniquely determined by any normal vector, e.g. $b(s_0)$.

Because γ is smooth, we can consider the *Taylor expansion* about the time $s = s_0$. We haven't introduced this at all so don't worry if this is new to you. The punchline is that the failure for $\gamma(s)$ to be in the osculating plane is controlled by $\gamma'''(s_0)$ (sometimes called the *jerk* vector). Any plane splits $\mathbb{R}^3$ into two parts. Applying this logic to the osculating plane, we can classify any vector $\mathbf{v} \in \mathbb{R}^3$ that is **not** in the osculating plane as follows:

(i) We call $\mathbf{v}$ *positive* if the dot product $\mathbf{v} \cdot b(s_0) > 0$.

(ii) We call $\mathbf{v}$ *negative* if the dot product $\mathbf{v} \cdot b(s_0) < 0$.

Thinking of the osculating plane as horizontal, the normal vector $b(s_0)$ points "upwards" (this is the standard picture to have in mind about any "normal" to a curve or plane or whatever). Therefore, the positive vectors are on the upside of the plane and the negative vectors are on the downside of the plane. Hence, we have this interpretation:

(i) If $\gamma'''(s_0) \cdot b(s_0) > 0$, the curve goes through the plane from below upwards.

(ii) If $\gamma'''(s_0) \cdot b(s_0) < 0$, the curve goes through the plane from above downwards.

What on Earth does any of this have to do with torsion?! Well, let's now find out.

Proposition 2.79 *Let $\gamma : I \to \mathbb{R}^3$ be a unit speed curve of non-vanishing curvature. Then,*

$$\tau(s) = \frac{\gamma'''(s) \cdot b(s)}{\kappa(s)}.$$

Proof: By definition, we know that $b(s) = u(s) \times n(s)$. Since γ is unit speed, this simplifies to

$$b(s) = \gamma'(s) \times \frac{\gamma''(s)}{\|\gamma''(s)\|}.$$

We can differentiate with the Product Rule (which is satisfied by the vector product) to get

$$\begin{aligned} b'(s) &= \gamma''(s) \times \frac{\gamma''(s)}{\|\gamma''(s)\|} + \gamma'(s) \times \left(\frac{\gamma'''(s)}{\|\gamma''(s)\|} - \frac{\gamma'''(s)\cdot\gamma''(s)}{\|\gamma''(s)\|^2}\gamma''(s) \right) \\ &= 0 + u \times \left(\frac{\gamma'''(s)}{\kappa(s)} - \left(\gamma'''(s)\cdot n(s)\right) n(s) \right) \\ &= u \times \frac{\gamma'''(s)}{\kappa(s)}, \end{aligned}$$

using the standard fact that the vector product of a vector with itself is zero, and substituting in $k(s) = \gamma''(s)$ into the denominator of the first fraction in the bracket and $n(s) = \frac{\gamma''(s)}{\|\gamma''(s)\|}$ into the second fraction in the brackets. Using $\tau(s) = -b'(s)\cdot n(s)$ from Exercise 26,

$$\tau(s) = \left(-u(t) \times \frac{\gamma'''(s)}{\kappa(s)}\right)\cdot n(s) = \left(\frac{\gamma'''(s)}{\kappa(s)} \times u(t)\right)\cdot n(s) = \left(u(t)\times n(t)\right)\cdot \frac{\gamma'''(s)}{\kappa(s)},$$

using **1.** and **5.** from the vector product reminder to get the second and third equalities. But this is exactly what we wanted to prove, because $b(t) = u(t)\cdot n(t)$ and the dot product is commutative (i.e. it doesn't matter which way around we do it). $\square$

Note: Hence, the torsion tells us about how the curve passes through the osculating plane:

(i) If $\tau(s_0) > 0$, the curve goes through the plane from below upwards.

(ii) If $\tau(s_0) < 0$, the curve goes through the plane from above downwards.

We now sketch the picture demonstrating this geometric intuition we have developed.

(a) The case where $\tau(s_0) > 0$. (b) The case where $\tau(s_0) < 0$.

Figure 12: The geometric intuition behind torsion.

At the moment, we have expressions for $u'(s)$ and $b'(s)$, but what about $n'(s)$? Well let's do it!

Lemma 2.80 *For all $s \in I$, we have $n'(s) = -\kappa(s)u(s) + \tau(s)b(s)$.*

Proof: As before, because $n'(s) \in \mathbb{R}^3$, it means there exist smooth functions λ, μ, ν such that

$$n'(s) = \lambda(s)u(s) + \mu(s)n(s) + \nu(s)b(s).$$

But $[u, n, b]$ is an orthonormal basis by Lemma 2.72, so the coefficients (the smooth functions) above can be extracted by virtue of taking dot products. Namely, we see that

$$\lambda(s) = n'(s) \cdot u(s), \qquad \mu(s) = n'(s) \cdot n(s), \qquad \nu(s) = n'(s) \cdot b(s).$$

The first is obtained by applying the Product Rule to $n(t) \cdot u(t) = 0$, that is

$$\begin{aligned} 0 &= n'(s) \cdot u(s) + n(s) \cdot u'(s) \\ &= n'(s) \cdot u(s) + n(s) \cdot \kappa(s)n(s) \\ &= \lambda(s) + \kappa(s), \end{aligned}$$

where we have used the alternate defining equation $u'(s) = \kappa(s)n(s)$ for the principal unit normal vector. Thus, we conclude that $\lambda(s) = -\kappa(s)$. The second coefficient is obtained by applying the Product Rule to $n(s) \cdot n(s) = 1$, but it is clear (write it out to double-check if you can't see it straight away) this implies $\mu(s) = 0$. The last coefficient is obtained by applying the Product Rule to $n(s) \cdot b(s) = 0$, that is

$$\begin{aligned} 0 &= n'(s) \cdot b(s) + n(s) \cdot b'(s) \\ &= n'(s) \cdot b(s) + n(s) \cdot \big(-\tau(s)n(s)\big) \\ &= \nu(s) - \tau(s), \end{aligned}$$

using Definition 2.76 in the second equality. Thus, we conclude that $\nu(s) = \tau(s)$. Substituting each of these into the expression at the start gives us precisely what we set out to prove. □

Exercise 27 (Continued) Verify that each of Proposition 2.79 and Lemma 2.80 are satisfied by the curve $\gamma : \mathbb{R} \to \mathbb{R}^3$ given by $\gamma(s) = \frac{1}{5}(4s, \cos 3s, \sin 3s)$.

So what is the point of these derivative formulae? It turns out that combining Lemmas 2.75 and 2.80 with the defining expression for n, we immediately obtain the proof of the following important theorem without any extra work.

Theorem 2.81 (Frenet Formulae) *Let $\gamma : I \to \mathbb{R}^3$ be a unit speed curve of non-vanishing curvature with Frenet frame $[u(s), n(s), b(s)]$. Then, the following formulae are satisfied:*

$$\begin{aligned} u'(s) &= \kappa(s)n(s), \\ n'(s) &= -\kappa(s)u(s) + \tau(s)b(s), \\ b'(s) &= -\tau(s)n(s). \end{aligned}$$

Proof: This has already been done implicitly. □

Exercise 28 (Harder) We call a curve $\gamma : I \to \mathbb{R}^3$ spherical if its image $\gamma(I)$ lies entirely on the surface of a sphere. Let γ be a spherical unit speed curve on the sphere $S \subseteq \mathbb{R}^3$ whose radius is r and is centred at the point $\mathbf{p} \in \mathbb{R}^3$; this means that $\left\|\gamma(s) - \mathbf{p}\right\|^2 = r^2$ for all $s \in I$. Prove that γ has a well-defined Frenet frame.

[**Hint:** This looks bizarrely complicated, but to show its Frenet frame is well-defined, it is enough to show that it has non-vanishing curvature. To do this, write $\left\|\gamma(s) - \mathbf{p}\right\|^2 = r^2$ in dot product form and differentiate it twice. Then, use some Frenet formulae to simplify it and conclude that $\kappa(s) \neq 0$.]

We will use the Frenet formulae to prove a result which establishes a link between torsion of a curve and its image (if you managed to read around Proposition 2.79, this may give you a hint already). We then show that scalar curvature and torsion uniquely (almost) define a space curve! First, we will define what it means for a curve to be planar in $\mathbb{R}^3$.

Definition 2.82 A plane is a subset $P \subseteq \mathbb{R}^3$ such that every $\mathbf{x} \in P$ satisfies the equation

$$\mathbf{B} \cdot \mathbf{x} = \nu,$$

where $\mathbf{B} \in \mathbb{R}^3$ is a vector determining the orientation of the plane and $\nu \in \mathbb{R}$ is a fixed constant such that $|\nu|$ is the distance of the plane from the origin $\mathbf{0} = (0, 0, 0)$.

Example 2.83 Let $\mathbf{B} = (2, -1, 5)$ and $\nu = 4$ define plane P. Then, $(x_1, x_2, x_3) \in P$ if and only if

$$2x_1 - x_2 + 5x_3 = 4.$$

Note: The plane described by $(\mathbf{B}, \nu)$ is identical to the plane described by $(-\mathbf{B}, -\nu)$.

Definition 2.84 A space curve $\gamma : I \to \mathbb{R}^3$ is called planar if its image $\gamma(I)$ is entirely contained in some plane, i.e. there exist $\mathbf{B} \in \mathbb{R}^3$ and $\nu \in \mathbb{R}$ with $\mathbf{B} \cdot \gamma(t) = \nu$ for all $t \in I$.

Without loss of generality, we can again assume our curve has unit speed; in the case it is planar, this means the containing plane is described by a **unit** vector $\mathbf{B} \in \mathbb{R}^3$ (meaning $\|\mathbf{B}\| = 1$). The real reason to introduce this notion is to give meaning to zero torsion.

Theorem 2.85 *Let $\gamma : I \to \mathbb{R}^3$ be a unit speed curve of non-vanishing curvature. Then, γ is planar if and only if its torsion is everywhere zero, that is $\tau(s) = 0$ for all $s \in I$.*

Proof: There are two things to prove in "if and only if" results: (i) that the curve being planar implies the torsion is zero, and (ii) that the curve having zero torsion implies it is planar.

($\Rightarrow$) Assume γ is planar. Then, there exists a unit vector $\mathbf{B} \in \mathbb{R}^3$ and $\nu \in \mathbb{R}$ as in Definition 2.82 such that $\mathbf{B} \cdot \gamma(s) = \nu$ for all $s \in I$. If we differentiate this equation twice, we see that

$$\mathbf{B} \cdot \gamma(s) = \nu \qquad \Rightarrow \qquad \mathbf{B} \cdot \gamma'(s) = 0 \qquad \Rightarrow \qquad \mathbf{B} \cdot \gamma''(s) = 0.$$

But the curve has unit speed, so $\gamma'(s) = u(s)$ and $\gamma''(s) = k(s) = \kappa(s)n(s)$. As the curvature is non-vanishing, we can divide through by it when we substitute this into the above equation:

$$\mathbf{B} \cdot n(s) = 0.$$

This tells us that $\mathbf{B}$ is a unit vector which is orthogonal to both $u(s)$ and $n(s)$ for all s; the only other vector with these properties is $b(s)$ either in the same direction or opposite direction. We therefore conclude that $b(s) = \pm\mathbf{B}$, which is constant. Hence, $b'(s) = \mathbf{0}$ and a Frenet formula tells us that $\tau(0) = 0$.

($\Leftarrow$) Assume $\tau \equiv 0$ (which is notation for $\tau(s) = 0$ for all s; we say it is "identically zero"). A Frenet formula implies that $b' \equiv \mathbf{0}$, so we know $b(s) = \mathbf{v}$ is a constant vector by the Mean Value Theorem (yes, it comes up again)! However, we see that

$$\frac{\mathrm{d}}{\mathrm{d}s}[\gamma(s) \cdot \mathbf{v}] = u(s) \cdot \mathbf{v} = u(s) \cdot b(s) = 0,$$

by orthonormality. So, $\gamma(s) \cdot \mathbf{b} = a$ for some $a \in \mathbb{R}$ is constant via the Mean Value Theorem once again. Consequently, we see that γ lies in the plane determined by $\mathbf{B} = \mathbf{v}$ and $\nu = a$. $\square$

Note: The proof of Theorem 2.85 actually only uses the first and last Frenet formulae.

Example 2.86 Consider the curve $\gamma : \mathbb{R} \to \mathbb{R}^3$ with $\gamma(s) = \frac{1}{2}(1 + 2\sin s, \sqrt{3} + \cos s, 1 - \sqrt{3}\cos s)$. This turns out to be planar, believe it or not. Showing this pre-Theorem 2.85 would be very difficult indeed. But now, we just need to very that it has unit speed, non-vanishing curvature and zero torsion. To that end, the unit speed-ness isn't too difficult. Namely,

$$\gamma'(s) = \frac{1}{2}(2\cos s, -\sin s, \sqrt{3}\sin s) \qquad \Rightarrow \qquad \|\gamma'(t)\| = \frac{1}{2}\sqrt{4\sin^2(s) + 4\cos^2(s)} = 1.$$

Now, the curvature vector is acceleration; we will then be able to get its scalar curvature easily:

$$k(t) = \frac{1}{2}(-2\sin s, -\cos s, \sqrt{3}\cos s) \qquad \Rightarrow \qquad \|k(t)\| = \frac{1}{2}\sqrt{4\sin^2(s) + 4\cos^2(s)} = 1.$$

The next task is to construct the Frenet frame. Indeed,

$$\begin{aligned} u(s) &= \frac{1}{2}(2\cos s, -\sin s, \sqrt{3}\sin s), \\ n(s) &= \frac{1}{2}(-2\sin s, -\cos s, \sqrt{3}\cos s), \\ b(s) &= \frac{1}{2}(0, \sqrt{3}, 1). \end{aligned}$$

Using the third Frenet formula, we find the torsion from the derivative of the binormal vector:

$$b'(s) = (0, 0, 0) \qquad \Rightarrow \qquad \tau(s) = 0.$$

Per Theorem 2.85, we know that γ is planar.

Exercise 29 Determine if the curve $\gamma : \mathbb{R} \to \mathbb{R}^3$ where $\gamma(s) = \frac{1}{5}(4s, \cos 3s, \sin 3s)$ is planar.

As with Theorem 2.50, we saw that specifying some conditions about a curve will uniquely determine a curve in $\mathbb{R}^2$ by its signed curvature. It turns out that a similar result holds for curves in $\mathbb{R}^3$, except the situation here is that scalar curvature and torsion uniquely determine it up to an initial position $\gamma(0)$, an initial velocity $\gamma'(0)$ **and** an initial normal vector $n(0)$.

Theorem 2.87 (Fundamental Theorem of Local Theory of Space Curves) *Given any smooth functions $\kappa : I \to (0, \infty)$ and $\tau : I \to \mathbb{R}$ with $0 \in I$ and $\mathbf{a}, \mathbf{u}, \mathbf{n} \in \mathbb{R}^3$, there exists a unique unit speed curve $\gamma : I \to \mathbb{R}$ with scalar curvature κ and torsion τ such that $\gamma(0) = \mathbf{a}$, $\gamma'(0) = \mathbf{u}$ and $n(0) = \mathbf{n}$.*

Proof: The existence of such a curve is a bit difficult, so we won't go through the gory details. However, we will mention a strategy which again uses *Taylor expansion* about $s = 0$; one can get an expression for $\gamma(s)$ in terms of all its derivatives $\gamma^{(k)}(0)$. Using the Frenet formulae, we can get expressions involving the scalar curvature κ, the torsion τ and the principal unit normal vector n. In particular, we can write the following approximation in the Frenet basis:

$$\gamma(s) - \gamma(0) \approx \left(s - \frac{\kappa(0)^2}{6}s^3\right)u(s) + \left(\frac{\kappa(0)}{2}s^2 + \frac{\kappa'(0)}{6}s^3\right)n(s) - \frac{\kappa(0)\tau(0)}{6}s^3 b(s).$$

We can then make a change of basis so that $u(s) = (1,0,0)$ and $n(s) = (0,1,0)$ and $b(s) = (0,0,1)$. This will then give us an expression for the above vector. If we add across $\gamma(0) = \mathbf{a} = (a_1, a_2, a_3)$ and define $\gamma(s) =: \big(x(s), y(s), z(s)\big)$, we get the *local canonical form* of the space curve:

$$\begin{aligned} x(s) &= a_1 + s - \frac{\kappa(0)^2}{6}s^3, \\ y(s) &= a_2 + \frac{\kappa(0)}{2}s^2 + \frac{\kappa'(0)}{6}s^3, \\ z(s) &= a_3 - \frac{\kappa(0)\tau(0)}{6}s^3. \end{aligned}$$

Thus, we have constructed a space curve with the correct scalar curvature and torsion; we can check that it satisfies the other properties in the statement of the theorem.

As for uniqueness, this is easier to prove. Suppose γ and $\widetilde{\gamma}$ are two such curves satisfying the hypotheses in the theorem, with respective Frenet frames $[u(s), n(s), b(s)]$ and $[\widetilde{u}(s), \widetilde{n}(s), \widetilde{b}(s)]$. As both curves have identical (non-vanishing) scalar curvature and torsion, their Frenet frames each satisfy the Frenet formulae. Now, let $f(s) := \big\|u(s) - \widetilde{u}(s)\big\|^2 + \big\|n(s) - \widetilde{n}(s)\big\|^2 + \big\|b(s) - \widetilde{b}(s)\big\|^2$. Then, omitting most of the "(s)" for ease of writing and reading, we see that

$$\begin{aligned} \frac{1}{2}f'(s) &= (u - \widetilde{u}) \cdot (u' - \widetilde{u}') + (n - \widetilde{n}) \cdot (n' - \widetilde{n}') + (b - \widetilde{b}) \cdot (b' - \widetilde{b}') \\ &= -\Big((u \cdot \widetilde{u}') + (u' \cdot \widetilde{u}) + (n \cdot \widetilde{n}') + (n' \cdot \widetilde{n}) + (b \cdot \widetilde{b}') + (b' \cdot \widetilde{b})\Big) \\ &= -\Big((u \cdot \kappa\widetilde{n}) + (\kappa n \cdot \widetilde{u}) + \big(n \cdot (-\kappa\widetilde{u} + \tau\widetilde{b})\big) + \big((-\kappa u + \tau b) \cdot \widetilde{n}\big) + (b \cdot (-\tau\widetilde{n})) + (-\tau n \cdot \widetilde{b})\Big) \\ &= 0. \end{aligned}$$

We conclude that $f \equiv K$ where $K \in \mathbb{R}$ is some constant. Written out fully, this is

$$\big\|u(s) - \widetilde{u}(s)\big\|^2 + \big\|n(s) - \widetilde{n}(s)\big\|^2 + \big\|b(s) - \widetilde{b}(s)\big\|^2 = K.$$

If we substitute $s = 0$, we see the only option is $K = 0$. But because each term on the left is non-negative, it must be that each of those is identically zero also. In particular, $\big\|u(s) - \widetilde{u}(s)\big\| = 0$

for all $s \in I$. This is equivalent to the following (using the Mean Value Theorem again):

$$\frac{\mathrm{d}}{\mathrm{d}s}[\gamma(s) - \widetilde{\gamma}(s)] = 0 \qquad \Rightarrow \qquad \gamma(s) - \widetilde{\gamma}(s) = \mathbf{v},$$

for $\mathbf{v} \in \mathbb{R}^3$ a constant vector. However, $\gamma(0) = \mathbf{a} = \widetilde{\gamma}(0)$, so we must have $\mathbf{v} = \mathbf{a} - \mathbf{a} = \mathbf{0}$. Hence, it follows that $\gamma(s) = \widetilde{\gamma}(s)$ for all $s \in I$; this means the curves are in fact exactly the same. □

Note: In words, we can state Theorem 2.87 as follows: a unit speed curve of non-vanishing curvature is uniquely determined by its scalar curvature and torsion up to *rigid motions* (meaning translations and rotations).

Example 2.88 (Continued) Consider $\gamma : \mathbb{R} \to \mathbb{R}^3$ with $\gamma(s) = \frac{1}{2}(1+2\sin s, \sqrt{3}+\cos s, 1-\sqrt{3}\cos s)$ from Example 2.86. We concluded earlier that its scalar curvature is constant and its torsion is zero. But a circle has constant curvature and is planar (see Problem **2.20** below), so we know from Theorem 2.87 that γ must be a circle too!

2.4 Problem Set 2

We now provide a number of problems to complement the notes in Section 2. Solutions can be found at the end of the notes, but do try to work things out as best as possible without looking the answers up. Of course, if you are very stuck, take a peak at the solutions and try to unravel what you were finding difficult.

Problems for Section 2.1

2.1. Determine, with justification, whether the following are regularly parametrised curves.

(a) The map $\alpha : \mathbb{R} \to \mathbb{R}^2$ given by $\alpha(t) = (\cos^2 t, \sin t)$.

(b) The map $\beta : \mathbb{R} \to \mathbb{R}^2$ given by $\beta(t) = (t^2, t^3 - t^2)$.

(c) The map $\gamma : \mathbb{R} \to \mathbb{R}^2$ given by $\gamma(t) = (t^2, t^3 - t)$.

(d) The map $\delta : \mathbb{R} \to \mathbb{R}^2$ given by $\delta(t) = (t \sin t, \cos t)$.

(e) The map $\varepsilon : \mathbb{R} \to \mathbb{R}^2$ given by $\varepsilon(t) = (\sin t, t \cos t)$.

2.2. Construct the tangent lines to the following regular curves at the time $t = 1$.

(a) The curve $\alpha : \mathbb{R} \to \mathbb{R}^3$ given by $\alpha(t) = (t, t^2, \frac{1}{3}t^3)$.

(b) The curve $\beta : \mathbb{R} \to \mathbb{R}^2$ given by $\beta(t) = (t, \cosh t)$.

(c) The curve $\gamma : \mathbb{R} \to \mathbb{R}^2$ given by $\gamma(t) = (t^3 + t, t^2)$.

2.3. Consider the regular curve $\gamma : \mathbb{R} \to \mathbb{R}^2$ given by $\gamma(t) = (t^3, t^2 - t)$. Construct its tangent line $\widehat{\gamma}_{-1}$ at $t_0 = -1$. Hence, determine the self-intersection points between γ and $\widehat{\gamma}_{-1}$.

2.4. Compute the arc length from $t = 0$ to $t = 4\pi$ along $\gamma : \mathbb{R} \to \mathbb{R}^3$ where $\gamma(t) = (2t, \cos t, \sin t)$.

2.5. Compute the total length of the curve $\gamma : (0, 2\pi) \to \mathbb{R}^2$ given by $\gamma(t) = (t + \sin t, 1 - \cos t)$.

[**Hint:** Pull out a factor of $\sqrt{2}$ from the integral and then use the identity $\cos(t) = 2\cos^2(\frac{t}{2}) - 1$. Be aware $\sqrt{x^2} = |x|$, **not** $\sqrt{x^2} = x$; use this to split the integral into a sum of two integrals over $(0, \pi)$ and $(\pi, 2\pi)$.]

2.6. Consider the regular curve $\gamma : \mathbb{R} \to \mathbb{R}^2$ given by $\gamma(t) = (e^t \cos t, e^t \sin t)$.

(a) Construct the arc length function σ_0 based at $t_0 = 0$.

(b) Construct the inverse function τ_0 based at $t_0 = 0$.

(c) Hence, find a unit speed reparametrisation $\widetilde{\gamma}$ of γ.

2.7. The *graph* of a function $f : D \to \mathbb{R}$ is the set of points $\big(t, f(t)\big) \in \mathbb{R}^2$. In the case that the domain $D = (0, \infty)$ and the function is given by $f(t) = at + b$ for some $a, b \in \mathbb{R}$, find the unit speed reparametrisation of the graph of f.

2.8. Find the curvature vectors of the following curves.

(a) The curve $\alpha : \mathbb{R} \to \mathbb{R}^2$ given by $\alpha(t) = (t, t^3)$.

(b) The curve $\beta : \mathbb{R} \to \mathbb{R}^4$ given by $\beta(t) = (\sin t, 2t, 2t, \cos t)$.

(c) The curve $\gamma : \mathbb{R} \to \mathbb{R}^2$ given by $\gamma(t) = \left(e^{t^2}, f(t)\right)$, where $f(t) := \displaystyle\int_0^t e^{u^2}\, \mathrm{d}u$.

(d) The curve $\delta : \mathbb{R} \to \mathbb{R}^3$ given by $\delta(t) = (t - \cos t, \sin t, t)$.

Problems for Section 2.2

2.9. (a) Consider the curve $\alpha : \mathbb{R} \to \mathbb{R}^2$ given by $\alpha(t) = (t, t^3)$.

i. Find the signed curvature of α by using your answer to Problem **2.8**.

ii. Hence, find any and all inflexion points of α.

(b) Consider the curve $\gamma : \mathbb{R} \to \mathbb{R}^2$ given by $\gamma(t) = \left(e^{t^2}, f(t)\right)$, where $f(t) := \displaystyle\int_0^t e^{u^2}\, \mathrm{d}u$.

i. Find the signed curvature of γ by using your answer to Problem **2.8**.

ii. Hence, find any and all inflexion points of γ.

2.10. Derive an expression for the signed curvature of the graph of $f : D \to \mathbb{R}$.

2.11. Find a unit speed curve $\gamma : \mathbb{R} \to \mathbb{R}^2$ such that $\gamma(0) = (0, 0)$ and $\gamma'(0) = (1, 0)$ which has signed curvature $\kappa(s) = \frac{1}{2}(s + 1)(s - 2)$ for every $s \in \mathbb{R}$ by using the formula (†).

2.12. Consider the curve $\gamma : \mathbb{R} \to \mathbb{R}^2$ given by $\gamma(t) = (t, \cosh t)$.

(a) Compute its signed curvature κ.

(b) Construct its evolute E_γ.

(c) Construct its involute I_γ.

[**Hint:** Recall that $\sinh$ and $\cosh$ are derivatives of each other, and $\cosh^2 \equiv 1 - \sinh^2$ and $\frac{\sinh}{\cosh} \equiv \tanh$.]

2.13. Assuming $\kappa'(t) < 0$, prove that the arc length along an evolute E_γ from t_1 to t_2 is given by

$$\frac{1}{\kappa(t_2)} - \frac{1}{\kappa(t_1)}.$$

[**Hint:** Make use of the formula for the derivative E_γ' you obtained in Exercise 20.]

Problems for Section 2.3

2.14. Consider the regular curve $\gamma : \mathbb{R} \to \mathbb{R}^3$ given by $\gamma(t) = (t, t^2, \frac{1}{3}t^3)$.

(a) Compute its curvature vector $k(1)$ at time $t = 1$.

(b) Hence, determine its scalar curvature $\kappa(1)$.

(c) Construct the Frenet frame $[u(1), n(1), b(1)]$ of this curve at time $t = 1$.

2.15. Consider a curve $\gamma : \mathbb{R} \to \mathbb{R}^3$ where $\gamma'(0) = \frac{1}{\sqrt{3}}(1, 1, -1)$ and $\gamma''(0) = (1, -2, 1)$.

(a) Construct the Frenet frame $[u(0), n(0), b(0)]$ of this curve at time $t = 0$.

(b) Is it possible to tell if this curve has unit speed? Justify your answer.

2.16. Find the scalar curvature and torsion of the curve $\gamma : \mathbb{R} \to \mathbb{R}^3$ given by $\gamma(t) = (at, bt^2, ct^3)$.

2.17. Let $\gamma : I \to \mathbb{R}^3$ be a unit speed curve of non-vanishing curvature and define $\Gamma : I \to \mathbb{R}^3$ by

$$\Gamma(t) = \gamma'(t).$$

(a) Show that $\left\|\Gamma'(t)\right\| = \kappa(t)$, where κ is the scalar curvature of γ.

(b) Is Γ a regularly parametrised curve? Justify your answer.

(c) Prove the curvature vector of Γ can be given in terms of geometric quantities of γ by

$$k^\Gamma(t) = -u(t) + \frac{\tau(t)}{\kappa(t)} b(t).$$

(d) What is the name of the (well-known) surface containing the image $\Gamma(I)$?

2.18. Let $\gamma : I \to \mathbb{R}^3$ be a spherical unit speed curve in the sense of Exercise 28 and suppose it has non-vanishing torsion. For some constant $r > 0$ (the radius of the sphere), prove that

$$\frac{1}{\kappa(s)^2} + \left(\frac{\kappa'(s)}{\tau(s)\kappa(s)^2}\right)^2 = r^2.$$

2.19. Consider the regular curve $\gamma : \mathbb{R} \to \mathbb{R}^3$ given by $\gamma(s) = (a\cos\frac{s}{c}, a\sin\frac{s}{c}, b\frac{s}{c})$ where $a, b, c > 0$ and $a^2 + b^2 = c^2$.

(a) Show that γ lies on the cylinder $c^2 + y^2 = a^2$.

(b) Show that γ is a unit speed curve.

(c) Determine the scalar curvature and torsion of γ.

(d) Derive an expression for the osculating plane.

[**Hint:** Use Definition 2.82, noting that the osculating plane is orthogonal to (and thus has orientation defined as) the binormal vector $b(s)$, and the plane contains γ so it is distance $\|\gamma(s)\|$ from $\mathbf{0} = (0, 0, 0)$.]

2.20. A *circle* with radius $r > 0$ and centre $\mathbf{c} = (c_1, c_2, c_3) \in \mathbb{R}^3$ is a curve $\gamma : \mathbb{R} \to \mathbb{R}^3$ given by

$$\gamma(t) = (a_1 r\cos t + b_1 r\sin t + c_1, a_2 r\cos t + b_2 r\sin t + c_2, a_3 r\cos t + b_3 r\sin t + c_3).$$

(a) Compute the scalar curvature κ of this circle.

(b) Determine the torsion τ of this circle.

(c) Hence, deduce that every circle is planar.

2.21. Let $\gamma : \mathbb{R} \to \mathbb{R}^3$ be the parametrised curve satisfying the second-order initial value problem

$$\gamma''(s) = \gamma(s) \times \gamma'(s), \qquad \gamma(0) = (1, 0, 0), \qquad \gamma'(0) = (0, 1, 0).$$

(a) Show that γ is a unit speed curve.

(b) Compute the scalar curvature $\kappa(0)$ and $t = 0$.

(c) Show that $\gamma(s) \cdot \gamma'(s) = s$.

(d) Show that $\|\gamma'(s)\|^2 = s^2 + 1$.

(e) Hence, deduce that the scalar curvature κ is constant.

(f) Given that $b(s) = \gamma(s) - s\gamma'(s)$, deduce a formula for the torsion τ of the curve.

3 Parametrised Surfaces

We are now going to move on to looking at *surfaces*, which are smooth two-dimensional subsets of $\mathbb{R}^3$. Much like how some smooth maps $\gamma : I \to \mathbb{R}^n$ did **not** product a 'nice' curve and we had to impose the regularity condition $\gamma'(t) \neq 0$, something similar will happen when we talk about surfaces. In that sense, we will only be discussing 'nice' surfaces.

3.1 General Surfaces

We will look at two-dimensional surfaces within three-dimensional space; this discussion actually generalises quite easily to $(n-1)$-dimensional surfaces within n-dimensional space but we won't discuss this... at least not yet. Keeping the dimensions low allows us to better visualise things.

Although we introduce this properly soon, we think of a surface as a smooth map $M : U \to \mathbb{R}^3$ where $U \subseteq \mathbb{R}^2$. Much like how a curve was given by a single parameter t, a surface is given by two parameters. Thus, we write the parametrisation of a surface as $M(x_1, x_2)$. We are now motivated to ensure that M is 'nice' and this is done by restricting what we allow U to be.

Definition 3.1 Let $\mathbf{x} \in \mathbb{R}^n$ and $r > 0$. The open ball of radius r centred at $\mathbf{x}$ is the subset

$$B_r(\mathbf{x}) \coloneqq \{\mathbf{y} \in \mathbb{R}^n : \|\mathbf{y} - \mathbf{x}\| < r\}.$$

We call this the *open* ball because the defining property includes "<" which is a sort-of 'open' relation; think back to when we looked at open intervals in Definition 1.4. Geometrically, we can think of an open ball as the set of points which are distance less than r away from the vector $\mathbf{x}$ the ball is centred at.

Note: When using Definition 3.1 for $n = 2$, we call it an open **disk** of radius r centred at $\mathbf{x}$.

Example 3.2 Let $\mathbf{x} = (0, 2) \in \mathbb{R}^2$ and $r = 3$. Then, the open disk $B_3\big((0, 2)\big)$ is sketched below.

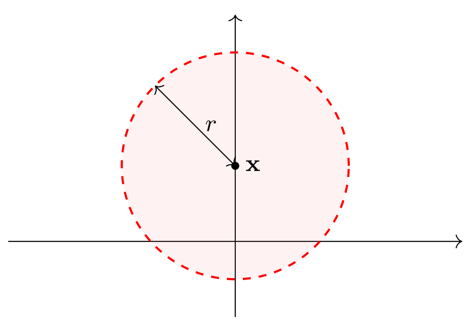


Figure 13: The open disk of radius 3 centred at $(0, 2)$.

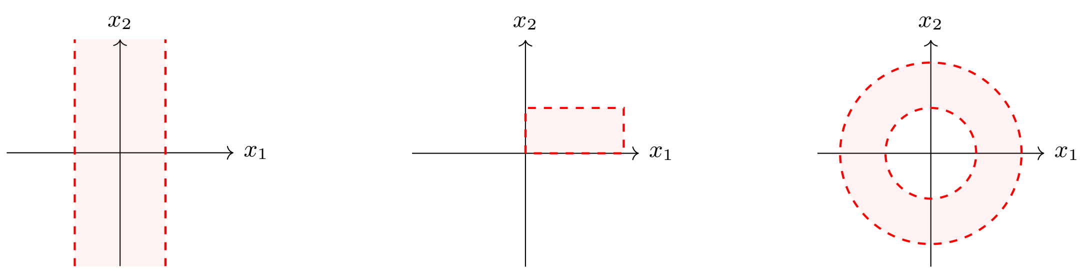
To be frank, open balls aren't the most important object to us. The benefit to introducing them is that they open up (no pun intended, but I'm not altering a gem like this) a world of possibilities for subsets of $\mathbb{R}^n$, in particular of $\mathbb{R}^2$.

Definition 3.3 A subset $U \subseteq \mathbb{R}^n$ is called open if for each $\mathbf{x} \in U$, there exists a positive number $r > 0$ such that $B_r(\mathbf{x}) \subseteq U$.

So what is the geometric picture to have in mind with this? The idea that no point in an open set can lie at the 'edge' of that set since every point can be surrounded by one of these open balls. This is the generalisation of openness that we promised when we looked at open intervals.

Note: We also say that a subset $U \subseteq \mathbb{R}^n$ is closed if its complement $\mathbb{R}^n \setminus U \subseteq \mathbb{R}^n$ is open.

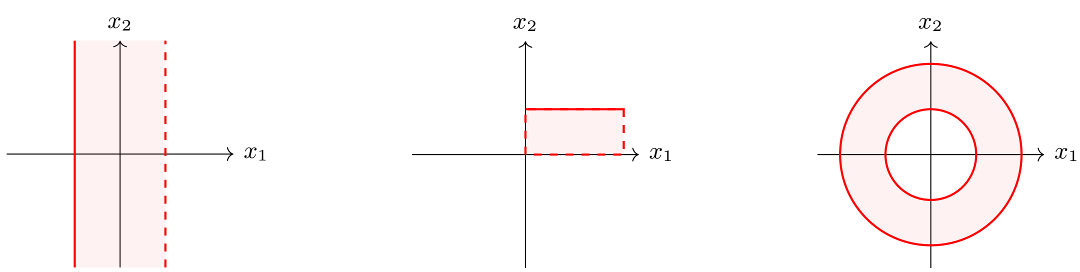
Example 3.4 We have the following examples of open subsets of $\mathbb{R}^2$.



(a) $(-1, 1) \times \mathbb{R}$. (b) $(0, \pi) \times (0, 1)$. (c) $\{\mathbf{x} \in \mathbb{R}^2 : 1 < \|\mathbf{x}\| < 2\}$.

Figure 14: Examples of open subsets of $\mathbb{R}^2$.

For comparison, the following are **not** open subsets of $\mathbb{R}^2$.



(a) $[-1, 1) \times \mathbb{R}$. (b) $(0, \pi) \times (0, 1]$. (c) $\{\mathbf{x} \in \mathbb{R}^2 : 1 \leq \|\mathbf{x}\| \leq 2\}$.

Figure 15: Examples of non-open subsets of $\mathbb{R}^2$.

We are almost ready to define the main object of study in this part of the module. However, one final thing we need is the concept of a *regular map*. Let's recall the notion of a partial derivative.

Reminder: Let $f : D \to \mathbb{R}$ be a function with domain $D \subseteq \mathbb{R}^n$, so $f = f(x_1, x_2, ..., x_n)$ is a function of n variables. The partial derivative of f with respect to one of these variables x_k is the usual derivative where we treat **all other** variables as constant. This is typically denoted by any of the following symbols: $\frac{\partial f}{\partial x_k}$ and f_{x_k} and $\partial_{x_k} f$.

Definition 3.5 Let $U \subseteq \mathbb{R}^2$ be open and $M : U \to \mathbb{R}^3$ be smooth. The following vector-valued functions $\varepsilon_1 : U \to \mathbb{R}^3$ and $\varepsilon_2 : U \to \mathbb{R}^3$ are called the coordinate basis vectors:

$$\varepsilon_1 := \frac{\partial M}{\partial x_1} \qquad \text{and} \qquad \varepsilon_2 := \frac{\partial M}{\partial x_2}.$$

A point $\mathbf{x} = (x_1, x_2) \in U$ is called a regular point of M if the vectors $\{\varepsilon_1(x_1, x_2), \varepsilon_2(x_1, x_2)\}$ are *linearly independent* (which in this context means they are **not** parallel; one isn't a multiple of the other). We say that M is regular if every $\mathbf{x} \in U$ is a regular point of M.

It is a fact that two vectors in $\mathbb{R}^3$ are parallel if and only if their vector product is zero; this is because there will be no other vector satisfying orthogonality since the two vectors travel in the same direction (up to sign). Therefore, an easy test to see that M is regular at $\mathbf{x} \in U$ is to show

$$\varepsilon_1(\mathbf{x}) \times \varepsilon_2(\mathbf{x}) \neq \mathbf{0}.$$

Definition 3.6 Let $U \subseteq \mathbb{R}^2$ be an open subset. A regularly parametrised surface is a smooth map $M : U \to \mathbb{R}^3$ that is both injective and regular.

Example 3.7 Let $M : \mathbb{R}^2 \to \mathbb{R}^3$ be given by $M(x_1, x_2) = (x_1, x_1^2 + x_2, x_1 - x_2^2)$; we will show that M is a regularly parametrised surface. First, it is clearly smooth because each member of the tuple is a polynomial. We now need to show that M is injective and regular. For injectivity,

$$\begin{aligned} M(x_1, x_2) = M(y_1, y_2) &\Leftrightarrow (x_1, x_1^2 + x_2, x_1 - x_2^2) = (y_1, y_1^2 + y_2, y_1 - y_2^2) \\ &\Rightarrow x_1 = y_1 \text{ and } x_1^2 + x_2 = y_1^2 + y_2 \\ &\Rightarrow x_1 = y_1 \text{ and } y_1^2 + x_2 = y_1^2 + y_2 \\ &\Rightarrow x_1 = y_1 \text{ and } x_2 = y_2, \end{aligned}$$

where we substitute $x_1 = y_1$ into the other equation. As for regularity, we compute the coordinate basis vectors and take their vector product; we then conclude M is indeed a regular surface:

$$\varepsilon_1 = (1, 2x_1, 1) \text{ and } \varepsilon_2 = (0, 1, -2x_2) \qquad \Rightarrow \qquad \varepsilon_1 \times \varepsilon_2 = (-4x_1x_2 - 1, 2x_2, 1) \neq \mathbf{0}.$$

Exercise 30 Prove that $M : \mathbb{R}^2 \to \mathbb{R}^3$ given below is a regularly parametrised surface:

$$M(x_1, x_2) = (x_2, x_1 - x_2^2, x_1^3 + x_1 x_2).$$

The next result isn't the most powerful but it will help us identify a lot of regular surfaces quickly just by looking at a smooth function of the form $\mathbb{R}^2 \to \mathbb{R}$.

Proposition 3.8 *Let* $f : \mathbb{R}^2 \to \mathbb{R}$ *be* **any** *smooth function. Then, the so-called* graph *of* f $M : \mathbb{R}^2 \to \mathbb{R}^3$ *given by* $M(x_1, x_2) = \big(x_1, x_2, f(x_1, x_2)\big)$ *is a regularly parametrised surface.*

Proof: It is clear that M is smooth because polynomials (the first two entries) are smooth and f is assumed to be smooth. It remains to check that M is both injective and regular. First then,

$$\begin{aligned} M(x_1, x_2) = M(y_1, y_2) &\Leftrightarrow \big(x_1, x_2, f(x_1, x_2)\big) = \big(y_1, y_2, f(y_1, y_2)\big) \\ &\Rightarrow x_1 = y_1 \text{ and } x_2 = y_2. \end{aligned}$$

As for regularity, we construct the coordinate basis vectors and then verify that their vector product is everywhere non-zero. Indeed, just use the vector product formula from earlier:

$$\varepsilon_1 = \left(1, 0, \frac{\partial f}{\partial x_1}\right) \text{ and } \varepsilon_2 = \left(0, 1, \frac{\partial f}{\partial x_2}\right) \qquad \Rightarrow \qquad \varepsilon_1 \times \varepsilon_2 = \left(-\frac{\partial f}{\partial x_1}, -\frac{\partial f}{\partial x_2}, 1\right) \neq \mathbf{0}. \quad \square$$

Note: Due to Proposition 3.8, we can now write down some rather complicated-looking regular surfaces with no further work required to prove regularity. For example, we know

$$M : \mathbb{R}^2 \to \mathbb{R}^3, \qquad M(x_1, x_2) = \left(x_1, x_2, \frac{x_2^2}{b^2} - \frac{x_1^2}{a^2}\right)$$

is a regular surface for any $a > 0$ and $b > 0$. This is because the third member of the tuple is the defining equation of a *hyperbola*; this is a smooth function. Fun fact: this is called a *hyperbolic paraboloid*, which is the mathematical name for the shape of a Pringle!

Let's relate the notion of curves to that of surfaces. Given a regular surface $M : U \to \mathbb{R}^3$, we can fix the second input x_2, say $x_2 = b$ for some constant $b \in \mathbb{R}$. By doing this, we get a curve

$$\alpha_b(t) := M(t, b)$$

along which the first input $x_1 =: t$ varies. Therefore, if we change the fixed constant b, we get a

family of curves that cover the entirety of our surface. We could equally do the same by fixing the first input $x_1 = a$ for some constant $a \in \mathbb{R}$. By doing this, we again obtain a curve

$$\beta_a(t) = M(a, t)$$

along which $x_2 =: t$ varies. Changing a gives us another **family** of curves covering the surface.

> **Exercise 31** For the curves α_b and β_a defined above, compute their derivatives α_b' and β_a'.

Example 3.9 Let $U = (0, \pi) \times (0, 2\pi) \subseteq \mathbb{R}^2$, which is open, and consider $M : U \to \mathbb{R}^3$ given by

$$M(x_1, x_2) = (\sin x_1 \cos x_2, \sin x_1 \sin x_2, \cos x_1).$$

This defines a regular surface (in fact, it is *almost* a sphere). We will not show injectivity because it can be a bit lengthy and uses a number of trigonometric identities that we haven't introduced. However, the fact it is regular is shorter:

$$\begin{aligned} \varepsilon_1 &= (\cos x_1 \cos x_2, \cos x_1 \sin x_2, -\sin x_1), \\ \varepsilon_2 &= (-\sin x_1 \sin x_2, \sin x_1 \cos x_2, 0) \\ \Rightarrow \qquad \varepsilon_1 \times \varepsilon_2 &= (\sin^2 x_1 \cos x_2, \sin^2 x_1 \sin x_2, \sin x_1 \cos x_1) \neq \mathbf{0}. \end{aligned}$$

To imaging what this surface looks like, we construct a family of curves; fix $x_1 =: \theta \in (0, \pi)$ and allow $x_2 \in (0, 2\pi)$ to vary as much as it wants. In this way, we think of $x_2 := t$ as time which gives us the associated curve $\beta_\theta : (0, 2\pi) \to \mathbb{R}^3$ given by

$$\beta_\theta(t) = M(\theta, t) = (\sin\theta \cos t, \sin\theta \sin t, \cos\theta).$$

This is a circle of radius $r = \sin(\theta)$ in the horizontal plane defined by $y_3 = \cos(\theta)$, where we use x's for the coordinates which define M and y's for the coordinates in the space $\mathbb{R}^3$ (we will say more about this soon). As θ varies, the plane containing the circle shifts between $y_3 = 1$ and $y_3 = -1$ all the while its radius varies from $r = 0$ up to $r = 1$ and shrinking back to $r = 0$. Each θ gives us a different "slice" of the surface. Note also that

$$\begin{aligned} \left\| M(\theta, t) \right\|^2 &= \sin^2(\theta)\cos^2(t) + \sin^2(\theta)\sin^2(t) + \cos^2(\theta) \\ &= \sin^2(\theta)(\cos^2 t + \sin^2 t) + \cos^2(\theta) \\ &= 1, \end{aligned}$$

so every point on M lies distance one away from the origin $(0, 0, 0)$.

Note: We can define a sphere with centre $\mathbf{c} \in \mathbb{R}^3$ and radius r by $M : U \to \mathbb{R}^3$ where

$$M(x_1, x_2) = (r \sin x_1 \cos x_2, r \sin x_1 \sin x_2, r \cos x_1) + \mathbf{c}.$$

Remark 3.10 In Example 3.9, the each curve β_θ actually misses out the point $(\sin\theta, 0, \cos\theta)$. Notice the *North pole* $(0, 0, 1)$ is precisely this point where $\theta = 0$ (and similarly for the *South pole* where $\theta = \pi$). The whole family of curves (β_θ), which is exactly the surface M, is a sphere without its poles and without a single semi-circular line between where the poles would be.

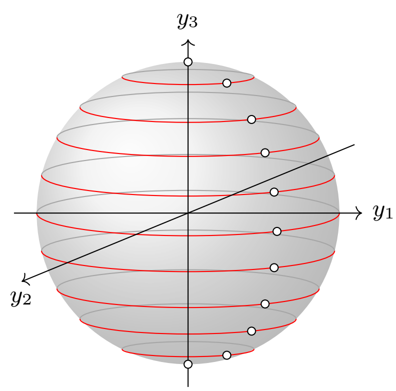


Figure 16: The parametrised surface $M(x_1, x_2) = (\sin x_1 \cos x_2, \sin x_1 \sin x_2, \cos x_1)$.

Definition 3.11 Let $M : U \to \mathbb{R}^3$ be a regularly parametrised surface and consider some arbitrary point $\mathbf{p} = (p_1, p_2, p_3) \in \mathbb{R}^3$ on M which is given by $M(x_1, x_2) = \mathbf{p}$.

(i) We call the pair (x_1, x_2) the local coordinates of $\mathbf{p}$.

(ii) We call the triple (p_1, p_2, p_3) the global coordinates of $\mathbf{p}$.

Per Definition 3.11, any point on a surface can be described by only two numbers (its local coordinates) rather than three (its global coordinates, which are just the usual coordinates in $\mathbb{R}^3$). We refer to $\mathbb{R}^3$ as the *ambient space* of the surface, that is the space in which the surface exists; you can visualise this as the space surrounding the "floating" surface of interest (e.g. actual space around Earth for instance).

Example 3.12 (Continued) Consider the sphere we introduced in Example 3.9, this time written

$$M(\theta, \varphi) = (\sin\theta \cos\varphi, \sin\theta \sin\varphi, \cos\theta).$$

Note that this is exactly the same surface; us changing the labels for the local coordinates is nothing more than to give our eyes a break from carefully looking at indices to ensure we don't muddle up x_1 with x_2 (and it is standard to use the Greek letters θ and φ for *spherical coordinates*, which is what these are). We will determine the local coordinates of $\mathbf{p} = (0,1,0)$; this means we want to find some $\theta \in (0,\pi)$ and $\varphi \in (0,2\pi)$ such that $M(\theta,\varphi) = (0,1,0)$. Well, this means that

$$\cos(\theta) = 0 \qquad \Rightarrow \qquad \theta = (2n-1)\frac{\pi}{2} \text{ where } n \in \mathbb{Z}.$$

The only solutions which ensure $0 < \theta < 2\pi$ are $\theta = \frac{\pi}{2}$ (where $n = 1$) and $\theta = \frac{3\pi}{2}$ (where $n = 2$).

(i) If $\theta = \frac{\pi}{2}$, we have $\sin(\theta) = 1$. Therefore, we must solve $\cos(\varphi) = 0$ and $\sin(\varphi) = 1$. Again, there are infinitely-many solutions but the only one with $0 < \varphi < \pi$ is $\varphi = \frac{\pi}{2}$.

(ii) If $\theta = \frac{3\pi}{2}$, we have $\sin(\theta) = -1$. Therefore, we must solve $\cos(\varphi) = 0$ and $-\sin(\varphi) = 1$. Again, there are infinitely-many solutions but **none** with $0 < \varphi < \pi$.

Therefore, there is one and only one pair of local coordinates for $\mathbf{p}$, namely $(\theta,\varphi) = \left(\frac{\pi}{2},\frac{\pi}{2}\right)$.

Note: Since we stipulate that $M : U \to \mathbb{R}^3$ is injective for a regular surface in Definition 3.6, this ensures that every point has one and only one pair of local coordinates!

Notation 3.13 If $\mathbf{p} \in \mathbb{R}^3$ is a point on the regular surface M, then we simply write $\mathbf{p} \in M$. This is another abuse of notation; because M is **not** a set (it is a vector-valued function remember), we really shouldn't be using the symbol "$\in$". However, we treat M and its image $M(U) \subseteq \mathbb{R}^3$ (which **is** a subset) as one in the same.

Exercise 32 Consider the map $M : \mathbb{R}^2 \to \mathbb{R}^3$ given by $M(x_1,x_2) = (x_1, x_1^2 + x_2, x_1 - x_2^2)$.

(i) Show that M is a regular surface.

(ii) Verify that the point $\mathbf{p} = (2,3,1) \in M$.

(iii) Find the local coordinates of $\mathbf{p}$.

As we saw with regular curves, every point had a well-defined tangent line. This idea can be generalised to fit the situation of regular surfaces; this leads to the idea of a *tangent space*, that is a collection of vectors that are all tangent to a point on the surface.

Definition 3.14 Let $M : U \to \mathbb{R}^3$ be a regular surface and $\mathbf{p} \in M$. A curve through $\mathbf{p}$ in M is a smooth map $\alpha : I \to M$ on an open interval $I \subseteq \mathbb{R}$ with $0 \in I$ such that $\alpha(0) = \mathbf{p}$. The tangent space to M at $\mathbf{p}$ is the set of vectors that are tangent to a curve through $\mathbf{p}$:

$$T_{\mathbf{p}}M := \{\mathbf{v} \in \mathbb{R}^3 : \text{there exists a curve } \alpha \text{ through } \mathbf{p} \text{ in } M \text{ where } \alpha'(0) = \mathbf{v}\}.$$

We then call any vector that lives within this space a tangent vector to M at $\mathbf{p}$.

Note: The curve α through a point on a regular surface is **not** always a regular curve.

Example 3.15 Consider the map $M : \mathbb{R}^2 \to \mathbb{R}^3$ given by $M(x_1, x_2) = (x_1, x_2, x_1^2 + x_2^2)$. Notice it fits the hypotheses of Proposition 3.8 where $f(x_1, x_2) = x_1^2 + x_2^2$ so we know immediately that it is a regularly parametrised surface. Suppose that $\mathbf{p} = (2, 1, 5)$ We will write down some curves through $\mathbf{p}$ in M. First, we will write $\mathbf{p}$ in local coordinates; this is easy because we can read-off immediately that $x_1 = 2$ and $x_2 = 1$. Hence, $\mathbf{p} = M(2, 1)$. Thinking in terms of local coordinates, any curve through $\mathbf{p}$ must satisfy $\widehat{\alpha}(0) = (2, 1)$ (recall the hat means it is a tangent curve). Then, applying M to this gives us a curve in M. Here are a number of examples:

$$
\begin{aligned}
\widehat{\alpha}_1(t) = (2, 1+t) \qquad &\Rightarrow \qquad \alpha_1(t) = M\left(\widehat{\alpha}_1(t)\right) \\
&\qquad\qquad = (2, 1+t, t^2 + 2t + 5), \\
\widehat{\alpha}_2(t) = (2+t, 1-t) \qquad &\Rightarrow \qquad \alpha_2(t) = M\left(\widehat{\alpha}_2(t)\right) \\
&\qquad\qquad = (2+t, 1-t, 2t^2 + 2t + 5), \\
\widehat{\alpha}_3(t) = (2+t^2, 1+\cos t) \qquad &\Rightarrow \qquad \alpha_3(t) = M\left(\widehat{\alpha}_3(t)\right) \\
&\qquad\qquad = (2+t^2, 1+\cos t, t^4 + 4t^2 + \cos^2 t + 2\cos t + 5).
\end{aligned}
$$

The corresponding tangent vectors (which are elements of the space $T_{\mathbf{p}}M$) are simply

$$
\alpha_1'(0) = (0, 1, 2), \qquad \alpha_2'(0) = (1, -1, 2) \qquad \alpha_3'(0) = (0, 0, 0).
$$

Reminder: A subspace is a non-empty subset $V \subseteq \mathbb{R}^n$ such that for all $\mathbf{v}, \mathbf{w} \in V$ and $\lambda \in \mathbb{R}$, we have *closure* under vector addition ($\mathbf{v} + \mathbf{w} \in V$) and scalar multiplication ($\lambda\mathbf{v} \in V$).

Be aware that "closure" here means we stay in this set if we add two vectors or multiply a vector by a real number; it has **nothing** to do with open/closed subsets from Definition 3.3. Now that we are reminded of subspaces, we can show that the tangent space is one.

Theorem 3.16 *The space $T_{\mathbf{p}}M$ is a two-dimensional vector space with basis $\{\varepsilon_1, \varepsilon_2\}$.*

Proof: Let $M : U \to \mathbb{R}^3$ be a regular surface with coordinate basis vectors ε_1 and ε_2, and fix $\mathbf{z} = (z_1, z_2) \in U$. If α is a curve through $\mathbf{p} := M(\mathbf{z})$, then $\alpha(t) = M(\widehat{\alpha}(t)) = M(\widehat{\alpha}_1(t), \widehat{\alpha}_2(t))$. Notice $\widehat{\alpha}(0) = \mathbf{z}$ by definition. We can write its derivative at $t = 0$ via the Chain Rule, namely

$$
\alpha'(0) = \widehat{\alpha}_1'(0) \left.\frac{\partial M}{\partial x_1}\right|_{\widehat{\alpha}(0)} + \widehat{\alpha}_2'(0) \left.\frac{\partial M}{\partial x_2}\right|_{\widehat{\alpha}(0)} = \widehat{\alpha}_1'(0)\varepsilon_1(\mathbf{z}) + \widehat{\alpha}_2'(0)\varepsilon_2(\mathbf{z}).
$$

This shows that every tangent vector can be written as a linear combination of the coordinate basis vectors. For the converse, suppose $\mathbf{v} = a\varepsilon_1(\mathbf{z}) + b\varepsilon_2(\mathbf{z})$ where $a, b \in \mathbb{R}$ is an arbitrary linear combination; we now prove that $\mathbf{v} \in T_{\mathbf{p}}M$, that is there exists a curve through $\mathbf{p}$ whose velocity vector at $t = 0$ is $\mathbf{v}$. Indeed, one such curve is

$$\alpha(t) = M(z_1 + at, z_2 + bt).$$

It is easy to see that its derivative at $t = 0$ produces $\alpha'(0) = a\varepsilon_1(\mathbf{z}) + b\varepsilon_2(\mathbf{z}) = \mathbf{v}$. We conclude

$$T_{\mathbf{p}}M = \{a\varepsilon_1 + b\varepsilon_2 : a, b \in \mathbb{R}\} = \operatorname{span}(\varepsilon_1, \varepsilon_2).$$

It is easy to see that this subset is closed under vector addition and scalar multiplication. This means that the sub*set* is actually a sub*space*. By the fact that M is regular, we know that $\{\varepsilon_1, \varepsilon_2\}$ are linearly independent (straight from Definition 3.5); combining this fact with the spanning above is precisely what it means to be a basis. □

Note: We see that Theorem 3.16 explains why ε_1 and ε_2 are called *coordinate basis* vectors!

We have developed the analogue of a tangent to a curve in the context of surfaces, so what about normals? In fact, we will now define the *normal space* to a point on some regular surface and develop a neat test for if vectors are in the tangent or normal spaces.

Definition 3.17 Let $M : U \to \mathbb{R}^3$ be a regular surface and $\mathbf{p} \in M$. The normal space to M at $\mathbf{p}$ is the set of vectors that are orthogonal to the tangent vectors at $\mathbf{p}$:

$$N_{\mathbf{p}}M := \{\mathbf{v} \in \mathbb{R}^3 : \mathbf{v} \cdot \mathbf{u} = 0 \text{ for all } \mathbf{u} \in T_{\mathbf{p}}M\}.$$

We then call any vector that lives within this space a normal vector to M at $\mathbf{p}$.

An important-yet-easy fact is that $N_{\mathbf{p}}M$ is one-dimensional! Anything living within the normal space is orthogonal to tangent vectors by definition; as $T_{\mathbf{p}}M$ is two-dimensional, there is only room for one further dimension. In particular, **any** non-zero normal vector is a basis for $N_{\mathbf{p}}M$.

Note: Even better, the fact $\varepsilon_1 \times \varepsilon_2$ is normal to $T_{\mathbf{p}}M$ means that this is a basis for $N_{\mathbf{p}}M$.

Remark 3.18 We also notice that the tangent space really just consists of all vectors that are orthogonal to the normal space (in particular, to any non-zero vector in $N_{\mathbf{p}}M$). In other words,

$$\mathbf{v} \in T_{\mathbf{p}}M \qquad \Leftrightarrow \qquad \mathbf{v} \cdot (\varepsilon_1 \times \varepsilon_2) = 0.$$

Exercise 33 Prove that $N_{\mathbf{p}}M$ is indeed a sub*space* of $\mathbb{R}^3$.

[**Hint:** You can show this directly by using the previous reminder. More efficiently, you can show these in one step: show that for arbitrary $\mathbf{v}, \mathbf{w} \in N_{\mathbf{p}}M$ and $a, b \in \mathbb{R}$, we have $a\mathbf{v} + b\mathbf{w} \in N_{\mathbf{p}}M$.]

We can now provide a really neat test for determining if a given vector is tangent/normal to a surface. Indeed, Remark 3.18 is precisely the test for membership of the tangent space. As for the normal space, we test to see that the vector is orthogonal to the basis vectors of the tangent space. This is all summarised below.

Method – Tangent or Normal Vector: Let M be a regular surface with point $\mathbf{p} \in M$ and $\mathbf{v} \in \mathbb{R}^3$. We have an easy test to determine if it is a tangent/normal vector or not.

1. We know $\mathbf{v} \in T_{\mathbf{p}}M$ if and only if $\mathbf{v} \cdot (\varepsilon_1 \times \varepsilon_2) = 0$.
2. We know $\mathbf{v} \in N_{\mathbf{p}}M$ if and only if both $\mathbf{v} \cdot \varepsilon_1 = 0$ and $\mathbf{v} \cdot \varepsilon_2 = 0$.
3. We know the vector is neither tangent nor normal if both of the above fail.

Note: The one and only one vector in **both** tangent and normal spaces is $\mathbf{0} = (0, 0, 0)$.

Example 3.19 Consider the surface $M : \mathbb{R}^2 \to \mathbb{R}^3$ given by $M(x_1, x_2) = (x_1, x_2, x_1x_2)$; this is another special case of Proposition 3.8 so we know for certain this is a regularly parametrised surface. We fix the point $\mathbf{p} = (3, -2, -6) \in M$ and notice that it has local coordinates $(3, -2)$. Hence, the coordinate basis vectors associated with the surface **at this point** are

$$\varepsilon_1(3, -2) = (1, 0, -2) \quad \text{and} \quad \varepsilon_2(3, -2) = (0, 1, 3) \qquad \Rightarrow \qquad \varepsilon_1 \times \varepsilon_2 = (2, -3, 1).$$

We will use the method above to determine if the below vectors are tangent or normal to M.

(i) Consider the vector $\mathbf{u} = (1, 2, 4)$:

- The dot product $\mathbf{u} \cdot (\varepsilon_1 \times \varepsilon_2) = 0$, which means $\mathbf{u} \in T_{\mathbf{p}}M$.

(ii) Consider the vector $\mathbf{v} = (2, 1, 1)$:

- The dot product $\mathbf{v} \cdot (\varepsilon_1 \times \varepsilon_2) = 2$, which means $\mathbf{v} \notin T_{\mathbf{p}}M$.
- The dot products $\mathbf{v} \cdot \varepsilon_1 = 0$ and $\mathbf{v} \cdot \varepsilon_2 = 4$, which implies $\mathbf{v} \notin N_{\mathbf{p}}M$.

(iii) Consider the vector $\mathbf{w} = (-4, 6, -2)$:

- The dot product $\mathbf{w} \cdot (\varepsilon_1 \times \varepsilon_2) = -28$, which means $\mathbf{w} \notin T_{\mathbf{p}}M$.
- The dot products $\mathbf{w} \cdot \varepsilon_1 = 0$ and $\mathbf{w} \cdot \varepsilon_2 = 0$, which implies $\mathbf{w} \in N_{\mathbf{p}}M$.

Exercise 34 (Revisited) Let $M : \mathbb{R}^2 \to \mathbb{R}^3$ given by $M(x_1, x_2) = (x_1, x_1^2 + x_2, x_1 - x_2^2)$ and fix the point $\mathbf{p} = (2, 3, 1) \in M$, as was done in Exercise 32. Determine whether the following are tangent vectors, normal vectors or neither. For those that are in $T_{\mathbf{p}}M$, write them as $a\varepsilon_1 + b\varepsilon_2$; for those that are in $N_{\mathbf{p}}M$, write them as $c(\varepsilon_1 \times \varepsilon_2)$, where $a, b, c \in \mathbb{R}$:

(i) $\mathbf{v}_1 = (-14, 2, 4)$.
(ii) $\mathbf{v}_2 = (1, 1, 5)$.
(iii) $\mathbf{v}_3 = (1, 4, -2)$.
(iv) $\mathbf{v}_4 = (-1, -4, 2)$.

We can now begin to discuss calculus on regularly parametrised surfaces. For this, we require the notion of *smoothness* for a map whose domain is a regular surface M. Recall that we know what it means for a function $D \to \mathbb{R}$ to be smooth where $D \subseteq \mathbb{R}$; see Definition 1.42.

Definition 3.20 Let $M : U \to \mathbb{R}^3$ be a regularly parametrised surface and $f : M \to \mathbb{R}$ be a function on M (so it assigns a real number to each point $\mathbf{p} \in M$). We call f smooth if the composition $f \circ M : U \to \mathbb{R}$ is smooth in the usual sense. We call the map $\widehat{f} := f \circ M$ the coordinate expression of f.

Example 3.21 (Continued) Consider again the sphere from Example 3.12 which is defined as

$$M : (0, \pi) \times (0, 2\pi) \to \mathbb{R}^3, \qquad M(\theta, \varphi) = (\sin\theta\cos\varphi, \sin\theta\sin\varphi, \cos\theta).$$

We consider two functions on M and write their coordinate expressions to see if they are smooth.

(i) Let $f : M \to \mathbb{R}$ be given by $f(y_1, y_2, y_3) = y_1 + y_2 + y_3$. Then, its coordinate expression is

$$\widehat{f}(\theta, \varphi) = f(M(\theta, \varphi)) = \sin(\theta)\cos(\varphi) + \sin(\theta)\sin(\varphi) + \cos(\theta).$$

This is smooth since sin and cos are smooth functions; f is a smooth map on the sphere.

(ii) Let $g : M \to \mathbb{R}$ be given by $g(y_1, y_2, y_3) = y_1^2 + y_2^2 + y_3^2$. Then, its coordinate expression is

$$\widehat{g}(\theta, \varphi) = g(M(\theta, \varphi)) = \sin^2(\theta)\cos^2(\varphi) + \sin^2(\theta)\sin^2(\varphi) + \cos^2(\theta) = 1.$$

This is smooth since constant functions are smooth; g is also a smooth map on the sphere.

Exercise 35 Let $M : \mathbb{R}^2 \to \mathbb{R}^3$ be the regular surface $M(x_1, x_2) = (x_1, x_1 + x_2, x_1 + x_2^2)$ and $f : M \to \mathbb{R}$ be given by $f(y_1, y_2, y_3) = y_1 y_2 y_3$. Determine the coordinate expression $\widehat{f}$ of this function. Hence, decide whether or not f is smooth function on M.

If we have a smooth function $f : M \to \mathbb{R}$ on a regular surface alongside a tangent vector $\mathbf{v} \in T_{\mathbf{p}}M$, we can ask "what is the rate of change of f at $\mathbf{p}$ in the direction of $\mathbf{v}$?". The answer to this question is encoded by the so-called *directional derivative* which we now introduce.

Definition 3.22 Let $f : M \to \mathbb{R}$ be a smooth function on a regular surface and $\mathbf{v} \in T_{\mathbf{p}}M$. The directional derivative of f at $\mathbf{p}$ along $\mathbf{v}$ is $\nabla_{\mathbf{v}} f(\mathbf{p}) \coloneqq (f \circ \alpha)'(0)$, where $\alpha : I \to M$ is a curve such that $\alpha(0) = \mathbf{p}$ and $\alpha'(0) = \mathbf{v}$. More explicitly, the directional derivative is

$$\nabla_{\mathbf{v}} f(\mathbf{p}) = \frac{\mathrm{d}}{\mathrm{d}t}\left[f(\alpha(t))\right]\Big|_{t=0}.$$

Note: The curve α is called a generating curve for $\mathbf{v}$; we know it exists by Definition 3.14. Also, be aware the direction derivative is a **number** associated with $\mathbf{p}$, not a function.

Example 3.23 Let $M : \mathbb{R}^2 \to \mathbb{R}^3$ be the regular surface $M(x_1, x_2) = (x_2, x_1 - x_2^2, x_1^3 + x_1 x_2)$ and fix the point $\mathbf{p} = (1, 0, 2) \in M$. We will compute the directional derivative of $f : M \to \mathbb{R}$ given by $f(y_1, y_2, y_3) = y_1 + y_2 + y_3$ at $\mathbf{p}$ along $\mathbf{v} = (1, -1, 5)$. To do this, we first have to choose a curve α that passes through $(1, 0, 2)$ in M and has initial velocity of $(1, -1, 5)$. We can proceed as we did in Example 3.15 to find a curve through $\mathbf{p}$. Indeed, $\mathbf{p} = M(1, 1)$ in local coordinates and so any curve through $\mathbf{p}$ must satisfy $\widehat{\alpha}(0) = (1, 1)$ for some $\widehat{\alpha}$. So one possibility is

$$\widehat{\alpha}(t) = (e^t, t + 1) \qquad \Rightarrow \qquad \alpha(t) = M(\widehat{\alpha}(t)) = (t + 1, e^t - t^2 - 2t - 1, e^{3t} + te^t + e^t).$$

If we substitute $t = 0$, we indeed see that $\alpha(0) = (1, 0, 2) = \mathbf{p}$. Now, the derivative of α is

$$\alpha'(t) = (1, e^t - 2t - 2, 3e^{3t} + 2e^t + te^t),$$

and we see that $\alpha'(0) = (1, -1, 5) = \mathbf{v}$. We have found ourselves a generating curve α. Next,

$$(f \circ \alpha)(t) = e^{3t} + (2 + t)e^t - t^2 - t \qquad \Rightarrow \qquad (f \circ \alpha)'(t) = 3e^{3t} + (2 + t + 1)e^t - 2t - 1.$$

Consequently, the directional derivative we are after is $\nabla_{\mathbf{v}} f(\mathbf{p}) = 3 + 3 - 1 = 5$.

Note: Before you ask "how on Earth did your choice of α produce the correct velocity?!", I will admit to cheating: I chose $\mathbf{p}$ and then found some α through it. I then differentiated it and set $t = 0$ to reverse engineer the 'correct' $\mathbf{v}$ to write at the start of the example.

Another possible problem that arises is that that are generally infinitely-many generating curves α. This should give us a bit of a headache at first because maybe the directional derivative isn't well defined; if $\nabla_{\mathbf{v}} f(\mathbf{p})$ changes depending on which α we choose, then Definition 3.22 is useless! Fortunately, we can rest our minds with the next result.

Proposition 3.24 *The directional derivative is independent of the generating curve used.*

Proof: Let $M : U \to \mathbb{R}^3$ be a regular surface and fix some $\mathbf{p} = M(\mathbf{z})$. For any $\mathbf{v} \in T_{\mathbf{p}}M$, take two generating curves $\alpha, \beta : I \to M$ for $\mathbf{v}$ through $\mathbf{p}$. As we have done previously, let's write

$$\alpha(t) = M(\widehat{\alpha}(t)) \qquad \text{and} \qquad \beta(t) = M(\widehat{\beta}(t))$$

for some functions $\widehat{\alpha}, \widehat{\beta} : I \to U$. Arguing similarly as in the proof of Theorem 3.16, we have

$$\begin{aligned}
\mathbf{0} &= \mathbf{v} - \mathbf{v} \\
&= \alpha'(0) - \beta'(0) \\
&= \frac{\mathrm{d}}{\mathrm{d}t}\left[M(\widehat{\alpha}(t)) - M(\widehat{\beta}(t))\right]\Big|_{t=0} \\
&= \frac{\mathrm{d}}{\mathrm{d}t}\left[M(\widehat{\alpha}_1(t), \widehat{\alpha}_2(t)) - M(\widehat{\beta}_1(t), \widehat{\beta}_2(t))\right]\Big|_{t=0} \\
&= \left(\widehat{\alpha}_1'(0) - \widehat{\beta}_1'(0)\right) \frac{\partial M}{\partial x_1}\Big|_{\mathbf{z}} + \left(\widehat{\alpha}_2'(0) - \widehat{\beta}_2'(0)\right) \frac{\partial M}{\partial x_2}\Big|_{\mathbf{z}} \\
&= \left(\widehat{\alpha}_1'(0) - \widehat{\beta}_1'(0)\right) \varepsilon_1(\mathbf{z}) + \left(\widehat{\alpha}_2'(0) - \widehat{\beta}_2'(0)\right) \varepsilon_2(\mathbf{z}).
\end{aligned}$$

As ε_1 and ε_2 are linearly independent (they are a basis for the tangent space by Theorem 3.16), we have $\widehat{\alpha}_1'(0) - \widehat{\beta}_1' = 0$ and $\widehat{\alpha}_2'(0) - \widehat{\beta}_2' = 0$, meaning $\widehat{\alpha}_1'(0) = \widehat{\beta}_1'(0)$ and $\widehat{\alpha}_2'(0) = \widehat{\beta}_2'(0)$. As such,

$$\begin{aligned}
\frac{\mathrm{d}}{\mathrm{d}t}\left[f(\alpha(t)) - f(\beta(t))\right]\Big|_{t=0} &= \frac{\mathrm{d}}{\mathrm{d}t}\left[f(M(\widehat{\alpha}(t))) - f(M(\widehat{\beta}(t)))\right]\Big|_{t=0} \\
&= \frac{\mathrm{d}}{\mathrm{d}t}\left[\widehat{f}(\alpha(t)) - \widehat{f}(\beta(t))\right]\Big|_{t=0} \\
&= \left(\widehat{\alpha}_1'(0) - \widehat{\beta}_1'(0)\right) \frac{\partial \widehat{f}}{\partial x_1}\Big|_{\mathbf{z}} + \left(\widehat{\alpha}_2'(0) - \widehat{\beta}_2'(0)\right) \frac{\partial \widehat{f}}{\partial x_2}\Big|_{\mathbf{z}} \\
&= 0.
\end{aligned}$$

□

Remark 3.25 The directional derivative isn't just the rate of change of f at $\mathbf{p}$ along $\mathbf{v}$; it depends on the length $\|\mathbf{v}\|$. Repeating Example 3.23 with $\mathbf{u} = 2\mathbf{v} = (2, -2, 10)$, this points in the same direction as $\mathbf{v}$ (they are parallel) but the corresponding directional derivative $\nabla_{\mathbf{u}} f(\mathbf{p}) = 2\nabla_{\mathbf{v}} f(\mathbf{p})$.

The problem with the directional derivative thus far is the need to conjure up a generating curve for the tangent vector given. However, we know that $T_{\mathbf{p}}M = \{a\varepsilon_1 + b\varepsilon_2 : a, b \in \mathbb{R}\}$ by Theorem 3.16. Therefore, if we can show that the directional derivative is linear with respect to both f and $\mathbf{v}$, then it turns out we will only need to ever compute $\nabla_{\varepsilon_1} f$ and $\nabla_{\varepsilon_2} f$ to find **any** $\nabla_{\mathbf{v}} f$.

Lemma 3.26 *The directional derivative is linear with respect to both the tangent vector and the function, i.e. for all $\mathbf{u}, \mathbf{v} \in T_{\mathbf{p}}M$ and $f, g : M \to \mathbb{R}$ with $a, b \in \mathbb{R}$, we have these:*

(i) $\nabla_{a\mathbf{u}+b\mathbf{v}} f = a(\nabla_{\mathbf{u}} f) + b(\nabla_{\mathbf{v}} f)$.

(ii) $\nabla_{\mathbf{v}}(af + bg) = a(\nabla_{\mathbf{v}} f) + b(\nabla_{\mathbf{v}} g)$.

Proof: Let $M : U \to \mathbb{R}^3$ be a regular surface and fix some $\mathbf{p} = M(\mathbf{z})$. For any $\mathbf{u}, \mathbf{v} \in T_{\mathbf{p}}M$, take two generating curves $\alpha, \beta : I \to M$ of $\mathbf{u}$ and $\mathbf{v}$ (respectively) through $\mathbf{p}$, so $\alpha'(0) = \mathbf{u}$ and $\beta'(0) = \mathbf{v}$. As usual, their coordinate expressions are denoted $\widehat{\alpha}, \widehat{\beta} : I \to U$ respectively. Then,

$$\widehat{\gamma}(t) = \widehat{\alpha}(at) + \widehat{\beta}(bt) - \mathbf{z}$$

is the coordinate expression of the curve $\gamma(t) = M(\widehat{\gamma}(t))$ through $\mathbf{p}$ with $\gamma'(0) = a\mathbf{u} + b\mathbf{v}$. Hence,

$$\begin{aligned}
\nabla_{a\mathbf{u}+b\mathbf{v}} f &= \frac{\mathrm{d}}{\mathrm{d}t}\left[f(\gamma(t))\right]\Big|_{t=0} \\
&= \frac{\mathrm{d}}{\mathrm{d}t}\left[\widehat{f}(\gamma(t))\right]\Big|_{t=0} \\
&= \widehat{\gamma}_1'(0)\,\frac{\partial \widehat{f}}{\partial x_1}\bigg|_{\mathbf{z}} + \widehat{\gamma}_2'(0)\,\frac{\partial \widehat{f}}{\partial x_2}\bigg|_{\mathbf{z}} \\
&= \left(a\widehat{\alpha}_1'(0) + b\widehat{\beta}_1'(0)\right)\frac{\partial \widehat{f}}{\partial x_1}\bigg|_{\mathbf{z}} + \left(a\widehat{\alpha}_2'(0) + b\widehat{\beta}_2'(0)\right)\frac{\partial \widehat{f}}{\partial x_2}\bigg|_{\mathbf{z}} \\
&= a\left(\widehat{\alpha}_1'(0)\,\frac{\partial \widehat{f}}{\partial x_1}\bigg|_{\mathbf{z}} + \widehat{\alpha}_2'(0)\,\frac{\partial \widehat{f}}{\partial x_2}\bigg|_{\mathbf{z}}\right) + b\left(\widehat{\beta}_1'(0)\,\frac{\partial \widehat{f}}{\partial x_1}\bigg|_{\mathbf{z}} + \widehat{\beta}_2'(0)\,\frac{\partial \widehat{f}}{\partial x_2}\bigg|_{\mathbf{z}}\right) \\
&= a(\nabla_{\mathbf{u}} f) + b(\nabla_{\mathbf{v}} f).
\end{aligned}$$

As for the second property, it follows straight from Definition 3.22:

$$\begin{aligned}
\nabla_{\mathbf{v}}(af + bg) &= \left((af + bg) \circ \beta\right)'(0) \\
&= (af \circ \beta + bg \circ \beta)'(0) \\
&= a(f \circ \beta)'(0) + b(g \circ \beta)'(0) \\
&= a(\nabla_{\mathbf{v}} f) + b(\nabla_{\mathbf{v}} g).
\end{aligned}$$

□

Note: We have achieved what we want: writing $\mathbf{v} \in T\mathbf{p}M$ as $\mathbf{v} = a\varepsilon_1 + b\varepsilon_2$, we know that

$$\nabla_{\mathbf{v}} f = a\nabla_{\varepsilon_1} f + b\nabla_{\varepsilon_2} f.$$

It remains to get neat expressions for the directional derivatives along the coordinate basis vectors. However, this can be done with a small amount of justification and is therefore left to justify in Exercise 36 (we present it as a 'show that' question since it will be convenient to refer to these expressions).

Exercise 36 Explain why the directional derivatives along coordinate vectors satisfy these:

$$\nabla_{\varepsilon_1} f = \frac{\partial \widehat{f}}{\partial x_1} \qquad \text{and} \qquad \nabla_{\varepsilon_2} f = \frac{\partial \widehat{f}}{\partial x_2}.$$

Example 3.27 Consider $M : \mathbb{R}^2 \to \mathbb{R}^3$ the regular surface $M(x_1, x_2) = (x_1 + x_2, x_1 x_2, x_1 - 2x_2)$. Given that $\mathbf{v} = (3, 1, 0)$ is a tangent vector to M at the point $\mathbf{p} = (1, 0, 1) = M(1, 0)$, we can compute the directional derivative of $f : M \to \mathbb{R}$ where $f(y_1, y_2, y_3) = y_1 y_2 + y_2 y_3$ at $\mathbf{p}$ along $\mathbf{v}$ by using Lemma 3.26 (specifically the above note that follows as a corollary). Indeed, labelling $\mathbf{z} := (1, 0)$, the coordinate basis vectors at $\mathbf{p} = M(\mathbf{z})$ are

$$\varepsilon_1(\mathbf{z}) = (1, 0, 1) \qquad \text{and} \qquad \varepsilon_2(\mathbf{z}) = (1, 1, -2).$$

We can use these to write $\mathbf{v} = a\varepsilon_1(\mathbf{z}) + b\varepsilon_2(\mathbf{z})$. Comparing each side tells us $a = 2$ and $b = 1$:

$$\mathbf{v} = 2\varepsilon_1 + \varepsilon_2.$$

Next, the coordinate expression of f is $\widehat{f}(x_1, x_2) = 2x_1^2 x_2 - x_1 x_2^2$. Using Exercise 36, we conclude

$$\nabla_{\mathbf{v}} f(\mathbf{p}) = 2 \left.\frac{\partial \widehat{f}}{\partial x_1}\right|_{\mathbf{z}} + \left.\frac{\partial \widehat{f}}{\partial x_2}\right|_{\mathbf{z}} = 2 \left[4x_1 x_2 - x_2^2\right]\Big|_{\substack{x_1=1\\x_2=0}} + \left[2x_1^2 - 2x_1 x_2\right]\Big|_{\substack{x_1=1\\x_2=0}} = 2.$$

Exercise 37 Let $M : \mathbb{R}^2 \to \mathbb{R}^3$ be given by $M(x_1, x_2) = (x_2, x_1 + x_2^2, x_2 + \cos(x_1 x_2))$.

(i) Show that M is a regularly parametrised surface.

(ii) Determine the local coordinates of the point $\mathbf{p} = (2, 4, 3) \in M$.

(iii) Show that $\mathbf{v} = (2, 5, 2) \in T_{\mathbf{p}}M$.

(iv) Show that $f : M \to \mathbb{R}$ given by $f(y_1, y_2, y_3) = y_1^3 - 2y_2$ is smooth.

(v) Compute the directional derivative $\nabla_{\mathbf{v}} f(\mathbf{p})$.

We can take the idea of the directional derivative further by introducing the notion of a *vector field* on a surface. We look at some theory and again reduce the calculations to simply finding partial derivatives.

Definition 3.28 A vector field on a regular surface M is a smooth map $V : M \to \mathbb{R}^3$, which means each component function $V_1, V_2, V_3 : M \to \mathbb{R}$ is smooth as in Definition 3.20. As with other functions, we call $\widehat{V} := V \circ M$ the coordinate expression of the vector field.

(i) We call V a tangent vector field if $V(\mathbf{p}) \in T_{\mathbf{p}}M$ for all $\mathbf{p} \in M$.

(ii) We call V a normal vector field if $V(\mathbf{p}) \in N_{\mathbf{p}}M$ for all $\mathbf{p} \in M$.

Remark 3.29 In the case that V is a tangent vector field, the coordinate expression has the form

$$\widehat{V}(x_1, x_2) = f(x_1, x_2)\varepsilon_1(x_1, x_2) + g(x_1, x_2)\varepsilon_2(x_1, x_2)$$

where $f, g : U \to \mathbb{R}$ are smooth functions. If V is a normal vector field, then we can write

$$\widehat{V}(x_1, x_2) = h(x_1, x_2)\varepsilon_1(x_1, x_2) \times \varepsilon_2(x_1, x_2)$$

for a smooth function $h : U \to \mathbb{R}$. These are a consequence of Theorem 3.16 and Remark 3.18, respectively, which give us explicit bases for $T_{\mathbf{p}}M$ and $N_{\mathbf{p}}M$.

Note: We often view vector fields and their coordinate expressions as one-and-the-same: the ε_1 and ε_2 are the tangent vector fields, and $\varepsilon_1 \times \varepsilon_2$ is the normal vector field.

Definition 3.30 Let $V : M \to \mathbb{R}^3$ be a vector field on a regular surface M and fix $\mathbf{v} \in T_{\mathbf{p}}M$. The directional derivative of V at $\mathbf{p}$ along $\mathbf{v}$ is $\nabla_{\mathbf{v}}V(\mathbf{p}) = (V \circ \alpha)'(0)$, where $\alpha : I \to M$ is a generating curve for $\mathbf{v}$ through $\mathbf{p}$.

In this way, the directional derivative takes a pair (consisting of a tangent vector and a vector field) and assigns to it a vector. There is no assumption on V: it does **not** have to be a tangent vector field (even if it is, there is no reason to expect $\nabla_{\mathbf{v}}V$ to be a tangent vector).

Lemma 3.31 *The directional derivative of V is independent of the generating curve used.*

Proof: If we write $V = (V_1, V_2, V_3)$ for component functions $V_i : M \to \mathbb{R}$, we can conclude

$$\nabla_{\mathbf{v}}V = \big((V_1 \circ \alpha)'(0), (V_2 \circ \alpha)'(0), (V_3 \circ \alpha)'(0)\big) = (\nabla_{\mathbf{v}}V_1, \nabla_{\mathbf{v}}V_2, \nabla_{\mathbf{v}}V_3), \tag{‡}$$

so the result is immediate from that for the usual directional derivative (Proposition 3.24). □

Example 3.32 Let $M : \mathbb{R}^2 \to \mathbb{R}^3$ be the regular surface given by $M(x_1, x_2) = (x_1, x_2, x_1x_2)$. Fix $\mathbf{p} = (1, 0, 0) = M(1, 0)$ and $\mathbf{v} = (1, -1, -1) \in T_{\mathbf{p}}M$. We compute the directional derivative of

$$V : M \to \mathbb{R}^3, \qquad V(y_1, y_2, y_3) = (-y_2, y_3, 0).$$

Indeed, we begin by verifying $\alpha(t) = (1 + t, -t, -t - t^2)$ is a generating curve for $\mathbf{v}$ through $\mathbf{p}$:

$$\alpha(0) = (1, 0, 0) = \mathbf{p} \qquad \text{and} \qquad \alpha'(0) = (1, -1, -1).$$

Now we have a generating curve (and we only need one by Lemma 3.31), we obtain

$$(V \circ \alpha)(t) = (t, -t^2 - t, 0) \qquad \Rightarrow \qquad (V \circ \alpha)'(t) = (1, -2t - 1, 0)$$

Consequently, the directional derivative we are after is $\nabla_{\mathbf{v}}V(\mathbf{p}) = (1, -1, 0)$.

Lemma 3.33 *The directional derivative of vector fields satisfies these properties, where V and W are vector fields and for every $\mathbf{u}, \mathbf{v} \in T_{\mathbf{p}}M$ and $f : M \to \mathbb{R}$ with $a, b \in \mathbb{R}$:*

(i) $\nabla_{a\mathbf{u}+b\mathbf{v}}V = a(\nabla_{\mathbf{u}}V) + b(\nabla_{\mathbf{v}}V)$.
(ii) $\nabla_{\mathbf{v}}(aV + bW) = a(\nabla_{\mathbf{v}}V) + b(\nabla_{\mathbf{v}}V)$.
(iii) $\nabla_{\mathbf{v}}(f \circ V)(\mathbf{p}) = \nabla_{\mathbf{v}}f(\mathbf{p})W(\mathbf{p}) + f(\mathbf{p})\nabla_{\mathbf{v}}W(\mathbf{p})$.
(iv) $\nabla_{\mathbf{v}}(V \cdot W)(\mathbf{p}) = \nabla_{\mathbf{v}}V(\mathbf{p}) \cdot W(\mathbf{p}) + V(p) \cdot \nabla_{\mathbf{v}}W(\mathbf{p})$.

Proof: (i) and (ii) both follow from Lemma 3.26 and (‡). Also, (iii) follows from Problem **3.XX** and (‡). As for (iv), we use both Lemma 3.26 and Problem **3.XX**. For ease of reading, we omit each "($\mathbf{p}$)" to give the reader a much easier time following what is going on:

$$\begin{aligned}
\nabla_{\mathbf{v}}(V \cdot W) &= \nabla_{\mathbf{v}}(V_1W_1 + V_2W_2 + V_3W_3) \\
&= \nabla_{\mathbf{v}}(V_1W_1) + \nabla_{\mathbf{v}}(V_2W_2) + \nabla_{\mathbf{v}}(V_3W_3) \\
&= (\nabla_{\mathbf{v}}V_1)W_1 + V_1(\nabla_{\mathbf{v}}W_1) + (\nabla_{\mathbf{v}}V_2)W_2 + V_2(\nabla_{\mathbf{v}}W_2) + (\nabla_{\mathbf{v}}V_3)W_3 + V_3(\nabla_{\mathbf{v}}W_3) \\
&= (\nabla_{\mathbf{v}}V_1, \nabla_{\mathbf{v}}V_2, \nabla_{\mathbf{v}}V_3) \cdot W + V \cdot (\nabla_{\mathbf{v}}W_1, \nabla_{\mathbf{v}}W_2, \nabla_{\mathbf{v}}W_3) \\
&= \nabla_{\mathbf{v}}V \cdot W + V \cdot \nabla_{\mathbf{v}}W.
\end{aligned}$$ □

Note: We use Lemma 3.33 to reduce the calculation of $\nabla_{\mathbf{v}}V$ to just $\nabla_{\varepsilon_1}V$ and $\nabla_{\varepsilon_2}V$.

Exercise 38 Compute the directional derivative from Example 3.32 **without** first writing a generating curve, that is use Lemma 3.33 to find $\nabla_{\mathbf{v}}V$ by calculating $\nabla_{\varepsilon_1}V$ and $\nabla_{\varepsilon_2}V$.

It is possible to define a new vector field from two other ones, as long as one of the two is a **tangent** vector field (there is nothing assumed about the second). In this way, we define the directional derivative along a tangent vector field.

> **Definition 3.34** Let U be a tangent vector field and V any other vector field. We define the directional derivative along a tangent vector field at $\mathbf{p} \in M$ as the directional derivative of V at $\mathbf{p}$ in the direction of the tangent vector $U(\mathbf{p})$. We denote this $\nabla_U V$.

Example 3.35 Consider the unit sphere $M(x_1, x_2) = (\sin x_1 \cos x_2, \sin x_1 \sin x_2, \cos x_1)$ in the notation of Example 3.9, and let N be the unit normal vector field, that is

$$N = \frac{\varepsilon_1 \times \varepsilon_2}{\|\varepsilon_1 \times \varepsilon_2\|}.$$

We will compute a number of directional derivatives along the tangent vector fields ε_1 and ε_2. These hint at the behaviours of ∇_{ε_i} and how they interact with the unit normal vector field (not just for the sphere M but **any** regular surface). Before we get ahead of ourselves, we see that

$$\begin{aligned}
\varepsilon_1 &= (\cos x_1 \cos x_2, \cos x_1 \sin x_2, -\sin x_1),\\
\varepsilon_2 &= (-\sin x_1 \sin x_2, \sin x_1 \cos x_2, 0)\\
\Rightarrow \qquad \varepsilon_1 \times \varepsilon_2 &= (\sin^2 x_1 \cos x_2, \sin^2 x_1 \sin x_2, \sin x_1 \cos x_1) \neq \mathbf{0}.
\end{aligned}$$

We already computed these in Example 3.9, but what we didn't compute was the vector field:

$$N = \frac{\varepsilon_1 \times \varepsilon_2}{\sqrt{\sin^4 x_1 \cos^2 x_2 + \sin^4 x_1 \sin^2 x_2 + \sin^2 x_1 \cos^2 x_1}} = (\sin x_1 \cos x_2, \sin x_1 \sin x_2, \cos x_1),$$

since the messy denominator simplifies to $\sin(x_1)$. We get the following directional derivatives:

$$\begin{aligned}
\nabla_{\varepsilon_1} \varepsilon_1 &= \frac{\partial \varepsilon_1}{\partial x_1} = (-\sin x_1 \cos x_2, -\sin x_1 \sin x_2, -\cos x_1),\\
\nabla_{\varepsilon_1} \varepsilon_2 &= \frac{\partial \varepsilon_2}{\partial x_1} = (-\cos x_1 \sin x_2, \cos x_1 \cos x_2, 0),\\
\nabla_{\varepsilon_2} \varepsilon_1 &= \frac{\partial \varepsilon_1}{\partial x_2} = (-\cos x_1 \sin x_2, \cos x_1 \cos x_2, 0),\\
\nabla_{\varepsilon_2} \varepsilon_2 &= \frac{\partial \varepsilon_2}{\partial x_2} = (-\sin x_1 \cos x_2, -\sin x_1 \sin x_2, 0),\\
\nabla_{\varepsilon_1} N &= \frac{\partial N}{\partial x_1} = (\cos x_1 \cos x_2, \cos x_1 \sin x_2, -\sin x_1),\\
\nabla_{\varepsilon_2} N &= \frac{\partial N}{\partial x_2} = (-\sin x_1 \sin x_2, \sin x_1 \cos x_2, 0).
\end{aligned}$$

Exercise 39 Show that $\nabla_{\varepsilon_1} N$ and $\nabla_{\varepsilon_2} N$ in Example 3.35 are **tangent** vector fields.

[**Hint:** You just need to demonstrate each dot product $N \cdot \nabla_{\varepsilon_i} N = 0$.]

Another interesting observation to come from Example 3.35 is that $\nabla_{\varepsilon_1}\varepsilon_2 = \nabla_{\varepsilon_2}\varepsilon_1$. Surely this can't be a coincidence, right? This is truly not a coincidence at all, as we can prove!

Proposition 3.36 *For each i and j, we have $\nabla_{\varepsilon_i}\varepsilon_j = \nabla_{\varepsilon_j}\varepsilon_i$.*

Proof: Of course this is trivial if $i = j$, so we only need to check the case where $i \neq j$, that is where $i = 1$ and $j = 2$ (without loss of generality). Writing the directional derivative in full then,

$$\nabla_{\varepsilon_1}\varepsilon_2 = \frac{\partial \varepsilon_2}{\partial x_1} = \frac{\partial}{\partial x_1}\left[\frac{\partial M}{\partial x_2}\right] = \frac{\partial^2 M}{\partial x_1 \partial x_2} = \frac{\partial^2 M}{\partial x_2 \partial x_1},$$

where the swapping of the partial derivatives is discussed in the below note. This means that

$$\nabla_{\varepsilon_2}\varepsilon_1 = \frac{\partial^2 M}{\partial x_2 \partial x_1} = \frac{\partial \varepsilon_2}{\partial x_1} = \nabla_{\varepsilon_1}\varepsilon_2. \qquad \square$$

Note: The fact we swap the order of the partial derivatives is a fundamental result which is always assumed in basic calculus courses known as the *Mixed Derivatives Theorem* (also known as *Schwarz's Theorem*, or *Clairaut's Theorem*, or *Young's Theorem*).

Exercise 40 Prove that $\nabla_{\varepsilon_1} N$ and $\nabla_{\varepsilon_2} N$ are tangent vector fields for **any** surface M.

3.2 Orientable Surfaces

For the remainder of this course, we focus on a specific class of surfaces known as *orientable*. It is on orientable surfaces that we can give rigorous mathematical meaning to curvature of surfaces.

Definition 3.37 An orientation on a regular surface $M : U \to \mathbb{R}^3$ is a choice of unit normal vector field N, i.e. a smooth assignment of a unit vector $\mathbf{v} \in N_{\mathbf{p}}M$ for each $y \in M$. We call a regular surface M with an orientation N an orientable surface.

From Remark 3.18, the unit normal $N(\mathbf{p})$ at a point $\mathbf{p} \in M$ determines the tangent space at $\mathbf{p}$:

$$T_{\mathbf{p}}M = \{\mathbf{v} \in \mathbb{R}^3 : \mathbf{v} \cdot N(\mathbf{p}) = 0\}.$$

In other words, we can obtain information about how the tangent space varies with $\mathbf{p}$ by calculating directional derivatives $\nabla_{\mathbf{u}} N$ where $\mathbf{u} \in T_{\mathbf{p}}M$. This is the analogue of how the tangent line along a curve varies. We must ensure that N is a **unit** normal vector field, otherwise $\nabla_{\mathbf{u}} N$ would contain information about the rate of change of the length of N, as well as the rate of change of direction. This is the surface analogue to Remark 2.32.

Note: Assuming that M is *connected*, there are only two possible orientations to choose:

$$N_+ = \frac{\varepsilon_1 \times \varepsilon_2}{\|\varepsilon_1 \times \varepsilon_2\|} \qquad \text{and} \qquad N_- = -\frac{\varepsilon_1 \times \varepsilon_2}{\|\varepsilon_1 \times \varepsilon_2\|} = \frac{\varepsilon_2 \times \varepsilon_1}{\|\varepsilon_2 \times \varepsilon_1\|}.$$

We call N_+ the *canonical orientation* on M; we always use this orientation (instead of N_-) unless otherwise stated, but this convention of course depends on the ordering of the local coordinates.

Lemma 3.38 *For $M : U \to \mathbb{R}^3$ an orientable surface with orientation N and $\mathbf{u} \in T_{\mathbf{p}}M$,*

$$\nabla_{\mathbf{u}} N \in T_{\mathbf{p}}M.$$

Proof: It suffices to show $(\nabla_{\mathbf{u}} N) \cdot N(\mathbf{p}) = 0$. To that end, N is a unit vector field which means $N \cdot N = 1$. Taking the directional derivative with respect to $\mathbf{u}$ using Lemma 3.33(iv) yields

$$(\nabla_{\mathbf{u}} N) \cdot N(\mathbf{p}) + N(\mathbf{p}) \cdot (\nabla_{\mathbf{u}} N) = 0.$$

But the left-hand side is simply $2(\nabla_{\mathbf{u}} N) \cdot N(\mathbf{p})$, from which the result follows. $\square$

Definition 3.39 Let $M : U \to \mathbb{R}^3$ be an orientable surface with orientation N. The shape operator (or Weingarten map) at $\mathbf{p} \in M$ is the map $S_{\mathbf{p}} : T_{\mathbf{p}}M \to T_{\mathbf{p}}M$ given by

$$S_{\mathbf{p}}(\mathbf{u}) = -\nabla_{\mathbf{u}} N.$$

The fact that the image of this map really is the tangent space $T_{\mathbf{p}}M$ follows from Lemma 3.38. A useful property of the shape operator is that it is a *linear map* between vector spaces: for all $\mathbf{u}, \mathbf{v} \in \mathbb{R}^3$ and $a, b \in \mathbb{R}$, it follows that

$$S_{\mathbf{p}}(a\mathbf{u} + b\mathbf{v}) = aS_{\mathbf{p}}(\mathbf{u}) + bS_{\mathbf{p}}(\mathbf{v}). \qquad (\diamond)$$

If you are familiar with abstract algebra, you may be thinking "but linear maps and matrices are in association with one another, so I wonder if we can write down a matrix for the shape operator". If you did think this, great! If not, do not worry; we will discuss this idea more soon.

Example 3.40 Let $M : \mathbb{R}^2 \to \mathbb{R}^3$ be given by $M(x_1, x_2) = (x_1, x_2, 0)$; this is a regular surface (a plane) which can be oriented by $N(\mathbf{p}) = (0, 0, 1)$, the unit vector in the y_3-direction. Note that this is a constant vector field, so for all $\mathbf{u} \in T_\mathbf{p}M$, we have

$$\nabla_\mathbf{u} N = \mathbf{0}.$$

Therefore, the shape operator $S_\mathbf{p} \equiv 0$, meaning it is zero when evaluated at every $\mathbf{u} \in T_\mathbf{p}M$.

Exercise 41 Determine the shape operator of the cylinder $M : (0, 2\pi) \times \mathbb{R} \to \mathbb{R}^3$ given by $M(x_1, x_2) = (x_2, r\cos x_1, r\sin x_1)$, where $r > 0$ is the radius of the cylinder.

Example 3.41 Let M be the unit sphere with its polar coordinate parametrisation from Example 3.9. Recall also that we computed the following directional derivatives in Example 3.35:

$$\nabla_{\varepsilon_1} N = \varepsilon_1 \qquad \text{and} \qquad \nabla_{\varepsilon_2} N = \varepsilon_2.$$

We can write any tangent vector $\mathbf{u} \in T_\mathbf{p}M$ as a linear combination of the coordinate basis vectors (Theorem 3.16), so $\mathbf{u} = a\varepsilon_1 + b\varepsilon_2$ for some $a, b \in \mathbb{R}$. Linearity ($\diamond$) of the shape operator implies

$$S_\mathbf{p}(\mathbf{u}) = S_\mathbf{p}(a\varepsilon_1 + b\varepsilon_2) = aS_\mathbf{p}(\varepsilon_1) + bS_\mathbf{p}(\varepsilon_2) = -a\nabla_{\varepsilon_1} N - b\nabla_{\varepsilon_2} N = -a\varepsilon_1 - b\varepsilon_2 = -\mathbf{u}.$$

Therefore, the shape operator $S_\mathbf{p} = -\,\mathrm{id}$, meaning it is minus the identity map (it flips the sign).

Reminder: Let V and W be vector spaces which have bases $\{\mathbf{v}_1, ..., \mathbf{v}_n\}$ and $\{\mathbf{w}_1, ..., \mathbf{w}_m\}$, respectively. Any linear map $L : V \to W$ can be expressed as $L(\mathbf{x}) = A\mathbf{x}$, where A is some $m \times n$ matrix. The matrix of L with respect to the W-basis $\{\mathbf{w}_1, ..., \mathbf{w}_m\}$ is obtained by writing each of $L(\mathbf{v}_1), ..., L(\mathbf{v}_n)$ in said basis, and putting the coefficients in **columns**.

Example 3.42 (Continued) We obtained the shape operator $S_\mathbf{p}(\mathbf{u}) = -\mathbf{u}$ for the unit sphere in Example 3.41. Because $S_\mathbf{p}$ is a linear map and $\{\varepsilon_1, \varepsilon_2\}$ is a basis for $T_\mathbf{p}M$, we can write the matrix of $S_\mathbf{p}$ with respect to the basis $\{\varepsilon_1, \varepsilon_2\}$. To do this, write $S_\mathbf{p}(\varepsilon_1)$ and $S_\mathbf{p}(\varepsilon_2)$ in this basis:

$$S_\mathbf{p}(\varepsilon_1) = -\varepsilon_1 = (-1)\varepsilon_1 + 0\varepsilon_2, \qquad \text{and} \qquad S_\mathbf{p}(\varepsilon_2) = -\varepsilon_2 = 0\varepsilon_1 + (-1)\varepsilon_2.$$

Consequently, the matrix of the shape operator of the unit sphere with respect to $\{\varepsilon_1, \varepsilon_2\}$ is

$$S_\mathbf{p} = \begin{pmatrix} -1 & 0 \\ 0 & -1 \end{pmatrix}.$$

Exercise 42 (Continued) Write the matrix of the shape operator found in Exercise 41.

It turns out that the matrix of a shape operator is not only a neat way to view $S_{\mathbf{p}}$, but its *eigenvalues* are what encode the notion of curvature we wish to study for surfaces. An important property of linear maps is now defined, and we will soon see that it is satisfied by $S_{\mathbf{p}}$.

Definition 3.43 A linear map $L : V \to V$ on a vector space is self-adjoint if, for all $\mathbf{u}, \mathbf{v} \in V$,

$$L(\mathbf{u}) \cdot \mathbf{v} = \mathbf{u} \cdot L(\mathbf{v}).$$

Remark 3.44 There is a more general concept called *adjointness*, which says that if we have a linear map $L : V \to W$ between vector spaces, there exists a linear map $L^* : W \to V$ such that

$$L(\mathbf{v}) \cdot \mathbf{w} = \mathbf{v} \cdot L^*(\mathbf{w}).$$

Notice that the dot product on the left happens in W, whereas the dot product on the right happens in V (but don't dwell on this too much; we are getting dangerously close to something too interesting, so we must stop ourselves before we overindulge). However, the point is that a **self**-adjoint is a special case where $V = W$ and $L^* \equiv L$. The basic idea is this: a self-adjoint allows us to move the linear map from one dot product factor to anther.

Theorem 3.45 *The shape operator is self-adjoint.*

Proof: Let $S_{\mathbf{p}} : T_{\mathbf{p}}M \to T_{\mathbf{p}}M$ be the shape operator at the point $\mathbf{p} \in M$ in some orientable surface M. Due to the fact that $\{\varepsilon_1, \varepsilon_2\}$ spans $T_{\mathbf{p}}M$ by Theorem 3.16, it is sufficient to show

$$S_{\mathbf{p}}(\varepsilon_i) \cdot \varepsilon_j = \varepsilon_i \cdot S_{\mathbf{p}}(\varepsilon_j)$$

for all i and j. To see this, we proceed as follows by using Lemma 3.33(iv) and Proposition 3.36:

$$\begin{aligned}
S_{\mathbf{p}}(\varepsilon_i) \cdot \varepsilon_j - \varepsilon_i \cdot S_{\mathbf{p}}(\varepsilon_j) &= -\nabla_{\varepsilon_i} N \cdot \varepsilon_j + \varepsilon_i \cdot \nabla_{\varepsilon_j} N \\
&= -\nabla_{\varepsilon_i}(N \cdot \varepsilon_j) + N \cdot \nabla_{\varepsilon_i}\varepsilon_j + \nabla_{\varepsilon_j}(\varepsilon_i \cdot N) - \nabla_{\varepsilon_j}\varepsilon_i \cdot N \\
&= -\nabla_{\varepsilon_i}(0) + N \cdot \nabla_{\varepsilon_i}\varepsilon_j + \nabla_{\varepsilon_j}(0) - \nabla_{\varepsilon_j}\varepsilon_i \cdot N \\
&= N \cdot \nabla_{\varepsilon_i}\varepsilon_j - \nabla_{\varepsilon_j}\varepsilon_i \cdot N \\
&= N \cdot (\nabla_{\varepsilon_i}\varepsilon_j - \nabla_{\varepsilon_j}\varepsilon_i) \\
&= 0.
\end{aligned}$$

□

Reminder: Let V be a vector space and $L : V \to V$ be a linear map. An eigenvalue of L is a **non-zero** number $\lambda \in \mathbb{R} \setminus \{0\}$ such that $L(\mathbf{v}) = \lambda\mathbf{v}$ for some $\mathbf{v} \in V$. The corresponding vector $\mathbf{v}$ is called an eigenvector corresponding to the eigenvalue λ.

Remark 3.46 We will try to give some intuition behind eigenvalues and eigenvectors. In general, a matrix A acts on a vector $\mathbf{v}$ by both rotating and rescaling it. Looking at the defining equation for an eigenvalue and an eigenvector (in matrix language), it says $A\mathbf{v} = \lambda\mathbf{v}$. In other words, the action of this matrix on an eigenvector is a **rescaling only**; the direction in which the vector points is unchanged by the matrix. Moreover, no other vectors will be stretched as much as the eigenvector by the action of A.

Looking at what we have introduced thus far and comparing to the above reminder (and Remark 3.46), it seems like we have an interest in the eigenvalues and eigenvectors of the shape operator. Indeed, we want to say that its eigenvalues measure curvatures of M, and its eigenvectors measure in what direction we think of these curvatures. But before we can do that, we must ensure that what we are talking about is well-defined. Let's begin by reminding ourselves on *how* we find eigenvalues and eigenvectors.

Method – Finding Eigenvalues and Eigenvectors: Let A be an $n \times n$ matrix.

1. Find the eigenvalues $\lambda_1, ..., \lambda_k$ by solving $\det(A - \lambda I_n) = 0$.
2. Choose one of the eigenvalues, say λ_1.
3. Solve the system $(A - \lambda_1 I_n)\mathbf{v}_1 = \mathbf{0}$ to get a family of solutions for $\mathbf{v}_1$.
4. Repeat this for each of the other eigenvalues $\lambda_2, ..., \lambda_k$ in Step **1**.

Example 3.47 Consider the linear map $L : \mathbb{R}^2 \to \mathbb{R}^2$ given by $L(\mathbf{x}) = A\mathbf{x}$ where

$$A = \begin{pmatrix} 2 & 2 \\ 5 & -1 \end{pmatrix}.$$

Using the above method, we must solve $\det(A - \lambda I_2) = 0$, which is precisely to say

$$\det \begin{pmatrix} 2-\lambda & 2 \\ 5 & -1-\lambda \end{pmatrix} = 0 \quad \Leftrightarrow \quad (2-\lambda)(-1-\lambda) - 10 = 0$$
$$\Leftrightarrow \quad -2 - 2\lambda + \lambda + \lambda^2 - 10 = 0$$
$$\Leftrightarrow \quad \lambda^2 - \lambda - 12 = 0.$$

Using the quadratic formula (or trying to factorise), we see that the two solutions to the above equation are $\lambda_1 = -3$ and $\lambda_2 = 4$. We now want to determine the associated eigenvectors.

(i) For $\lambda_1 = -3$, consider the system of equations $(A - \lambda_1 I_2)\mathbf{v}_1 = \mathbf{0}$, which can be written as

$$\begin{pmatrix} 5 & 2 \\ 5 & 2 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}.$$

Both the top and the bottom equations read $5x_1 + 2x_2 = 0$. We rearrange this to make one of the variables the subject, say $x_1 = -\frac{2}{5}x_2$. In words, the first entry is $-\frac{2}{5}$ times the second entry. Therefore, all solutions $\mathbf{v}_1$ to this system of equations have the form

$$\mathbf{v}_1 = t \begin{pmatrix} 2 \\ -5 \end{pmatrix}, \qquad \text{for any } t \in \mathbb{R}.$$

(ii) For $\lambda_2 = 4$, consider the system of equations $(A - \lambda_1 I_2)\mathbf{v}_1 = \mathbf{0}$, which can be written as

$$\begin{pmatrix} -2 & 2 \\ 5 & -5 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}.$$

The top equation reads $-2x_1 + 2x_2 = 0$, whereas the bottom reads $5x_1 - 5x_2 = 0$. Note that both of these rearrange to $x_1 = x_2$. In words, the first entry is equal to the second entry. Therefore, all solutions $\mathbf{v}_2$ to this system of equations have the form

$$\mathbf{v}_2 = t \begin{pmatrix} 1 \\ 1 \end{pmatrix}, \qquad \text{for any } t \in \mathbb{R}.$$

Note: In practice, when asked to find *the* eigenvector associated to an eigenvalue, we often mean the **unit** eigenvector; we select t to be the reciprocal of the norm of the vector part.

Exercise 43 Find the eigenvalues and associated eigenvectors of the following matrix:

$$A = \begin{pmatrix} 0 & 1 \\ -2 & -3 \end{pmatrix}.$$

There is no general guarantee the eigenvalues are real numbers, e.g. if $\det(A - \lambda I_2) = \lambda^2 + 1$, then the corresponding eigenvalues are $\lambda_1 = i$ and $\lambda_2 = -i$, where $i^2 = -1$ is the *imaginary unit*. However, it turns out that self-adjointness of a linear map does guarantee this.

Theorem 3.48 *Let $L : V \to V$ be a self-adjoint linear map. Then, all eigenvalues of L are real and the corresponding eigenvectors form an orthonormal basis of V.*

Proof: Let $\{\mathbf{e}_1, \mathbf{e}_2\}$ be an orthonormal basis for V, and suppose the matrix of L in this basis is

$$A = \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix}.$$

Using the matrix form of the linear map, we see that

$$L(\mathbf{e}_1) \cdot \mathbf{e}_2 = A\mathbf{e}_1 \cdot \mathbf{e}_2 = (a_{11}\mathbf{e}_1 + a_{21}\mathbf{e}_2) \cdot \mathbf{e}_2 = a_{21}$$

and

$$\mathbf{e}_1 \cdot L(\mathbf{e}_2) = \mathbf{e}_1 \cdot A\mathbf{e}_2 = \mathbf{e}_1 \cdot (a_{12}\mathbf{e}_1 + a_{22}\mathbf{e}_2) = a_{12}.$$

But self-adjointness tells us that $L(\mathbf{e}_1) \cdot \mathbf{e}_2 = \mathbf{e}_1 \cdot L(\mathbf{e}_2)$, that is the two equations above are equal. In other words, $a_{21} = a_{12}$. Therefore, the characteristic equation $\det(A - \lambda I_2) = 0$ is precisely

$$\det(A - \lambda I_2) = \det \begin{pmatrix} a_{11} - \lambda & a_{12} \\ a_{\mathbf{12}} & a_{22} - \lambda \end{pmatrix} = \lambda^2 - (a_{11} + a_{22})\lambda + a_{11}a_{22} - a_{12}^2.$$

The discriminant of this quadratic tells us about the type of solutions, and in this case it is

$$(a_{11} + a_{22})^2 - 4a_{11}a_{22} + 4a_{12}^2 = (a_{11} - a_{22})^2 + 4a_{12}^2 \geq 0.$$

The fact the discriminant is non-negative guarantees that all solutions λ (i.e. the eigenvalues) are real. As for the second part of this statement, let λ_1 and λ_2 be the (real) eigenvalues, and denote by $\mathbf{v}_1$ and $\mathbf{v}_2$ the corresponding respective eigenvectors. There are two cases to consider.

(i) If $\lambda_1 = \lambda_2$ $(= \lambda)$, then the characteristic equation has a repeated root. This corresponds to the discriminant being exactly zero, so $a_{11} = a_{22} = \lambda$ and $a_{12} = 0$. In this case, we see that $A = \lambda I_2$ and therefore $L \equiv \lambda\,\mathrm{id}$; **every** vector in V is an eigenvector with corresponding eigenvalue λ. Thus, we can choose $\mathbf{v}_1 = \mathbf{e}_1$ and $\mathbf{v}_2 = \mathbf{e}_2$ to be the orthonormal basis.

(ii) If $\lambda_1 \neq \lambda_2$, then we can use self-adjointness to see that

$$0 = L(\mathbf{v}_1) \cdot \mathbf{v}_2 - \mathbf{v}_1 \cdot L(\mathbf{v}_2) = (\lambda_1 - \lambda_2)\mathbf{v}_1 \cdot \mathbf{v}_2.$$

But since $\lambda_1 - \lambda_2 \neq 0$, it must be that $\mathbf{v}_1 \cdot \mathbf{v}_2 = 0$, i.e. they are orthogonal. Because we can rescale our eigenvectors, we can ensure they have unit length, giving orthonormality. □

Definition 3.49 Let $M : U \to \mathbb{R}^3$ be an orientable surface and $S_\mathbf{p} : T_\mathbf{p}M \to T_\mathbf{p}M$ be its shape operator. The principal curvatures of M at $\mathbf{p}$ are the eigenvalues of $S_\mathbf{p}$, denoted κ_1 and κ_2. The principal curvature directions of M at $\mathbf{p}$ are the corresponding unit eigenvectors.

Example 3.50 Consider the regular surface $M : \mathbb{R}^2 \to \mathbb{R}^3$ given by $M(x_1, x_2) = (x_1, x_2, \frac{1}{2}x_2^2)$ (which is certainly regular by Proposition 3.8). We will calculate the principal curvatures at a general point on this surface. We begin by constructing the unit normal vector field N. For this, we first require the coordinate basis vectors, which are

$$\varepsilon_1 = \frac{\partial M}{\partial x_1} = (1, 0, 0) \qquad \text{and} \qquad \varepsilon_2 = \frac{\partial M}{\partial x_2} = (0, 1, x_2).$$

As a consequence, we have $\varepsilon_1 \times \varepsilon_2 = (0, -x_2, 1)$. From it, we may construct the unit normal:

$$N = \frac{\varepsilon_1 \times \varepsilon_2}{\|\varepsilon_1 \times \varepsilon_2\|} = \frac{(0, -x_2, 1)}{\sqrt{1 + x_2^2}}.$$

Following this, the definition of the shape operator means we have

$$S_\mathbf{p}(\varepsilon_1) = -\nabla_{\varepsilon_1} N = -\frac{\partial N}{\partial x_1} = (0, 0, 0) = 0\varepsilon_1 + 0\varepsilon_2$$

and

$$S_\mathbf{p}(\varepsilon_2) = -\nabla_{\varepsilon_2} N = -\frac{\partial N}{\partial x_2} = \frac{(0, 1, x_2)}{(1 + x_2^2)^{\frac{3}{2}}} = 0\varepsilon_1 + \frac{1}{(1 + x_2^2)^{\frac{3}{2}}}\varepsilon_2.$$

If we construct the corresponding shape operator matrix, we can read-off the eigenvalues:

$$S_\mathbf{p} = \begin{pmatrix} 0 & 0 \\ 0 & (1 + x_2^2)^{-\frac{3}{2}} \end{pmatrix},$$

from which it follows that $\kappa_1 = 0$ and $\kappa_2 = (1 + x_2^2)^{-\frac{3}{2}}$. The principal curvature directions are

$$\mathbf{v}_1 = \frac{\varepsilon_1}{\|\varepsilon_1\|} = (1, 0, 0) \qquad \text{and} \qquad \mathbf{v}_2 = \frac{\varepsilon_2}{\|\varepsilon_2\|} = \frac{(0, 1, x_2)}{\sqrt{1 + x_2^2}}.$$

Exercise 44 Compute the principal curvatures and the principal curvature directions of the regular surface $M : \mathbb{R}^2 \to \mathbb{R}^3$ given by $M(x_1, x_2) = (x_1, x_2, x_1 x_2)$ at the point $\mathbf{p} = (1, 0, 0)$.

Due to Theorem 3.48, we immediately know that the principal curvatures $\kappa_1, \kappa_2 \in \mathbb{R}$ and the principal curvature directions form an orthonormal basis for $T_{\mathbf{p}}M$, for any orientable surface M.

Note: We often want to discuss tangent vectors of length one, so the unit tangent space is

$$U_{\mathbf{p}}M := \{\mathbf{v} \in T_{\mathbf{p}}M : \|\mathbf{v}\| = 1\}.$$

Geometrically, this is the unit circle inside $T_{\mathbf{p}}M$ and it is **not** itself a vector space.

This gives us our first taste of the notion of curvature for surfaces. We will soon define some more notions in terms of the principal curvatures, one of which is an *intrinsic* property of whatever surface it is we work with. First, we will introduce a function that encodes κ_1 and κ_2.

Definition 3.51 Let $M : U \to \mathbb{R}^3$ be an orientable surface and $S_{\mathbf{p}}$ its shape operator. The normal curvature function of M at $\mathbf{p}$ is the map $k_{\mathbf{p}} : U_{\mathbf{p}}M \to \mathbb{R}$ given by

$$k_{\mathbf{p}}(\mathbf{u}) = \mathbf{u} \cdot S_{\mathbf{p}}(\mathbf{u}).$$

Before we look at how this function encodes the principal curvatures, we should justify why it is called the "normal curvature" function. For this, we use the notion of curvature for curves, and we recall the notion of a generating curve that we introduced in Definition 3.22.

Lemma 3.52 *Let M be an orientable surface with orientation N, $\mathbf{u} \in U_{\mathbf{p}}M$ and let $\alpha : I \to M$ be any unit speed curve in M that generates $\mathbf{u}$. Then, the component of the curvature vector k of α in the direction of $N(\mathbf{p})$ is precisely $k_{\mathbf{p}}(\mathbf{u})$.*

Proof: By definition, we know that $\alpha(0) = \mathbf{p}$ and $\alpha'(0) = \mathbf{u}$. Because α is a unit speed curve, we know its curvature vector at $\mathbf{p}$ is $k(0) = \alpha''(0)$, by Definition 2.33. Since α stays in M, its velocity is always tangent to M. It follows from this that

$$\alpha'(s) \cdot N(\alpha(s)) = 0$$

for all $s \in I$. Differentiating the above equation wit respect to s, and setting $s = 0$, we get

$$\alpha''(0) \cdot N(\alpha(0)) + \alpha'(0) + (N \circ \alpha)'(0) = 0 \qquad \Rightarrow \qquad k(0) \cdot N(\mathbf{p}) + \mathbf{u} \cdot (N \circ \alpha)'(0) = 0.$$

By Definition 3.22, we know that $(N \circ \alpha)'(0) = \nabla_{\mathbf{u}} N$. Consequently, the above says that

$$k(0) \cdot N(\mathbf{p}) = -\mathbf{u} \cdot \nabla_{\mathbf{u}} N = u \cdot S_{\mathbf{p}}(\mathbf{u}) = k_{\mathbf{p}}(\mathbf{u}).$$ □

Theorem 3.53 *For M an orientable surface with principal curvatures $\kappa_1 \leq \kappa_2$ at $\mathbf{p} \in M$,*

$$\kappa_1 = \min\{k_{\mathbf{p}}(\mathbf{u}) : \mathbf{u} \in U_{\mathbf{p}}M\} \qquad \textit{and} \qquad \kappa_2 = \max\{k_{\mathbf{p}}(\mathbf{u}) : \mathbf{u} \in U_{\mathbf{p}}M\}.$$

Proof: Let $\mathbf{u}_1$ and $\mathbf{u}_2$ be the orthonormal eigenvectors corresponding to κ_1 and κ_2, respectively (as a result of Theorem 3.48). Since they span $T_{\mathbf{p}}M$, we can write any $\mathbf{u} \in U_{\mathbf{p}}M$ as

$$\mathbf{u} = \cos(\theta)\mathbf{u}_1 + \sin(\theta)\mathbf{u}_2$$

for some number $\theta \in [0, 2\pi)$ because it has unit length and thus lies on the circle (see the previous note). We can use this to express the normal curvature function in the following helpful way:

$$\begin{aligned} k_{\mathbf{p}}(\mathbf{u}) &= (\cos\theta\mathbf{u}_1 + \sin\theta\mathbf{u}_2) \cdot S_{\mathbf{p}}(\cos\theta\mathbf{u}_1 + \sin\theta\mathbf{u}_2) \\ &= \kappa_1 \cos^2(\theta) + \kappa_2 \sin^2(\theta) \\ &= \kappa_1 + (\kappa_2 - \kappa_1)\sin^2(\theta). \end{aligned}$$

Because $\sin^2(\theta) \in [0, 1]$, the above function clearly attains its minimum when $\alpha = 0$ and maximum when $\alpha = 1$,l which give κ_1 and κ_2, respectively. □

Example 3.54 Consider $M : (-\pi, \pi) \times \mathbb{R} \to \mathbb{R}$ given by $M(\theta, z) = (\sqrt{1+z^2}\cos\theta, \sqrt{1+z^2}\sin\theta, z)$, which is an orientable surface. We will compute the normal curvature function at the point $\mathbf{p} = (1, 0, 0) = M(0, 0)$. First, the coordinate basis vectors are

$$\varepsilon_\theta = (-\sqrt{1+z^2}\sin\theta, \sqrt{1+z^2}\cos\theta, 0) \qquad \text{and} \qquad \varepsilon_z = \left(\frac{z}{\sqrt{1+z^2}}\cos\theta, \frac{z}{\sqrt{1+z^2}}\sin\theta, 1\right).$$

Hence, we can compute the unit normal vector field in the usual way to obtain

$$N = \frac{(\sqrt{1+z^2}\cos\theta, \sqrt{1+z^2}\sin\theta, -z)}{\sqrt{1+2z^2}}.$$

At $\mathbf{p}$, we have $\varepsilon_\theta = (0, 1, 0)$ and $\varepsilon_z = (0, 0, 1)$; they are both elements of $U_{\mathbf{p}}M$. We can compute the shape operator and the normal curvature function at each of these coordinate basis vectors:

$$S_{\mathbf{p}}(\varepsilon_\theta) = -\nabla_{\varepsilon_\theta} N = -\left.\frac{\partial N}{\partial \theta}\right|_{(0,0)} = -(0, 1, 0) = -\varepsilon_\theta \qquad \Rightarrow \qquad k_{\mathbf{p}}(\varepsilon_\theta) = \varepsilon_\theta \cdot (-\varepsilon_\theta) = -1,$$

$$S_{\mathbf{p}}(\varepsilon_z) = -\nabla_{\varepsilon_z} N = -\left.\frac{\partial N}{\partial z}\right|_{(0,0)} = -(0, 0, -1) = \varepsilon_\theta \qquad \Rightarrow \qquad k_{\mathbf{p}}(\varepsilon_z) = \varepsilon_z \cdot \varepsilon_z = 1.$$

Note: The *sign* of the normal curvature function tells us whether the surface is curving towards its unit normal ($k_{\mathbf{p}}(\mathbf{u}) > 0$) or curving away from its unit normal ($k_{\mathbf{p}}(\mathbf{u}) < 0$).

If we know the principal curvatures κ_1 and κ_2, we may be able to use Theorem 3.53 to determine whether or not there exists a unit vector with a prescribed normal curvature value, that is if there exists $\mathbf{u} \in U_{\mathbf{p}}M$ such that $k_{\mathbf{p}}(\mathbf{u}) = c$ for whatever $c \in \mathbb{R}$ we want. Of course, if c does **not** lie between κ_1 and κ_2, then **no such u exists** immediately from the quoted theorem. The key equation to use, which we derived in the proof of this theorem, is

$$k_{\mathbf{p}}(\mathbf{u}) = \kappa_1 \cos^2(\theta) + \kappa_2 \sin^2(\theta).$$

Example 3.55 (Continued) Let M be the surface in Example 3.54 and $\mathbf{p} = (1, 0, 0) = M(0, 0)$ once again. We will construct a unit tangent vector whose image under the normal curvature function is zero. Indeed, recall that $S_{\mathbf{p}}(\varepsilon_\theta) = -\varepsilon_\theta$ and $S_{\mathbf{p}}(\varepsilon_z) = \varepsilon_z$. It follows from this that the principal curvatures are $\kappa_1 = -1$ and $\kappa_2 = 1$, with respective directions $\mathbf{u}_1 = \varepsilon_\theta$ and $\mathbf{u}_2 = \varepsilon_z$ (notice they are unit vectors already). To find such a unit vector $\mathbf{u}$ where $k_{\mathbf{p}}(\mathbf{u}) = 0$, we solve

$$0 = k_{\mathbf{p}}(\mathbf{u}) = \kappa_1 \cos^2(\theta) + \kappa_2 \sin^2(\theta) = -\cos^2(\theta) + \sin^2(\theta) \qquad \Rightarrow \qquad \cos^2(\theta) = \sin^2(\theta).$$

There are four solutions, namely $\cos(\theta) = \pm\frac{1}{\sqrt{2}}$ and $\sin(\theta) = \pm\cos(\theta)$. Therefore, we obtain

$$\mathbf{u} = \frac{1}{\sqrt{2}}(\varepsilon_\theta \pm \varepsilon_z) = \frac{1}{\sqrt{2}}(0, 1, \pm 1), \qquad \mathbf{u} = \frac{1}{\sqrt{2}}(-\varepsilon_\theta \pm \varepsilon_z) = \frac{1}{\sqrt{2}}(0, -1, \pm 1).$$

Exercise 45 Let M be an orientable surface which, at a certain point $\mathbf{p}$, has principal curvatures $\kappa_1 = -1$ and $\kappa_2 = 2$ with directions $\mathbf{u}_1 = (1, 0, 0)$ and $\mathbf{u}_2 = (0, \frac{1}{\sqrt{2}}, -\frac{1}{\sqrt{2}})$. Construct a vector $\mathbf{u} \in U_{\mathbf{p}}M$ with the stated property, or explain why no such $\mathbf{u}$ exists:

(i) $k_{\mathbf{p}}(\mathbf{u}) = 0$.

(ii) $k_{\mathbf{p}}(\mathbf{u}) = -2$.

(iii) $k_{\mathbf{p}}(\mathbf{u}) = -1$.

It is now that we can meet two more notions of curvature, given in terms of κ_1 and κ_2, the second of which is fundamentally important in differential geometry (see Part II of the module).

Definition 3.56 Let $M : U \to \mathbb{R}^3$ be an orientable surface and $\mathbf{p} \in M$.

(i) The mean curvature of M at $\mathbf{p}$ is $H(\mathbf{p}) = \frac{1}{2}(\kappa_1 + \kappa_2)$.

(ii) The Gauss curvature of M at $\mathbf{p}$ is $K(\mathbf{p}) = \kappa_1\kappa_2$.

The “easy” way to compute these is to solve the eigenvalue problem for a given shape operator $S_{\mathbf{p}}$ to obtain κ_1 and κ_2. However, it turns out there is a much simpler way to obtain both the mean and Gauss curvatures directly from the matrix of the shape operator.

Reminder: The trace $\mathrm{tr}(A)$ of an $n \times n$ matrix A is the sum of its diagonal entries.

Proposition 3.57 *Let M be an orientable surface. Then, for all $\mathbf{p} \in M$, we have*

$$H(\mathbf{p}) = \frac{1}{2}\,\mathrm{tr}(S_{\mathbf{p}}) \qquad \textit{and} \qquad K(\mathbf{p}) = \det\big(S_{\mathbf{p}}\big).$$

Proof: Denote by s_{ij} the entries of the matrix of the shape operator, and consider the polynomial

$$p(\lambda) = \det\big(S_{\mathbf{p}} - \lambda I_2\big) = \lambda^2 - (s_{11} + s_{22})\lambda + s_{11}s_{22} - s_{12}s_{21} = \lambda^2 - \mathrm{tr}(S_{\mathbf{p}})\lambda + \det\big(S_{\mathbf{p}}\big).$$

We know that κ_1 and κ_2 are roots, that is $p(\kappa_1) = 0$ and $p(\kappa_2) = 0$, so we can write

$$p(\lambda) = (\lambda - \kappa_1)(\lambda - \kappa_2) = \lambda^2 - (\kappa_1 + \kappa_2)\lambda + \kappa_1\kappa_2.$$

Of course, we could multiply the above by a constant and it would still have principal curvatures as roots, but the first equation for it is *monic* (i.e. the coefficient of λ^2 is one). By comparing coefficients in each expression for p, we see that $\mathrm{tr}(S_{\mathbf{p}}) = \kappa_1 + \kappa_2$ and $\det\big(S_{\mathbf{p}}\big) = \kappa_1\kappa_2$, from which the result is immediately clear. □

Example 3.58 Let M be the unit sphere as in Example 3.41, with shape operator given by

$$S_{\mathbf{p}} = \begin{pmatrix} -1 & 0 \\ 0 & -1 \end{pmatrix},$$

found in Example 3.42. By Proposition 3.57, at any $\mathbf{p} \in M$, the mean and Gauss curvatures are

$$H(\mathbf{p}) = \frac{1}{2}(-1 - 1) = -1 \qquad \text{and} \qquad K(\mathbf{p}) = (-1)(-1) = 1.$$

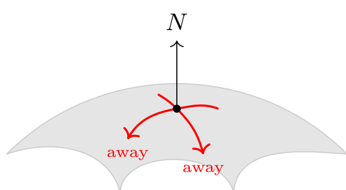
Note: If we change orientation $N \mapsto -N$, the shape operator $S_{\mathbf{p}} \mapsto -S_{\mathbf{p}}$. As such, the principal curvatures $\kappa_i \mapsto -\kappa_i$. A consequence of this is that the mean curvature $H \mapsto -H$, but the Gauss curvature $K \mapsto K$ does **not** change. In fact, the Gauss curvature is *intrinsic*; it is totally independent of the orientation we choose on the surface.

Gauss curvature is something you can see with your own two eyes when encountering an orientable surface in the wild. It is precisely the intrinsicity of the Gauss curvature that allows us to simply look at a point on some surface and determine whether it has positive, negative or zero Gauss curvature. We will conclude by explaining this and practising it by way of an example.

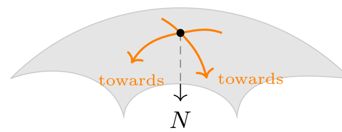
Method – Seeing Gauss Curvature: Let M be an orientable surface with orientation N.

- The Gauss curvature $K(\mathbf{p}) > 0$:
  - The principal curvatures κ_1 and κ_2 have the same sign.
  - The normal curvature $k_{\mathbf{p}}(\mathbf{u}) > 0$ everywhere **or** $k_{\mathbf{p}}(\mathbf{u}) < 0$ everywhere.
  - All curves curve **either** towards $N(\mathbf{p})$ **or** away from $N(\mathbf{p})$.
- The Gauss curvature $K(\mathbf{p}) < 0$:
  - The principal curvatures κ_1 and κ_2 differ in sign.
  - The normal curvature $k_{\mathbf{p}}(\mathbf{u})$ outputs **both** positive and negative values.
  - There exist curves which curve both towards and away from $N(\mathbf{p})$.

Positive curvature can be pictured in Figure 17 below, for the two cases in which they arise.



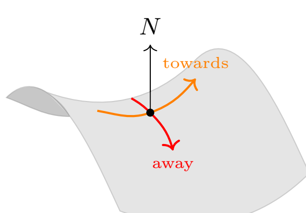
(a) The case where $k_{\mathbf{p}} < 0$.



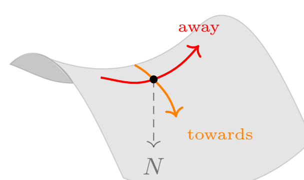
(b) The case where $k_{\mathbf{p}} > 0$.

Figure 17: Visualising positive Gauss curvature.

As for negative curvature, this can be pictured in Figure 18 below, similarly for two cases.



(a) The case of upward normal.



(b) The case of downward normal.

Figure 18: Visualising negative Gauss curvature.

The final case is zero Gauss curvature. This occurs if and only if (at least) one of the principal curvatures is zero; there is a curve through $\mathbf{p}$ which neither curves towards nor away from N.

3.3 Problem Set 3

We now provide a number of problems to complement the notes in Section 2. Solutions can be found at the end of the notes, but do try to work things out as best as possible without looking the answers up. Of course, if you are very stuck, take a peak at the solutions and try to unravel what you were finding difficult.

Problems for Section 3.1

3.1. (a)

(b)

3.2.

Problems for Section 3.2

3.3. (a)

(b)

3.4.

4 Exercise Solutions

We provide detailed solutions to the exercises interwoven within each section of the module. Hopefully you have given these questions a try whilst on your learning journey with the module. But mathematics is difficult, so don't feel disheartened if you had to look up an answer before you knew where to begin (we have all done it)!

Solutions to Exercises in Section 2

Exercise 1 Determine if the curve $\gamma : \mathbb{R} \to \mathbb{R}^2$ given by $\gamma(t) = (t^3, t^6)$ is regular.

Solution: The velocity $\gamma'(t) = (3t^2, 6t^5)$ is such that $\gamma'(0) = \mathbf{0}$. Hence, γ is **not** regular. □

Exercise 2 Prove that the tangent line $\widehat{\gamma}_{t_0}$ to a regular curve γ at any t_0 is itself regular.

Solution: Recall the tangent line to γ at t_0 is given by $\widehat{\gamma}_{t_0}(t) = \gamma(t_0) + t\gamma'(t_0)$. Differentiating, we see that $\widehat{\gamma}'_{t_0}(t) = \gamma'(t_0)$, recalling that $\gamma(t_0)$ is just a constant (vector) and therefore its derivative is $\mathbf{0}$. But we assume γ is regular, so we know $\gamma(t_0) \neq \mathbf{0}$. In other words, $\widehat{\gamma}'_{t_0}(t) \neq \mathbf{0}$ for any t. □

Exercise 3 Find any self-intersection points, if they in fact exist, of the parametrised curve $\gamma : \mathbb{R} \to \mathbb{R}^3$ given by $\gamma(t) = (t^2, 4t^3, t-1)$.

Solution: Suppose $\gamma(t_1) = \gamma(t_2)$, which is equivalent to $(t_1^2, 4t_1^3, t_1 - 1) = (t_2^2, 4t_2^3, t_2 - 1)$. By comparing the third coordinates, we see that $t_1 - 1 = t_2 - 2 \Rightarrow t_1 = t_2$. By Definition 2.9, a self-intersection only occurs when $t_1 \neq t_2$. This means that γ has **no** self-intersection points. □

Exercise 4 Compute the arc length of the regularly parametrised curve $\gamma : (0, \infty) \to \mathbb{R}^2$ given by $\gamma(t) = (t, \frac{2}{3}t^{\frac{3}{2}})$ from $t = 3$ to $t = 15$.

Solution: We appeal to Definition 2.11 to do this computation. Indeed, $\gamma'(t) = (1, t^{\frac{1}{2}})$ from which we see the speed is $\|\gamma'(t)\| = \sqrt{1+t}$. Therefore, substituting everything into the formula produces

$$s = \int_3^{15} \sqrt{1+t}\,\mathrm{d}t = \left[\frac{2}{3}(1+t)^{\frac{3}{2}}\right]_3^{15} = \frac{2}{3}(16^{\frac{3}{2}} - 4^{\frac{3}{2}}) = \frac{112}{3}.$$ □

Exercise 5 Let $\sigma_{t_0}(t)$ be the arc length function associated to some regular curve γ.

(i) Write down the value $\sigma_{t_0}(t_0)$ of the arc length function at the basepoint.

(ii) Write down the derivative $\sigma'_{t_0}(t)$ of the arc length function.

Solution: (i) We have $\sigma_{t_0}(t_0) = 0$ since the upper and lower limits of integration are equal.

(ii) From (another version of) the Fundamental Theorem of Calculus, we see that

$$\sigma'_{t_0}(t) = \frac{\mathrm{d}}{\mathrm{d}t}\left[\int_{t_0}^{t} \|\gamma'(u)\| \,\mathrm{d}u\right] = \|\gamma'(t)\|. \qquad \square$$

Exercise 6 Let τ_{t_0} be the inverse of σ_{t_0} as above. Write down the value $\tau_{t_0}(0)$.

Solution: Because τ_{t_0} is the inverse of σ_{t_0}, we know this means that $\tau_{t_0}\left(\sigma_{t_0}(t)\right) = t$ for all $t \in I$. But from Exercise 5, we know $\sigma_{t_0}(t_0) = 0$. Thus, $\tau_{t_0}(0) = t_0$ by substituting $t = t_0$ into the formula we just wrote down. $\square$

Exercise 7 Consider the regular curve $\gamma : (0, \infty) \to \mathbb{R}^2$ given by $\gamma(t) = (t, \frac{2}{3}t^{\frac{3}{2}})$ from Exercise 4 and fix the basepoint $t_0 = 1$. Find expressions for both $\tau_1(s)$ and $\tau_1'(s)$.

[**Hint:** You are free to use the expression $\sigma_1(t)$ we obtained in Example 2.15. You should write down the range $J = \mathrm{im}(\sigma_1)$ in your answer, which can be found by considering the limits $t \to 0$ and $t \to \infty$ of $\sigma_1(t)$.]

Solution: Recall in Example 2.15 we obtained the following arc length function based at $t_0 = 1$:

$$\sigma_1(t) = \frac{2}{3}\left((1+t)^{\frac{3}{2}} - 2^{\frac{3}{2}}\right).$$

The inverse function τ_1 has domain $J = \mathrm{im}(\sigma_1)$, but the domain of σ_1 is precisely the domain of the curve γ, that is $(0, \infty)$. Since σ_{t_0} is strictly increasing (Corollary 2.18), it is sufficient to look at its behaviour when $t \in (0, \infty)$ is "close" to each endpoint:

$$a := \lim_{t\to 0} \sigma_1(t) = \frac{2}{3}\left(1 - 2^{\frac{3}{2}}\right) \qquad \text{and} \qquad b := \lim_{t\to\infty} \sigma_1(t) = \infty.$$

Therefore, $J = (a, b)$ where a and b are given by the above limits. To actually find a formula for $\tau_1(s)$, start with $s = \sigma_{t_0}(t)$ and rearrange to make t the subject. With this in mind, we have

$$s = \frac{2}{3}\left((1+t)^{\frac{3}{2}} - 2^{\frac{3}{2}}\right) \quad \Rightarrow \quad 1 + t = \left(\frac{3}{2}s + 2^{\frac{3}{2}}\right)^{\frac{2}{3}} \quad \Rightarrow \quad t = \left(\frac{3}{2}s + 2^{\frac{3}{2}}\right)^{\frac{2}{3}} - 1.$$

Now that we have the expression $\tau_1(s) = \left(\frac{3}{2}s + 2^{\frac{3}{2}}\right)^{\frac{2}{3}} - 1$, we can carefully compute its derivative:

$$\tau_1'(s) = \frac{3}{2} \cdot \frac{2}{3}\left(\frac{3}{2}s + 2^{\frac{3}{2}}\right)^{\frac{2}{3}-1} = \left(\frac{3}{2}s + 2^{\frac{3}{2}}\right)^{-\frac{1}{3}}. \qquad \square$$

Note: In the proof of Lemma 2.21, we used the Chain Rule in another way to see that

$$\tau_{t_0}'(s) = \frac{1}{\|\gamma'(\tau_{t_0}(s))\|}.$$

We can check this agrees exactly with the (pretty nasty) expression in the above solution.

Exercise 8 Find a unit speed reparametrisation of the regular curve $\gamma : \mathbb{R} \to \mathbb{R}^4$ where

$$\gamma(t) = (t, \cos(2t), \sin(2t), 7).$$

Solution: The first task is to compute the arc length function $\sigma_0(t)$ based at $t_0 = 0$. Well then, $\gamma'(t) = (1, -2\sin(2t), 2\cos(2t), 0)$ and so $\|\gamma'(t)\| = \sqrt{1^2 + 4\sin^2(2t) + 4\cos^2(2t)} = \sqrt{5}$. Hence,

$$\sigma_0(t) = \int_0^t \sqrt{5}\,\mathrm{d}u = \sqrt{5}t.$$

It is now obvious to see that $\tau_0(s) = \frac{s}{\sqrt{5}}$. Finally, the unit speed reparametrisation is simply

$$\gamma(\tau_0(s)) = \left(\frac{s}{\sqrt{5}}, \cos\left(\frac{2s}{\sqrt{5}}\right), \sin\left(\frac{2s}{\sqrt{5}}\right), 7\right). \qquad \square$$

Exercise 9 Find the curvature vector of $\gamma : \mathbb{R} \to \mathbb{R}^3$ with $\gamma(s) = (\frac{1}{\sqrt{2}}\cos s, \frac{1}{\sqrt{2}}\sin s, \frac{1}{\sqrt{2}}s)$.

Solution: We can find the curvature vector in general by using Proposition 2.35. However, recall the much easier form in Definition 2.33 in the case that our curve is of unit speed. Well then,

$$\gamma'(s) = \left(-\frac{1}{\sqrt{2}}\sin s, \frac{1}{\sqrt{2}}\cos s, \frac{1}{\sqrt{2}}\right) \quad \Rightarrow \quad \|\gamma'(s)\| = \frac{1}{2}(\sin^2 s + \cos^2 s + 1) = 1.$$

Consequently, the curvature vector is

$$k(s) = \gamma''(s) = \left(-\frac{1}{\sqrt{2}}\cos s, -\frac{1}{\sqrt{2}}\sin s, 0\right). \qquad \square$$

Exercise 10 Find the curvature vector of $\gamma : \mathbb{R} \to \mathbb{R}^2$ with $\gamma(t) = (2\cos t, \sin t)$.

Solution: The curve γ is **not** a unit speed curve (this will become apparent very soon); we will have to use the full power of Proposition 2.35 to obtain the curvature vector. To that end then,

$$\gamma'(t) = (-2\sin t, \cos t) \qquad \Rightarrow \qquad \left\|\gamma'(t)\right\|^2 = 3\sin^2(t) + 1 \quad \text{and} \quad \gamma''(t) = (-2\cos t, -\sin t).$$

The dot product $\gamma'(t) \cdot \gamma''(t) = 3\sin(t)\cos(t)$. Hence, we can now write the curvature vector:

$$k(t) = \frac{1}{3\sin^2 t + 1}\left((-2\cos t, -\sin t) - \frac{3\sin t\cos t}{3\sin^2 t + 1}(-2\sin t, \cos t)\right). \qquad \square$$

Exercise 11 (Harder) Let $f : \mathbb{R} \to \mathbb{R}$ be smooth and consider $\gamma : \mathbb{R} \to \mathbb{R}^3$ given by

$$\gamma(t) = \left(t^2, f(t), \int_0^t e^{f(u)}\,\mathrm{d}u\right).$$

(i) Show that γ is a regularly parametrised curve.
(ii) Given that $f(1) = \log(2)$, $f'(1) = 0$ and $f''(1) = 1$, find the curvature vector $k(1)$.

[**Hint:** Write γ' and γ'' in terms of f and its derivatives and then substitute $t = 1$.]

Solution: (i) Because f is assumed smooth, and the integral of the smooth function $\exp \circ f$ is smooth, it follows that γ is smooth. It remains to verify that $\gamma'(t) \neq \mathbf{0}$ for any $t \in \mathbb{R}$. By the Fundamental Theorem of Calculus, we can differentiate the integral and this allows us to write

$$\gamma'(t) = (2t, f'(t), e^{f(t)}).$$

Because exp is a positive function, there is **no** $t \in \mathbb{R}$ with $e^{f(t)} = 0$, so it is true that γ is regular.

(ii) We are working specifically with $t = 1$, which will simplify some of the calculations. Similar to Exercise 10, we will use Proposition 2.35 to find $k(1)$. As usual, we begin the calculation with

$$\gamma'(1) = (2, 0, 2) \qquad \Rightarrow \qquad \left\|\gamma'(1)\right\|^2 = 8 \quad \text{and} \quad \gamma''(1) = (2, f''(1), f'(1)e^{f(1)}) = (2, 1, 0).$$

The dot product $\gamma'(1) \cdot \gamma''(1) = 4$. Hence, we can now write the curvature vector:

$$k(1) = \frac{1}{8}\left((2, 1, 0) - \frac{4}{8}(2, 0, 2)\right) = \frac{1}{8}(1, 1, -1). \qquad \square$$

Exercise 12 Compute the signed curvature for $\gamma : \mathbb{R} \to \mathbb{R}^2$ given by $\gamma(t) = (t, t^2)$.

Solution: We obtain $\kappa(t)$ via Lemma 2.45. Indeed, we first need to construct the unit tangent, and hence unit normal, vector. Well, $\gamma'(t) = (1, 2t)$ and so $\|\gamma'(t)\| = \sqrt{1+4t^2}$. Hence, we get

$$u(t) = \frac{1}{\sqrt{1+4t^2}}(1, 2t).$$

Consequently, the unit normal vector is obtained by swapping entries and negating the first:

$$n(t) = \frac{1}{\sqrt{1+4t^2}}(-2t, 1).$$

Next, we require the acceleration vector and its dot product with the unit normal. Indeed,

$$\gamma''(t) = (0, 2) \qquad \Rightarrow \qquad \gamma''(t) \cdot n(t) = \frac{2}{\sqrt{1+4t^2}}.$$

Dividing this by the square root of the speed gives us the signed curvature:

$$\kappa(t) = \frac{2}{(1+4t^2)^{\frac{3}{2}}}.$$

□

Exercise 13 Find all inflexion points of $\gamma : \mathbb{R} \to \mathbb{R}^2$ given by $\gamma(t) = (t, t + \sin t)$.

Solution: We begin by finding $\kappa(t)$, again using Lemma 2.45. Indeed, $\gamma'(t) = (1, 1 + \cos t)$ and thus $\|\gamma'(t)\| = \sqrt{2 + 2\cos t + \cos^2 t}$. We see that the unit tangent, and hence normal, vectors are

$$u(t) = \frac{1}{\sqrt{2+2\cos t+\cos^2 t}}(1, 1+\cos t) \qquad \Rightarrow \qquad n(t) = \frac{1}{\sqrt{2+2\cos t+\cos^2 t}}(-1-\cos t, 1).$$

Taking the dot product of the unit normal vector with the acceleration $\gamma''(t) = (0, -\sin t)$ and substituting into the formula for the signed curvature, we obtain the expression

$$\kappa(t) = -\sin(t)(2 + 2\cos t + \cos^2 t)^{-\frac{3}{2}}.$$

We are interested in inflexion points, i.e. solutions to $\kappa(\tau) = 0$ such that κ changes sign at $t = \tau$. In this case, $\kappa(t) = 0$ if and only if $t = n\pi$ for $n \in \mathbb{Z}$. But does κ change sign at these values? Yes it does, since the denominator is always positive. We have determined that γ has infinitely-many inflexion points $\gamma(n\pi) = (n\pi, n\pi)$.

□

Exercise 14 Compute the signed curvature for $\gamma : \mathbb{R} \to \mathbb{R}^2$ given by $\gamma(t) = (t, t^4)$.

Solution: Again, use Lemma 2.45 for this. Indeed, $\gamma'(t) = (1, 4t^3)$ and so $\|\gamma'(t)\| = \sqrt{1+16t^6}$. We can now write the unit tangent vector, from which we get the unit normal vector, namely

$$u(t) = \frac{1}{\sqrt{1+16t^6}}(1, 4t^3) \qquad \Rightarrow \qquad n(t) = \frac{1}{\sqrt{1+16t^6}}(-4t^3, 1).$$

We now require the acceleration vector as to compute its dot product with the unit normal:

$$\gamma''(t) = (0, 12t^2) \qquad \Rightarrow \qquad \gamma'(t) \cdot n(t) = \frac{12t^2}{\sqrt{1+16t^6}}.$$

Consequently, dividing the above dot product by the square root of the speed gives us our answer:

$$\kappa(t) = 12t^2(1+16t^6)^{-\frac{3}{2}}.$$ □

Exercise 15 (Harder) Prove the above solution to $\frac{\mathrm{d}\varphi}{\mathrm{d}s} = \kappa(s)$ with $\varphi(0) = \theta$ is unique.

[**Hint:** Assume there is another solution $f \neq \varphi$ satisfying $f(0) = \theta$ and apply the Mean Value Theorem to the function $g(s) := \varphi(s) - f(s)$ with the aim of obtaining a contradiction.]

Solution: Assume to the contrary that $f \neq \varphi$ is another solution with initial value θ. Because these two functions are **not** the same, there exists a value $x \in I$ in their domain (an interval) such that $f(x) \neq \varphi(x)$. Define the function

$$g(s) := \varphi(s) - f(s).$$

It is clear that $g(0) = \varphi(0) - f(0) = \theta - \theta = 0$ and that $g(x) \neq 0$ (since if it was, this would tell us that $f(x) = \varphi(x)$, a contradiction). Because f and φ are differentiable on the interval I, so too is this new function g. All of this is to say that the Mean Value Theorem (Theorem 2.16) applies to g, meaning there exists a value $y \in I$ lying between 0 and x such that

$$g'(y) = \frac{g(x) - g(0)}{x - 0} = \frac{g(x)}{x} \neq 0.$$

But because both f and φ solve the differential equation in the statement of the exercise, we know that $g'(s) = \varphi'(s) - f'(s) = \kappa(s) - \kappa(s) = 0$ for **all** $s \in I$, a contradiction. □

Exercise 16 Find a unit speed curve $\gamma : \mathbb{R} \to \mathbb{R}^2$ such that $\gamma(0) = (0,0)$ and $\gamma'(0) = (1,0)$ which has signed curvature $\kappa(s) = 0$ for every $s \in \mathbb{R}$ by using the formula (†).

Solution: Suppose $\kappa(s) \equiv 0$. Given $\gamma'(t) = (\cos\varphi(t), \sin\varphi(t))$ and $\gamma'(0) = (1,0)$, we see that $\varphi(0) = 0$. But from (†), using the fact the signed curvature is identically zero, it must be that $\varphi(s) = \theta + 0$. We therefore conclude $\varphi(s) = 0$ for all $s \in \mathbb{R}$. Substituting into (†) yields

$$\gamma(s) = \left(a + \int_0^s 1\,\mathrm{d}u, b + \int_0^s 0\,\mathrm{d}u\right) = (a+s, b).$$

Using the assumption $\gamma(0) = (0,0)$, we conclude that $a = b = 0$ and so our unit speed curve is precisely $\gamma(s) = (s, 0)$. Geometrically, this is a horizontal straight line. □

Exercise 17 Consider $\kappa(s) = s^2 - 1$. Determine the symmetries of the corresponding unit speed curve γ via Proposition 2.52 and find any inflexion points per Definition 2.47.

Solution: First, $\kappa(-s) = (-s)^2 - 1 = s^2 - 1 = \kappa(s)$, so the signed curvature is an even function. By Proposition 2.52, it follows that γ has reflective symmetry across the y-axis. Furthermore, note that $\kappa(s) = 0$ has solutions $s = 1$ and $s = -1$; there are two inflexion points since κ has a change of sign at these values. □

Exercise 18 Find the centre of curvature $c(1)$ of $\gamma : \mathbb{R} \to \mathbb{R}^2$ where $\gamma(t) = (t^2 - t, t + t^3)$.

Solution: From Definition 2.53, we see that $c(1) = \gamma(1) + \frac{1}{\kappa(1)} n(1)$. It remains to compute the unit normal vector and the signed curvature of γ (evaluated at $t = 1$). As such, we see that

$$\gamma'(t) = (2t - 1, 1 + 3t^2) \qquad \Rightarrow \qquad \gamma'(1) = (1,4) \quad \text{and} \quad n(1) = \frac{1}{\sqrt{17}}(-4, 1).$$

As for the signed curvature, one can compute

$$\gamma''(t) = (2, 6t) \qquad \Rightarrow \qquad \gamma''(1) = (2,6) \quad \text{and} \quad \kappa(1) = \frac{\gamma''(1) \cdot n(1)}{17} = \frac{-2}{17^{\frac{3}{2}}}.$$

Substituting everything into the equation for the centre of curvature, we obtain

$$c(1) = (0,2) - \frac{17^{\frac{3}{2}}}{2} \frac{1}{\sqrt{17}}(-4, 1) = \left(34, -\frac{13}{2}\right).$$

□

> **Exercise 19** Follow the *above* proof and derive the expressions we got for $\alpha(t)$ and $\beta(t)$.

Solution: Recall that $u(t)$ and $n(t)$ form an orthonormal basis of $\mathbb{R}^2$. Hence, we may write

$$u'(t) = \alpha(t)u(t) + \beta(t)n(t)$$

Taking the dot product of this equation with $u(t)$, we obtain

$$u'(t) \cdot u(t) = \alpha(t)\underbrace{u(t) \cdot u(t)}_{1} + \beta(t)\underbrace{n(t) \cdot u(t)}_{0} = \alpha(t).$$

If we consider the orthogonality equation and differentiate it using the Product Rule, we see that

$$u(t) \cdot u(t) = 1 \qquad \Rightarrow \qquad u'(t) \cdot u(t) + u(t) \cdot u'(t) = 0 \qquad \Rightarrow \qquad u'(t) \cdot u(t) = 0.$$

This tells us that $\alpha(t) = 0$, as expected. On the other hand, we consider the dot product

$$u'(t) \cdot n(t) = \alpha(t)\underbrace{u(t) \cdot n(t)}_{0} + \beta(t)\underbrace{n(t) \cdot n(t)}_{1} = \beta(t).$$

We again apply the Product Rule to the following dot product:

$$u(t) \cdot n(t) = 1 \qquad \Rightarrow \qquad u'(t) \cdot n(t) + u(t) \cdot n'(t) = 0 \qquad \Rightarrow \qquad u'(t) \cdot n(t) = -u(t) \cdot n'(t).$$

Comparing this to the previous expression for $u'(t) \cdot n(t)$, and using the expression for $n'(t)$ that was derived in the first part of the proof of Lemma 2.55, the result follows immediately:

$$\begin{aligned}
\beta(t) &= -u(t) \cdot n'(t) \\
&= -u(t) \cdot (-\kappa(t)\gamma'(t)) \\
&= u(t) \cdot \kappa(t)\gamma'(t) \\
&= \kappa(t)\frac{\gamma'(t) \cdot \gamma'(t)}{\|\gamma'(t)\|} \\
&= \kappa(t)\|\gamma'(t)\|.
\end{aligned}$$

□

> **Exercise 20** (Harder) Is the evolute E_γ itself always regular? Prove it or explain why not.
>
> [**Hint:** Although this is the standard thing to do, I'll write out what must be done: look at the derivative E'_γ and analyse when this is $(0, 0)$; the key is to look at the expression involving κ and κ'.]

Solution: Recall that the evolute is $E_\gamma(t) = \gamma(t) + \frac{1}{\kappa(t)}n(t)$. Differentiating this produces

$$E'_\gamma(t) = \gamma'(t) - \frac{\kappa'(t)}{\kappa(t)^2}n(t) + \frac{1}{\kappa(t)}n'(t).$$

But we know from Lemma 2.55 that $n'(t) = -\kappa(t)\gamma'(t)$. Substituting this yields

$$E'_\gamma(t) = \gamma'(t) - \frac{\kappa'(t)}{\kappa(t)^2}n(t) - \frac{1}{\kappa(t)}\kappa(t)\gamma'(t) = -\frac{\kappa'(t)}{\kappa(t)^2}n(t).$$

Therefore, we see that $E'_\gamma(t) = \mathbf{0}$ if and only if the derivative of the signed curvature $\kappa'(t) = 0$. In other words, the evolute is **not** regular if the signed curvature κ itself has critical points. □

> **Exercise 21** Construct the evolute of the regular curve $\gamma : \mathbb{R} \to \mathbb{R}^2$ given by $\gamma(t) = (t, t^2)$.

Solution: Recall again the evolute is $E_\gamma(t) = \gamma(t) + \frac{1}{\kappa(t)}n(t)$. We computed the signed curvature of γ in Exercise 12, wherein we also determined the unit normal vector, so there is little more to do than a substitution here. As a reminder, we have

$$n(t) = \frac{1}{\sqrt{1+4t^2}}(-2t, 1) \qquad \text{and} \qquad \kappa(t) = \frac{2}{(1+4t^2)^{\frac{3}{2}}}.$$

Consequently,

$$E_\gamma(t) = (t, t^2) + \frac{1+4t^2}{2}(-2t, 1) = \left(-4t^3, \frac{1}{2} + 3t^2\right).$$ □

> **Exercise 22** (Continued) Write the unit tangent vector, unit normal vector and arc length function of the evolute E_γ for the curve $\gamma : \mathbb{R} \to \mathbb{R}^2$ where $\gamma(t) = (t, t^2)$ from Exercise 21.

Solution: This is an application of Theorem 2.60. To that end, we see that

$$u^E(t) = n(t) = \frac{1}{\sqrt{1+4t^2}}(-2t, 1) \qquad \text{and} \qquad n^E(t) = -u(t) = -\frac{1}{\sqrt{1+4t^2}}(1, 2t).$$

As for the arc length function based at some arbitrary $t_0 \in I$, this is

$$\sigma^E_{t_0}(t) = \frac{(1+4t^2)^{\frac{3}{2}} - (1+4t_0^2)^{\frac{3}{2}}}{2}.$$ □

> **Exercise 23** Construct the involute of $\gamma : \mathbb{R} \to \mathbb{R}^2$ given by $\gamma(t) = (\cos t, \sin t)$ at $t_0 = 0$.

Solution: Recall that the involute at t_0 is $I_\gamma(t) = \gamma(t) - \sigma_{t_0}(t)u(t)$. Here, $t_0 = 0$ and the arc length function based at this time is precisely

$$\sigma_0(t) = \int_0^t \left\|\gamma'(u)\right\| \mathrm{d}u = \int_0^t \left\|(-\sin t, \cos t)\right\| \mathrm{d}u = \int_0^t 1 \, \mathrm{d}u = t.$$

As seen above, this is a unit speed curve, so $u(t) = \gamma'(t)$. So, we find that the involute at zero is

$$I_\gamma(t) = (\cos t, \sin t) - t(-\sin t, \cos t) = (\cos t + t\sin t, \sin t - t\cos t).$$

□

Exercise 24 Suppose $\gamma : \mathbb{R} \to \mathbb{R}^2$ is given by $\gamma(t) = (t, t^2)$ as in Example 2.69. Continuing from the example, deduce which values of λ ensure that the parallel curve γ_λ is regular.

[**Hint:** You have already found the signed curvature κ in Exercise 12; just use this with Lemma 2.68.]

Solution: Recall that a parallel curve is $\gamma_\lambda(t) = \gamma(t) + \lambda n(t)$ for some $\lambda \in \mathbb{R}$. Fortunately, we constructed the unit normal vector in the solution to Exercise 12:

$$n(t) = \frac{1}{\sqrt{1+4t^2}}(-2t, 1).$$

Therefore, the parallel curve to this here the parabola is

$$\gamma_\lambda(t) = (t, t^2) + \frac{\lambda}{\sqrt{1+4t^2}}(-2t, 1) = \left(t - \frac{2t\lambda}{\sqrt{1+4t^2}}, t^2 + \frac{\lambda}{\sqrt{1+4t^2}}\right).$$

Now, we know from Lemma 2.68 that γ_λ is regular if and only if $\kappa(t) \neq \frac{1}{\lambda}$ for any $t \in \mathbb{R}$. Recall also that Exercise 12 had us find the signed curvature of the parabola, namely

$$\kappa(t) = \frac{2}{(1+4t^2)^{\frac{3}{2}}}.$$

Notice that $0 < \kappa(t) \leq 2$ for all t, so γ_λ is regular if and only if $\frac{1}{\lambda} < 0$ or $\frac{1}{\lambda} > 2$. In other words, the parallel curve is regular if and only if $\lambda < \frac{1}{2}$. □

Exercise 25 Consider the curve $\gamma : \mathbb{R} \to \mathbb{R}^3$ given by $\gamma(s) = \frac{1}{5}(4s, \cos 3s, \sin 3s)$.

(i) Show that γ is a unit speed curve of non-vanishing curvature.

(ii) Construct the Frenet frame $[u(s), n(s), b(s)]$ of this curve.

Solution: (i) First, we see that $\gamma'(s) = \frac{1}{5}(4, -3\sin 3s, 3\cos 3s)$ from which we deduce $\left\|\gamma'(s)\right\| = 1$ for all $s \in \mathbb{R}$. Hence, γ is a unit speed curve. In this case, $k(s) = \gamma''(s)$ from Definition 2.33.

Assume to the contrary that this curve has somewhere-vanishing curvature, meaning

$$\gamma''(s) = 0 \quad \Leftrightarrow \quad -\frac{9}{5}(0, \cos 3s, \sin 3s) = \mathbf{0} \quad \Rightarrow \quad \cos(3s) = \sin(3s) = 0,$$

contradicting the trigonometric identity $\sin^2(x) + \cos^2(x) = 1$. Hence, $k(s) \neq 0$ for all $s \in \mathbb{R}$.

(ii) Because γ is a unit speed curve, we immediately know the unit tangent vector is the velocity:

$$u(s) = \gamma'(s) = \frac{1}{5}(4, -3\sin 3s, 3\cos 3s).$$

From the definition of the Frenet frame (Definition 2.71), the principal unit normal vector is

$$n(s) = \frac{k(s)}{\|k(s)\|} = \frac{\gamma''(s)}{\|\gamma''(s)\|} = (0, -\cos 3s, -\sin 3s).$$

Last but not least, the binormal vector is the vector product of the previous two. Explicitly,

$$b(s) = u(s) \times n(s) = \frac{1}{5}(3, 4\sin 3s, -4\cos 3s).$$

□

Exercise 26 (Continued) Let $\gamma : \mathbb{R} \to \mathbb{R}^3$ with $\gamma(s) = \frac{1}{5}(4s, \cos 3s, \sin 3s)$ from Exercise 25.

(i) Determine its torsion τ.

(ii) Verify that you get the same answer if you instead compute $-b'(s) \cdot n(s)$.

(iii) Does $\tau(s) = -b'(s) \cdot n(s)$ for **any** unit speed curve of non-vanishing curvature?

Solution: (i) Recall that torsion is defined by $b'(s) = -\tau(s)n(s)$. In this case, we have

$$b'(s) = \frac{12}{5}(0, \cos 3s, \sin 3s) = -\frac{12}{5}n(s) \quad \Rightarrow \quad \tau(s) = \frac{12}{5}.$$

(ii) On the other hand, we can in fact obtain the torsion via the following calculation:

$$-b'(s) \cdot n(s) = -\frac{12}{5}(0, \cos 3s, \sin 3s) \cdot (0, -\cos 3s, -\sin 3s) = \frac{12}{5}(\cos^2 3s + \sin^2 3s) = \frac{12}{5}.$$

(iii) It turns out that **yes**, $\tau(s) = -b'(s) \cdot n(s)$ for any unit speed curve of non-vanishing curvature. To justify this, simply take the dot product of the defining equation $b'(s) = -\tau(s)n(s)$ with the principal unit normal vector $n(s)$. Since the Frenet frame is orthonormal (Lemma 2.72), we get

$$b'(s) \cdot n(s) = -\tau(s) \underbrace{n(s) \cdot n(s)}_{1} = -\tau(s) \quad \Rightarrow \quad \tau(s) = -b'(s) \cdot n(s).$$

□

Exercise 27 (Continued) Verify that each of Proposition 2.79 and Lemma 2.80 are satisfied by the curve $\gamma : \mathbb{R} \to \mathbb{R}^3$ given by $\gamma(s) = \frac{1}{5}(4s, \cos 3s, \sin 3s)$.

Solution: (Proposition 2.79) The first thing we are asked to check is that

$$\tau(s) = \frac{\gamma'''(s) \cdot b(s)}{\kappa(s)},$$

where $\kappa(s) = \|k(s)\|$ is the *scalar* curvature (see Notation 2.74). We have computed $b(s)$ and $k(s) = \gamma''(s)$ in the solution to Exercise 25, and even the scalar curvature (implicitly). Hence, most of the heavy-lifting is complete and it is straightforward to obtain the following:

$$\begin{aligned}\gamma'''(s) \cdot b(s) &= -\frac{27}{5}(0, -\sin 3s, \cos 3s) \cdot \frac{1}{5}(3, 4\sin 3s, -4\cos 3s)\\ &= -\frac{27}{25}(-4\sin^2 3s - 4\cos^2 3s)\\ &= \frac{108}{25}\end{aligned}$$

and $\kappa(s) = \|\gamma''(s)\| = \frac{9}{5}$. Combining these two pieces together, Proposition 2.79 tells us that

$$\tau(s) = \frac{108/25}{9/5} = \frac{12}{5}.$$

(Lemma 2.80) The second thing we aim to verify is $n'(s) = -\kappa(s)u(s) + \tau(s)b(s)$. The left-hand side of this equation is obtained by differentiating the principal normal vector directly. Indeed,

$$n'(s) = (0, 3\sin 3s, -3\cos 3s).$$

As for the right-hand side, we calculate that it is

$$\begin{aligned}-\kappa(s)u(s) + \tau(s)b(s) &= -\frac{9}{5}\left(\frac{1}{5}(4, -3\sin 3s, 3\cos 3s)\right) + \frac{12}{5}\left(\frac{1}{5}(3, 4\sin 3s, -4\cos 3s)\right)\\ &= \left(-\frac{36}{25} + \frac{36}{25}, \frac{27}{25}\sin 3s + \frac{48}{25}\sin 3s, -\frac{27}{25}\cos 3s - \frac{48}{25}\cos 3s\right)\\ &= (0, 3\sin 3s, -3\cos 3s)\\ &= n'(s),\end{aligned}$$

which is exactly what we were asked to demonstrate. □

Exercise 28 (Harder) We call a curve $\gamma : I \to \mathbb{R}^3$ spherical if its image $\gamma(I)$ lies entirely on the surface of a sphere. Let γ be a spherical unit speed curve on the sphere $S \subseteq \mathbb{R}^3$ whose radius is r and is centred at the point $\mathbf{p} \in \mathbb{R}^3$; this means that $\left\|\gamma(s) - \mathbf{p}\right\|^2 = r^2$ for all $s \in I$. Prove that γ has a well-defined Frenet frame.

[**Hint:** This looks bizarrely complicated, but to show its Frenet frame is well-defined, it is enough to show that it has non-vanishing curvature. To do this, write $\left\|\gamma(s) - \mathbf{p}\right\|^2 = r^2$ in dot product form and differentiate it twice. Then, use some Frenet formulae to simplify it and conclude that $\kappa(s) \neq 0$.]

Solution: Start with $\left\|\gamma(s) - \mathbf{p}\right\|^2 = (\gamma(s) - \mathbf{p}) \cdot (\gamma(s) - \mathbf{p})$. Using the Product Rule, we obtain

$$\frac{\mathrm{d}}{\mathrm{d}s}\left[\left\|\gamma(s) - \mathbf{p}\right\|^2\right] = u(s) \cdot (\gamma(s) - \mathbf{p}) + (\gamma(s) - \mathbf{p}) \cdot u(s) = 2(\gamma(s) - \mathbf{p}) \cdot u(s),$$

since γ is unit speed (so $\gamma' = u$). Hence, differentiating the equation $\left\|\gamma(s) - \mathbf{p}\right\|^2 = r^2$ yields

$$2(\gamma(s) - \mathbf{p}) \cdot u(s) = 0 \qquad \Rightarrow \qquad (\gamma(s) - \mathbf{p}) \cdot u(s) = 0.$$

Differentiating the latter equation once more and again applying the Product Rule, we obtain

$$u(s) \cdot u(s) + (\gamma(s) - \mathbf{p}) \cdot u'(s) = 0.$$

The orthonormality of the Frenet frame implies $u(s) \cdot u(s) = 1$. The Frenet formulae (Theorem 2.81) also tells us that $u'(s) = \kappa(s)n(s)$. Substituting these into the above produces this:

$$1 + (\gamma(s) - \mathbf{p}) \cdot \kappa(s)n(s) = 0.$$

This actually concludes the proof, because if the curve did **not** have non-vanishing curvature, i.e. there exists $s \in I$ such that $\kappa(s) = 0$, then substituting that particular s into the above equation would result in $1 = 0$, which is of course a huge contradiction. □

Exercise 29 Determine if the curve $\gamma : \mathbb{R} \to \mathbb{R}^3$ where $\gamma(s) = \frac{1}{5}(4s, \cos 3s, \sin 3s)$ is planar.

Solution: We will use Theorem 2.85, which says that a curve is planar if and only if its torsion is zero everywhere. Well, we computed the torsion of this curve in Exercise 26: we found that $\tau(s) \equiv \frac{12}{5} \neq 0$. In a sense, this is the "worst-case" scenario when it comes to planarity, given that the torsion is zero exactly nowhere. Thus, we conclude that γ is **not** planar. □

Solutions to Exercises in Section 3

5 Problem Solutions

We provide detailed solutions to the problems found at the end of each section of the module. Hopefully you have given these questions a try whilst on your learning journey with the module. But mathematics is difficult, so don't feel disheartened if you had to look up an answer before you knew where to begin (we have all done it)!

Solutions to Problems in Section 1

1.1.

Solutions to Problems in Section 2

2.1.

Solutions to Problems in Section 3

3.1.